



Instituto Politécnico de Leiria  
Escola Superior de Tecnologia e Gestão  
Departamento de Engenharia Informática  
Mestrado em Ciência de Dados

MODELLING GAMIFIED EXERCISE DATA TO SUPPORT  
THERAPEUTIC MONITORING OF FRAILTY IN OLDER  
ADULTS

ESTUDANTE JOÃO MIGUEL ANTÓNIO

Leiria, Março de 2024





ESCOLA SUPERIOR  
DE TECNOLOGIA  
E GESTÃO

Instituto Politécnico de Leiria  
Escola Superior de Tecnologia e Gestão  
Departamento de Engenharia Informática  
Mestrado em Ciência de Dados

MODELLING GAMIFIED EXERCISE DATA TO SUPPORT  
THERAPEUTIC MONITORING OF FRAILTY IN OLDER  
ADULTS

ESTUDANTE JOÃO MIGUEL ANTÓNIO  
Número: 2210786

Dissertação realizada sob orientação do Professor Doutor Ricardo Malheiro ([ricardo.malheiro@ipleiria.pt](mailto:ricardo.malheiro@ipleiria.pt)).

Leiria, Março de 2024



## ACKNOWLEDGMENTS

---

The completion of this master's dissertation had important support and incentives without which it would not have become a reality. To all addressed below, I am eternally grateful.

I express my deepest gratitude to my professor, supervisor and element of my committee, Prof. Doc. Ricardo Malheiro, for his invaluable guidance and prompt availability, which were crucial to my success.

Additionally, I thank Techframe, for providing me with an interesting and challenging project canvas, as well as the resources and funding needed for its completion.

I am grateful to my partner, Rosa, for her undying support. As her nursing degree lends credibility to my work, so did her acts of encouragement motivate me to strive for excellence. To her, I dedicate this work.

I also thank my professor and friend Sandra Jardim, with whom I have previously written and published multiple works. The experience I was awarded over the last few years has helped tremendously in organizing this work.

To my close friends, including Luke Davies, José Roque, among others, who provided companionship and much needed distractions during more difficult times, I thank you all.

Finally, I extend a special acknowledgement to my parents, for their unconditional support throughout the entire endeavour, and to my brother, who will always be my best friend.



## ABSTRACT

---

Research indicates that video games can provide mental stimulation and stress relief, which may contribute to the well-being of gamers. With the continued growth of the video game industry and its demographic, attempts to evoke similar feelings of enjoyment in non-game environments have been proposed, gradually defining the concept of 'gamification'. Over the last two decades, gamification has shown positive effects on user engagement and participation in tasks across several fields of research, such as education and workplace productivity. In the area of healthcare, specifically in subjects of physical therapy and rehabilitation exercise, the implementation of game design choices and social features is evident in mobile applications and smart devices, however, despite the outlined benefits, the adoption of these strategies by the elderly population faces methodological difficulties that bar its progress. Research highlights a lack of process standardization, the employment of inefficient game concepts, and improper design choices for the referred demographic. In this thesis, the complications associated with the employment of gamification strategies in senior healthcare are explored and discussed, leading up to the proposal of a methodological approach to improving the success of such strategies, through the conception and development of a data mining project applied to a real clinical scenario. The use of this approach has been shown to effectively improve the quality of gamification and aid in making better business decisions based on the underlying data. Moreover, five distinct data models and processes were developed, acting on improving the platform with feedback collection, automated game suggestions and logical patient profiling based on physical factors.



## RESUMO

---

Estudos indicam que os videojogos podem proporcionar estímulo mental e alívio de stress, contribuindo assim para o bem-estar dos jogadores. À medida que a indústria e alcance demográfico dos videojogos aumenta, surgem também tentativas de evocar sensações semelhantes em ambientes não relacionados com os jogos, constituindo assim a formação do conceito de "gamificação". Nas últimas duas décadas, a gamificação demonstrou efeitos positivos no envolvimento e participação dos utilizadores em tarefas de várias vertentes da indústria, tais como na educação e na produtividade em local de trabalho. Na área da saúde, mais concretamente, em aplicações de fisioterapia e exercícios de reabilitação, a implementação de conceitos de jogos e funcionalidades sociais é evidente, por exemplo, em aplicações móveis e dispositivos inteligentes. No entanto, apesar dos benefícios apresentados, a adoção de estratégias de gamificação pela população idosa enfrenta dificuldades metodológicas que impedem o seu progresso. Em particular, destaca-se a falta de padronização de processos, uma ineficiência dos conceitos de videojogos utilizados, e a implementação de funcionalidades e escolhas de design inadequadas para o grupo demográfico referido anteriormente. Nesta dissertação, é feita uma exploração e discussão das complicações associadas ao emprego de estratégias de gamificação nos cuidados de saúde sénior, levando à proposta de uma abordagem metodológica padronizada para melhorar o grau de sucesso de tais estratégias, e a sua aplicação prática com o planeamento e desenvolvimento de um projeto de *Data Mining* aplicado a um cenário clínico verídico. A utilização desta abordagem demonstrou ser eficaz na melhoria da qualidade da gamificação, e no suporte à decisão comercial com base nos dados subjacentes. Finalmente, foram também desenvolvidos cinco modelos de dados que, em conjunto, atuam na melhoria de uma plataforma de jogos, com base em princípios de recolha de *feedback*, mecanismos de recomendação, e agrupamentos lógicos de pacientes com base nos seus perfis clínicos.



## CONTENTS

---

Acknowledgments	i
Abstract	iii
Contents	vii
List of Figures	xi
List of Tables	xiii
List of Acronyms	xvii
1 Introduction	1
2 Related Work	5
2.1 Gamification - Origins and Modern Day Applications . . . . .	5
2.2 Data Mining in Healthcare . . . . .	11
2.3 Critical analysis . . . . .	13
3 Context	21
3.1 Frailty Syndrome and Sarcopenia . . . . .	21
3.1.1 Assessing Frailty . . . . .	22
3.2 GameVitae . . . . .	23
3.2.1 Relationship with gamification . . . . .	23
3.2.2 Limitations . . . . .	25
3.3 Motivation . . . . .	26
4 Methods	29
4.1 Business Understanding . . . . .	30
4.1.1 Business Objectives . . . . .	30
4.1.2 Business Assessment . . . . .	35
4.1.3 Data Mining Goals . . . . .	43
4.1.4 Project Planning . . . . .	46
4.2 Data Understanding . . . . .	47
4.2.1 Data Collection . . . . .	49
4.2.2 Data Description . . . . .	51
4.2.3 Data Exploration . . . . .	54
4.2.4 Data Quality Assessment . . . . .	55

4.3	Data Preparation . . . . .	56
4.3.1	Data Selection . . . . .	57
4.3.2	Data Cleaning . . . . .	58
4.3.3	Data Engineering . . . . .	58
4.3.4	Data Integration . . . . .	59
4.3.5	Formatting . . . . .	59
4.4	Modelling . . . . .	60
4.4.1	Modelling Techniques . . . . .	60
4.4.2	Test Design . . . . .	63
4.4.3	Model Building . . . . .	63
4.4.4	Model Assessment . . . . .	69
4.5	Evaluation . . . . .	73
4.5.1	Results . . . . .	73
4.5.2	Data Mining Process Revision . . . . .	74
4.5.3	Next Steps . . . . .	75
4.6	Deployment . . . . .	76
4.6.1	Plan Deployment . . . . .	76
4.6.2	Monitoring and Maintenance . . . . .	79
4.6.3	Report . . . . .	79
4.6.4	Project Revision . . . . .	79
5	Conclusions and Future Work	83
	Bibliography	85
	<b>Appendices</b>	
A	Appendix A - Project Glossary	99
A.1	Glossary . . . . .	99
A.1.1	Project-specific Terms . . . . .	99
A.1.2	Healthcare Terms . . . . .	100
A.1.3	Data Mining Terms . . . . .	101
B	Appendix B - Project Plan	103
B.1	Project Schedule . . . . .	103
B.2	Risk Scheduling . . . . .	103
B.3	Communication Plan . . . . .	105
B.3.1	Publication Vehicle . . . . .	105

B.3.2	Financing . . . . .	105
B.3.3	Publication Time Window . . . . .	106
B.4	Future Events . . . . .	106
B.4.1	Model Assessment Trials . . . . .	106
B.4.2	Model Incubation . . . . .	107
c	Appendix C - Data Understanding Report . . . . .	109
c.1	Data Loading . . . . .	109
c.1.1	First Look at the Data . . . . .	109
c.1.2	Additional Considerations . . . . .	110
c.2	Data Description . . . . .	110
c.2.1	Patient Description . . . . .	110
c.2.2	Activity Description . . . . .	112
c.3	Data Exploration . . . . .	112
c.3.1	Patient Data Exploration . . . . .	113
c.3.2	Activity Data Exploration . . . . .	123
c.3.3	Merged Exploration . . . . .	128
c.4	Data Quality . . . . .	131
c.4.1	Structural Analysis . . . . .	131
c.4.2	Missing Values . . . . .	134
c.4.3	Outliers . . . . .	134
D	Appendix D - Data Cleaning Report . . . . .	137
D.1	Structural Errors . . . . .	137
D.1.1	Renaming Attributes . . . . .	137
D.1.2	Categorical Correction . . . . .	138
D.1.3	Data Type Correction . . . . .	139
D.2	Missing Values . . . . .	139
D.3	Deviations . . . . .	140
E	Appendix E - Data Engineering Report . . . . .	141
E.1	Attribute Engineering . . . . .	141
E.1.1	Year of Birth . . . . .	141
E.1.2	FRIED Count . . . . .	142
E.1.3	Engagement Score . . . . .	142
E.2	Dataset Preprocessing . . . . .	143
E.2.1	Clustering Dataset . . . . .	143
E.2.2	Recommendation System Dataset . . . . .	144

CONTENTS

F	Appendix E - Test Design	145
F.1	Test Data . . . . .	145
F.2	Supervised Model Testing . . . . .	145
F.2.1	Quality Metrics . . . . .	146
F.2.2	Test Process . . . . .	146
F.2.3	Failing Criteria . . . . .	147
F.3	Unsupervised Model Testing . . . . .	147
F.3.1	Quality Metrics . . . . .	147
F.3.2	Test Process . . . . .	148
F.3.3	Failing Criteria . . . . .	148
G	Appendix G - Project Evaluation	151
G.1	Business Success Evaluation . . . . .	151
G.1.1	Result Presentation Meeting . . . . .	151
G.1.2	Rationale for Objective Completion . . . . .	154
G.2	Data Mining Process Summary . . . . .	156
G.2.1	Business Understanding Phase . . . . .	156
G.2.2	Data Understanding Phase . . . . .	157
G.2.3	Data Preparation Phase . . . . .	158
G.2.4	Data Modelling Phase . . . . .	159
G.2.5	Critical Remarks . . . . .	160
H	Appendix H - Deployment Plan	161
H.1	Model Deployment . . . . .	161
H.1.1	GameVitae API . . . . .	162
H.2	Deploying Findings . . . . .	166
H.2.1	Conclusive Findings . . . . .	167
H.2.2	Other Findings . . . . .	168
H.3	Deployment Schedule . . . . .	168
	Statement	171

## LIST OF FIGURES

---

Figure 1	Number of steam releases games by year. 2004-2005 data: (statista, 2023). 2006-2023 data: (SteamDB, 2023) . . . . .	14
Figure 2	Number of article publications in three popular publishers: 2005-2018 (left), 2019-2023 (right). Data collected by configuring similar searches on Springer, ScienceDirect and IEEE search engines. . . . .	15
Figure 3	The senior population faces many barriers that deter the adoption of new technologies. (a) (Smith, 2014). (b) (Eurostat, 2018). . . . .	16
Figure 4	Number of publications containing the keywords "Data Mining", "Healthcare", and a combination of the two. Data collected by configuring searches on Springer's search engine . . . . .	17
Figure 5	Versions of the GripWise dynamometer. Adapted from (Fernandes, 2022) . . . . .	22
Figure 6	Deconstruction of the CRISP-DM Methodology . . . . .	29
Figure 7	Organizational chart of available indirect human resources . . . . .	32
Figure 8	Breakdown of subjects related to project . . . . .	33
Figure 9	Process of elimination throughout trial stages . . . . .	50
Figure 10	Distortion Elbow Method . . . . .	64
Figure 11	Cluster Profiling . . . . .	70
Figure 12	KNN Parameter Optimization Results . . . . .	71
Figure 13	Project schedule . . . . .	104
Figure 14	Patient demographic variable exploration . . . . .	114
Figure 15	Self-reported variable exploration . . . . .	115
Figure 16	Substance consumption variable exploration . . . . .	116
Figure 17	Rest and exercising variable exploration . . . . .	117
Figure 18	Audition and vision variable exploration . . . . .	117
Figure 19	Key frailty factor variable exploration . . . . .	118
Figure 20	FRIED variable exploration . . . . .	119
Figure 21	Demographic and anthropometric factors in FRIED assessment	120
Figure 22	Effects of weight loss and low grip strength on frailty assessment	120
Figure 23	Effects of sedentary behaviour and exhaustion on frailty assessment	121
Figure 24	Effects of balance on frailty assessment . . . . .	122
Figure 25	Presence of multiple risk factors on frailty assessment . . . . .	123

LIST OF FIGURES

Figure 26	Distribution of sessions played per game characteristics . . . . .	124
Figure 27	Difficulty range per game . . . . .	125
Figure 28	Range of therapeutic targets for activities . . . . .	126
Figure 29	Range of therapeutic scores for activities . . . . .	126
Figure 30	Success rate of activity sessions . . . . .	127
Figure 31	Score distribution per game . . . . .	127
Figure 32	Patient satisfaction . . . . .	128
Figure 33	Activity success rate per FRIED level . . . . .	129
Figure 34	Risk factor impact on activity targets . . . . .	129
Figure 35	Remaining risk factors and overall activity scores . . . . .	130
Figure 36	Share of sessions played by FRIED level . . . . .	131
Figure 37	Risk factor impact on activity targets . . . . .	135
Figure 38	Outlier assessment for activity scores . . . . .	136
Figure 39	Count of risk factor coexistence . . . . .	142
Figure 40	Split of patients with satisfactory engagement scores . . . . .	143
Figure 41	Result deployment schedule . . . . .	169

## LIST OF TABLES

---

Table 1	Business Understanding Breakdown . . . . .	30
Table 2	Business objective reference table . . . . .	33
Table 3	Hardware Specification Table . . . . .	36
Table 4	Software and web-based resources for the project . . . . .	37
Table 5	Risk documenting and contingency planning table . . . . .	40
Table 6	Retroactive, recurrent and future cost evaluation for the project.	42
Table 7	Data Mining Goals . . . . .	44
Table 8	Success criteria for previously defined data mining goals . . . .	47
Table 9	Data mining techniques and roles . . . . .	48
Table 10	Data Understanding Breakdown . . . . .	49
Table 11	Categories for self-assessed measures of depression, anxiety and pain. *The GAD-7 score maxes out at 21. . . . .	53
Table 12	Data Preparation Breakdown . . . . .	57
Table 13	Removed attributes by model . . . . .	58
Table 14	Modelling Breakdown . . . . .	60
Table 15	Evaluation Breakdown . . . . .	73
Table 16	Business objective completion summary . . . . .	74
Table 17	Data Mining revision summary . . . . .	75
Table 18	Deployment Breakdown . . . . .	76
Table 19	Summary of deployable project results . . . . .	77
Table 20	Project phases and associated risks . . . . .	104
Table 21	Patient data technical description . . . . .	111
Table 22	Activity data technical description . . . . .	112
Table 23	FRIED category distribution by number of risk factors . . . . .	123
Table 24	Structural deficiencies in both datasets . . . . .	133
Table 25	Missing values in both datasets . . . . .	134
Table 26	Attribute renaming process . . . . .	138
Table 27	Categorical correction process . . . . .	139
Table 28	Data type correction process . . . . .	139
Table 29	Missing value replacement strategy . . . . .	140
Table 30	Model deployment overview sheet . . . . .	162
Table 31	GameVitae <a href="#">API</a> technical information . . . . .	163

LIST OF TABLES

Table 32	Model monitoring overview . . . . .	165
Table 33	Technical difficulty contingency planning . . . . .	167





## LIST OF ACRONYMS

---

ADHD	Attention-Deficit/Hyperactivity Disorder. , 7
AI	Artificial Intelligence. , 10, 12, 18, 19
ANN	Artificial Neural Network. , 12
APC	Article Processing Charge. , 105
API	Application Interface. <a href="#">xiii</a> , , 37, 42, 72, 77, 79, 107, 162–167
ASD	Autism Spectrum Disorder. , 7
BIA	Bio Electrical Impedance Analysis. , 22
BMI	Body Mass Index. , 113, 114, 119, 120, 134
BNN	Bayesian Belief Network. , 13
CFS	Clinical Frailty Scale. , 22
CRISP-DM	Cross Industry Standard Process for Data Mining. , 2, 29, 30, 36, 41, 43, 46, 54, 58, 73–76, 78, 80, 103, 106, 137, 145, 151, 153, 156, 159–161
CT	Computerized Tomography. , 12
CUDA	Compute Unified Device Architecture. , 36
DDoS	Distributed Denial of Service. , 163
DT	Decision Tree. , 11
DXA	Dual Energy Xray Absorptiometry. , 22
ED	Estimation of Distribution. , 12
ER	Emotional Regulation. , 7
ESA	Entertainment Software Association. , 13

## List of Acronyms

FS	Frailty Syndrome. , 21, 22
GDPR	General Data Protection Regulation. , 36, 39
GPU	Graphical Processing Unit. , 36
GV	GameVitaе. , 23, 25-27, 30-32, 35, 38, 41-44
IIS	Internet Information Services. , 163, 167
IT	Information Technologies. , 23, 42
KDD	Knowledge Discovery in Data. , 17
KNN	K-Nearest Neighbors. , 61, 62, 64, 65, 67, 68, 143, 145-147
LDA	Linear Discriminant Analysis. , 12
LGR	Logistic Regression. , 11
LR	Linear Regression. , 12
MDPI	Multidisciplinary Publishing Institute. , 105, 106
ML	Machine Learning. , 12
MTS	Mahalanobis Taguchi System. , 11
MUD <sub>1</sub>	Multiplayer Dungeon 1. , 6
NRS	Numerical Rating Scale. , 53
OA	(Knee) Osteoarthritis. , 11, 15
PNN	Probabilistic Neural Network. , 12
PTSD	Post-Traumatic Stress Disorder. , 7
RA	Rheumatoid Arthritis. , 10, 15
REFS	Reported Edmonton Frail Scale. , 22

RFMI	Rockwood Mitnitski Frailty Index. , 22
SQL	Structured Query Language. , 35, 42, 50, 163–166
SVDD	Support Vector Data Description. , 11
SVM	Support Vector Machine. , 11
UWP	Universal Windows Platform. , 26
VPN	Virtual Private Network. , 167
VR	Virtual Reality. , 9, 10, 22



## INTRODUCTION

---

As the digital era gains traction, efforts regarding the usage of gamification approaches to a broad range of health and lifestyle subjects have been constantly growing. Generally speaking, gamification aims to provide the means to increase an individual's fun, engagement and compliance to a specific task (Lenihan, 2012) by introducing interactivity and satisfaction factors to otherwise mundane or uninteresting chores, creating experiences similar to those observed when playing games (J. Hamari, 2019). Usually, this is achieved through thoughtful application of game-design knowledge and principles in non-game contexts (Deterding, 2011). Earlier gamification strategies use rewards to encourage the player to reach a predetermined set of goals or to engage in specific activities. These rewards vary from the more traditional point systems (Sutter, 2010) to more in-depth achievement badges (Hamari, 2011), though recent shifts in the industry and the introduction of blockchains have inevitably made their impact on modern gamification efforts (Parizi, 2018). In some areas of the industry, such as education and workplace productivity, gamification has been accepted and used to great benefit for a long time, with shown improvements in the learning ability (Abu-dawood, 2016; Dong et al., 2012; Kerfoot et al., 2012; Nevin et al., 2014) and decision making of students (Lin et al., 2015), as well as perceived psychological improvements in office workers (Coonradt and Nelson, 1985).

The application of gamification in fields related to population health has been gradually researched and tested for its proposed benefits in improving patient engagement and motivation, with some preliminary research suggesting improvements in psychological behaviour (Derksen et al., 2020; Lau et al., 2017; Villani et al., 2018), and supporting the addition of game features in exercise programs (Brauner et al., 2013; Dias, Tibes, et al., 2017; Lentelink et al., 2013; Munson and Consolvo, 2012). Ailments of the musculoskeletal system, in particular, are closely associated with this work, what with them being much more common in older adults and thus affecting a population that is less familiar with the concepts employed in gamified programs (J. Clement, 2022). Given the specificity of this subject, the amount of available research supporting its pairing with gamification is extremely limited, with only some efforts reporting on its safety aspects (Özlu et al., 2023; Varga et al., 2021).

In contrast, the healthcare industry is massively potentiated by data mining, an umbrella term that covers a wide range of processes and methods for extracting valuable information from data. One approach commonly used by hospitals and clinical businesses is anomaly detection, where atypical changes in datasets are analyzed with the aid of data models for enabling better decisions about patient conditions (B. Liu et al., 2013; Šabić et al., 2021). There are many relevant works on the effects of data mining techniques in improving healthcare processes, from calculating disease incidence (Das et al., 2009; Er et al., 2010; Mamiya et al., 2015; Samanta et al., 2009) and recurrence (Khan et al., 2008), to revealing risk factors for particular conditions (Curiac et al., 2009; Haraty et al., 2015; Moon et al., 2012). More specifically, pattern discovery in clinical scenarios is closely related to this work, and has been used to integrate clustering-based approaches for fraud detection (Hillerman et al., 2017; Q. Liu and Vasarhelyi, 2013), decision support (Elbattah and Molloy, 2017), and patient classification (Fuente-Tomas et al., 2019; Grant et al., 2020; Silitonga, 2017). Moreover, the specific sub-field of machine learning has seen relevance in methodological attempts to personalize healthcare to the individual, with the emergence of recommendation systems for automating several areas of health, such as medication planning (Suryadevara, 2020), diagnosis (Bhatti et al., 2019) and dieting (Kim et al., 2009) by extracting features from patient data, although achieving a significant degree of data representation is often difficult due to quality and access restrictions found in clinical data. Often, datasets generated from clinical records contain noise and are prone to incur in potential privacy concerns, which ultimately elevates analysis costs while reducing model accuracy (Fang et al., 2016; Vimark, 2023).

Despite the benefits that either discipline brings to healthcare scenarios, the usage of data mining to drive a more effective gamification effort is a seemingly unexplored approach, constituting a clear gap in research and a fault in the industry. In this thesis, I propose the development and deployment of a data mining project about a gamified platform for the monitoring of frailty in older adults, with the goal of revealing insights that lead to a conscious and effective proposal of data models for the improvement of the platform. More specifically, a standardized process (CRISP-DM) will be used to guide the development of the individual components of a data mining project, from the identification of business requirements to the conception and deployment of the models themselves.

The results brought on by this approach set up a positive precedent in the way gamification is employed in healthcare. From a methodological standpoint, the intricacies of following an objective conscious process allow for a closer understanding of the targeted medical conditions and how patients respond to the employed gaming concepts. Similarly, the thoroughness of the data related tasks that happen prior to modelling (i.e.: exploration,

cleaning and preparation of the data) help counteract the identified deficiencies regarding a lack of standardized procedures, inefficient game design choices, and poor data quality.



## RELATED WORK

---

At the core of this work stand two disciplines that both play an important role in the way modern systems are conceptualized and implemented. The creation and use of systems that use engagement mechanisms generally associated with video games is still seen as a novelty, although certain studies and their results demonstrate a clear aptitude for this 'gamification' to develop further. On the other hand, the collection and analysis of data from these systems is familiar with data mining tasks, an area which is already well established and displays a trend to continue growing significantly. As a means to aid with thought organization and better structuring, this section will be divided in two topics: The first one dedicated to the the concept of gamification and its uses, specially in the healthcare scene, and a second section for documenting descriptive and predictive tasks and general issues faced when data mining efforts are applied to clinical data.

### 2.1 GAMIFICATION - ORIGINS AND MODERN DAY APPLICATIONS

Evaluating the effects of applying game mechanics and concepts on different aspects of the commercial and business sectors stands at the root of studies regarding gamification. As the idea that games employ important design choices designed to increase enjoyment and motivation began to be accepted and adopted by more organizations as a potential enabler of high productivity, the clear line that once divided entertainment driven systems and productivity based products has been blurred, with added benefits and raised concerns on both sides of the spectrum. At a conceptual level, information systems can be categorized according to their tasks and goals (Heijden, 2004). Traditionally, utilitarian systems follow a logical framework designed to maximize productivity in regards to a specific task, with minimal effort into rewarding the usage of the system itself, but with a clear goal of aiding the user in reaching external objectives more efficiently. In contrast, entertainment systems aim to reward the user with simulated experiences and recreational benefits, with the sole purpose of entertainment. In more recent years, the rise of applications designed to meet more varied user motivations and objectives has given origin to a slew of implementations that differ from the traditional categorization of information systems, and given way for broader spectrum of orientations and thought-processes behind modern

technologies. The term acknowledged since 2008 as 'Gamification' (Bret ,T., 2008) fits neatly in this new era of technological advances, being defined as "an informal umbrella term which defines the use of video game elements in non-gaming systems in order to improve user experience and engagement." (Deterding, 2011). Although this meaning is generally accepted, some authors credit other game-related terms, such as "serious games" and "alternate-reality games", to this definition (Bonsignore et al., 2013; Susi et al., 2007; Szulborski, 2005). Efforts to gamify tasks have been observed decades before the term was recognized at a global scale, with its first appearance dating back to 1978, when Richard Bartle and Roy Trubshaw, developers at University of Essex, developed a project code-named 'MUD1', short for 'Multi-Player Dungeon 1' (Slator et al., 2007), a text-based multiplayer system deployed on the university's computer network, in which students could, for the first time, share a single virtual world.

Since then, the concept grew steadily and solidified itself as an effective strategy showcasing positive psychological and behavioral outcomes when employed in the context of education (Hamari, 2013; Malone, 1981) and workplace productivity (Coonradt and Nelson, 1985). Many studies have been carried out in the sense of evaluating the many different motivational elements typically contained in a gamified system and how they impact user satisfaction and engagement with the activity. In education-related scenarios, the employment of point systems and competitive elements such as leaderboards and timed challenges demonstrates improved levels of participation, enjoyment, and motivation (Cheong et al., 2013; Denny, 2013; Domínguez et al., 2013; Hakulinen et al., 2015). Narrative driven approaches which introduce clear goals and mapped out levels are shown to improve the learning curve of students (Abu-dawood, 2016; Dong et al., 2012). Regarding workplace environments, game design components contributed greatly in increasing the workers' productivity and motivation (Farzan and Brusilovsky, 2011; Farzan, Joan M DiMicco, et al., n.d.; Farzan, Joan M. DiMicco, et al., 2008). A particular counter-experiment by Thom et al in 2012 (Thom et al., 2012), in which gamified components were removed from an enterprise social networking system, displayed adverse results, with overall reduced participation and less contribution being made to these systems by the company's constituents.

Relatively speaking, the healthcare sector is still at an early stage of adoption of game mechanics, and while a respectable amount of literature studies and reports showcase positive results when such concepts are applied to healthcare, many efforts still rely on theoretical assumptions rather than concrete experimenting and documenting. Naturally, this could be credited to the opinion that certain sub-sectors of this domain are more compatible with the addition of gaming components, such as physical activity and mental health. However, the specific area of healthcare education seems to have benefitted greatly

with the usage of game design elements in a medical context, with a number of efforts for improving learning (Kerfoot et al., 2012; Nevin et al., 2014) and reinforcing correct decision making (Lin et al., 2015) in clinical scenarios showcasing swift adoption by the learning population, increasing information retention and engagement, and even facilitated assessment by educators.

When analyzing the effects of gamification in a population's general health, some studies raise important observations and reveal gaps in the state-of-art. Derksen et al. (Derksen et al., 2020) analyzed the impact of 14 serious games on key indicators of smoking addiction. The study highlights occasional reminders and penalty systems as the most relevant game-like features for smoking avoidance, and indicates generally beneficial impact in key factors leading to smoking cessation and better patient behaviour, but reveals necessary improvements in game design methodology to verify the quality of the results. Villani et al. (Villani et al., 2018) conducted a preliminary evaluation on the effects of using video games for emotional regulation (ER) purposes, and found that limited gameplay time reserved to serious games reduced the ability for ER improvement, when compared with more frequent game sessions with entertainment-oriented games. The study concludes that the fun gameplay and fictitious properties of commercial games are key factors for these results, but highlights that future evaluations should include clinical populations. Lau et al. (Lau et al., 2017) presented a meta-analysis of serious games and their effectiveness in helping address symptoms of depression, PTSD, ADHD, ASD, cognitive functioning and alcoholism. Findings support that games might be effective in improving the related symptoms, but claims that additional research is necessary to include more mental diseases and assess long-term impacts.

Overweight and obesity, which account for a large portion of premature mortality and affect an estimated over 1 billion adults worldwide (WHO, 2013), are also important motivators for efforts regarding the gamification of health platforms which promote fitness and dietary adjustments. When game design concepts are employed as an extra layer of motivation for doing physical activity, studies show people favour the short-term reward systems provided by gamified systems rather than the long-term benefits innate to exercising itself. According to a study conducted by Hamari and Koivisto in 2015 (Hamari and Koivisto, 2015), positive social recognition and reciprocity have a positive impact on how much people are willing to exercise. Additionally, the study claims the more social contacts an individual has in a digital service, the larger the aforementioned effects are. In 2013, Lentelink et al. (Lentelink et al., 2013) evaluated a series of scientifically backed game design concepts with the goal of implementing a simple proof-of-concept serious game that positively reinforces healthy weight management via the use of in-game rewards (currency and cosmetic items), social features (multiplayer exercise sessions), and systems

aimed to avoid over-stimulating the user, such as limiting the acquisition of in-game currency to the proposed weekly goals. Intuitively, the concept of beneficial goal-setting and exploration of reward systems in physically motivated games requires developers to have thorough understanding of potential setbacks and barriers with certain user demographics. In 2012, Munson and Consolvo (Munson and Consolvo, 2012) conducted a four-week field study to evaluate how certain game mechanics designed to encourage activity may affect user interaction positively and negatively. During their experiment, they found that the most beneficial components were non-judgmental exercise reminders and primary / secondary goal-setting strategies, allowing the user to have a more activity-focused approach, as opposed to in-game rewards (such as trophies and ribbons) and social features (milestone Facebook sharing), which many subjects found little to no interest in. In 2013, Brauner et al. (Brauner et al., 2013) carried on a practical experiment to assess the role of gamer types, technical expertise and personality factors in the enjoyment and effort put into gamified exercises. The study, conducted primarily on young and elderly players, helped conclude that game performance is heavily influenced by gamer type, and not as much the outstanding factors. Additionally, a side conclusion of the experiment indicates physical activity in the form of gamified exercise has a positive impact on perceived pain levels, specially on the older population. In 2017, Dias et al. (Dias, Tibes, et al., 2017) performed a review on the effects of serious games in treating or preventing childhood obesity. Out of the contemplated 17 research pieces, the study concludes that serious games might have a positive effect on the promotion of healthy behaviours and dietary changes. Moreover, the young population's frequent usage of technology may induce a behavioural change in the public's opinion on juvenile obesity. A follow-up study evaluated the behavioural improvements of an overweight/obese young population by introducing serious games to promote healthy habits. The included studies reveal that positive reinforcement of frequent physical exercise and improved eating habits are a potential strategy to overcome childhood obesity (Dias, Domingues, et al., 2018). More recently, in 2020, Chow et al. (Chow et al., 2020) determined whether gamified techniques were an effective motivator to improve the eating habits of a younger population. Their evaluation included 43 evidence-based research papers and concludes that introducing positive reinforcement feedback, challenge and progress systems helped develop good eating habits, but games promoting unhealthy snacking behaviours negatively impact the younger population just as much.

Mobile fitness applications are one of the main vehicles of delivering gamified experiences to a large population, serving as an optimal platform to introduce social features. A user study report developed by Yu and Pearl (Y. Chen and Pu, 2014) explores social interactions in fitness apps on a deeper level. Across the two week study, it is concluded that

users have a positive bias towards cooperative and hybrid gameplay sessions rather than competitive exercises, and that, on average, social sessions produce enhanced physical activity when compared to exercising alone.

Given the benefits outlined, it is easy to overlook the fact that a large percentage of the population enrolled in fitness applications is relatively young and technologically savvy. With this, come many problems for the older, less technical, but arguably more vulnerable to issues, set of individuals that strive to engage in exercise to correct preexisting ailments or delay new ones from forming. A heuristic evaluation on fitness applications in 2014 by Silva et al. (Silva et al., 2014) argues that several popular physical activity monitors available on smartphones contain design choices and systems that do not comply with the standard needs for the successful use by the older population, a demographic that suffers greatly from the lack of effective exercising. Additionally, another study presented by (Hoffmann et al., 2017) assesses the implementation of gamification techniques on existing stress management mobile applications, by performing methodological evaluation based on efficient gamification strategies and behavioral therapy techniques on 62 different stress treating applications. The study shows no correlation between the use of gamification techniques and behavioural change techniques or stress management methods, concluding that stress management application designers do not implement gamification strategies to an efficient degree.

More recently, the rise of wearable technology has motivated efforts to hybridize gamified exercise with on-body activity trackers. In 2016, Zhao Zhao et al (Zhao, A. Etemad, et al., 2016) reviewed the technical capabilities and availability of several body movement trackers, dedicating a section of the paper to the discussion of the implications such devices can have on the gamification of health and fitness, in which the response was generally favorable. Later in the same year, the authors have also investigated whether the usage of wearable-based exercise systems can produce positive results in long-term scenarios (usage for at least 70 days). The study concludes that based on preliminary data, the gradual addition of features or changes in said systems (i.e.: content updates) is effective in maintaining users interested in engaging in regular physical exercise (Zhao, S. A. Etemad, et al., 2016). Virtual Reality (VR), a recent evolution of the wearable technology, has already seen some usage in healthcare scenarios, despite its novelty. Ayed et al. (Ayed et al., 2019) conducted a review on the usage of vision-based serious games for motor rehabilitation programs. The study reveals that while many efforts focused on efficient posturing and upper-limb workouts, with generally favorable results, there is a lack of result clarity and standardized methodologies for VR-based evaluation. Another literary study by Jin, Pillozzi and Huang (Jin et al., 2020) supports that full immersion VR (with headsets and hand motion controls) positively impacts in-game learning acquisition and information

retention, as well as clinical rehabilitation results, but presents uncontrollable constraints such as motion sickness and inaccessibility to VR devices.

Another particularly pronounced area of research highlights the negative effects of prolonged game sessions on the musculoskeletal system (Borecki et al., 2013; Ha et al., 2020; Hodgetts et al., 2021; Jalink et al., 2014), with some of these studies reporting that players participating in excessive daily playtime (more than three hours a day) were up to 5.2 times as likely to develop neck, back, shoulder or hand complications. As such, video-gaming is seen as a sedentary behaviour (Tremblay et al., 2017) and is often classified as a purely entertainment focused activity. For this reason, associating games with muscular disorder management requires cautionary approaches to other key game characteristics, such as average session length, required level of physical effort, and ideal posturing, since all of these aspects may easily become counter-therapeutic if not properly adjusted to the individual. Furthermore, the main demographic affected by musculoskeletal disorders is the elder population, which in turn makes up for only around 6% of the gamer population in the United States (J. Clement, 2022). Such discrepancy in the population sizes makes this area of interest relatively harder to research, and thus have a slower development rate. Still, some studies have been conducted to evaluate the effects of gamified components in the perceived pain levels, treatment reception, medication use and motivational levels of patients diagnosed with such complications.

In 2015, Allam et al (Allam et al., 2015) performed a randomized controlled trial for 157 patients diagnosed with rheumatoid arthritis (RA), with the goal of evaluating the effect of social support features in web-based gamification approaches. Results from this study indicate the population appears less reliant on healthcare utilization and medication overuse when presented with social support options, with an increased tendency towards self-management empowerment and positive behavioral changes. In 2020, Liu et al. (Y. Liu et al., 2020) evaluated the acceptability of 110 patients afflicted by upper extremity weakness to a custom electromyography device designed to be integrated with gaming platforms. Most patients, even with a previously manually recorded strength score of below 40%, were quickly able to understand the game environment and engage in reliable gameplay. Answers to surveys submitted post-evaluation indicate higher motivation, enjoyment, ease-of-use, and even safety levels as opposed to manual strength assessment with traditional EMG devices. In the following year, Varga et al. (Varga et al., 2021) propose a neural network based serious game for evaluating and rehabilitating hand-movement at different stages of RA. The application introduces motion integration and AI assistance for aiding in performing correct, therapeutic movements. Experimental results show that mental stress, hand-fatigue and physical exertion were insignificant in most-cases. More recently, Özlü et al. (Özlü et al., 2023) conducted a randomized controlled study on the

effects of a virtual reality gamified rehabilitation program for managing knee osteoarthritis (OA) symptoms, namely pain, lack of functionality and loss of balance. The achieved results show that disease-specific gamification is effective and safe in knee OA, but that long-term and side-effects need to be properly managed to avoid counter-therapeutic scenarios.

## 2.2 DATA MINING IN HEALTHCARE

Given the data-forward nature of the healthcare sector, most activities carried out by medical organizations give origin to enormous amounts of data, making traditional storage and analysis approaches quickly unviable. Data mining is a wide term described as the process of uncovering patterns and other valuable information from large sets of data (IBM, 2023), and one that aims to solve problems associated with dealing with such massive amounts of patient information. Typically, data mining tasks can be either descriptive or predictive (Bellazzi and Zupan, 2008), with the former being generally associated with the drawing of interpretable conclusions from a series of data, and the latter, as the name suggests, characterizing tasks which generate a useful prediction. In healthcare, the majority of interest and research effort is centered around predictive tasks encompassing classification, regression and categorization (Jothi et al., 2015). One relevant sub-discipline of mining medical data is anomaly detection, in which researchers aim to find significant changes in datasets as a means to identify underlying conditions in patients. Sabic et al. (Šabić et al., 2021) perform a comparative study between several algorithms tuned for anomaly detection in patient heart-rate data, achieving a high degree of precision in simulated datasets with 0.5% and 2.5% anomalies. Bo Liu (B. Liu et al., 2013) et al. proposed a SVDD-based approach for detecting anomalies in datasets with uncertain data, including liver disorder data. Clustering, a task that aims to identify a finite set of categories to describe data (Fayyad et al., 1996), can also be useful in the context of healthcare. Veloso et al (Veloso et al., 2014) propose a vector quantization method for building baseline patient profiles and predict re-admissions to intensive care units. Haraty et al (Haraty et al., 2015) outline an enhanced K-Means clustering algorithm for discovering relevant patterns in medical data. The proposed approach, referred to as G-Means, utilizes a greedy approach to generating  $t = 0$  centroids and outperforms regular K-Means clustering in relevance to entropy and F-Score values.

Multivariate statistical analysis is another area of great prevalence amongst the medical sector, as data is usually not limited to one outcome variable and, more often than not, is also non-parametric. Su et al. (Su et al., 2012) combine the Mahalanobis Taguchi System (MTS), Support Vector Machines (SVM), Decision Trees (DT) and Logistic Regression (LGR)

techniques to achieve a high degree of efficiency in filtering high relevance attributes in a class-unbalanced dataset for determining the frequency of pressure ulcers. Khan et al. (Khan et al., 2008) propose a system based on fuzzy decision trees to estimate breast cancer recurrence and patient survival, averaging 85% accuracy. Their study concludes that further improvements could be achieved with the introduction genetic algorithms and adequate rule weighing. Moon et al (Moon et al., 2012) propose a decision tree algorithm to characterize the smoking patterns of the senior population of the US, with the goal of identifying key explanatory variables for predicting the number of cigarettes smoked per day.

Approaches to classifying unlabeled observations have also emerged, with Armañanzas et al. introducing five variations of an estimation of distribution (ED) algorithm for predicting the severity of non-motor Parkinson's Disease symptoms, with accuracy values ranging from 72% to 92% (Armañanzas et al., 2013), and Wang et al (Jen et al., 2012) presenting a Linear Discriminant Analysis (LDA) based method for establishing early-warning criteria for common chronic diseases with a corrected rate of 80%.

Statistical modelling (or logit modelling), in opposition to multivariate analysis, is normally used to predict the percentile likelihood of a binary event occurring. In the scope of medical data and in view of recent events, a statistical model could be used to determine the chance of a patient's infection with a coronavirus. The body of literature regarding the usage of linear regression (LR) models in healthcare is sizeable. Numerous studies introduce LR approaches to predict the development of diseases, such as Periventricular Leukomalacia (Samanta et al., 2009) and Tuberculosis (Mamiya et al., 2015), and patients' adherence to some types of cancer treatment (Thompson et al., 2014).

Machine Learning (ML), a branch of Artificial Intelligence (AI), is an area that has seen unprecedented discovery and growth over the past years. Predictive models powered by ML are, in many ways, more powerful, accurate, and robust than the traditional counterparts. As an area-of-research with many critical systems in play, healthcare has taken educated and cautionary approaches towards implementing ML-based systems to predict disease severity, incidence and recovery rates, and even perform full-on diagnosis in some cases. Chang and Chen (Chang and C.-H. Chen, 2009) pair decision trees and neural networks to potentiate a dermatological condition diagnosis system with very high (92.62%) prediction accuracy. Er et al. (Er et al., 2010) present several artificial neural network (ANN) models for classifying multiple respiratory system diseases with very high success rates. On average, they deduced that probabilistic neural networks (PNN) produced the best results (92.1% classification accuracy) across the problem. On the topic of lung diseases, Gunasundari and Baskar (Gunasundari and Baskar, 2009) apply image transformation techniques to improve the quality of chest computerized tomography (CT)

scans, before classifying images with an artificial neural network. They achieve upwards of 90% accuracy by adjusting the input images' inter-pixel distances. Another research paper by Das et al (Das et al., 2009) proposes a neural network ensemble method for diagnosing heart disease with 89% accuracy. Their approach is based on the principle of iteratively creating new networks by combining previous models' prediction values. Bayesian Belief Networks (BNN) are probabilistic models that do not have any dependency on attributes. They can be exceptionally useful in medical use-cases due to their ability to reflect causality relations, for example, when assessing the probability of diagnosing certain diseases given the symptoms provided. Liu and Lu (K. F. Liu and Lu, 2009) propose a BNN as a decision support for health risk estimation versus several air pollutants, with respect to the diseases they cause, and the associated cancer development risks upon exposure to such toxic substances. Curiac et al. (Curiac et al., 2009) outline a BNN-supported analysis on psychiatric patient data, which led to the identification of the most significant factors that link psychiatric symptoms to diseases.

### 2.3 CRITICAL ANALYSIS

It is no surprise that improving users' enjoyment of otherwise mundane activities can contribute positively to the outcome quality of those tasks. Video games have quickly risen in popularity due to their unique fun-factor and engagement, and therefore leveraging the elements that contribute to their success, and applying them to other strands of life, is a natural evolution, and one that has come to be greatly appreciated. Nowadays, video games are a massive part of the modern world culture, with roughly 65% of the entire USA's population claiming to be gamers, according to a survey conducted by the Entertainment Software Association (ESA) in 2023 (ESA, 2023). Moreover, so much as 96% of the survey's respondents consider video games to be beneficial to people's lives. Not only are people more invested in playing games, but the video game industry is growing extremely fast. Steam, a widely known digital game distribution platform, reports 27,345 video game releases between 2004 and 2018, which is impressive if not for the fact that between 2019 and 2023 year-to-date, this number reaches upwards for 52,981 (see graph depicted in Figure 1).

Curiously, this same spike in popularity matches the growth of the bibliographic corpus for a search of the term 'Gamification' in several popular publishers' online libraries. Figure 2 reveals that the total number of articles, research papers, or other publications containing the keyword 'gamification', increased substantially in recent years (2019 to

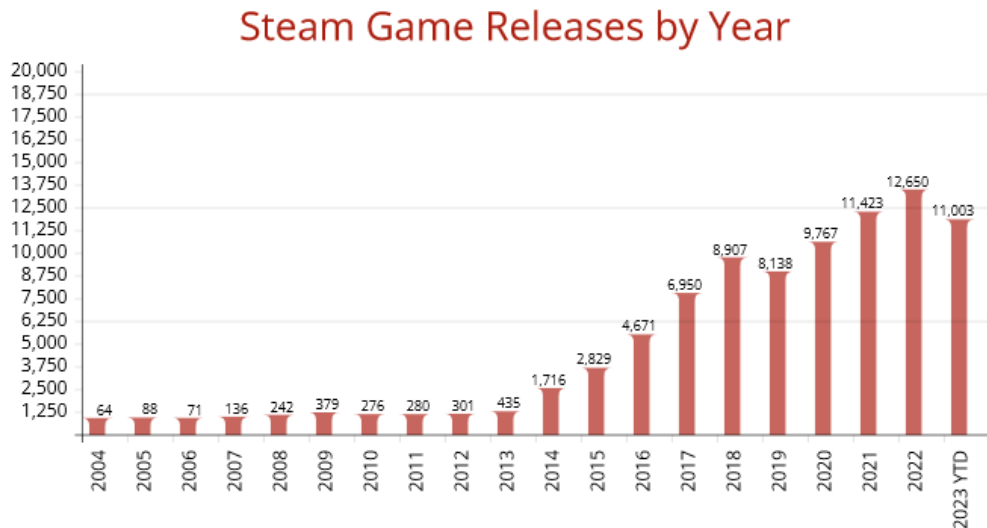


Figure 1: Number of steam releases games by year. 2004-2005 data: (statista, 2023). 2006-2023 data: (SteamDB, 2023)

present), when compared with a broader, although less technologically mature time period (2005-2018).

Reasonable inference from these two analyses concludes that the world's video game industry shows unprecedented growth in the last half-decade, perhaps aided by *force majeure* events such as the COVID-19 pandemic, which inevitably led the population to become homebound (Cetin and Kokalan, 2021) and more inclined to invest in indoor activities. Nevertheless, it is only logical to assume this upwards trend will continue: As the industries that support video gaming directly (electronics, in particular) create powerful new devices and ways to play, so do other industries (e.g.: medical, sport) discover interesting means to improve user engagement with their related activities, such as motion sensors and virtual reality headsets. This mutual growth is evidence that cross-industry developments are key to significant breakthroughs in the areas of user enjoyment and productivity, leading to overall improved adoption of procedures that have been revamped with game-related mechanisms. Despite a promising future, the current state-of-art still showcases significant gaps between proposed solutions and their perceived impact in the related industry, mostly due to the perceived technical difficulty that comes with employing gamified components, terms and concepts in populations with little tech know-how (ie: the senior population).

This problem is most evident in healthcare, an area-of-research where a large percentage of its end customers (patients) are elderly, and therefore generally less accustomed to newer technologies. When we look at the disparity between different sub-areas of healthcare, this

## Number of Article Publications by Year Range

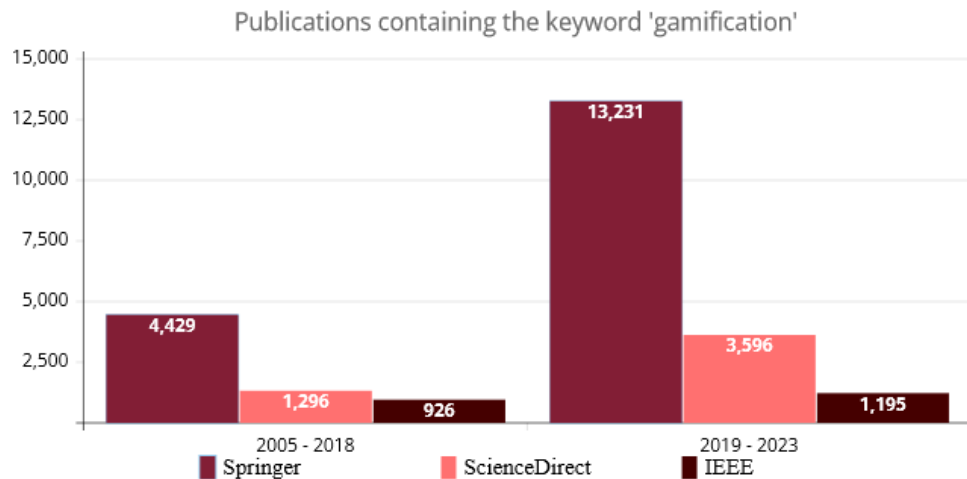


Figure 2: Number of article publications in three popular publishers: 2005-2018 (left), 2019-2023 (right). Data collected by configuring similar searches on Springer, ScienceDirect and IEEE search engines.

becomes even clearer: Exercising and fitness programs, which are naturally compatible with social features and reward systems, are abundant among the younger population and appear to greatly improve motivation and effort. In opposition, musculoskeletal rehabilitation tasks (physical therapy), which aim to minimize the effects of diseases prominent in the elderly, are still treated, for the most part, with routine appointments with little emphasis on the patients' engagement factor. The reasoning behind this issue can come across as paradoxical, after all, why is it no good use drilling these drastic, vanguardist changes to standard healthcare procedures in a population that perceives these "improvements" as nonsensical and unjustified, even if they are proven to vastly improve performance, recovery rates and motivation in other populations? The problem isn't that the elderly disapprove of new technology, this technology simply isn't designed in a way that is intuitive or user-friendly to patients in an advanced age, as demonstrated by the statistics in Figure 3.

When looking at the current state-of-art in this regard, there is a wide range of tests and sometimes contradicting results. For instance, there does not seem to be a clear consensus on whether social features (milestone sharing, friend system) or in-game rewards (trophies) are the more effective motivator for exercise-based applications, which likely mirrors the preferences of the reduced control populations used in these experiments, as opposed to the general population. In terms of state-of-art maturity, the subject of natural compatibility is also raised, since there is an abundance of studies revolving around physical activity and weight management in recent years, but very little progress regarding [RA](#), [OA](#), or

## The elderly and new technologies

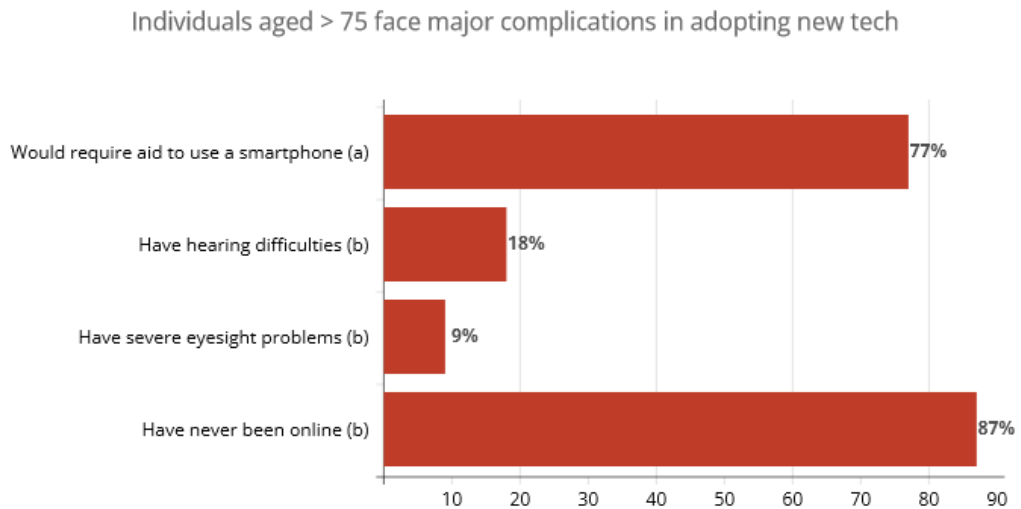


Figure 3: The senior population faces many barriers that deter the adoption of new technologies. (a) (Smith, 2014). (b) (Eurostat, 2018).

other musculoskeletal disorders. While the latter can be considered a niche topic, it is worth noting that results presented are generally short-term, inconclusive, or related to a broader psychological concepts (improved self-management, motivation), lacking positive evidence pertaining to the physiological issue being treated. This puts the current state-of-art in gamified musculoskeletal rehabilitation in a very early stage, where significant contributions are still few and far between, but the results of existing proofs-of-concept are sufficiently motivating for a small number of dedicated research papers to be frequently published.

One general research gap in the current corpus relates to the lack of effective standardization in the proposed evaluation methods, since most existing studies do not conform with predetermined rules for neither the development of gamified components, nor the assessment of results, barring effective comparison between studies. Additionally, the current understanding of effective design choices in gamified products for specific population groups (Silva et al., 2014) and specific disorders (Hoffmann et al., 2017) seems to be sub-optimal, which can affect integration of such products into existing healthcare systems or deter their adoption by a broader population. From an implementation point-of-view, research also shows some dissonance between the most effective game mechanics (Hamari and Koivisto, 2015; Lentelink et al., 2013; Munson and Consolvo, 2012), though some studies already aim to minimize this issue by providing a thorough analysis on the

relationships between different game components and their perceived effectiveness on the population (Baptista and Oliveira, 2019).

Dialing back to broader areas-of-research, both Knowledge Discovery in Data (KDD) and Healthcare are subjects where the existing corpora can be considered mature, with millions of published articles, chapters and books supporting a vast number of theories in either case. While there are constantly newer approaches and improvements both topics, the sheer scale of the pre-existing contributions make it so the available knowledge-base is almost immutable, and newer papers don't always have much impact because due to their increased specificity being necessary to constitute a novelty contribution. Over extended periods of time, some of these offshoots become relevant enough that they form a relatively comprehensive sub area-of-research, such as the subset of literature dedicated to hybridizing Data Mining and Healthcare approaches (see Figure 4 for breakdown of state-of-art density).

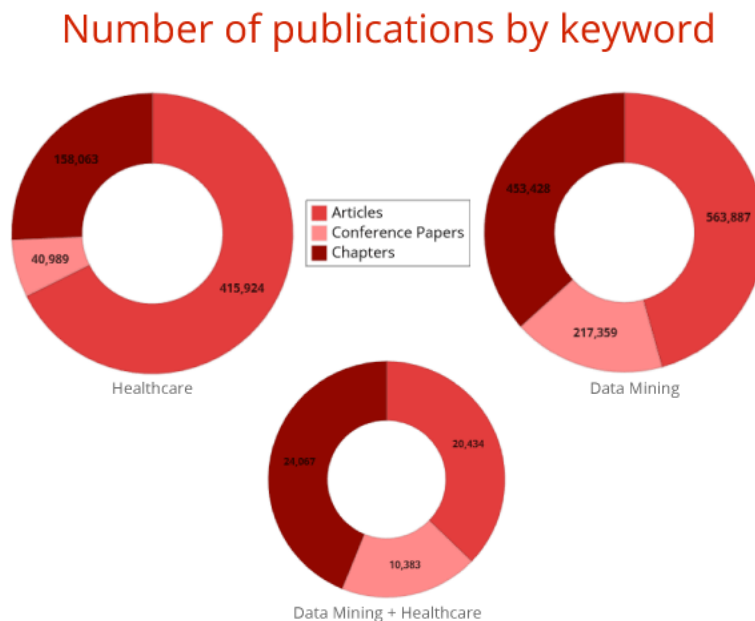


Figure 4: Number of publications containing the keywords "Data Mining", "Healthcare", and a combination of the two. Data collected by configuring searches on Springer's search engine

Even though the corpus abridging this hybridization of disciplines is rather small, the fact is most healthcare facilities have either fully transitioned to digital approaches, or are in the process of doing so, since the massive amounts of data generated nowadays can quickly overwhelm paper-based systems. This makes the assessment of issues related to data mining all the more pertinent, and raises questions regarding the viability, confidentiality and quality of the data used in current approaches.

One key issue with performing any sort of mining on healthcare data relates to ensuring the quality of the datasets used. As medical data can come from many sources (records, clinical trials, sensors, medical imaging), its naturally prone to noise, bias and other irregularities (Vimark, 2023) that may affect the studies performed on it, leading to lower accuracy, higher analysis costs (Fang et al., 2016) and raising the importance of a thorough cleaning phase. Additionally, many clinical records contain extremely sensible information that should be properly privatized while keeping the data as true to origin as possible, which presents a significant challenge (Sarwar et al., 2022; Zainab and Kechadi, 2019). One recent literature review goes in great detail to highlight the privacy concern in medical data, compiling approaches that include cryptography, blockchain, randomization and other strategies for keeping critical information safeguarded from malicious access, and underlining the lack of efficient testing of these techniques in face of the huge amount of medical data collected daily (Vimark, 2023). Additionally, some hospitals are still majorly paper-based, with an alarming number of EU-based hospitals showing signs of insufficient security and data privacy measures shortly after shifting to paperless (Doarn, 2019; Uwizeyemungu et al., 2019), making the implementation of digital solutions unviable.

As technology continues to lean heavily into the digital, most industries are experiencing some sort of "migration" from traditional, hands-on implementations to ones where AI takes the reins of repetitive or high-precision tasks. For some, the thought of letting a computer sift through insurmountable amounts of information in a fraction of the time it takes them to achieve similar results, is a promising step in the direction of freeing technicians from a major portion of the tedious, data-heavy tasks they have been accustomed to before. In fact, the introduction of AI in healthcare is already tested and proven to improve clinical decision support, health record system modernization, and patient adherence to treatments (Davenport and Kalakota, 2019), even potentiating the growth of sub-fields of medicine, such as precision medicine (Lee et al., 2018). While the benefits are clear in domains related to patient care, one often overlook area-of-application is the administrative and paper work that healthcare professionals do for a significant portion of their work time (up to 25% for US nurses) (Commins, 2010). AI, in particular robotic process automation, is relevant in automating the processing of medical claims, clinical documentation, and revenue cycles (Utermohlen, 2018). Automated Support Systems such as chatbots are already being explored in some healthcare organizations, though their efficacy is still unproven, as patients often feel "disconnected" to machine helpers (UserTesting, 2019).

Despite the use-cases outlined above, the usage of AI in the industry is often seen as concerning, with studies estimating that the automation of jobs could potentially deprecate up to 35% of human roles in the next decade (LLP, 2015), while others argue that the social

and regulatory factors could pose a significant barrier to job automation (Manyika et al., 2017). Specifically in healthcare, there are also a number of ethical factors that denounce the use of AI, with the biggest offender being the lack of transparency of deep learning algorithms used for clinical image analysis, which often leads to a lack of explanation of the model's outputs, causing confusion to patients and professionals alike (Char et al., 2018). The social factor plays a big role against the wide adoption of deep learning in healthcare: Even if a model is successful in the early prediction of disease in thousands of patients, the first error that it inevitably commits instantly fuels the demonization of AI, causing fear and distrust among the population.

The deployment of AI in medical contexts is a sensitive topic, raising questions of privacy, ethics, and replacement of human roles. Despite the benefits of precision medicine, the general population is still not comfortable in allowing machines to make decisions about their health, especially when risk of death is involved. For a clinical environment that is potentiated by AI to be considered safe, the data must be absolutely secure to external threats, kept private, and the models as transparent as possible within technical reason.



## CONTEXT

---

Given that the proposed methods are focused almost entirely on data mining tasks, some additional context regarding the motivation, origin and positioning of the underlying project is needed to create a logical bridge between the approached research concepts, and to define GameVitae in a clear but concise manner. As such, the current section is intended to position my work logically in both its relevance to the state-of-art and its value to the enveloping project.

### 3.1 FRAILTY SYNDROME AND SARCOPENIA

The continuous wear and degradation of cells in the human body after adulthood is a natural process associated with aging, and is one that affects all functions of life, from physical to psychological. While leading a healthy life can help mitigate or even delay the negative effects of aging, these are inevitable and will, at some point, affect every human that lives to a certain age. One particular effect of aging translates into the gradual loss of muscle tissue, defined by the deterioration of cells in the musculoskeletal system, and aggravated by sedentary behaviour, nutrient deficiency or other pre-existing conditions, constituting an ailment referred to as frailty syndrome (FS). Geriatric patients diagnosed with FS present a higher risk of falling, becoming immobile, delirium, incontinence, and ultimately death (Yeolekar and Sukumaran, 2014), and its diagnosis is usually determined by the presence of at least three diagnostic criterion: (a) Involuntary weight loss; (b) Exhaustion; (c) Loss of grip strength; (d) Slow movement speed; or (e) Decreased physical activity (Fried et al., 2001). The existence of at least one criterion can also be a indicator of pre-frailty (PF), which is an aggravating factor for FS (Xue, 2011). FS is also closely linked with sarcopenia, a syndrome characterized by a progressive and generalized loss of limb strength and function, and muscular mass (Lima et al., 2020), which affects over 73% of the geriatric population (80 or older) in Europe (Sipers et al., 2019).

### 3.1.1 Assessing Frailty

The wide effects of musculoskeletal ailments on the older population, namely **FS** and sarcopenia, raise the importance for effective means of diagnosing, preventing, and monitoring to be made available. One of the better-validated tools for the assessment of these conditions is the Clinical Frailty Scale (**CFS**) proposed by the Canadian Study of Health and Aging in 2005 (Rockwood, Song, et al., 2005), which aims to classify an individual in regards to their perceived fitness level in relation to their age group, and helps improve quality of life in geriatric patients (Mendiratta and Latif, 2020). Other widely used screening tools are the **FRAIL** questionnaire (Morley et al., 2012), the **RFMI** (Rockwood Mitnitski Frailty Index) (Rockwood and Mitnitski, 2007), and the **REFS** (Reported Edmonton Frail Scale) (Rolfson et al., 2006). The assessment of frailty through medical imaging is also possible, with dual energy X-ray absorptiometry (**DXA**) (Andreoli et al., 2009) and bio-electrical impedance analysis (**BIA**) (Fuller et al., 1999) being the most commonly used practices, though the cost and availability of these tests deter its wide adoption by hospitals. Due to this, small devices in the form of wearable technology (i.e. sensors) have been suggested as an optimal way for monitoring symptoms and collecting day-to-day data that can be useful in the early detection of adverse health situations (Anabitarte-García et al., 2021). Moreover, some of these devices are already well established in other industries like videogames, and can passively collect important data as a side-effect of their intended main use, constituting a non-intrusive method of frailty evaluation. This is the case for depth-perceptive cameras, like the Kinect system proposed by Microsoft (Zhang, 2012), or most modern **VR** headset and controller setups. One such device that is closely linked with this study is Gripwise, a medically-oriented reiteration of a traditional dynamometer, designed and proposed by a spin-off of the University of Porto, WISIFY. It features two charge cells that measure a patient's grip strength and timing, and is engineered to allow seamless connection to most modern systems and easy integration with video game engines (see Figures 5a and 5b for prototype and finalized versions of the Gripwise controller).



Figure 5: Versions of the GripWise dynamometer. Adapted from (Fernandes, 2022)

### 3.2 GAMEVITAE

One limitation of gaming-oriented medical devices such as Gripwise is that they require a dedicated set of compatible experiences (clinically-approved games or activities), which means their initial impact on the industry can be unremarkable. Moreover, if patients or medical staff who adopt these technologies feel restricted in the number of diagnostic tools they have access to, they will simply revert to traditional methods with larger scientific backing. For this reason, a good way to improve an experimental technology's life cycle and adoption is to ensure its conformity with industry standards, thus making data backwards compatible with a larger number of diagnostic or assessment methods.

To ensure that any and all data collected with Gripwise is explainable, accessible and usable by clinical data professionals, WISIFY established a consortium with Techframe, Sa., a company dedicated to the development and commercialization of complex IT systems. Gamevitae (GV) is the product of this partnership, and consists on a centralized mini-game platform focused on the field of healthcare and well-being, in which medical data, collected through various activities and devices, can be stored and accessed, constituting a professionally maintained pool of knowledge. Active development on GV began in 2022, through conceptualization of the necessary infrastructure, proof-of-concept mini-games, and establishment of partnerships with companies dedicated to health research.

#### 3.2.1 *Relationship with gamification*

Improving user-activity engagement is one of the main premises behind the platform, and gamification plays a big role in achieving this ever since the earlier conceptual phases of the project. In July 2022, with the development of a series of proof-of-concept game projects meant to promote the integration of the Gripwise controller, gamification made its debut to GV, and its benefits were highlighted, kick-starting a follow-up movement to explore the possibilities of this concept in the medical area.

##### 3.2.1.1 *The minigame sideproject*

When the Gripwise controller was firstly introduced as a medical device for improved rehabilitation practice, its presentation was complemented by a demonstration, which included footage of a gamified exercise in the form of a farm-life simulator game. Upon the establishment of the consortium between WISIFY and Techframe, the latter found the included game to be suitable for demo purposes, but uninteresting and lackluster for the

population that it targets. The "Gripwise Gamification" project (Fernandes, 2022) aims to improve upon user satisfaction and participation in games where the Gripwise controller is the main source of input, and includes three distinct game prototypes, following the trends of rhythm games, endless runners, and incremental progression games. All of the following concepts were developed in Unreal Engine 4, with access to external plugins to ease the communication between game sessions and the underlying data structure.

**GRIPMUSIC** aims to leverage the emotional connection of people and music as an effective motivator for exercising. It consists in a rhythm game where different input indicators approach at a cadence that matches the background music, and the user's goal is to successfully apply sufficient grip pressure at the correct timings, building a score that is presented after a song's duration. To add to the therapeutic potential, this game also includes a level creation tool, which allows loading a custom song and adding input indicators along its duration, promoting:

- Motivation, by supporting virtually unlimited different combinations of levels, with their own difficulty level.
- Emotional connection, by letting users play songs they enjoy, rather than confining them to a limited set of predetermined choices.
- Therapeutic accompaniment, by having specialized professionals construct levels designed for each patient's needs, they can keep a close look at the progression / deterioration of the muscular capacity and perform adjustments accordingly.

**GRIPRUN** plays on a popular sub-genre of mobile games called "endless runners", where the main goal is to perform "jumps" between differently sized platforms by gripping the controller with the intended pressure and timing. The gameplay cycle for this proof-of-concept is intended to be short, and includes the generation of a level with a configurable number of easy, medium, and hard tasks (jumps), meant as a therapeutic plan for the patient. Afterwards, a character is shown in a side-scroller view, and a random disposition of platforms is gradually presented to the user, whose task is to cue the character to jump between each platform, until they cross a finish line. This game is a more direct approach to therapy, given that:

- Jumps of easy, medium and hard difficulties required lighter/firmer grip squeezes and exercise the patient through repeated flexing of the forearm and hand muscle groups.

- Levels are preconfigured by a qualified person, to avoid excessive exercising and counter-therapy due to excessive exertion.

GRIPLAND is an incremental progression game where the user must grip the controller adequately to acquire a virtual currency, used to purchase different assets that compose a fictitious country, providing a continuous sense of progression. The development stage of this game is earlier than the others, but its portability and unique approach to motivation makes it important nonetheless:

- The game is conceptualized for android devices, and connects with the Gripwise controller via Bluetooth, allowing for portability and on-the-move exercising.
- It is also the only game where an actual reading of force (measured in Kgf) is measured frequently, which allows for real-time monitoring of a patient's grip strength. Naturally, a constant increase/decrease in grip strength, over a period of time, can lead professional therapists to draw certain conclusions about the evolution of frailty.
- The perceived "improvement" of the fictitious country plays an important role in keeping users motivated to autonomously continue the game, making it a more "self-therapeutic" approach to gamifying the Gripwise controller.

### 3.2.2 Limitations

Logically, introducing gamification into every device that is supported by GameVitae is not a straightforward task, and previous work helped raise a substantial amount of difficulties and gaps in knowledge, but also key strengths in the way the platform is conceptualized. These topics involve technical conflicts, increases in scope, and architectural advantages, making their compilation and presentation a crucial step in preambing my project. Note that the following list contains observations raised in previous work (Fernandes, 2022), in preliminary meetings with Techframe superiors, and during the assessment of my project's viability and value to the business:

- The technologies used for the development of the three proofs-of-concept are heavily reliant on plugins for communicating with the [GV](#) database. While these plugins may be sufficient for smaller experiments, they have many limitations that, ultimately, prove too troublesome for commercial usage. Moving forward, a native solution must be developed to ensure it remains supported and efficient throughout the lifespan of the platform.

- Bluetooth connectivity to Windows-based systems is also a more delicate matter, though an initial assessment of solutions revealed that writing a native C++ wrapper for Universal Windows Platform (UWP) might be possible. Given the close connection to the WISIFY, an alternative communication protocol could also be engineered into the device (for example, Serial).
- While Gripwise is directly linked with assessing frailty through upper limb evaluation, it is not the only device envisioned for support within the GV platform. As the project matures, a wide range of tools and accessories needs to be supported, and thus the challenge of producing engaging gamified experiences (while keeping high-quality data collection a priority), is always present.
- The previous issue has a positive counterpart, however, because the platform was engineered with the gradual addition of features and devices in mind, making its scalability one of its most attractive qualities for partners. Subsequently, this also means that the constant development of gamified experiences can be powered by a continuous cycle of R&D, which brings both academic and scope benefits, increasing the platform accessibility to a greater number of populations.
- The centralized nature of the platform opens up interesting ways to explore future work, such as the inclusion of intelligent data tools (this project), dedicated reward systems, and social features. As a side-note, it is relevant to point out that all of the aforementioned motivational affordances were used successfully in some of the works referenced in the state-of-art review.

### 3.3 MOTIVATION

As assessed in the first chapter of this work, gamification is already leveraged to a great extent in improving the engagement of a population to an (usually) otherwise mundane task. In the fields of education and office work, for example, motivated participants are observed to showcase better academic or professional performance, accordingly. In contrast, the adoption of these mechanisms in healthcare has not been as clear, and one of the key rationales for this could be attributed to the technological difficulties experienced by the elderly population in understanding complex systems.

With GameVitae, the overarching goal is to provide a centralized integration tool where patients, therapists, and companies dedicated to medical devices can work together in making the adoption of gamification a smooth and enjoyable process. Naturally, a project of this ambition takes significant research, time, funding, and testing to come

into fruition, and **GV** is very much still in one of its earliest phases, characterized by a dedication to researching the viability of its components. The implementation of Gripwise support constitutes a significant advancement in this regard, and opened several new doors that motivate the involved parties to assign more resources to subsequent studies and sub-projects.

As a long-term developer for Techframe, I quickly learned about what GameVitae is striving to achieve, and was presented with an academic proposal for carrying out one significant follow-up project, originally dedicated to the 'knowledge discovery in gamified exercise activity data, for the assessment of frailty syndrome and sarcopenia in the older population'. With time, the definition and scope was refined and better aligned with the needs of GameVitae, and my project was better suited as an open-ended data mining task, whereby the extraction of insights related to frailty syndrome is still encouraged, but more incentive is provided to explore activity suggestion systems and perform patient profiling, aiming at improving the platform from a usage and commercial point of view.



## METHODS

To achieve high-quality results within strict time constraints, I have decided to adopt the Cross-Industry Standard Process for Data Mining ([CRISP-DM](#)) model, which not only deconstructs the common tasks for completing a data mining project in six comprehensive phases, but also provides flexibility for adjusting and fine-tuning to specific industries or goals. The [CRISP-DM](#) methodology is a decomposition of tasks described at four levels of abstraction, ranging from general to specific (see [Figure 6](#)).

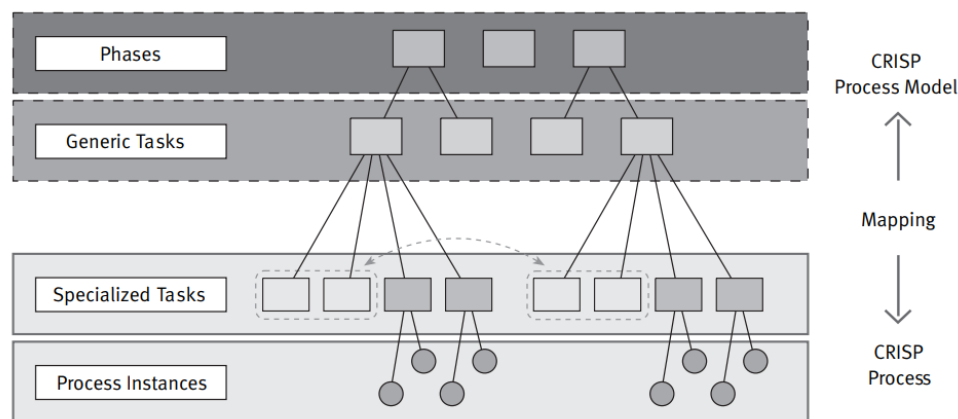


Figure 6: Deconstruction of the CRISP-DM Methodology

At the highest level of abstraction, there is a breakdown of the process into phases that consist of several second-level generic tasks designed to be adjusted into all data mining situations. The stability of the process is one of the driving forces behind the abstraction layers of [CRISP-DM](#), ensuring any activities developed will be valid for yet unexpected developments in the data mining industry. The third level of specification occurs when a generic task is decomposed into smaller, specialized chores that are specific to the data mining process at hand, such as tasks that are carried out differently depending on the format of data. Finally, the fourth level records any actions or results of actual data mining processes, specific to a particular engagement, rather than the general output or action of a given task.

The following sections and corresponding subsections compose the full documentary corpus for the application of the [CRISP-DM](#) methodology, to complete a large scale data

mining project over data collected for the purpose of experimenting with the GameVitae platform. As a means to provide guidance and synthesize the expected phases of this project, a breakdown of the proposed [CRISP-DM](#) generic tasks, along with any expected project outputs resulting from said tasks, will be provided after the introductory statement of each section. Furthermore, several small-form deliverables (small tables, listings and figures) will be presented directly within their respective subsections, whereas larger sub-documents and auxiliary reports will each be represented by their own appendix chapter, to be consulted at the end of the main document.

#### 4.1 BUSINESS UNDERSTANDING

The main goal of this first section is to provide an overall cohesion and understanding of the project objectives and requirements. This is achieved through two distinct phases: The first one assumes a business point-of-view to raise key concerns and provide measurable goals for the project. This knowledge is then used to define an equivalent data mining problem and generate a preliminary plan aimed to achieve the proposed goals. [Table 1](#) contains the foreseen tasks, deliverables and related appendix chapters for this phase.

Generic Tasks	Deliverables	Appendices
Business Objectives	Background Assessment, Objectives, Success Criteria	N/A
Business Assessment	Resource Inventory, Requirements, Constraints, Risk Analysis, Terminology Dictionary, Cost / Benefit Assessment	<a href="#">Appendix A</a>
Data Mining Goals	Data Mining Goals, Data Mining Success Criteria	N/A
Project Planning	Project Plan, Assessment of Tools and Techniques	<a href="#">Appendix B</a>

Table 1: Business Understanding Breakdown

##### 4.1.1 *Business Objectives*

The first action is to compile every relevant insight on business goals for the data mining project. Given the unusually early development stage of the underlying project ([GV](#)), this

can be a task with relatively mutable deliverables. Regardless, it is extremely important to document every possible piece of information associated with goals, their risks and the available resources to achieve success, and thus most of the information in the following subsections was directly obtained, or derived from, personal interviews with coworkers related to **GV**, i.e.: qualified personnel from Techframe, SA.

#### 4.1.1.1 *Background*

A simple background check of the underlying organization and its motives is the first general task to ensure proper knowledge of the project. The aim of this task is to record relevant information regarding the overall business situation before any engagement happens with the data. The data mining project's viability and contribution to the business is also highlighted at this stage, since referencing the current 'as-is' and reflecting on the pretended end goal will help position my tasks in a way that inches **GV** closer to its finalized state.

**PROJECT ORGANIZATIONAL STRUCTURE** The first topic to touch on is how Techframe is organized internally to work on **GV**, and create a pseudo-structure that can clearly identify each member's responsibility and areas-of-expertise. This task will help understand the indirect human resource inventory available for the project, and sets a point of contact between myself and coworkers that can provide technical expertise not directly related to the data. Below, I have constructed a simple organizational chart (Fig. 7) that illustrates the adjacent personnel which have and will be contacted throughout the data mining task for enquiries related to their areas of expertise. This chart has been reviewed and approved by the project proposal party, with the condition that any names under the highest position be omitted to preserve team confidentiality.

**PROBLEM AREA** As mentioned in the context chapter, **GV** broaches the topic of **Healthcare Data**, but has significant ties with *Gamification*, given the method by which data is retrieved from patients. To further illustrate relevant subjects, a mindmap-style visualization is present in Fig. 8 The project I have been assigned to relates to the modernization of the **GV** platform through the development of a set of data models aimed at extracting insights, predictions, and any other valuable information from the data stored within it. Purposely open-ended, this project doubles as an exercise in creative and critical thinking, and intentionally leads me to conduct my own research regarding the types of data confidentiality issues, counter-therapeutic risks and expected model quality associated with advancements in healthcare.

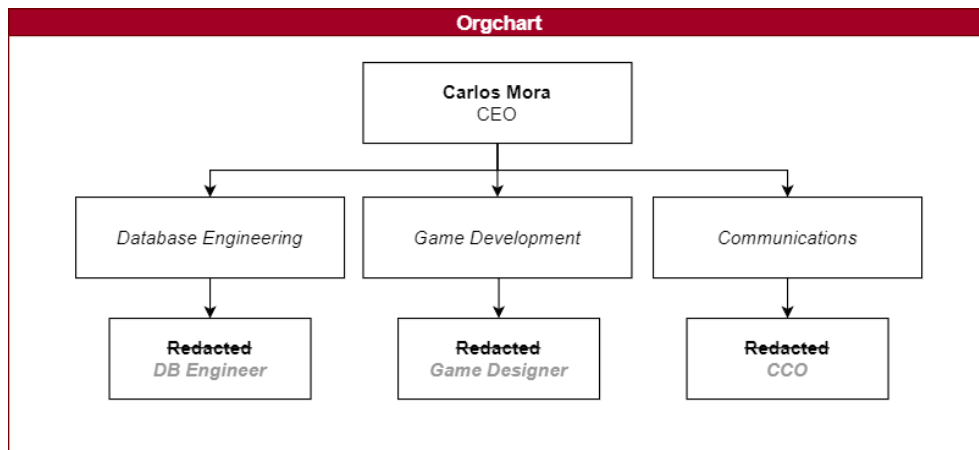


Figure 7: Organizational chart of available indirect human resources

**CURRENT STATE** I have been approached by Techframe superiors regarding the development of this project with the precognition that it would become a significant advancement of *GV* as a whole. During the initial communications, I had the opportunity to understand the current state of the business regarding the problem to be addressed, and whether any preexisting plans or solutions exist. As it stands, there have been no previous data mining efforts related to *GV*, nor has any solution been proposed previous to my acceptance of this project. Prior to my contribution, Techframe has carried out a significant portion of development on the data integration platform, and has successfully efficient data retrieval with the Gripwise controller on a series of proof-of-concept minigames. As such, the basic logic loop for *GV* is functional, but no valuable information can be generated from the raw data, hence the relevance of my project. While this situation means that any and all contributions will be novel, it also indicates the nonexistence of well defined structures, workflows, frameworks, or internal protocols to guide data-related tasks, which may complicate the deployment and evaluation stages.

#### 4.1.1.2 Business Objectives

Setting business objectives is crucial in any Data Mining project, since they will be used to gauge the degree of success of the project as a whole, and steer the development of other tasks towards tangible goals. The following objectives were carefully revised to be considered aligned with the stakeholders' expectations for the final models, as well as reasonable within the scope of the project as an academic-level responsibility.

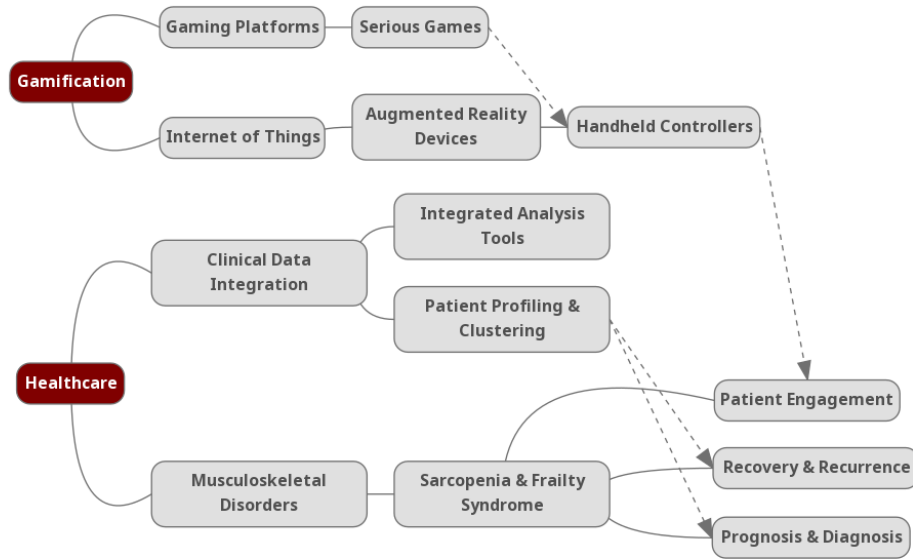


Figure 8: Breakdown of subjects related to project

Alias	Type	Description
BO.01	Engagement	Improve patient retention in suggested activities
BO.02	Engagement	Improve enjoyment of patients submitted to gamified activities
BO.03	Efficiency	Reduce counter-therapeutic or counter-productive activities
BO.04	Efficiency	Reduce time required for therapists to analyze patient activity data
BO.05	Efficiency	Reduce time required for therapists to set therapeutic goals for activities
BO.06	Therapy	Stabilize the scores of patients across different physical ability levels
BO.07	Technical	Standardize the process of assigning a difficulty level to an activity
BO.08	Technical	Personalize the activity experience per user

Table 2: Business objective reference table

#### 4.1.1.3 Business Success Criteria

Since the previously defined business objectives may not be easily classified as complete / incomplete through perception alone, the definition of quantifiable criteria can help formulate conclusions in the later stages of the project. As observable in several upcoming tasks, criterion for this project can be objective or subjective. This is also applicable

in defining measurements that determine the success of the project from a business perspective. Below, objective and subjective business criteria are documented in their respective paragraphs.

**OBJECTIVE CRITERIA** These are usually simpler metrics, measurable in practical terms, related to certain improvements that will be felt in the business operations some time after deployment is completed. Most of the objective criteria for my project relate to the business' aspiration to make the platform as agile and engaging for patients as possible, and providing direct improvements to the tools therapists use for setting up and monitoring activities is a way of achieving this. Since the platform is currently being used in small and controlled set of individuals, the criteria defined here are actually envisioned for a wider audience (i.e.: When the platform is widely distributed). The following criteria will determine with success of the data mining project in measurable form:

- Improve patient engagement by 15% within 12 months of model deployment (currently business reported engagement is 68%).
- Reduce therapist setting up time for new patients (first time they participate in an activity) by 90%.
- Increase the frequency of patient reporting to daily (currently bi-weekly).
- Improve average activity rating score for new patients to 4.5 out of 5 (currently 4.1) within 6 months of model deployment.

**SUBJECTIVE CRITERIA** As the name suggests, these are other benefits that can not be easily measurable, but bring value to the business in general from a knowledge perspective. Subjective criteria are more prevalent in my project due to the early development stage at which the underlying platform presently is. The business has stated in an objective-setting interview that the most important criteria for my project are related to the knowledge discovery, that is, the **findings** that underline the model construction process:

- Find groups of patients with similar physical attributes and activity performance.
- Find motivating factors for patient enjoyment.
- Improve patient enjoyment through automated activity suggestions.
- Reduce subjectivity of activity difficulty rating.
- Determine adequate metrics for patient performance analysis.
- Discover most efficient exercises for a particular patient typology.

#### 4.1.2 *Business Assessment*

This section is, essentially, a compilation of fact-finding subtasks related to any non-data forward factors that should be considered to produce a better informed definition of the data analysis goal. In a sense, the main goal of this task is to expand upon the shallow background check documented earlier, by performing in-depth inventorying, assessing all foreseen constraints and risks, and closing the bridge between the conceptual idea of the project and the corresponding means to make it viable. Each subsequent division is dedicated to one of the expected deliverables for this enveloping task, and will contain a more detailed description about its relevance to the project.

##### 4.1.2.1 *Inventory of Resources*

The purpose of this short section is to produce individual and consolidated documentation of the human, data, hardware and software resources available for the completion of the project.

**HUMAN RESOURCES** A human resource is any data specialist (analyst, scientist) qualified to perform the data mining tasks and any adjacent subtasks planned for the project's completion. As per the unusual scenario englobing this project, and with it being an academic level responsibility designed for one individual, I, the author of this document, am the only human resource assigned for carrying out all activities regarding the project. My qualifications for this task are equivalent to that of a professional data scientist. Additionally, I have permanent access to several contacts that do not directly influence the data mining tasks, but aid the knowledge retrieval portions of the project through their individual qualifications and expertise. They include a Game Designer responsible for the [GV](#) mini-game proofs-of-concept, a Database Engineer who helped build the underlying [SQL](#) structure, a Communications Officer with partnership entity knowledge, and Techframe's CEO, the project manager for [GV](#).

**DATA RESOURCES** For the completion of the data mining tasks for this project, the business (Techframe) has provided two curated sets of data. For ease of explanation, these will be referred to as **patient data** and **activity data** and are characterized below. Note that section [4.2](#) goes into much greater detail to provide a clear definition to the data.

- Patient Data is a comma separated values (.csv) file extracted from the integration of several tables, containing detailed patient characteristics and behavioural data. Due

to the sensitive nature of some attributes in this set, a process of anonymization has been carried out, making data compliant with protection regulations. The final version of this set contains 873 observations and 30 attributes.

- Activity Data is, similarly, a CSV file generated from selecting only the activity data of the users included in Patient Data. Each row describes the activity carried out by one patient, its therapeutic targets, and the real scores achieved by the player. Activity data is not deemed sensitive, so the final version has not had any external operations done to it. The final version of this set contains 3196 rows and 14 attributes.

Both datasets are compliant with the European General Data Protection Regulation ([GDPR](#)) and been approved for academic use by Techframe.

**HARDWARE RESOURCES** There are a total of 3 (three) workstations available for the development of the various activities regarding the project. A consolidated view of hardware specifications can be consulted in [table 3](#). All workstations available also contain Compute Unified Device Architecture ([CUDA](#)) processing units, enabling [GPU](#)-oriented model training, if necessary.

Specification	Laptop 1	Desktop 1	Virtual Desktop 1
OS	Windows 11	Windows 11	Windows 10 Pro
<a href="#">GPU</a>	RTX 2060 Max-Q	RTX 3060 Ti	RTX 2060
CPU	Intel Core i7-10750H	Ryzen 5 5600X	Intel Core i5-8500
RAM	8GB	32GB	16GB
Storage	512GB SSD	2TB HDD + 1.5TB SSD	256GB HDD

Table 3: Hardware Specification Table

**SOFTWARE** A preliminary assessment of the available software resources has been performed to provide an initial idea of the advantages and drawbacks of the tools used. Here, a compilation of known solutions for the tasks previewed for upcoming [CRISP-DM](#) phases is presented in [Table 4](#). Given that all workstations run on Windows operating systems, we can also assume that all tools assessed for the development of data mining tasks must support this OS. Additionally, due to familiarity with the language and the availability of useful add-ons (packages), the programming language of choice for all data mining tasks is Python 3.9 (October 2020). I have not chosen a more recent version of Python due to have experienced incompatibilities with certain packages in previous

projects. To the extent of my knowledge and research, Python 3.9 is the most stable and sufficient version for the intended purposes.

Type	Name	Function	Notes
Software	Draw.io	Flowchart Designer	N/A
Software	PyCharm Community Edition	Python IDE	Free edition has limited built-in API support, may affect framework choice
Software	Adobe Illustrator	Graphic designing	N/A
Software	Microsoft Teams	Communication	Contact point with external HR
Software	Trello	Task manager	Note-taking functionalities and prioritization of tasks is included
Web	Overleaf Student	Web-based $\LaTeX$ Editor	N/A
Web	Google Colab	Web-based Python notebooks	Cloud-based notebooking is useful for developing and sharing reproducible code
Web	Google Drive	Cloud storage of datasets and notebooks	Interoperability with Colab
Web	Microsoft DevOps	Versioning tool	Included in business toolbox

Table 4: Software and web-based resources for the project

#### 4.1.2.2 Project Requirements, Assumptions and Constraints

As with most projects, its objectives contain underlying requirements that must be satisfied to avoid problems in future steps. These requirements usually come in the form of potential liabilities, and must be raised so that they are constantly explicit throughout the project and do not incur in unexpected legal, funding, scheduling, or data-related issues.

**REQUIREMENT ANALYSIS** One of the more pertinent requirements for this project is related to the manner in which results (models) are to be used in the future. The deployment plan for this project foresees the development of small, standalone plugins that work directly on the data at a predetermined recurrence period. Moreover, any references to patients' living habits, gender identification, sexual preferences or other sensitive information must ensure absolute confidentiality and not include any identifying traits,

be it in model outputs or designed visualizations (graphs, tables, charts) throughout the project activities. In relation to the team scheduling, all indirect collaborators (mentioned in section 4.1.2.1) have agreed to be contacted frequently by me, should the need arise, for the entire duration of the project.

**ASSUMPTIONS** At the stage **GV** currently is, it makes sense to disregard most economic factors, since no solution for this problem is currently available to the business. However, the matter of competitive markets can still make or break the success of the project, and is where a deep knowledge of the state-of-art is a valuable asset to enable smart positioning choices in the market. As averred, the state-of-art in this subject is still underdeveloped and a first screening revealed no competing products with either **GV** as a whole, or my contribution to it. We can reduce assumptions down to the data itself, for instance, in regards to its expected quality. Because data quality can be assessed through several measurable dimensions, and given the limited contact I had with the data at this stage, the following list simply brushes on superficial expectations and establishes a point of reference for some of the dimensions stated.

- **Completeness** - Missing values are tolerated and disregarded up to 0.2% of the dataset dimension. Completeness violations over this amount require case-by-case analysis and decision.
- **Timeliness** - All data is to be no longer than 18 months older than the start date of the project.
- **Uniqueness** - Duplicate values are not expected in any dataset table. Uniqueness is expected in the game session registry tables.
- **Validity** - Rationale for all data formatting choices will be requested, validated, and, should need be, corrected in a future task.
- **Accuracy** - Data accuracy is a given assumption, given its direct influence on the resulting models.
- **Availability** - All data must be freely available at all times through direct, secure connections to the source server, or through previously downloaded, localized sample files.

Finally, Techframe superiors have established their interest in the project results rather than model understanding, ensuring that explainable modelling is appreciated, but not expected.

**CONSTRAINTS** Due to the sensitive nature of clinical information that is kept within the system, the greatest constraint is ensuring all data usage legality is guaranteed, and that no unauthorized access to critical data happens at any point of the project development or result presentation. As a preventive measure taken by the business to avoid data spills, the project dataset is isolated and password protected in an internal network, to which only I will have access at all times during the expected duration of the project. Additionally, there are special indications to omit personal contact information from all work tables, and as such, the dataset is fully stripped of characterization, and conforming with [GDPR 2016/679](#).

#### 4.1.2.3 *Risk Analysis*

The assessment of risks for this project differs from typical data mining projects, since the financial counterpart is to be disregarded from most tasks. To ensure adequate cover of all possible setbacks, I have constructed a table of risk documentation and contingency with unique identifiers for each possibility, which can be consulted below (Tab. 5) and will be referenced should any of its contents be relevant in future parts of the project.

Alias	Description	Contingency
RC.01	The non-compliance of any project tasks with the proposed project scheduling for a period larger than 15 days.	Revise the proposed project plan and identify the effect of the delay on the later stages. Propose the disposal of non-essential tasks to make up for lost scheduling.
RC.02	Data quality assumptions regarding completeness are not verified.	Verify the faulting criteria and propose missing value replacement strategies.
RC.03	Data quality assumptions regarding timeliness are not verified.	Request a sample of data from another collection period and evaluate all data quality assumptions. If verified, propose its usage instead. If not verified, clarify timeliness issues in the results portion of the project.
RC.04	Data quality assumptions regarding uniqueness are not verified.	Propose duplicate value removal and validate its consistency with a control dataset.
RC.05	Data quality assumptions regarding validity are not verified.	Propose a dataset engineering sub-project to solve issues.
RC.06	Data quality assumptions regarding accuracy are not verified. This issue is critical.	Propose a short-term data collection strategy and verify its accuracy, utilizing data expansion techniques to account for dimension issues. Replace the project dataset if proposed solution is valid.
RC.07	Data quality assumptions regarding availability are not verified.	Utilize the most recent localized (offline) version of the data available, clearly stating its usage.
RC.08	One (1) or more final models do not comply with the proposed quality standards (i.e. model quality assessment fails).	Determine if the source of the issue can be identified. If the issue can be solved, assess the resolution's impact on project scheduling. If the issue is unsolvable or incurs in irrecoverable delays, do not proceed with model deployment.
RC.09	One (1) or more final models could not be successfully deployed.	Explore alternative deployment methods documented in section 4.6.1.
RC.10	The economic conditions for the development of the project change.	Perform a new cost-assessment and determine whether the previewed budget allows for project completion. Suggest cost-effective alternatives for the biggest financial commitments.
RC.11	Infrastructural (hardware & network) incompatibilities with the proposed project techniques	Consider a cloud-based computing system as a temporary fix. Document relevant software and add-on versioning for troubleshooting.

Table 5: Risk documenting and contingency planning table

#### 4.1.2.4 *Terminology Dictionary*

This task aims to close the semantic distance between the areas of study involved in the project, as some terms specific to data mining may be detached from the subject matter where it is being implemented, and vice-versa. The way this task was approached in the current project is by means of a terminology glossary supporting, a document containing subject-specific terms from a medical and data mining point of view, as well as terms that can be unique to this project's environment. It can be consulted on appendix [A](#).

#### 4.1.2.5 *Cost-Benefit Assessment*

The nature of this data mining project is heavily reliant on decisions which carry potential benefits, and their respective cost, whether it be financial, chronological or non-tangible (impact on results). As such, performing a basic cost-benefit analysis is a requirement for most projects following a in-depth methodology such as [CRISP-DM](#). In a cost-benefit analysis tailored for this specific data mining project, a degree of priority is given to choices that may affect the overall quality of the final model(s), as the financial component of this work is not substantial enough to justify its prioritization over model quality. This results in an atypical situation for this subsection, as any costs related to data collection and model deployment are negligible at this stage. Even still, Techframe superiors have requested an assessment of retrospective, recurrent and future costs associated with the data mining project, which is provided in [Table 6](#). Note that the total project cost defined in this table assumes the compliance with the proposed project scheduling, as per section [4.1.4](#).

The counterpart to the cost table is the reflected benefits this project can have on the business, whether through the direct achievement of the proposed objectives, or through secondary means. Contrary to costs, benefits are not strictly monetary, and can be felt along several faces of the business, hence the importance of their documentation. For this, I have prepared two introspections related to direct and secondary benefits, both of which have been reviewed and validated by my superiors.

**PRIMARY BENEFITS** These are positive changes and outcomes felt shortly after successfully achieving the project's main objectives. In this case, they are directly linked with the successful deployment of the final models.

- The most pertinent financial benefit that stems from this project's success is the approximation of [GV](#) to its commercial state. The model-assisted functionalities

Non-Recurring Costs			
Material / Task	Description	Cost (€)	Acquired?
Hardware Servers	Virtual Desktop 1	2500	Yes
	SQL Secure Server	3400	Yes
Recurring Costs			
Material / Task	Description	Monthly (€)	Yearly (€)
Communication Services	VPN	12,5	150
Software (SQL)	SQL Server	81,72	899
Software (Others)	Additional Supporting Software	50	600
IT Staff	HR for the project	1732,5	24555
Future Costs (est.)			
Material / Task	Description	Cost (€)	Recurrent?
API	Monthly Server Costs	99	Yes
Deployment	Assuming 4 models	8000	No
Hardware	Dedicated API Server	3400	No
User Training	IT Staff	2000	No
User Training	Deployment Phase	10.000	No
Communication	Communication Plan	1800	No
<b>Total Project Cost (€)</b>			<b>44.300</b>

Table 6: Retroactive, recurrent and future cost evaluation for the project.

make the platform marketable as "AI-Powered", which falls in line with modern trends, and attracts investors.

- With the backbone of [GV](#) being its partnership program, this project's success also plays an important role in providing a better definition to the system, captivating small companies focused on healthcare device integration, and larger partners interested in clinical data. In turn, these partnerships produce value to [GV](#), by increasing the number of supported activities and data sources in the long-term.
- The proposed models can be seen as modular additions to the platform, and follow instructions that have been derived specifically for this problem. However, the methods tap into an area of the scientific corpus that is still underdeveloped, as assessed previously. Techframe agreed to make this project and its results public, and not only does it constitute a significant contribution to the way healthcare systems

can be hybridized to improve engagement, but it can also serve as a fully realized example on the application of the [CRISP-DM](#) methodology in such scenarios.

**SECONDARY BENEFITS** Indirect benefits include all other minor, but relevant improvements related to intermediate tasks carried out throughout the project. While they don't necessarily determine the project's degree of success, they may provide valuable intelligence for follow-up tasks.

- Given [GV](#)'s early stage, the business and data understanding sections of the work provide valuable rationalization on the project's direction, a better grasp at associated difficulties, and how to mitigate risks. This sort of induced reflection provides the business with know-how that can ease the early stages of follow-up projects.
- Any processes carried out in the data preparation section can be seen as valuable knowledge for preventing similar gaps in quality from occurring in the future. The business might refer to this project's data cleaning report (see section [4.3.2](#)) to identify recurring mistakes and prevent common issues from slowing down developments.
- The technical know-how necessary for deploying models is also documented throughout section [4.6](#), and details the structures, technologies, and network requisites in a manner that is specific to this business. Not only is a foundation laid out for subsequent modelling efforts, but the preparation to allow follow-up deployments should be much more linear.

A honorable mention could be made towards my own professional and technical skills being benefited greatly with the project, since I am working on it while under an employment contract with the business. Moreover, this opens up future opportunities to continue the technological development of [GV](#). With this, my position is clearly privileged in the sense that the perceived benefits far outweigh the related costs, should the project succeed.

#### 4.1.3 *Data Mining Goals*

Here, a technical translation was carried out to convert the previously established business objectives into ones that are measurable in data mining terms. This provides me with a view of objectives that need to be achieved through contact with the dataset, and allows for a better understanding of the evaluation tools and modelling techniques I can assess, and the overall functions and impact each proposed model should have. The resulting

table (table 7 aims to respond to every question raised in table 2, through the definition of individual data mining problems and proposed benchmarks for some of the issues they aim to solve.

Alias	Type	Description
DM.G.01	Modelling	The resolution of a clustering problem to identify patterns in patient characterization and behaviour.
DM.G.01.1	Technical	The assignment of cluster labels that characterize each of the proposed groups.
DM.G.02	Model	The resolution of a classification task aligned with DM-01, for the labeling of new patients into preexisting clusters.
DM.G.02.1	Technical	A 12 week validity on cluster assignments.
DM.G.03	Modelling	Development of a rule system for constraining activity difficulty targets
DM.G.04	Modelling	Development of a recommendation system for the automatic suggestion of new activities
DM.G.04.1	Technical	Interactivity with the user / therapist by means of a feedback module
DM.G.05	Modelling	Development of a performance analysis tool for the analysis of patient evolution
DM.G.05.1	Technical	Comparative results with the clustered observations

Table 7: Data Mining Goals

#### 4.1.3.1 Data Mining Success Criteria

To provide a thorough way to measure the technical success of the project, this section needs to encompass several aspects related to the methods used for evaluating the models, the specific targets for each measure, and even set some rules about the deployment of the models. To keep the information as clutter-free as possible, I have decided to elaborate the considered factors with short subsections, and provide full criteria listing with a detailed table (Table 8).

**DEPLOYMENT** In some projects, the model deployment phase is a major checkpoint and an integral part of the success of the project as a whole. Given the benefits outlined in section 4.1.2.5, it's clear the models will greatly impact the future viability of *GV*, making it unreasonable to not dedicate all the time and resources needed to ensure deployment is properly planned, documented, and evaluated. For this project, the deployment of each model will be considered part of the project success.

**MODEL EVALUATION** If the modelling phase goes according to plan, there will be several models to evaluate. Depending on whether the methods are supervised / unsupervised, and the usage goals for each model, the underlying metrics for evaluation will vary. As a starting point for evaluating the models, the following metrics will be considered:

- **Accuracy:** A baseline indicator for supervised method assessment (i.e.: Patient Classification). Indicates the percentage of correct classifications over N samples.
- **Time:** The total elapsed time per iteration can be measured as an indicator of performance regardless of model type. All models will have steep performance targets, since most mechanisms are meant to be used in real-time.
- **Interpretation:** A subjective measurement particularly important for models that will be commonly interfaced with human users. In my project, it relates to the ease with which a user can understand the model's outputs.
- **Deployment:** A measurement that relates to the ease with which I can deploy a certain model in the business with as little modifications as possible. It's a measure of particular importance in my project since the platform (GameVita) foresees a modular approach for its intelligent functions.

Naturally, these will be better defined in the test design for model assessment, which can be found in Appendix F.

**BENCHMARKS** The numbers provided here are indicative of good performance, but are not necessarily hard constraints for approving or refusing models, as these are specified in the test design (Appendix F) and applied during model assessment (Section 4.4.4). Since subjective measurements are impossible to quantify, they will be left out of this portion of the document.

- Supervised methods should consistently exceed 80% classification accuracy.
- Supervised methods should take no longer than 5 seconds per iteration, from start to finish.
- Methods for therapist interface (recommendation system and performance analysis tools) should take no longer than 2 seconds per iteration.
- Auxiliary files should require no more than 1/5<sup>th</sup> of a second for reading operations.
- Feedback collection methods should take no longer than 2 seconds for weight adjustment operations.
- The number of patient clusters should be no smaller than 3 and no bigger than 10.

Any model that does not immediately conform with these benchmarks will not be submitted to assessment, and will need to be continuously adjusted for parameters or replaced until the applicable criteria are met. When a specific model meets with these benchmarks, then a more specialized test design is devised and applied for proper assessment.

**SUBJECTIVE MEASUREMENTS** These are perceived success-driven improvements that can not be precisely measured, but are practical when the definition of an exact metric is difficult. For my project, these measurements are especially handy in defining the impact the models have on the end-users. From a subjective perspective, the models achieve success if:

- Patients enjoy partaking in activities more.
- Patients report participating in more activities per given time-period.
- Patients feel like the activity suggestions are adequate and engaging.
- Therapists draw adequate conclusions from the patient performance indicators.
- Setting up activities becomes less demanding for therapists.
- Setting up activity targets becomes less subjective.

#### 4.1.4 *Project Planning*

Often, mismanagement of a project leads to funding issues, missed deadlines, and scope creep, creating an environment unsuitable for its success. I consider project planning one of the most important tasks, and have carried out a thorough assessment of the planning assets necessary for this project. Due to the somewhat extensive contents of this section, I deemed an appendix to be the optimal way of transmitting the information. As such, to consult all planning tasks previewed for this project, one should redirect to appendix [B](#).

##### 4.1.4.1 *Technique Assessment*

As stated in the beginning of this chapter, the tool I have adopted to measure data mining success is the [CRISP-DM](#) methodology. This leaves the remainder of this section as the ideal place to discuss what data mining techniques are better aligned with goals for the project. It is important to note that this section will not include references to the programming languages or any other type of supporting software (see section [4.1.2.1](#) for this topic), and will only serve as an overview on technical approaches that may prove useful during the

Alias	Metric	Criterion
DM.C.01	Deployment	Successful clustering of similar patients (Deployment of model DM.G.01).
DM.C.02	Technical	Creation of logical cluster profiles
DM.C.03	Deployment	Successful deployment of classification model (model DM.G.02)
DM.C.03.1	Evaluation	Achieved 85% accuracy on classifications
DM.C.03.2	Evaluation	Less than 1 second of total classification time
DM.C.04	Deployment	Creation of difficulty constraints for activity targets (Deployment of model DM.G.03)
DM.C.04.1	Evaluation	File query time under $1/10^{\text{th}}$ of a second.
DM.C.05	Deployment	Successful deployment of the recommendation system (Deployment of model DM.G.04)
DM.C.05.1	Evaluation	Generation of first recommendations in under 1 second.
DM.C.05.2	Deployment	Successful collection and parsing of user feedback (Achievement of technical goal DM.G.04.1)
DM.C.05.3	Evaluation	Generation of second and subsequent recommendations in under $2/10^{\text{ths}}$ of a second.
DM.C.06	Deployment	Successful deployment of the performance assessment tool (model GM.G.05)
DM.C.06.1	Evaluation	Presentation of performance results in under 1 second.
DM.C.07	Engagement	Increased patient engagement with activities
DM.C.08	Subjective	Increased patient-reported frequency of activity participation
DM.C.09	Subjective	Satisfactory adoption of the activity suggestion module by the patients and therapists
DM.C.10	Subjective	Satisfactory adoption of the performance assessment tool

Table 8: Success criteria for previously defined data mining goals

data related tasks. Table 9 contains a list of techniques used throughout the data mining project, with a short description on their function.

#### 4.2 DATA UNDERSTANDING

Understanding the goal of the project is, arguably, entirely depending on understanding the data that will be worked on. From the collection process of that data to the steps necessary to fully assess its flaws, drawbacks and potential subsets, this section will

Technique	Function
Clustering	Identify and act upon specific patient data distributions, aiming to partition the existing patients into a reduced number of groups for profiling.
Outlier Detection	A crucial part of the patient and activity cleaning process. Identify patients that stray much further away from the norm, which could mean a severe disparity in physical state or game performance.
Decision Trees	For modelling a feedback system based on a modular questionnaire applied to the user post-activity. Internally, guides the program to take specific actions based on the patient's answers.
Regression Analysis	During exploration, will be used extensively to analyze relationships between variables. Similarly, could be a useful metric in predicting the improvement or decline of a patient's activity scores.
Prediction	In my project, will be mainly used as a follow-up to clustering, in which new patients can be classified into existing clusters.
Insight Analysis	A generic process that will be applied throughout the entire project, with the goal of identifying patterns or changes in the data that could potentiate better decision making for the business, or for the modelling tasks.
Visualization	The main form of result presentation used in my project. Graphs and tables will constantly coexist with observations, takeaways and conclusions about the data at any point in development.

Table 9: Data mining techniques and roles

document all of the preliminary data-forward tasks previous to the actual data mining work.

While the outputs expected for this phase are quite straightforward, mainly consisting on additional reports documenting each of the generic tasks to greater detail, they are very important artifacts that will be referred to throughout the entire data mining project, since they document the vast majority of the knowledge regarding the origin, meaning, and quality status of the data. Table 10 contains the foreseen tasks, deliverables and related appendix chapters for this phase.

Generic Tasks	Deliverables	Appendices
Data Collection	Collection Process	N/A
Data Description	Data Description	Section C.2 of Appendix C
Data Exploration	Exploration Process	Section C.3 of Appendix C
Data Quality Assessment	Quality Assessment	Section C.4 of Appendix C

Table 10: Data Understanding Breakdown

#### 4.2.1 Data Collection

Data can come from a variety of sources. Depending on the situation, data can be as simple as a single spreadsheet, or as complex as multiple real-time streams. Nevertheless, it is important to produce a document detailing the initial data collection process without necessarily assessing its usability at this point. Here, the focus is explaining the method by which the data was collected by the business, and also how it was provided to me for this project. This first task also includes dealing with loading the data onto the modelling tool, which is a process that is documented in section C.1 of appendix C

##### 4.2.1.1 Collection Trial

The data for this project was collected by the business at a date prior to my involvement with GameVita. The process of collection was a preliminary demonstration of the proof-of-concept for assessing their adequability to users will real therapeutic needs. Participation in the trial was proposed to a total of 1500 patients chosen to represent the population that the platform caters to. Participants who accepted were then submitted to three short stages of data collection. Out of the initial pool of patients, 874 completed the trial successfully and their data are the focus of this project. The remaining 626 (41.7% of total) patients were excluded from the trial due to participation denial, data screenings, or other factors, explained in the graph below (Fig. 9).

The names "Stage 1", "Stage 2" and "Stage 3" refer to the three distinct phases that were carried out for data collection, detailed below in order.

**STAGE 1 - PERSONAL FORMS** In the first stage, participants were asked to fill a digital form for the collection of personal and demographic data. An automatic screening process eliminated participants aged under 65, those with over 3 other comorbidities, and those with conditions that would make it impossible to participate in games (upper limb

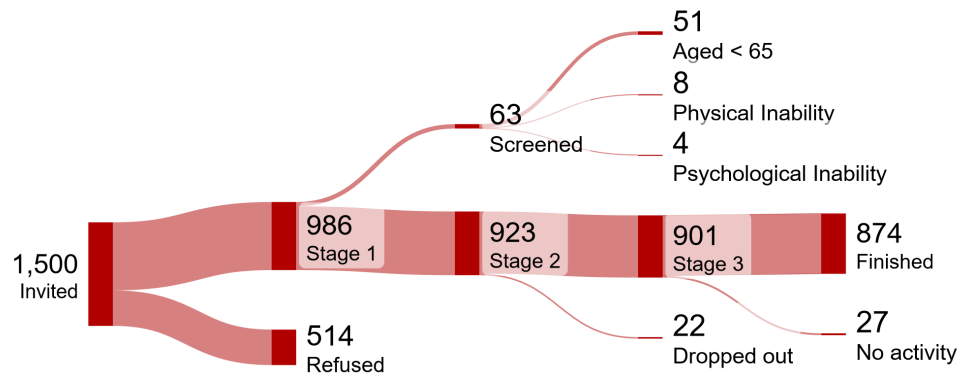


Figure 9: Process of elimination throughout trial stages

amputations, palsy, neurological conditions related to cognition). The final data from form filling was validated and inserted to an [SQL](#) Server Table.

**STAGE 2 - SUPERVISED SESSIONS** Participants who moved on to the second phase were submitted to a short (1 to 3 days, depending on availability) training stage in which the mechanics of the games were explained. Patients were able to participate in a controlled number of game sessions under supervision, and were free to drop out of the study at any time. At this point, session data was not collected since a lack of understanding could influence the scores in a way that is unrelated to the patient’s physical capabilities.

**STAGE 3 - UNSUPERVISED SESSIONS** Finally, patients that reached the third stage of the trial were provided with temporary access to the platform, and could engage in real activities over the course of 9 days. During this time, patients could engage in a maximum of one activity a day, having to wait for the next day to engage in another game. This measure was implemented to prevent overexertion, and to give therapists time to configure new game settings. All data pertaining to session characterization (game played, difficulty level, targets), scoring (measured performance) and rating (engagement time, satisfaction forms) were recorded to an [SQL](#) Server Table. Patients who did not engage in any activity were, naturally, removed from the study. In total, 3525 relevant game sessions were present at the end of this stage, and the data pertaining to these activities constitutes one of the datasets for this project.

#### 4.2.1.2 *Data Acquisition*

Access to the data has been granted in two distinct phases. First, I was provided with an administrator-level account to perform an initial assessment of the data structure, which allowed for a pre-evaluation of the attributes that could bring value to the project. Afterwards, processes to extract the data and censor the sensitive attributes were performed by Techframe, who provided the final two files themselves. No additional external data was purchased or sourced for this project.

#### 4.2.2 *Data Description*

Describing the data is usually a task of documenting the quantity and quality of the sources, and includes taking notes of important characterizing factors that may affect the way I will approach the preparation and modelling phases. This section will contain a description of each relevant field's purpose. For information about formatting, data types and basic measures about each attribute, refer to section C.2 of appendix C. Starting with the attributes present in the patient data set:

- **fried**: Patient frailty classification according to the FRIED assessment.
- **gender**: Binary gender assigned at birth.
- **age**: Age of the patient at the time of data collection.
- **hospitalization one year**: Number of times the patient was hospitalized in the 12 months past data collection.
- **vision**: Categorical evaluation for the patient's quality of sight.
- **hearing**: Categorical evaluation for the patient's quality of hearing.
- **weight loss**: Whether the patient has unwillingly lost weight.
- **exhaustion score**: Whether the patient reports an aggravated feeling of tiredness and exhaustion.
- **balance single**: Whether the patient can balance themselves on one leg for a duration superior to 5 seconds.
- **grip strength abnormal**: Whether the patient's grip strength falls severely short of the normality.
- **low physical activity**: Classifies a patient's behaviour as sedentary.
- **bmi score**: The patient's body mass index.

- **sleep:** Patient's self reported frequency of sleeping issues.
- **depression total score:** Patient's self reported measure of depressive stage.
- **anxiety perception:** Patient's self reported state of anxiety
- **pain perception:** Whether the patient feels pain chronically, and how severe it is.
- **activity regular:** Measured activity time per week.
- **smoking:** Whether the patient has ever smoked, or is a current smoker.
- **alcohol units:** Number of alcohol units consumed in a week.
- **medication count:** Number of prescription medication being taken by the patient.
- **birthplace:** Country of birth

There is a lot of potential in this data for finding subsets and correlated attributes. Given the number and range of attributes present, the task of finding clusters of similarly characterized patients remains logical, which means the data exploration phase for this set will be dedicated to achieving a deeper understanding on the relationships between attributes.

Moving over to activity data:

- **game id:** Identifies the game that was played in the current session.
- **therapeutic difficulty:** The relative difficulty of the task, given the target scores.
- **target grip force:** The average amount of force that needs to be registered during the activity, in kilograms of force.
- **target grip timing:** The target accuracy and timeliness of grip presses.
- **target instances:** How many times the controller needs to be successfully gripped during the activity.
- **grip force score:** The actual average force the patient registered during the activity.
- **grip timing score:** The actual percentage of instances hit at the correct timing.
- **grip instances hit:** The number of times the controller was successfully gripped.
- **score:** Measures the target scores against real scores, providing a numeric performance measure for the current activity.
- **session target duration:** Number of seconds foreseen for the activity.
- **session real duration:** Real duration of the activity.
- **satisfaction query:** Self reported enjoyment of the activity by the user.

Here, the data is more technical. There are no real inferences that can be made just by looking at a numerical data sample, however, the presence of both satisfaction queries and game identifiers is a positive indicator that a game recommendation model can be constructed. Also, it would be interesting to deconstruct the measure of therapeutic difficulty and figure out how it affects the target scores for each session.

#### 4.2.2.1 Self-reported symptoms

Some of the data included in the patient dataset originates from self-assessments done at regular intervals. Because of this, the verity of some attributes may be question. To clarify how these symptoms were measured, the following list includes the evaluation scale for self-assessed variables:

- **Exhaustion Score:** Users were asked whether they feel some degree of exhaustion. Respondents answered 1 for **No exhaustion** and 2 for **Exhaustion of any degree**.
- **Depression:** Users were asked to rank their level of depression in a scale from 0 to 27 according to the PHQ-9 test (Williams, 2014a).
- **Anxiety:** Users were asked to rank their perceived anxiety in a scale from 0 to 21, according to the GAD-7 test (Williams, 2014b).
- **Pain:** Users were asked to rank their pain in a Numerical Rating Scale (NRS) (Haefeli and Elfering, 2006) from 0 to 20.

Despite the usage of clinically-proven methods and scales, the accuracy of these factors can not be guaranteed beyond doubt. Still, for this project, these attributes will be considered sufficiently accurate, and will be evaluated, studied and included in all data mining phases, if proven significant. Below, in Tab. 11, the categorical scale for each assessment test is provided. Note that the PHQ-9 and GAD-7 share categories, so the range scale column has been merged for the two assessments.

PHQ-9 / GAD-7 Score	Depression	Anxiety	NRS-20	Pain
0 - 4	None/Minimal	Minimal	0 - 1	Negligible
5 - 9	Mild	Mild	2 - 6	Mild
10 - 14	Moderate	Moderate	7 - 11	Moderate
15 - 19 (21)*	Moderately Severe	Severe	11 - 15	Severe
20 - 27	Severe	N/A	16 - 20	Unbearable

Table 11: Categories for self-assessed measures of depression, anxiety and pain. \*The GAD-7 score maxes out at 21.

#### 4.2.2.2 *Therapeutic targets*

Each session played has a series of objectives meant to encourage an appropriate exercising of the hand and forearm muscles needed for gripping. These goals are not completely random, vary tailored according to each patient's overall physical well-being, and related to:

- **Grip Force:** The average level of measured grip force throughout the activity, measured in Kilograms (Kgf).
- **Grip Timing:** The rate of grip presses being registered at the correct cadence. For example, pressing the controller in rhythm with a song.
- **Grip Instances:** The total number of instances where a grip press is required throughout the activity.

Alongside these values is a categorical, overall level of difficulty for the session, which can range from 0 to 10. Naturally, sessions measured lower on this scale have easier therapeutic targets, and are meant for users in a worse physical shape, or more advanced stage of frailty. Finally, the score of each session is calculated by summing all target values ( $T$ ), and subtracting the sum of all achieved scores ( $S$ ), as per equation 1. Scores equal to or higher than 0 mean a pass on a therapeutic level, where as scores lower than 0 signify a fail.

$$FS = (T_f + T_t + T_i) - (S_f + S_t + S_i) \quad (1)$$

Lastly, sessions have proposed duration times of 45, 60 and 90 seconds, and although the patient is encouraged to complete the activity, they can stop at any time if their energy is too low to proceed or they are not enjoying the game's difficulty.

#### 4.2.3 *Data Exploration*

This is a relatively open-ended task, where I can perform as shallow or deep of a dive into the data as needed, depending on my familiarity with relevant attributes, interesting subsets, and its compatibility with the proposed data mining goals. This is one of the most time consuming tasks in the [CRISP-DM](#) methodology, with it being a stage where the data is still unfamiliar, making the construction of adequate views and graphs a challenge. Moreover, the exploration phase should help in assessing the viability of certain data

mining goals, thus making the identification of interesting subsets, promising attributes and gaps in information a priority. The bulk of the content for this task is available in section C.3 of Appendix C. Here, a short summary of findings is provided to contextualize the results of exploration.

In both datasets, exploration revealed that some attributes are more promising than others, showing stronger relationships and interesting distributions. I focused attributes that seem to directly correlate with a patient's frailty assessment, and ended up confirming that all five frailty risk factors present in the data (abnormal grip strength, low balance, unwanted weight loss, exhaustion and sedentism) have the most potential for modelling users according to their risk. Moreover, exhaustion and abnormal grip strength also appear to directly affect a patient's grip timing and grip force scores in-game, respectively. Other attributes, such as demographic indicators, mental assessments and measures related to habits, while not as impactful for classifying patients, can be considered as secondary risk factors, providing additional depth to clustering tasks. It is also important to note that merging both datasets provided interesting insights about the population's game preferences, which could help with tasks related to increasing patient engagement with the activities.

When reviewing data mining goals, one of the main things to discuss is whether the exploration process revealed the need to engineer new variables to work with due to a lack of important subsets. One way to find an answer to this question is by reviewing the exploration process with the data mining goals in mind. First and foremost, the patient data has been verified to be sufficient in both quality and quantity to permit a clustering task to be carried out, creating subsets of similarly disposed patients regarding their general well-being and frailty severity. The attributes present in the data contain interesting distributions that will, hopefully, reduce intra-cluster variation and increase inter-cluster variation. By consequence, I also validated the existence of important variables for adjusting patient cluster over time. Secondly, both datasets have been found to contain significant variables that support the creation of a prediction model which, depending on the level of their physical capacity, foresees the assignment of risk and enjoyment variables to patients, helping therapeutic professionals to make better informed decisions on the activities users should partake in.

#### 4.2.4 *Data Quality Assessment*

To fully understand the data, this final task deals with the assessment of quality to a higher degree of specificity, creating an important listing of what issues need to be dealt with

during preparation. The quality assessment process is detailed in section C.4 of appendix C. Some additional considerations and a statement about the data quality can be consulted below.

If possible, determining the meaning behind quality gaps can be helpful in making the decision to keep or exclude it from the data. Following the same order as the contents in the appendix section, the first set of problems to analyze are structural issues with the data. There are ten irrelevant columns to be removed, and five columns with non-critical issues that can be fixed in a simple manner, out which one requires the engineering of a new attribute (birth year), and another implies modifying the value range (exhaustion). No critical structural issues were found. Missing values are present in two columns of patient data (sleep and activity regularity), and one column of activity data (satisfaction query), though more notably in the latter. Given this situation, missing values in either column of the patient dataset will instead be replaced with the mode for the corresponding attribute, whereas missing values in the activity dataset column will be kept as a new categorical option (For patients who did not answer the satisfaction form). Lastly, I searched both datasets for outlying values and found negligible appearances on either one. In patient data, there are some deviations in the number of alcohol units and concurrent medications consumed, whose corresponding patient rows will be kept as to not remove potential cluster information from the set. There are also 8 rows of outlier performance in activity data, though only in the attribute that measures grip timing score. Since eliminating these rows translates into a dataset loss well under .01%, this will be the most optimal approach to avoid skewing the models.

Having finished the quality assessment task, I can safely conclude that the data quality is adequate for data mining, and that the data preparation steps that were raised from this process are minimal and should not lead to unforeseen delays for the project completion.

#### 4.3 DATA PREPARATION

At this point in the project's development, a clear understanding of the data characteristics marks the beginning of this section. Here, several activities related to the improving the quality of datasets take place to construct the version of the data that will be modelled. This task is wildly variable depending on the project's goals, the data formatting and modelling tools available, and as such may be revised several times. Of course, the main output from this section is the data itself, through which the modelling phase is enabled. Table 12 contains the foreseen tasks, deliverables and related appendix chapters for this phase.

Generic Tasks	Deliverables	Appendices
Data Selection	Inclusion / Exclusion Rationales	N/A
Data Cleaning	Data Cleaning Report	Appendix D
Data Engineering	Derived Attributes, Generated Records	Appendix E
Data Integration	Merged Dataset	N/A
Formatting	N/A <sup>1</sup>	N/A

Table 12: Data Preparation Breakdown

<sup>1</sup> No additional formatting tasks were carried out.

#### 4.3.1 Data Selection

Prior to fixing the issues found in the last process of the previous phase, there are some important decisions to be made about the data I wish to use during modelling. Selecting subsets must be a conscious and justified process, as the removal of important attributes can preclude the validity of the results.

Initially, every attribute that was simultaneously (a) not empty, (b) not tagged for replacement and (c) not unnamed was kept for data cleaning. The removal of attributes was only considered afterwards, when deciding on the required engineering operations to make the data compliant with model requirements. To assess whether a given attribute should be kept or removed, I looked to answer to following questions:

- Is the attribute relevant to a particular model?
- Is this attribute's quality acceptable?
- Does the attribute have any particularly interesting relationships with other attributes?
- Are there any legal or moral constraints in using this attribute?

In Tab. 13, I have compiled the main attribute removal decisions, relating them to particular models. Other models not included in this table make use of all attributes that remained after cleaning.

Model	Removed Attributes	Rationale
Clustering	fried, fried count, gender, hospitalizations vision, audition, bmi, sleep issues, smoking habits, alcohol units, medication count, id, birthplace, birth year	During exploration, most of these attributes were found to have little correlative properties, while others were removed for being classification mechanisms themselves.
Recommendation	force score, timing score, instances hit, final score, session duration	The recommendation algorithm makes decisions based on the proposed activity targets, and is agnostic to the actual scores achieved by the patients.

Table 13: Removed attributes by model

#### 4.3.2 Data Cleaning

During the Data Understanding phase, the quality of both sets was assessed, revealing the issues that would need solving during the current phase. Thankfully, there were no major gaps in quality, resulting in a relatively simple cleaning task, which involved:

- **Renaming Attributes:** Mostly for clarity, since some attributes' names did not initially match the purpose of the stored values.
- **Removing Invalid Categories:** Ensuring all categorical possibilities for attributes are valid.
- **Dealing with Missing Values:** By means of replacement with a central tendency value, as decided previously.
- **Eliminating Outliers:** In this case, only activity data outliers were removed.

The tasks carried out are detailed in Appendix D of the document. None of the issues identified were left unresolved, and the resulting datasets contain no liabilities for data quality.

#### 4.3.3 Data Engineering

Due to this project having multiple modelling tasks, the data engineering process is the most time consuming preparation step. Here, not only is new data created, but all manner of transformations to the existing fields will be performed to ensure that the final datasets will fit the modelling techniques. Despite the CRISP-DM methodology not suggesting a dedicated appendix for this task, I have deemed that its creation is vital to enable a good understanding of the process.

As a result, the tasks carried out are detailed in Appendix E of the document. Note that the process of engineering was revisited multiple times as new modelling tools were tested, and thus it does not necessarily take place immediately after Data Cleaning.

#### 4.3.4 *Data Integration*

To align the data with the specific goals related to clustering, I performed a short data integration process by **merging** both sets with a key attribute. The goal of this task was to expand the number of significant variables in the patient dataset by reflecting the average activity scores each patient achieved. For this, the **874** row patient data was first merged with the **3525** row activity data on the patient identifier attribute, and all rows pertaining to the same user were aggregated with the mean function. Due to some patients not having registered activities, the final merged dataset contains **861** rows. The merged data will be used extensively in data modelling, and has been saved alongside the other data resources.

#### 4.3.5 *Formatting*

After having carried out extensive tests regarding the modelling tools to be used, the task of identifying necessary formatting for the data was finally possible. As it stands, out of all of the proposed systems, only the methods surrounding the clustering algorithm require specific data formats. On one hand, this makes data formatting very simple, since there are no remaining tasks to ensure the data fits in the proposed models. On the other hand, this is happening because the majority of the systems are developed from scratch, hence the modelling phase having a much steeper complexity in exchange. To stay in conformity with the methodology, the following models are to be used. These modelling approaches have been evaluated for viability in section [4.4.1](#), and will be used as a "blueprint" to solve the data mining goals.

- **K-Prototypes** for clustering merged data.
- **K-NN Classifier** for predicting clusters on data.
- **Collaborative Filtering**, the basis for the recommendation system.
- **Polynomial Curve Fitting**, the basis for the performance analysis system.
- **Markov Decision Process**, the template for the feedback module.

None of the aforementioned models require additional formatting operations to be done in advance, justifying the non-necessity for the current task.

## 4.4 MODELLING

The fourth phase includes the selection and calibration of one or more modelling techniques. Depending on the models applied and their data constraints, several iterations of the previous phase (data preparation) phase may be necessary. At the end of this section, most of the actual technical side of the Data Mining project is complete, and an overall idea of the success of the project becomes clear. Table 14 contains the foreseen tasks, deliverables and related appendix chapters for this phase.

Generic Tasks	Deliverables	Appendices
Modelling Techniques	List of Modelling Techniques	N/A
Test Design	Model Assessment Process	Appendix F
Model Building	Parameters, Models, Model Descriptions	N/A
Model Assessment	Assessment Results	N/A

Table 14: Modelling Breakdown

4.4.1 *Modelling Techniques*

The choice of techniques used for data modelling depends on several criteria, like the available data types, the size of the dataset, and whether or not the results need to be easily presentable. For this project, the overarching data mining goal is to produce a series of tools to help the therapists make better decisions about their patients' activities. For this reason, the ease-of-interpretation is highly prioritized, and the models will be built with user friendliness in mind. To make modelling as related to each of the proposed data mining goals as possible, this short section will be used to raise the modelling techniques that will be used, and document any additional data requirements that may need solving before proceeding. In short, the objectives I have raised can be boiled down to six distinct tasks, each of them requiring the application of different techniques and methods.

4.4.1.1 *Patient Clustering*

The creation of groups of similarly disposed patients creates a reference upon which many of the upcoming tasks will be based. When building clusters, we're looking for the least possible *intra-cluster* distance, while maximising *inter-cluster* distance, which raises

two major questions. The first is deciding what attributes should be considered when running the clustering iterations, since having too many irrelevant attributes can have repercussions on all aspects of the task (more time to converge, clustering being affected by attributes with virtually no impact, harder usage of the elbow method). Secondly, deciding how many clusters to keep is rarely a straightforward process: Having too many (or too few) clusters detracts from the meaningfulness of patterns in the original data, which can often become underrepresented. Apart from this, there are questions regarding the dimensions and formats within the data, which may greatly influence the choice of technique. For the problem at hand, a K-Means based approach appears to be adequate given the low dimensionality of the data, and the coexistence of both categorical and numerical attributes. To work with vector quantization algorithms such as K-Means, the data must abide to some formatting constraints, such as the scaling of numerical attributes and the encoding of categorical attributes (either one-hot encoding or label encoding). The preprocessing of merged data, detailed in section E.2 of Appendix E, ensures the data is conforming with some of these specifications. For handling categorical encoding, a variation of K-Means, aptly named K-Prototypes, is preferred for its ability to automatically generate and include dummy variables from specified categorical variables into the clustering process.

#### 4.4.1.2 *Patient Classification*

As a follow-up to the previous task, classifying new patients into pre-existing K-Means clusters can be achieved with a **K-Nearest Neighbors (KNN)** method, requiring minimal processing of new observations. For this, the clustered data is analyzed and a measure of distance is defined between a new observation and all existing observations. Afterwards, a configurable number of neighbors (most similar patients) is used to set a label for the new observation. This can be considered a supervised model, since we will have ground-truth labels to measure concrete performance metrics. As such, it could make sense to split the data into testing and validation sets when evaluating the results, which is a factor to consider when developing the test design for the project.

#### 4.4.1.3 *Activity Difficulty Reverse Engineering*

As the name of this task suggests, I won't target any specific technique for achieving the underlying goal of standardizing the difficulty labels assigned to activities. Instead, this is a relatively open-ended task to solve an undefined problem related to the level of subjectivity adjacent to ranking an activity from one to ten in terms of its difficulty. My approach will mainly involve the analysis of central tendency values of each therapeutic

target, for all games and difficulty labels, hoping to set concrete ranges and deviations that therapists can consult when defining the targets for a specific activity.

#### 4.4.1.4 *Activity Recommendation*

I will base the logic for the recommendation system on the initial premise that similar patients will share similar interest, which is an approach commonly used in media streaming platforms. The process, referred to as **Collaborative Filtering**, will use the findings of the previous tasks to search similar patients, and then recommend unique activity parameters based on the preferences of the neighboring patients. As most calculations are based on the **KNN** algorithm, no additional data quality requirements are foreseen for this model.

#### 4.4.1.5 *Feedback System*

This task can be understood as an add-on to the previous model, whereby the therapist can guide the patient through some feedback collection questions that capture their sentiment towards a recently recommended activity. The premise for the reinforcement model I aim to produce is influenced by the **Markov Decision Process**, where a mathematical framework is defined to model choices controlled by a decision maker, which sequentially and iteratively makes decisions over time to maximize a reward. Here, my approach will be to define a set of weights referring to the different components of a recommendation (the game, its duration, each of its therapeutic targets...) and, through a series of choice paths, adjust the weights to reflect the enjoyment of patient. In turn, subsequent iterations of the recommendation system will analyze the weights and perform changes to the recommended parameters, aiming to satisfy the likes of patients while still providing a challenging environment. This model will not directly manipulate any of the original data sources, and instead will perform parameter management within a separate file. No additional data constraints are foreseen for the construction or evaluation of this model.

#### 4.4.1.6 *Patient Performance Analysis*

To close out the list of modelling techniques for the project, I aim to leverage **Polynomial Curve fitting** to discover trends in patient performance, providing therapists with a simple tool to track the evolution of a user across activities. Two distinct insights will be provided with this tool: The first refers to a patient's isolated performance, that is, how they have grown in terms of physical well-being since the they've begun partaking in gamified exercises. The other refers to a patient's relative performance, where they will be indirectly compared with other patients in the same cluster, to highlight discrepancies

that may symbolize significant improvement or deterioration of their physical ability. Again, this will be an analytic process with no added data requirements.

#### 4.4.2 *Test Design*

The test design is a specification document that contains a step-by-step breakdown on how to properly assess the quality of the models I intend to develop. It is a crucial deliverable, containing measurable performance metrics, subjective indicators, and thresholds that indicate a premature non-viability of the models. For the current project and the several underlying models, I have devised an auxiliary document detailing the test designs to be employed on supervised and unsupervised models, which can be consulted in Appendix F. The instructions in this document are the basis for which section 4.4.4 will be based on.

#### 4.4.3 *Model Building*

In this section, a description of each model, alongside additional context, insights, patterns and issues found during building, will be presented. This section directly complements the modelling notebook section with the same name. Whereas the latter is purely a code-oriented approach, the former details some of the thought processes that went into solving the problems found along the way, serving as a minimal development log for the final models.

##### 4.4.3.1 *K-Prototypes Clustering: U.01*

The baseline model for the project is a clustering process that aims to partition  $N$  observations into  $K$  groups of similar disposed patients. With this model, my goal is to abstract the data from the preexisting frailty labels, which take in consideration an outdated recording of the patient's physical state, and in turn discover new logical groups that also reflect a patient's ability to play the games within GameVitae.

**OPTIMIZING  $K$**  In iterative clustering processes like **K-Means** or **K-Prototypes**, the number of clusters,  $K$ , to build is predetermined. The task of choosing the best candidate number for  $K$  is rarely straightforward, but several cost-analysis techniques exist to aid with this choice. For this particular task, I employed a distortion score elbow method, which, for a range of values for  $K$ , aims to minimize the average euclidean squared distance between centroids of clusters. Typically, this point in a graph represents an 'elbow' curve,

after which the decreases in distortion are less prevalent. I settled on  $K = 4$  for my clustering task, supported by the graph present in Fig. 10.

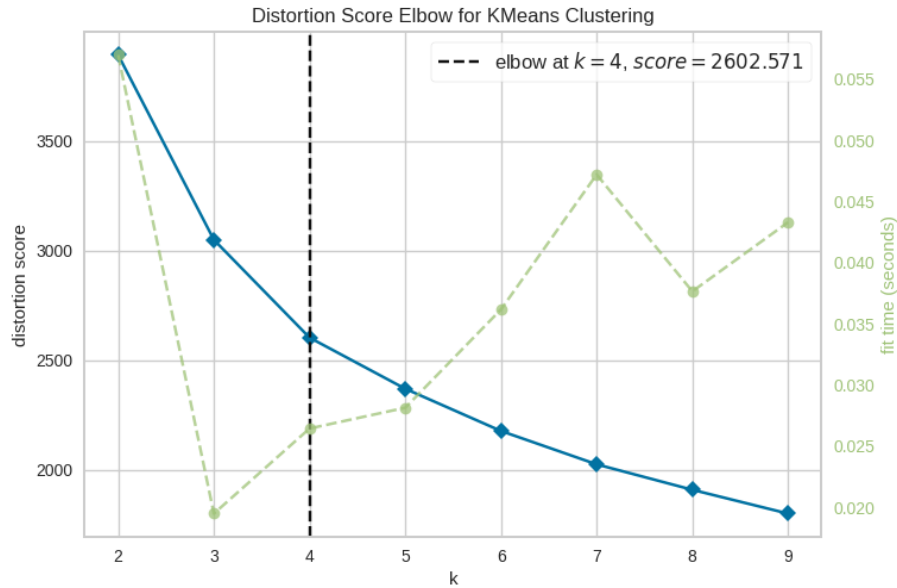


Figure 10: Distortion Elbow Method

**CATEGORICAL ENCODING** **K-Means** expects all attributes to be numerical and properly scaled prior to clustering. While the scaling of attributes is something that I prepared for previously (See Section E.2 of Appendix E, the usage of categorical attributes had initially proven difficult to implement. The migration from typical **K-Means** to the variant **K-Prototypes** resolves this issue by inherently generating one-hot encoded labels for the specified categorical variables. One-Hot Encoding is a type of categorical encoding that splits each possible category from an attribute into  $N$  binary attributes with one mutually exclusive high bit (True) observation.

#### 4.4.3.2 KNN Classifier: S.01

The second model is a natural progression of the clustering function, and aims to classify new patients into preexisting clusters, by analyzing the most common label in the  $K$  most similar observations. The classification method is generally known as **K-Nearest Neighbors**.

**DISTANCE METRICS** In the unusual situation that my data presents (clustered on data of two data types), there is an added difficulty in selecting the metrics to be used in calculating neighbor distances. To ensure there is no mismatch in the data, the **KNN** process

must consider the exact same attributes used in clustering, which, in this case, means numerical and one-hot encoded categorical attributes. As such, I propose the hybridization of **two distance metrics** in my **KNN** approach, where the *euclidean* distance (Eq. 2) is used to evaluate vectors of continuous numerical variables, and a *modified Jaccard index* (Eq. 3) enables the comparison of binary vectors (i.e.: one-hot encoded categorical variables). Both values are combined in a weighted average, where, to keep the level of importance consistent with the prevalence of each variable type in the data rows (10 numerical and 6 categorical), the weights 0.625 and 0.375 are applied, respectively.

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (2)$$

$$j(p, q) = \frac{\sum p \cap q}{\sum p \cup q} \quad (3)$$

**PREPROCESSING DIFFICULTIES** Given the substantial amount of prerequisites and preprocessing needed for this model to function, I developed an auxiliary function that simplifies the process of normalizing the input row (that is, the patient to be classified). When developing the logic for this function, I faced significant setbacks related to construction of dummy variables on the input row, since there is no information about the available categories for categorical attributes when the sample size is 1. To overcome this issue, the input row is appended to original data, and dummy variables are generated for the entire dataset. Thereafter, all rows except the input row are removed from the data, resulting a one-hot encoded dataset with 1 row. In turn, this approach has the benefit of enabling pairwise operations between the data, which are vastly more efficient than the employment of loops. Numerical data on the input row also originated an issue, due to the need of scaling attributes in exactly the same proportion as the clustered data. For this, when generating the **clustering dataset** (see Section E.2 of Appendix E), the weights for the scaling function were saved in a file that is critical for the operation of the **KNN** function, and enable consistent scaling of numerical attributes. Finally, since clustering was performed on a mix of activity and patient information, it is impossible to do **KNN** classification on a patient with no history of activity data. Because there is no workaround for this requirement, one limitation of this model is its inability to classify fresh patients that still haven't participated in an activity.

#### 4.4.3.3 *Therapeutic Difficulty Reverse Engineering: P.01*

The third modelling task is, essentially, a set of reverse engineering operations that aim to define specific constraints for the different therapeutic difficulty levels found in activity data rows. Due to the lack of a definition for this classification, my limited understanding is that the *therapeutic difficulty* attribute is simply a subjective measurement used by therapists to assess the overall steepness of the targets for a particular session, thus having no real meaning or range of values underlying the evaluation. Here, I analyze rows on a game-by-game and difficulty-by-difficulty basis, finding theoretical lower and upper bounds for each of the three therapeutic targets (force, timing, instances), and determining the standard deviation for these metrics. The file that results from this task is an important dependency for the upcoming recommendation model, and alleviates the subjective nature of the difficulty attribute by providing a concrete definition for it.

**BOUNDARY SETTING STRATEGY** In its current stage, ten difficulty levels are observable in the activity data, though the distribution of therapeutic targets is not consistent across the three available games, which means there are 30 distinct combinations of **game / difficulty** to define. My approach was to aggregate rows that share a particular game and difficulty, and extract the minimum, maximum, and average values observed for each of the three difficulty parameters (an activity's duration was not considered part of difficulty assessment). The difference between the minimum and maximum extremes represents the standard deviation (sd) for a particular target, and determines the level of divergence that is allowed from the average value. For example, if the minimum, maximum and average force targets for a specific game and difficulty are 18.5Kgf, 20.0Kgf and 19.0Kgf, respectively, then  $sd = 20.0 - 18.5 = 1.5$ . In turn, the allowed range for this metric is given by  $19.0 \pm 1.5$ , or [17.5, 20.5].

#### 4.4.3.4 *Recommendation Algorithm: U.02*

The fourth model is the most complex, and helps the therapist in setting up parameters for activities that not only require a sufficient level of physical exertion, but is also to the patient's liking, helping them stay engaged with the platform. Here, I have prepared a recommendation system that works in three distinct moments, detailed below. The premise is that as long as the patient has at least one activity recorded in their profile, this mechanism can continuously generate new activities and suggest adjustments based on frequent user interaction.

**FIRST RECOMMENDATION** The process for generating the first recommendation for a user takes most of its logic from the [KNN](#) method described earlier, following a collaborative filtering method (similar patients have similar interests) in which K neighbors are evaluated. The main difference, however, is that the process cuts off before a label is assigned, and an ordered list of K neighbor rows with aggregated activity results (average values per game) is used instead. The generation of the first recommendation follows several steps in a specific order: First, games are ordered by overall rating (given by the satisfaction queries), and the best received game amongst neighbor data is picked. Secondly, ratings are separated by difficulty level for that particular game, which in turn provides information about the range of therapeutic targets that is best received by similar patients. Finally, the auxiliary file provided by the previous model (specified in [Section 4.4.3.3](#)), is queried, and the corresponding difficulty ranges are used to generate randomized targets within the constraints.

**FEEDBACK COLLECTION** Ever since the first general ideas for the modelling phase were drafted, I envisioned the implementation of a method of feedback collection that would work iteratively to improve the quality of activity recommendations. The purpose of this mechanism is to act as an intermediate step after the first and all subsequent recommendations, applying a series of questions that aim to understand the user's general level of satisfaction towards various aspects of the recommended activity (game, duration, difficulty). To record answers, I implemented a weight shifting system, where a 5-position vector is initialized as  $[0, 0, 0, 0, 0]$ , and assigned to the latest recommended activity, symbolizing a neutral feedback status (no answers have been collected yet). Throughout the questioning process, answers that indicate the need for positive change in an activity's parameters (longer duration, harder targets) trigger a weight shift to 1 in the corresponding vector position. On the contrary, answers that represent a need to reduce an activity parameter trigger a weight shift to  $-1$ . The vector positions are as such:

- 1: Game recommended
- 2: Duration of the activity
- 3: Grip force target (in Kgf)
- 4: Grip timing target (in %)
- 5: Grip instance target

As an example, if, after feedback collection, the weight vector is  $[-1, 0, 1, 1, 0]$ , it means the patient wants to play a different game, and that they felt both grip force and grip timing targets need to be a little higher.

**ITERATIVE RECOMMENDATIONS** Through the adjustment of weight vectors, subsequent recommendations become iterative in the sense that after each one, the vector is reset to  $[0, 0, 0, 0, 0]$  until the user provides more feedback about the latest recommended activity. When generating activities based on user feedback, the process becomes individual to the current patient and thus the **KNN** method is no longer required. Instead, the difficulty range sheet is simply consulted whenever an adjustment in difficulty needs to be made, and new difficulty targets are gradually tailored to the user's liking (always overruled by a therapist's decision).

#### 4.4.3.5 *Performance Analysis Tool: U.03*

The fifth and final model is a therapist focused tool that quickly provides a summarized view of how a particular patient is performing. To achieve this, the following plots are generated and presented:

- **Performance Graphs:** Across all activities recorded for the patient, plots the proposed activity targets and actual scores for force, timing, and instances.
- **Comparative Graphs:** Across all activities recorded for the patient, plots their real achieved scores for force, timing and instances against the average score achieved by patients in the same cluster.

Additionally, auxiliary performance metrics (trend values and score coefficients) are calculated, providing a way to determine the physical evolution of the patient, and whether the activities they participate in are adequately challenging. To make the matter of data digestion and interpretation easier, I have also implemented a simple conditional logic that automatically provides contextually relevant conclusions about each of the metrics presented.

**POLYNOMIAL CURVES** To determine whether a patient is evolving or deteriorating physically over a period of time, I use polynomial fitting of the first degree, given by  $y = ax + b$ . Because first degree polynomials are given by a line, and a line connects any two points, then the slope of the function,  $a$ , can be interpreted as the trend of patient performance. In that sense, a positive slope indicates an general improvement in a patient's therapeutic scores, whereas a negative slope signifies a decrease in a patient's therapeutic scores.

#### 4.4.4 *Model Assessment*

In this section, the test design detailed in Appendix F is applied to each model, with the goal of deciding if:

- The models in question meet with the quality criteria.
- The results of the models are deployable. If so, what would be a logical way of deploying?
- The models achieve the goals proposed in Section 4.1.3.

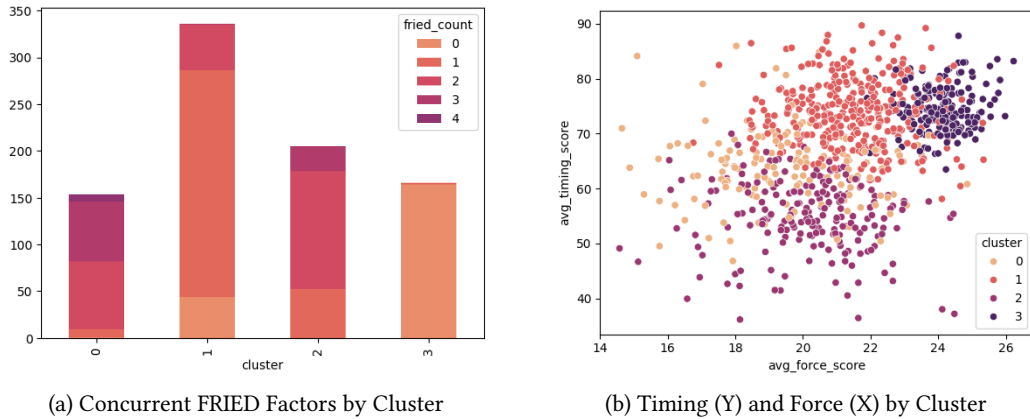
The entire assessment task is shared in the Data Modelling notebook. Here, I dedicate a small subsection to answering the previous questions for each of the developed models.

##### 4.4.4.1 *K-Prototypes Clustering*

This first model directly answers to objective **DM.G.01** and sub-objective **DM.G.01.1** of the project, related to the identification of logical groups of patients with similar therapeutic dispositions. Here, the main criterion under assessment is whether the results actually make sense from a profiling point-of-view (i.e.: if the assigned clusters can be segregated to showcase different distributions). To answer this, I performed a small exploratory analysis in which I aimed to assign a description to each of the four clusters, characterizing the average patient within that group (a profiling task). The profiles, slightly altered from the notebook section, are:

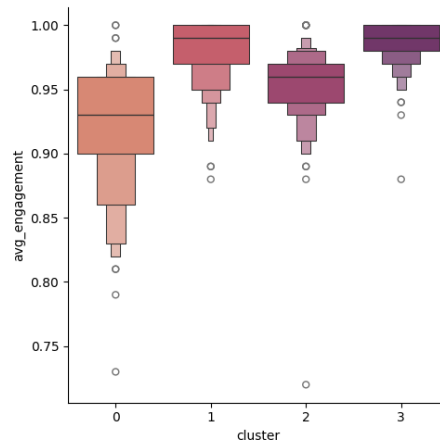
- Cluster 0 - High occurrence of simultaneous risk factors. These patients are in a less favorable physical state and tend to stay on the lower range of activity scores. Out of all groups, they show the least engagement, often failing to complete the ' suggested duration.
- Cluster 1 - Low incidence of simultaneous risk factors, but very likely to have at least one. This population is relatively healthy and their scores quite high. They have the second best adherence to the full program duration.
- Cluster 2 - Medium incidence of simultaneous risk factors, but never exempt of their presence (always at least one, more often still to have two). Shows similar scoring patterns to Cluster 0, but are clearly much more engaged, often completing at least 95% of the activity.
- Cluster 3 - These are patients that can, at this point, be considered healthy. The presence of even one risk factor is rare, and their scores are consistently higher. Patients in this group also show near perfect adherence to the activity's duration.

For context, the data visualizations contained in Fig. 11 are the basis for the profiling descriptions provided above. They represent the number of risk factors for frailty (Fig. 11a), the average scores achieved in activities (Fig. 11b), and the degree of engagement of patients (Fig. 11c), per cluster.



(a) Concurrent FRIED Factors by Cluster

(b) Timing (Y) and Force (X) by Cluster



(c) Engagement by Cluster

Figure 11: Cluster Profiling

#### 4.4.4.2 KNN Classifier

As mentioned in the test design, this is the only supervised method under assessment, with the respective ground truth labels being defined by the previous model, and whose task is to resolve the issue of classifying new observations into the most adequate preexisting cluster, concluding objectives **DM.G.02** and **DM.G.02.1** of the project. In this case, the assessment can be much less reliant on subjective interpretations, and my goal is to prioritize accuracy and performance by optimizing the K parameter, for the number of neighbors to use when calculating a new label. As specified in the test design, two tasks were carried out. First, the baseline accuracy of the method ( $K = 14$ ) was measured at

80.2%. To accompany this, the preprocessing steps and neighbor similarity calculations require just over half a second (0.502 seconds total) to produce a label. With this in mind, and the perspective that the model is **denied** if accuracy does not reach at least 85% at any point, I extensively tested all values for K in the interval [5, 86] (86 as a stopping point represents 10% of the entire dataset size, which is far too much for calculating a label, since quantity bias becomes a problem), and found that accuracy peaks at 88.9% when  $K = 31$ , as shown in Fig. 12a. Moreover, I found that the time spent for classification does not suffer great deviations with the increase / decrease of K, consistently staying around the half-second mark throughout the entire experiment (See Fig. 12b).

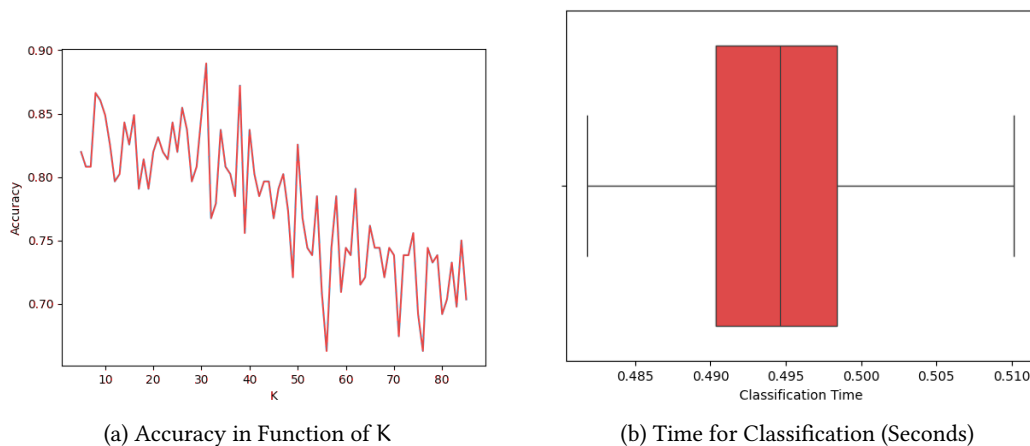


Figure 12: KNN Parameter Optimization Results

#### 4.4.4.3 Therapeutic Difficulty Reverse Engineering

This task, while directly related to a modelling goal (**DM.G.03**), would be better understood as an intermediate process that supports the upcoming model. Regardless, its assessment is not any less important, since the misunderstanding of its results may undermine the quality of the suggested activities. Additionally, this process is of great value to the business itself, since it reduces the subjectivity attached to assigning a therapeutic difficulty score to activities. Since the result of this process is a supporting file, the test design determines that its ease-of-understanding is the most important measure of assessment. Additionally, to avoid adding significant waiting times to other processes, the file loading times were also evaluated. Regarding measurable metrics, out of 1000 iterations, the average loading time was consistently measured under  $1/200^{\text{th}}$  of a second. Since this number is almost negligible, no additional file formats were considered, though this may be necessary in the future if the number of games increases significantly.

#### 4.4.4.4 *Recommendation Algorithm*

Internally, this singular system actually leverages two modelling techniques that work in unison to constitute a semi-intelligent algorithm, answering to objectives **DM.G.04** and **DM.G.04.1** simultaneously. Since this module foresees very frequent use and interpretation by therapists and patients alike, the test design I have devised is unquestionably insufficient for an accurate assessment of its quality, since it does not foresee the data collection from a clinical trial of the system, aimed at collecting feedback regarding the adequability of the recommendations. For more details about this shortcoming, please refer to Section [B.4](#) of Appendix [B](#).

Regardless, the test design states the importance of all matters related to ease-of-interpretation, performance and deployment to be considered. To achieve this, the usage of this tool was extensively demonstrated in the notebook, and the system is shown to be capable of producing first recommendations in 1.5 seconds, whereas subsequent recommendations are adjusted in less than  $1/10^{\text{th}}$  of a second, after collecting relevant feedback. Additionally, the functions supporting this model are modular and deal with data requirements dynamically, which is a design choice specifically made for the deployment of the model in [API](#) form.

#### 4.4.4.5 *Performance Analysis Tool*

In line with the previous model, the performance analysis tool is also meant to be used by therapeutic professionals, providing the functionality to quickly analyze a particular patient's evolution, and answering directly to objectives **DM.G.05** and **DM.G.05.1**. Similarly to the recommendation system, this model can only be fully assessed when it is put into practical use by GameVitae therapists, falling into the same situation of incomplete assessment, since I alone can not guarantee that the reports show information that is pertinent for professional analysis.

As with the other systems for therapeutic use, all subjective factors are equally important, hence a demonstration of the system's functionality was my go-to approach in the respective notebook section. With respect to performance, iterations were consistently completed in under 1 second, with the six underlying graphs also being saved with local files. To aid in interpreting the slope (trend) values, the module has a built-in function that customizes the output messages with small concluding statements about the data it presents. Moreover, necessary files are loaded dynamically when needed, and, in its current version, only one parameter is needed (the identifier of the patient), making the module ideal for its deployment in [API](#) form.

## 4.5 EVALUATION

The fifth and penultimate phase is used for evaluating the data mining results using the previously defined Business Criteria as a benchmark for success. In general, the main goal here is to assess whether the business can make use of the results (models and data findings) of the project. Table 15 contains the foreseen tasks, deliverables and related appendix chapters for this phase.

Generic Tasks	Deliverables	Appendices
Results	Result Assessment, Final Models	Section G.1 of Appendix G
Process Revision	Data Mining Process Review	Section G.2 of Appendix G
Next Steps	Possible Actions, Decisions	N/A

Table 15: Evaluation Breakdown

## 4.5.1 Results

To precede the first task in evaluating the success of my project, a clear definition of what is a result should be formally represented in the text. In data mining projects such as this one, there are two types of results:

- Each of the **models** obtained in the previous phase is a result.
- Any and all conclusions, inferences and **findings** drawn from the entire data mining process (phases 2 through 4) are results.

Results for my project are stated throughout its various sections and appendices, however, they are not documented in a way that is easily accessible and digestible by readers due to formatting choices and conformity with the CRISP-DM methodology. In this sense, the result dissemination plan that is envisioned for the very near-future will be solid contribution to making the results more accessible, and highlighting the significant findings of my project.

To provide a summarized version of the evaluation between the results and the proposed business goals, I have compiled Table 16, which serves as an overview for this subsection. In it, the business goals and criteria are reiterated and classified as either achieved or not achieved.

Business Objective	Criteria	Related Models	Achieved?
BO.01	Improve engagement by 15% Generate automatic activities	Activity Recommendation Model (U.02)	Yes (Sub.)
BO.02	Improve ratings to 4.5 Generate automatic activities Find motivating factors	Activity Recommendation Model (U.02)	Yes (Sub.)
BO.03	Daily performance reporting Define metrics for reporting	Performance Analysis Tool (U.03)	Yes (Sub.)
BO.04	Daily performance reporting	Performance Analysis Tool (U.03)	Yes
BO.05	Set-up time reduced by 90%	Activity Recommendation Model (U.02)	Yes
BO.06	Find clusters Discover efficient exercises	K-Prototypes Clustering (U.01) Activity Recommendation Model (U.02) Performance Analysis Tool (U.03)	Partly
BO.07	Reduce subjectivity of difficulty ratings	Therapeutic Difficulty Reverse Engineering (P.01)	Yes
BO.08	Generate automatic activities	Activity Recommendation Model (U.02)	Yes

Table 16: Business objective completion summary

At this stage, some of the proposed business objectives are hard to assess due to their subjectivity. Still, an attempt has been made to provide a definite answer to all goals, and thus arises the need to provide rationale for the answers in the table, as to properly fund the model approval decisions below. All models and processes were informally presented to the project manager for GameVitae in a functionality showcase, in which the business goals were constantly used as a metric for assessing the adequacy of the project results. This meeting was especially useful in solidifying the answers for the table. To keep this section short and direct, the main takeaways from this meeting, as well as the resulting rationale for defining success can be consulted in Section G.1 of Appendix G. Lastly, the previous information is sufficient to motivate the decision that **all models** are approved for deployment, albeit with the incentive to carry out testing phases for the better assessment of some of the criteria that are subjective as of now.

#### 4.5.2 Data Mining Process Revision

The second part of evaluation involves taking a second look at every CRISP-DM activity and decision made up until this point. For each phase, the following questions must be considered:

- Was the stage a contributor to the value of the final results?
- Are there ways of streamlining or improving this stage?
- Were there mistakes or failures in this stage? If so, how can they be avoided in the future?
- Were there any surprises (positive or negative)? How did they affect the project?

CRISP-DM Phase	Difficulties	Surprises	Alternate Decisions / Improvements	Valuable to results?
Business Understanding	Subjectivity in goal setting	Project maturity	Better prior investigation of project maturity Tighter scheduling	Yes
Data Understanding	Unfamiliarity with the medical context Difficulty in producing documentation	Delays due to additional researching phases	Improve the logical flow of exploration, documentation and reporting	Yes
Data Preparation	File "juggling" of many dataset versions	N/A	Better file organization planning	Yes
Data Modelling	Cyclic revisions made adjustments convoluted and messy Methodological shortcomings in model assessment	N/A	Anticipate model requirements, opening up schedule for better depth of assessment	Yes

Table 17: Data Mining revision summary

- Could there have been alternative strategies or decisions during this stage?

Following a similar procedure as before, the final considerations from this subsection can be found in tabular form (Table 17), while the entire revision process is present in Section G.2 of Appendix G. Overall, all data mining phases were completed successfully with no major issues that would put the quality of the project in question. Difficulties were faced and overcome in every stage of the process, but this also gave room for improving my critical thinking abilities and decision making.

#### 4.5.3 Next Steps

Upon compiling a summary of the previous subsections, I organized a short follow-up meeting with the project manager, with the sole finality of understanding whether the project can go ahead with deployment, or if the models require additional refinements to meet the business' expectations.

At this point, **all models were green-lit** for deployment by myself and the business, and the latter's reasoning is the following:

- The models meet with the business objectives and adjacent data mining objectives.
- The results of the data mining project (namely insights about the data) reveal patterns that have been adequately integrated in the model building methodology.
- The models are novel to the business and add value to its processes.

## 4.6 DEPLOYMENT

The final phase of the [CRISP-DM](#) methodology involves using the insights gained from the data mining project to elicit change in the organization. In this case, it also relates to documenting the procedure by which the project results will be integrated in the GameVitae platform. Particularly due to it being the last phase of the project, special care needs to be attributed to planning a smooth deployment, since the presence of major errors here can be disastrous. Table 18 contains the foreseen tasks, deliverables and related appendix chapters for this last phase.

Generic Tasks	Deliverables	Appendices
Plan Deployment	Deployment Plan, Contingency Plan	Appendix <a href="#">H</a>
Monitoring and Maintenance	Monitoring Plan	
Report	Final Report, Public Result Report, Auxiliary Presentations	N/A
Project Revision	Revision Process	N/A

Table 18: Deployment Breakdown

4.6.1 *Plan Deployment*

To ensure a thorough approach to the way results are to be deployed in the business, this first section will contain a general description of the deployment plan. In short, the following aspects will be accounted for:

- Summarizing the results
- Planning the model deployment procedure
- Devising a plan to share major findings with the business and the scientific community.
- Planning for contingencies or unexpected requirements

The result of this section is the deployment plan, which I have documented in the form of an Appendix at the end of this document (Appendix [H](#)), and the majority of actions and decisions for deployment are present in that file. Other choices or tasks that are unrelated to the deployment methodology can be consulted in the following subsections.

4.6.1.1 *Result Summary*

Boiling down the entire data mining project into a list of results is a useful way of assessing the impact of the project, and understanding the business aspects that will be influenced by the models. I have decided to separate results into two major categories: **Findings** are particularities revealed throughout the data mining process, including results from specific models. **Models** are finalized products, and the vehicle by which the business can achieve the goals they defined. Note that all models are deployable results, and the majority of findings will be accessible as a byproduct of model deployment. Below is a table of project results (Table 19), categorized by type and classified by how they will be deployed in the business. Results that stem from the usage of models (such as reduced activity setup times) and are not directly deployable were omitted from this table.

Result	Type	Deployment
Patient clustering function	Model (U.01)	<a href="#">API</a>
Assignment of cluster labels to new patients	Model (S.01)	<a href="#">API</a>
Normalization of activity targets	Model (P.01)	<a href="#">API</a>
Automatic activity configuration	Model (U.02)	<a href="#">API</a>
Streamlined patient performance analyses	Model (U.03)	<a href="#">API</a>
Cluster profiles	Finding	Presentation
Discovery of risk factors with low incidence on the presence of frailty	Finding	Presentation
Discovery of high-impact factors on the gravity of frailty	Finding	Presentation
Discovery of preferred games for patients with varying degrees of frailty	Finding	Presentation
Discovery of optimal activity settings for particular patient groups	Finding	Presentation

Table 19: Summary of deployable project results

Evidently, there are two distinct methods I intend to use to deploy the results of this project. While the models and their respective results will compose an application interface ([API](#)) specific to the GameVitae platform, the remaining findings will be directly presented

to a board of Techframe officials, with the goal of defining further directions and follow-up interactions. Both of these processes are detailed in Appendix H.

#### 4.6.1.2 *Deployment and Integration*

I have decided to approach this task slightly differently to what is proposed in a barebones CRISP-DM implementation, by combining every individual model's deployment strategy into a single, standardized process. The reasoning for this is extensive, and mostly relates to avoiding redundancy, saving time in deployment, and simplifying the integration process as much as possible, leaving fewer room for errors. To keep this section concise, the entire deployment and integration plan for the results of the project is documented in Appendix H.

#### 4.6.1.3 *Communication Plan*

Despite having similar names, this task is distinct to the result communication plan outlined in Appendix B. Instead of sharing the results publicly, here, the dissemination of findings is reduced to only the business, and thus the scale is much smaller. In my project, there is the particularity of me being an integrant of the business with close connections to the remaining coworkers, which simplifies this task immensely.

The communication of results within the business will be achieved through an on-line meeting with all team members specified in section 4.1.1.1. This meeting will be accompanied by a presentation that follows the following agenda:

- **Scope and Objectives** Present a short critical analysis of the state-of-art, introduce the context of my project, and list the business objectives.
- **Data Findings** Sumrise the major findings of the data understanding phase and the state of the data quality before and after the preparation phase.
- **Models** Describe each of the models and their purpose, interactively showcase their functioning.
- **Deployment and Maintenance** Discuss the deployment method, the model life-cycle and the data that will be monitored throughout deployment.

After the introduction of each topic, the meeting will take form of a discussion in which the coworker's ideas and feedback are to be collected as indicators for a follow-up project or improvements.

#### 4.6.1.4 *Contingencies*

Similar to the risk analysis performed for the data mining project itself, this task involves expanding the deployment plan with contingencies for unexpected situations or additional requirements that may present themselves at any point during deployment of the [API](#). I see this task as complementary to the deployment plan, and have included its development in Section [H.1.1.2](#) of Appendix [H](#).

#### 4.6.2 *Monitoring and Maintenance*

After deploying the results, there will be a time period during which the business will need to be closely monitored for anomalies, model performance, and other factors that may cause the results to deviate from the expected. Naturally, this won't be a permanent situation, since this task also deals with setting "expiration" criteria for the models, after which they need to be revised, updated, or completely replaced. I have considered this entire step as an add-on to Section [4.6.1.2](#), which is included in Appendix [H](#).

#### 4.6.3 *Report*

To close out deployment, one or more reports are expected for documenting the entire process from start to finish in great detail. For my project, there are two major documents that pertain to the reporting of the data mining project and the results:

- This thesis, as the final project report
- The technical article specified in the communication plan ([B.3](#) of Appendix [B](#)).

#### 4.6.4 *Project Revision*

Finally, the project itself needs to be revised. This task no longer has an effect on the success of the project, but is equally important as a way to document how it went and what was learned. In this revision, I saw it necessary to have both a self-evaluation and a third-party evaluation pertaining to the different aspects of the project's success, allowing me to compare similarities and differences in the way the project is perceived by different people related to it. The revision was centered around finding answers to the following questions:

- What are the overall impressions of the project? Does it bring value to the business?
- What was learned about the data mining process? And what about the data?
- Which parts of the project went well? Where did difficulties arise?

#### 4.6.4.1 *Business Interview*

I approached the external revision by conducting a short interview with the business decision makers (in particular, the project manager and the CCO assigned to GameVitae) to ask for their overall impressions on the project's success and impactfulness. In general, I found the business to be favorable towards the quality of project results and the thoroughness of the data mining process. From a data standpoint, my project solidifies the business' preexisting beliefs that the data obtained from gamified exercises is valuable and can be explored for insightful information about underlying health conditions. The marketable aspect of the platform was also discussed, with the communications officer mentioning that the project positions the platform from a commercial standpoint, highlighting advantages and deficiencies of the market, and relating the technologies used in the context of my project to the current trends of commercial buzzwords (such as AI-Powered, Smart, and others).

Moreover, the decisions makers have unanimously agreed on the importance of my project as a validating process of the GameVitae platform from a clinical point of view, encouraging a series of future projects and developments related to the integration of more devices, biometric sensors and gamified exercises, which in turn can bring in data partners, encourage medical studies trials, and contribute to the state-of-art regarding gamification in healthcare.

#### 4.6.4.2 *Self Reflection*

To finalize the revision, I carried out a self-reflection to understand my perspective on the project's success and impact. Alongside the previous section, this task can be seen as a conclusive statement dedicated to the [CRISP-DM](#) methodology.

First and foremost, I noticed a stark difference in the thoroughness and pacing of the project in comparison with my previous short projects using the [CRISP-DM](#) methodology. In fact, my past experience did not adequately prepare me for the number of needed revisits to past CRISP phases, and thus the first significant difficulty I faced was organizing the timeline and list of tasks for the initial stages. The fact that this project carries real repercussions and its success brings tangible value to a business was, at first, an added

factor of stress, which alleviated gradually as my grasp on the project tightened and my ideas solidified into concrete activities.

One added benefit of the business understanding phase was the frequent and proactive communication with the related company, which helped form good habits and avoid misunderstandings that could aggravate into serious issues in the later stages. Due to my inexperience with the pacing of business operations, this first phase took much longer than initially anticipated, and while its results are valuable beyond doubt, some of the scheduling for the project was slightly displaced. Despite these difficulties, I now have the confidence and means to accelerate this task in future projects, allowing more time of involvement with the data.

Speaking of which, I'm particularly satisfied with the way I approached data-forward phases, with phases 2 (Data Understanding) through 4 (Data Modelling) being my personal favourites. The one flaw that is constantly reiterated in my mind and reflected throughout the project is the lack of proper model testing prior to deployment. While I am conscious that this was simply impossible due to tight scheduling, I also believe that future GameVitaе projects must be carefully planned in terms of practical model testing, since acting upon insufficiently tested clinical models can be costly. To counteract this issue, not only were additional testing phases proposed for post-deployment, but the actual deployment stage was expanded to be as complete as possible, with feedback collection and decision making stages that will approve or disapprove model usage in short-term.

Above all, I see this data mining project as an invaluable learning experience and a success from a business perspective, as well as a source of practical knowledge which I will carry over to future projects. I believe both by technical skills and my critical thinking have significantly improved since the beginning of the project, and my newly-acquired understanding of the inner workings of data collection, business procedures and model deployment in real-world scenarios will be of great value in upcoming projects. Each difficulty and setback I have faced was a lesson to promote better preparation, which is one of my main takeaways from this project into the future.



## CONCLUSIONS AND FUTURE WORK

---

The practice of incorporating design choices typically associated with video games into otherwise mundane activities is called gamification, and it has been shown to improve user engagement, information retention, and enjoyment when employed in a variety of fields of industry. Despite the logical compatibility between games and health, the combination of the two is still weighed down by a lack of standardized methodologies, efficient gamification strategies, and inclusive design choices, deterring its wide adoption and significantly slowing down research.

In this thesis, I propose the employment of data mining as a way to potentiate knowledge discovery in gamified exercise platforms, leading to a better understanding of the factors that affect the targeted medical conditions, and enabling an objective-driven approach to providing patients with more enjoyable activities, and therapists with better monitoring tools.

In specific, I have succeeded in utilizing the CRISP-DM methodology to carry out an extensive data mining project on GameVitae, a gamified exercise platform that currently targets the maintenance of frailty syndrome in older adults through the integration of a special grip controller capable of measuring grip force. This project resulted in the proposal of five distinct data models, aimed at improving the quality and efficiency of the exercise sessions through gradual feedback collection and improved monitoring of patient scores. The models were successful in forming logical patient groups based on similarities in physical ability, generating recommended activities with respect to a patient's personal preferences, reducing the subjectivity of setting activity goals, and providing therapists with tools for quickly assessing patient evolution.

Intermediate results achieved through the exploration of patient health and activity data have also helped tremendously in defining the scope for the models and guiding the decision making process towards achieving as many business objectives as possible. An example of this can be seen in the way a certain game is preferred among a group of patients, which indicates a clue as to whether newer patients with similarly disposed health conditions could also prefer that particular game. This was reflected in the recommendation system, through the analysis of neighbors for generating relevant activities based on group similarities.

The main contribution of my work can be described as a methodological proof that following an objective-driven approach to data mining can predetermine better success conditions for the quality and efficiency of gamification, specially in business areas in which its adoption has been slower due to social-cultural factors, with the healthcare being a prime example with good natural compatibility but challenging practical application. With this work, I strived to directly counteract the deficiencies identified in the state-of-art, pertaining to lackluster standardization, inefficient game design strategies, and introduce a conscious approach to developing systems that place the patient's feedback and preferences first.

Despite having met or surpassed most of the the proposed data mining goals, the underlying methods behind the testing of some of the models lack thoroughness in terms of assessing scalability and usability, due to restrictions in time and project scope. This opens up a clear direction for the future, with practical testing of the models through clinical trials for understanding the efficiency and practicality of the solutions in a real-use scenario. Moreover, the clinical adequacy of the models needs to be reviewed to ensure therapeutic benefits. More specifically, the questionnaire presented by the feedback collection system may not yet capture the full range of feelings of the patients, and the metrics used in the performance analysis tool could still be insufficient to allow a proper assessment of a patient's evolution.

## BIBLIOGRAPHY

---

- Abu-dawood, Sumayah (July 2016). «The Cognitive and The Social Motivational Affordances of Gamification in E-Learning Environment». In: DOI: [10.1109/ICALT.2016.126](https://doi.org/10.1109/ICALT.2016.126).
- Allam, Ahmed et al. (Jan. 2015). «The Effect of Social Support Features and Gamification on a Web-Based Intervention for Rheumatoid Arthritis Patients: Randomized Controlled Trial». EN. In: *Journal of Medical Internet Research* 17.1. Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada, e3510. DOI: [10.2196/jmir.3510](https://doi.org/10.2196/jmir.3510). URL: <https://www.jmir.org/2015/1/e14> (visited on 12/13/2022).
- Anabitar-García, Francisco et al. (2021). «Early diagnosis of frailty: Technological and non-intrusive devices for clinical detection». In: *Ageing Research Reviews* 70, p. 101399. ISSN: 1568-1637. DOI: <https://doi.org/10.1016/j.arr.2021.101399>. URL: <https://www.sciencedirect.com/science/article/pii/S156816372100146X>.
- Andreoli, A et al. (2009). «Body composition assessment by dual-energy X-ray absorptiometry (DXA)». In: *La radiologia medica* 114.2, pp. 286–300.
- Armañanzas, Rubén et al. (July 2013). «Unveiling relevant non-motor Parkinson's disease severity symptoms using a machine learning approach». In: *Artificial Intelligence in Medicine* 58.3, pp. 195–202. ISSN: 0933-3657. DOI: [10.1016/j.artmed.2013.04.002](https://doi.org/10.1016/j.artmed.2013.04.002). URL: <https://doi.org/10.1016/j.artmed.2013.04.002> (visited on 03/10/2023).
- Ayed, Ines et al. (2019). «Vision-based serious games and virtual reality systems for motor rehabilitation: A review geared toward a research methodology». In: *International journal of medical informatics* 131, p. 103909.
- Baptista, Gonçalo and Tiago Oliveira (2019). «Gamification and serious games: A literature meta-analysis and integrative model». In: *Computers in human behavior* 92, pp. 306–315.
- Bellazzi, Riccardo and Blaz Zupan (Feb. 2008). «Predictive data mining in clinical medicine: current issues and guidelines». eng. In: *International Journal of Medical Informatics* 77.2, pp. 81–97. ISSN: 1386-5056. DOI: [10.1016/j.ijmedinf.2006.11.006](https://doi.org/10.1016/j.ijmedinf.2006.11.006).

- Bhatti, Uzair Aslam et al. (2019). «Recommendation system using feature extraction and pattern recognition in clinical care systems». In: *Enterprise information systems* 13.3, pp. 329–351.
- Bonsignore, Elizabeth et al. (2013). «Alternate reality games as platforms for practicing 21st-century literacies». In.
- Borecki, Lukasz, Katarzyna Tolstych, and Mieczyslaw Pokorski (2013). «Computer Games and Fine Motor Skills». en. In: *Respiratory Regulation - Clinical Advances*. Ed. by Mieczyslaw Pokorski. Advances in Experimental Medicine and Biology. Dordrecht: Springer Netherlands, pp. 343–348. ISBN: 978-94-007-4546-9. DOI: [10.1007/978-94-007-4546-943](https://doi.org/10.1007/978-94-007-4546-943).
- Brauner, Philipp et al. (Jan. 2013). «Increase Physical Fitness and Create Health Awareness through Exergames and Gamification. The Role of Individual Factors, Motivation and Acceptance». In: pp. 349–362.
- Bret ,T. (2008). *My Coverage of Lobby of the Social Gaming Summit*. Blog.
- Cetin, Mehmet and Ozgür Kokalan (2021). «A multilevel analysis of the effects of indoor activities on psychological wellbeing during COVID-19 pandemic». In: *Annals of Psychology* 37.3, pp. 500–507.
- Chang, Chun-Lang and Chih-Hao Chen (Mar. 2009). «Applying decision tree and neural network to increase quality of dermatologic diagnosis». en. In: *Expert Systems with Applications* 36.2, Part 2, pp. 4035–4041. ISSN: 0957-4174. DOI: [10.1016/j.eswa.2008.03.007](https://doi.org/10.1016/j.eswa.2008.03.007). URL: <https://www.sciencedirect.com/science/article/pii/S0957417408002005> (visited on 03/22/2023).
- Char, Danton S, Nigam H Shah, and David Magnus (2018). «Implementing machine learning in health care—addressing ethical challenges». In: *The New England journal of medicine* 378.11, p. 981.
- Chen, Yu and Pearl Pu (2014). «HealthyTogether: Exploring Social Incentives for Mobile Fitness Applications». In: *Proceedings of the Second International Symposium of Chinese CHI*. Chinese CHI '14. event-place: Toronto, Ontario, Canada. New York, NY, USA: Association for Computing Machinery, pp. 25–34. ISBN: 978-1-4503-2876-0. DOI: [10.1145/2592235.2592240](https://doi.org/10.1145/2592235.2592240). URL: <https://doi.org/10.1145/2592235.2592240>.
- Cheong, C., F. Cheong, and J. Filippou (2013). «Quick Quiz: A Gamified Approach for Enhancing Learning». In: URL: <https://www.semanticscholar.org/paper/Quick-Quiz%3A-A-Gamified-Approach-for-Enhancing-Cheong-Cheong/438e976185003c4ff0658a4313922aaf86154> (visited on 12/13/2022).
- Chow, Ching Yue et al. (2020). «Can games change children’s eating behaviour? A review of gamification and serious games». In: *Food Quality and Preference* 80, p. 103823.

- Commins, John (2010). «Nurses say distractions cut bedside time by 25%». In: *Health Leaders*.
- Coonradt, C.A. and L. Nelson (1985). *The Game of Work: How to Enjoy Work as Much as Play*. Shadow Mountain. ISBN: 978-0-87747-771-6. URL: <https://books.google.co.in/books?id=6dZh%5C40Pu7sC>.
- Curiac, Daniel-Ioan et al. (June 2009). «Bayesian network model for diagnosis of psychiatric diseases». In: *Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces*. ISSN: 1330-1012, pp. 61–66. DOI: [10.1109/ITI.2009.5196055](https://doi.org/10.1109/ITI.2009.5196055).
- Das, Resul, Ibrahim Turkoglu, and Abdulkadir Sengur (May 2009). «Effective diagnosis of heart disease through neural networks ensembles». en. In: *Expert Systems with Applications* 36.4, pp. 7675–7680. ISSN: 09574174. DOI: [10.1016/j.eswa.2008.09.013](https://doi.org/10.1016/j.eswa.2008.09.013). URL: <https://linkinghub.elsevier.com/retrieve/pii/S095741740800657X> (visited on 03/23/2023).
- Davenport, Thomas and Ravi Kalakota (2019). «The potential for artificial intelligence in healthcare». In: *Future healthcare journal* 6.2, p. 94.
- Dayhoff, Nancy E et al. (1998). *Balance and muscle strength as predictors of frailty among older adults*.
- Denny, Paul (2013). «The effect of virtual achievements on student engagement». In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. New York, NY, USA: Association for Computing Machinery, pp. 763–772. ISBN: 978-1-4503-1899-0. DOI: [10.1145/2470654.2470763](https://doi.org/10.1145/2470654.2470763). URL: <https://doi.org/10.1145/2470654.2470763> (visited on 12/13/2022).
- Derksen, ME et al. (2020). «Serious games for smoking prevention and cessation: A systematic review of game elements and game effects». In: *Journal of the American Medical Informatics Association* 27.5, pp. 818–833.
- Deterding, Sebastian (2011). «From game design elements to gamefulness: Defining "Gamification"». In: *Proceedings of the 15th international Academic Mindtrek Conference* Blackwell Encyclopedia of Sociology, Malden, pp. 9–15. DOI: [doi: 10.1145/2181037.2181040](https://doi.org/10.1145/2181037.2181040).
- Dias, Jéssica David, Aline Natalia Domingues, et al. (2018). «Serious games as an educational strategy to control childhood obesity: a systematic literature review». In: *Revista latino-americana de enfermagem* 26, e3036.
- Dias, Jéssica David, Chris Mayara dos Santos Tibes, et al. (2017). «Use of serious games for coping with childhood obesity: integrative literature review». In: *Texto & Contexto-Enfermagem* 26.

- Doarn, Charles R (2019). «Advanced Technologies: Paperless Hospital, the Cost and the Benefits». In: *The Modern Hospital: Patients Centered, Disease Based, Research Oriented, Technology Driven*, pp. 145–155.
- Domínguez, Adrián et al. (Apr. 2013). «Gamifying learning experiences: Practical implications and outcomes». en. In: *Computers & Education* 63, pp. 380–392. ISSN: 0360-1315. DOI: [10.1016/j.compedu.2012.12.020](https://doi.org/10.1016/j.compedu.2012.12.020). URL: <https://www.sciencedirect.com/science/article/pii/S0360131513000031> (visited on 12/13/2022).
- Dong, Tao et al. (May 2012). «Discovery-based games for learning software». In: pp. 2083–2086. DOI: [10.1145/2207676.2208358](https://doi.org/10.1145/2207676.2208358).
- Elbattah, Mahmoud and Owen Molloy (2017). «Clustering-aided approach for predicting patient outcomes with application to elderly healthcare in Ireland». In: *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*.
- Er, Orhan, Nejat Yumusak, and Feyzullah Temurtas (2010). «Chest diseases diagnosis using artificial neural networks». In: *Expert Systems with Applications* 37.12, p. 7648. ISSN: 0957-4174. URL: [https://www.academia.edu/4569693/Chest\\_diseasesdiagnosisusingartificialneuralnetworks](https://www.academia.edu/4569693/Chest_diseasesdiagnosisusingartificialneuralnetworks) (visited on 03/23/2023).
- ESA (2023). *Video Games Remain America’s Favorite Pastime With More Than 212 Million Americans Playing Regularly*. Website. URL: <https://www.theesa.com/news/video-games-remain-americas-favorite-pastime-with-more-than-212-million-americans-playing-regularly/>.
- Eurostat (2018). *Eurostat*. Website. URL: <https://ecas.ec.europa.eu/>.
- Fang, Ruogu et al. (2016). «Computational health informatics in the big data age: a survey». In: *ACM Computing Surveys (CSUR)* 49.1, pp. 1–36.
- Farzan, Rosta and Peter Brusilovsky (Jan. 2011). «Encouraging user participation in a course recommender system: An impact on user behavior». In: *Computers in Human Behavior* 27.1, pp. 276–284. ISSN: 0747-5632. DOI: [10.1016/j.chb.2010.08.005](https://doi.org/10.1016/j.chb.2010.08.005). URL: <https://doi.org/10.1016/j.chb.2010.08.005> (visited on 12/13/2022).
- Farzan, Rosta, Joan M DiMicco, et al. (n.d.). «When the experiment is over: Deploying an incentive system to all the users». en. In: ().
- Farzan, Rosta, Joan M. DiMicco, et al. (2008). «Results from deploying a participation incentive mechanism within the enterprise». en. In: *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI ’08*. Florence, Italy: ACM Press, p. 563. ISBN: 978-1-60558-011-1. DOI: [10.1145/1357054.1357145](https://doi.org/10.1145/1357054.1357145). URL: <http://portal.acm.org/citation.cfm?doid=1357054.1357145> (visited on 12/13/2022).

- Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth (Mar. 1996). «From Data Mining to Knowledge Discovery in Databases». en. In: *AI Magazine* 17.3. Number: 3, pp. 37–37. ISSN: 2371-9621. DOI: [10.1609/aimag.v17i3.1230](https://doi.org/10.1609/aimag.v17i3.1230). URL: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1230> (visited on 03/08/2023).
- Fernandes, João (2022). «Gamificação Gripwise». In.
- Fried, Linda P et al. (2001). «Frailty in older adults: evidence for a phenotype». In: *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 56.3, pp. M146–M157.
- Fuente-Tomas, Lorena de la et al. (2019). «Classification of patients with bipolar disorder using k-means clustering». In: *PloS one* 14.1, e0210314.
- Fuller, NJ et al. (1999). «Predicting composition of leg sections with anthropometry and bioelectrical impedance analysis, using magnetic resonance imaging as reference». In: *Clinical Science* 96.6, pp. 647–657.
- Gaddey, Heidi L and Kathryn Holder (2014). «Unintentional weight loss in older adults». In: *American family physician* 89.9, pp. 718–722.
- Grant, Richard W et al. (2020). «Use of latent class analysis and k-means clustering to identify complex patient profiles». In: *JAMA network open* 3.12, e2029068–e2029068.
- Gunasundari, S. and S. Baskar (Dec. 2009). «Application of Artificial Neural Network in identification of lung diseases». In: *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, pp. 1441–1444. DOI: [10.1109/NABIC.2009.5393702](https://doi.org/10.1109/NABIC.2009.5393702).
- Ha, Park Jung et al. (2020). «Sedentary Lifestyle: Overview of Updated Evidence of Potential Health Risks». In: *Korean J Fam Med* 41.6, pp. 365–373. DOI: [10.4082/kjfm.20.0165](https://doi.org/10.4082/kjfm.20.0165). eprint: <http://www.kjfm.or.kr/journal/view.php?number=4549>. URL: <http://www.kjfm.or.kr/journal/view.php?number=4549>.
- Haefeli, Mathias and Achim Elfering (2006). «Pain assessment». In: *European spine journal* 15, S17–S24.
- Hakulinen, Lasse, Tapio Auvinen, and Ari Korhonen (Feb. 2015). «The Effect of Achievement Badges on Students' Behavior: An Empirical Study in a University-Level Computer Science Course». In: *International Journal of Emerging Technologies in Learning (iJET)* 10. DOI: [10.3991/ijet.v10i1.4221](https://doi.org/10.3991/ijet.v10i1.4221).
- Hamari, Juho (2011). «Framework for designing and evaluating game achievements». In: *Proceedings of Digra 2011 Conference Think Design Play*, pp. 14–17.
- (July 2013). «Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service». en. In: *Electronic Commerce Research and Applications. Social Commerce- Part 2* 12.4, pp. 236–

245. ISSN: 1567-4223. DOI: [10 . 1016 / j . elerap . 2013 . 01 . 004](https://doi.org/10.1016/j.elerap.2013.01.004). URL: <https://www.sciencedirect.com/science/article/pii/S1567422313000112> (visited on 12/13/2022).
- Hamari, Juho and Jonna Koivisto (2015). «Working out for likes”: An empirical study on social influence in exercise gamification». In: *Computers in Human Behavior* 50, pp. 333–347. ISSN: 0747-5632. DOI: <https://doi.org/10.1016/j.chb.2015.04.018>. URL: <https://www.sciencedirect.com/science/article/pii/S0747563215003052>.
- Haraty, Ramzi A., Mohamad Dimishkieh, and Mehedi Masud (June 2015). «An Enhanced  $k$ -Means Clustering Algorithm for Pattern Discovery in Healthcare Data». en. In: *International Journal of Distributed Sensor Networks* 11.6, p. 615740. ISSN: 1550-1477, 1550-1477. DOI: [10 . 1155 / 2015 / 615740](https://doi.org/10.1155/2015/615740). URL: <http://journals.sagepub.com/doi/10.1155/2015/615740> (visited on 03/08/2023).
- Heijden, Hans van der (2004). «User Acceptance of Hedonic Information Systems». In: *MIS Quarterly* 28.4, pp. 695–704. ISSN: 02767783. URL: <http://www.jstor.org/stable/25148660> (visited on 11/21/2022).
- Hillerman, Tiago et al. (2017). «Applying clustering and AHP methods for evaluating suspect healthcare claims». In: *Journal of computational science* 19, pp. 97–111.
- Hodgetts, Christopher J. et al. (Nov. 2021). «Shoulder pain prevalence by age and within occupational groups: a systematic review». In: *Archives of Physiotherapy* 11.1, p. 24. ISSN: 2057-0082. DOI: [10 . 1186 / s40945 - 021 - 00119 - w](https://doi.org/10.1186/s40945-021-00119-w). URL: <https://doi.org/10.1186/s40945-021-00119-w> (visited on 12/21/2022).
- Hoffmann, Alexandra, Corinna A Christmann, Gabriele Bleser, et al. (2017). «Gamification in stress management apps: a critical app review». In: *JMIR serious games* 5.2, e7216.
- IBM (2023). *What is Data Mining? – IBM*. en-us. URL: <https://www.ibm.com/topics/data-mining> (visited on 02/26/2023).
- J. Clement (Oct. 2022). *U.S. video gaming audiences 2022, by age group*. en. URL: <https://www.statista.com/statistics/189582/age-of-us-video-game-players/> (visited on 12/21/2022).
- J. Hamari (2019). «Gamification». In: *Blackwell Pub Blackwell Encyclopedia of Sociology*, Malden, pp. 1–3. DOI: <https://doi.org/10.1002/9781405165518.wbeos1321>.
- Jalink, Maarten B et al. (2014). «Nintendo related injuries and other problems: review». In: *BMJ* 349. Publisher: BMJ Publishing Group Ltd. DOI: [10 . 1136 / bmj . g7267](https://doi.org/10.1136/bmj.g7267). URL: <https://www.bmj.com/content/349/bmj.g7267>.
- Jani, Bhautesh Dinesh et al. (2021). «Association between patterns of alcohol consumption (beverage type, frequency and consumption with food) and risk of adverse health outcomes: a prospective cohort study». In: *BMC medicine* 19.1, pp. 1–14.

- Jen, Chih-Hung et al. (2012). «Application of classification techniques on development an early-warning system for chronic illnesses». In: *Expert Systems with Applications: An International Journal* 39.10, pp. 8852–8858. ISSN: 0957-4174. DOI: [10.1016/j.eswa.2012.02.004](https://doi.org/10.1016/j.eswa.2012.02.004). URL: <https://doi.org/10.1016/j.eswa.2012.02.004> (visited on 03/10/2023).
- Jin, Roger, Alexander Pillozzi, and Xudong Huang (2020). «Current cognition tests, potential virtual reality applications, and serious games in cognitive assessment and non-pharmacological therapy for neurocognitive disorders». In: *Journal of Clinical Medicine* 9.10, p. 3287.
- Jothi, Neesha, Nur'Aini Abdul Rashid, and Wahidah Husain (Jan. 2015). «Data Mining in Healthcare – A Review». en. In: *Procedia Computer Science. The Third Information Systems International Conference 2015* 72, pp. 306–313. ISSN: 1877-0509. DOI: [10.1016/j.procs.2015.12.145](https://www.sciencedirect.com/science/article/pii/S1877050915036066). URL: <https://www.sciencedirect.com/science/article/pii/S1877050915036066> (visited on 02/26/2023).
- Kerfoot, B. Price et al. (Oct. 2012). «An online spaced-education game to teach and assess medical students: a multi-institutional prospective trial». eng. In: *Academic Medicine: Journal of the Association of American Medical Colleges* 87.10, pp. 1443–1449. ISSN: 1938-808X. DOI: [10.1097/ACM.0b013e318267743a](https://doi.org/10.1097/ACM.0b013e318267743a).
- Khan, Muhammad Umer et al. (Aug. 2008). «Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare». In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. ISSN: 1558-4615, pp. 5148–5151. DOI: [10.1109/IEMBS.2008.4650373](https://doi.org/10.1109/IEMBS.2008.4650373).
- Kim, Jong-Hun et al. (2009). «Design of diet recommendation system for healthcare service based on user information». In: *2009 Fourth International Conference on Computer Sciences and Convergence Information Technology*. IEEE, pp. 516–518.
- Kıskaç, Muharrem et al. (2022). «What is the optimal body mass index range for older adults?» In: *Annals of geriatric medicine and research* 26.1, p. 49.
- Ko, F and Jeremy D Walston (2013). «What are the special needs of patients with frailty». In: *Evidence-based practice of palliative medicine. Philadelphia, PA: Elsevier Saunders*, pp. 371–375.
- Lau, Ho Ming et al. (2017). «Serious games for mental health: are they accessible, feasible, and effective? A systematic review and meta-analysis». In: *Frontiers in psychiatry* 7, p. 209.
- Lee, Su-In et al. (2018). «A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia». In: *Nature communications* 9.1, p. 42.
- Lenihan, D. (2012). «Health Games: A Key Component for the Evolution of Wellness Programs». In: *Games for Health Journal* 1 (3), pp. 233–235.

- Lentelink, Stefan et al. (May 2013). «Healthy weight game!: Lose weight together». In: pp. 1–8. DOI: [10.1109/SeGAH.2013.6665311](https://doi.org/10.1109/SeGAH.2013.6665311).
- Lima, Danielle Pessoa et al. (2020). «Clinical correlates of sarcopenia and falls in Parkinson’s disease». In: *PLoS One* 15.3, e0227238.
- Lin, Dana T. et al. (Jan. 2015). «Validity evidence for Surgical Improvement of Clinical Knowledge Ops: a novel gaming platform to assess surgical decision making». eng. In: *American Journal of Surgery* 209.1, pp. 79–85. ISSN: 1879-1883. DOI: [10.1016/j.amjsurg.2014.08.033](https://doi.org/10.1016/j.amjsurg.2014.08.033).
- Liu, Bo et al. (Mar. 2013). «SVDD-based outlier detection on uncertain data». en. In: *Knowledge and Information Systems* 34.3, pp. 597–618. ISSN: 0219-3116. DOI: [10.1007/s10115-012-0484-y](https://doi.org/10.1007/s10115-012-0484-y). URL: <https://doi.org/10.1007/s10115-012-0484-y> (visited on 03/02/2023).
- Liu, Kevin F.R. and Che-Fan Lu (Aug. 2009). «BBN-Based Decision Support for Health Risk Analysis». In: *2009 Fifth International Joint Conference on INC, IMS and IDC*, pp. 696–702. DOI: [10.1109/NCM.2009.187](https://doi.org/10.1109/NCM.2009.187).
- Liu, Qi and Miklos Vasarhelyi (2013). «Healthcare fraud detection: A survey and a clustering model incorporating geo-location information». In: *29th world continuous auditing and reporting symposium (29WCARS), Brisbane, Australia*.
- Liu, Yusha et al. (Apr. 2020). «Abstract 11: Surface Electromyography-Based Gamification Therapy For Rehabilitation Of Upper Extremity Weakness: An Acceptability Study». en-US. In: *Plastic and Reconstructive Surgery – Global Open* 8.4S, p. 8. DOI: [10.1097/01.GOX.0000667108.93983.47](https://doi.org/10.1097/01.GOX.0000667108.93983.47). URL: [https://journals.lww.com/prsgo/fulltext/2020/04001/abstract11\\_surfaceelectromyographybased.12.aspx](https://journals.lww.com/prsgo/fulltext/2020/04001/abstract11_surfaceelectromyographybased.12.aspx) (visited on 12/13/2022).
- LLP, Deloitte (2015). «From brawn to brains: the impact of technology on jobs in the UK». In.
- Malone, Thomas (Jan. 1981). «What Makes Things Fun to Learn? A Study of Intrinsically Motivating Computer Games». In: *Pipeline* 6.
- Mamiya, Hiroshi et al. (Feb. 2015). «Towards probabilistic decision support in public health practice: Predicting recent transmission of tuberculosis from patient attributes». en. In: *Journal of Biomedical Informatics* 53, pp. 237–242. ISSN: 1532-0464. DOI: [10.1016/j.jbi.2014.11.006](https://doi.org/10.1016/j.jbi.2014.11.006). URL: <https://www.sciencedirect.com/science/article/pii/S1532046414002366> (visited on 03/15/2023).
- Manyika, James et al. (2017). «A future that works: AI, automation, employment, and productivity». In: *McKinsey Global Institute Research, Tech. Rep* 60, pp. 1–135.
- Mendiratta, Priya and Rafay Latif (2020). «Clinical frailty scale». In.
- Moon, Sung Seek et al. (Jan. 2012). «Decision tree models for characterizing smoking patterns of older adults». In: *Expert Systems with Applications* 39.1, pp. 445–451.

- ISSN: 0957-4174. DOI: [10.1016/j.eswa.2011.07.035](https://doi.org/10.1016/j.eswa.2011.07.035). URL: <http://www.scopus.com/inward/record.url?scp=81855218753&partnerID=8YFLogxK> (visited on 03/22/2023).
- Morley, John E, TK Malmstrom, and DK Miller (2012). «A simple frailty questionnaire (FRAIL) predicts outcomes in middle aged African Americans». In: *The journal of nutrition, health & aging* 16, pp. 601–608.
- Munson, Sean and Sunny Consolvo (May 2012). «Exploring Goal-setting, Rewards, Self-monitoring, and Sharing to Motivate Physical Activity». In: *2012 6th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, PervasiveHealth 2012*. DOI: [10.4108/icst.pervasivehealth.2012.248691](https://doi.org/10.4108/icst.pervasivehealth.2012.248691).
- Nevin, Christa R. et al. (Dec. 2014). «Gamification as a tool for enhancing graduate medical education». eng. In: *Postgraduate Medical Journal* 90.1070, pp. 685–693. ISSN: 1469-0756. DOI: [10.1136/postgradmedj-2013-132486](https://doi.org/10.1136/postgradmedj-2013-132486).
- Özlü, Aysun et al. (Jan. 2023). «The Effect of a Virtual Reality-Mediated Gamified Rehabilitation Program on Pain, Disability, Function, and Balance in Knee Osteoarthritis: A Prospective Randomized Controlled Study». In: *Games for Health Journal*. Publisher: Mary Ann Liebert, Inc., publishers. ISSN: 2161-783X. DOI: [10.1089/g4h.2022.0130](https://doi.org/10.1089/g4h.2022.0130). URL: <https://www.liebertpub.com/doi/abs/10.1089/g4h.2022.0130> (visited on 01/26/2023).
- Parizi, R. M. (2018). «On the understanding of Gamification in Blockchain Systems». In: *2018 6th International Conference on Future Internet of Things and Cloud Workshops*. Vol. 6. FiCloudW, pp. 214–219. DOI: [10.1109/W-FiCloud.2018.00041](https://doi.org/10.1109/W-FiCloud.2018.00041).
- Rockwood, Kenneth and Arnold Mitnitski (2007). «Frailty in relation to the accumulation of deficits». In: *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 62.7, pp. 722–727.
- Rockwood, Kenneth, Xiaowei Song, et al. (2005). «A global clinical measure of fitness and frailty in elderly people». In: *Cmaj* 173.5, pp. 489–495.
- Rolfson, Darryl B et al. (2006). «Validity and reliability of the Edmonton Frail Scale». In: *Age and ageing* 35.5, pp. 526–529.
- Šabić, Edin et al. (Mar. 2021). «Healthcare and anomaly detection: using machine learning to predict anomalies in heart rate data». en. In: *AI & SOCIETY* 36.1, pp. 149–158. ISSN: 1435-5655. DOI: [10.1007/s00146-020-00985-1](https://doi.org/10.1007/s00146-020-00985-1). URL: <https://doi.org/10.1007/s00146-020-00985-1> (visited on 03/02/2023).
- Samanta, Biswanath et al. (July 2009). «Prediction of periventricular leukomalacia. Part I: Selection of hemodynamic features using logistic regression and decision tree algorithms». en. In: *Artificial Intelligence in Medicine* 46.3, pp. 201–215. ISSN: 0933-3657. DOI: [10.1016/j.artmed.2008.12.005](https://doi.org/10.1016/j.artmed.2008.12.005). URL: <https://www>.

- [sciencedirect.com/science/article/pii/S0933365708001851](https://www.sciencedirect.com/science/article/pii/S0933365708001851)  
(visited on 03/15/2023).
- Sarwar, Tabinda et al. (2022). «The secondary use of electronic health records for data mining: Data characteristics and challenges». In: *ACM Computing Surveys (CSUR)* 55.2, pp. 1–40.
- Silitonga, Parasian (2017). «Clustering of patient disease data by using K-means clustering». In: *International Journal of Computer Science and Information Security (IJCSIS)* 15.7, pp. 219–221.
- Silva, Paula Alexandra, Kelly Holden, and Aska Nii (June 2014). «Smartphones, Smart Seniors, But Not-So-Smart Apps: A Heuristic Evaluation of Fitness Apps». In: ISBN: 978-3-319-07526-6. DOI: [10.1007/978-3-319-07527-333](https://doi.org/10.1007/978-3-319-07527-333).
- Sipers, Walther MWH et al. (2019). «Sarcopenia is related to mortality in the acutely hospitalized geriatric patient». In: *The journal of nutrition, health & aging* 23, pp. 128–137.
- Slator, Brian M. et al. (2007). «From Dungeons to Classrooms: The Evolution of MUDs as Learning Environments». In: *Evolution of Teaching and Learning Paradigms in Intelligent Environment*. Ed. by Lakhmi C. Jain, Raymond A. Tedman, and Debra K. Tedman. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 119–159. ISBN: 978-3-540-71974-8. DOI: [10.1007/978-3-540-71974-86](https://doi.org/10.1007/978-3-540-71974-86). URL: <https://doi.org/10.1007/978-3-540-71974-86>.
- Smith, Aaron (2014). «Attitudes, Impacts, and Barriers to Adoption». In: *Older adults and technology use*. URL: <https://www.pewresearch.org/internet/2014/04/03/attitudes-impacts-and-barriers-to-adoption/>  
(visited on 10/11/2023).
- statista (2023). *Number of Games released on Steam worldwide from 2004 to 2023 YTD*. Website. URL: <https://www.statista.com/statistics/552623/number-games-released-steam/>.
- SteamDB (2023). *Steam Game Releases by Year*. Website. URL: <https://steamdb.info/stats/releases/>.
- Su, Chao-Ton et al. (Aug. 2012). «Data mining techniques for assisting the diagnosis of pressure ulcer development in surgical patients». eng. In: *Journal of Medical Systems* 36.4, pp. 2387–2399. ISSN: 0148-5598. DOI: [10.1007/s10916-011-9706-1](https://doi.org/10.1007/s10916-011-9706-1).
- Suryadevara, Chaitanya Krishna (2020). «TOWARDS PERSONALIZED HEALTHCARE-AN INTELLIGENT MEDICATION RECOMMENDATION SYSTEM». In: *IEJRD-International Multidisciplinary Journal* 5.9, p. 16.
- Susi, Tarja, Mikael Johannesson, and Per Backlund (2007). «Serious games: An overview». In.

- Sutter, John D. (2010). «Browse the web, earn points and prizes». In: *CNN*. URL: <https://web.archive.org/web/20141009151929/https://edition.cnn.com/2010/TECH/web/09/30/>.
- Szulborski, Dave (2005). *This is not a game: A guide to alternate reality gaming*. Incunabula.
- Thom, Jennifer, David Millen, and Joan DiMicco (2012). «Removing gamification from an enterprise SNS». In: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. CSCW '12. New York, NY, USA: Association for Computing Machinery, pp. 1067–1070. ISBN: 978-1-4503-1086-4. DOI: [10.1145/2145204.2145362](https://doi.org/10.1145/2145204.2145362). URL: <https://doi.org/10.1145/2145204.2145362> (visited on 12/13/2022).
- Thompson, Vetta L Sanders et al. (Nov. 2014). «Identifying key variables in African American adherence to colorectal cancer screening: the application of data mining». In: *BMC Public Health* 14, p. 1173. ISSN: 1471-2458. DOI: [10.1186/1471-2458-14-1173](https://doi.org/10.1186/1471-2458-14-1173). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4256804/> (visited on 03/15/2023).
- Tremblay, Mark S. et al. (June 2017). «Sedentary Behavior Research Network (SBRN) – Terminology Consensus Project process and outcome». In: *International Journal of Behavioral Nutrition and Physical Activity* 14.1, p. 75. ISSN: 1479-5868. DOI: [10.1186/s12966-017-0525-8](https://doi.org/10.1186/s12966-017-0525-8). URL: <https://doi.org/10.1186/s12966-017-0525-8> (visited on 12/21/2022).
- UserTesting, H (2019). *Chatbot apps are on the rise but the overall customer experience (cx) falls short according to a UserTesting report*.
- Utermohlen, Karl (2018). «Four robotic process automation (RPA) applications in the healthcare industry». In: *Medium*.
- Uwizeyemungu, Sylvestre, Placide Poba-Nzaou, Michael Cantinotti, et al. (2019). «European hospitals' transition toward fully electronic-based systems: do information technology security and privacy practices follow?» In: *JMIR medical informatics* 7.1, e11211.
- Varga, Gabriela, Lăcrămioară Stoicu-Tivadar, and Stelian Nicola (May 2021). «Serious Gaming and AI Supporting Treatment in Rheumatoid Arthritis». en. In: *Studies in Health Technology and Informatics*. Ed. by John Mantas et al. IOS Press. ISBN: 978-1-64368-184-9 978-1-64368-185-6. DOI: [10.3233/SHTI210262](https://doi.org/10.3233/SHTI210262). URL: <https://ebooks.iospress.nl/doi/10.3233/SHTI210262> (visited on 01/26/2023).
- Veloso, Rui et al. (Dec. 2014). «A Clustering Approach for Predicting Readmissions in Intensive Medicine». In: *Procedia Technology* 16, pp. 1307–1316. DOI: [10.1016/j.protcy.2014.10.147](https://doi.org/10.1016/j.protcy.2014.10.147).

- Villani, Daniela et al. (2018). «Videogames for emotion regulation: a systematic review». In: *Games for health journal* 7.2, pp. 85–99.
- Vimark, Sara (2023). «Data mining in healthcare: A security and privacy perspective». In: Walston, Jeremy D (2012). «Sarcopenia in older adults». In: *Current opinion in rheumatology* 24.6, pp. 623–627.
- WHO (2013). *Obesity and Overweight*. Website.
- Williams, Nerys (2014a). «PHQ-9». In: *Occupational medicine* 64.2, pp. 139–140.
- (2014b). «The GAD-7 questionnaire». In: *Occupational medicine* 64.3, pp. 224–224.
- Xue, Qian-Li (2011). «The frailty syndrome: definition and natural history». In: *Clinics in geriatric medicine* 27.1, pp. 1–15.
- Yeolekar, ME and Sushija Sukumaran (2014). «Frailty syndrome: a review». In: *J Assoc Physicians India* 62.11, pp. 34–8.
- Zainab, Syeda Sana e and Tahar Kechadi (2019). «Sensitive and private data analysis: A systematic review». In: *Proceedings of the 3rd International Conference on Future Networks and Distributed Systems*, pp. 1–11.
- Zhang, Zhengyou (2012). «Microsoft kinect sensor and its effect». In: *IEEE multimedia* 19.2, pp. 4–10.
- Zhao, Zhao, Ali Etemad, and Ali Arya (Nov. 2016). «Gamification of Exercise and Fitness using Wearable Activity Trackers». In: 392. ISBN: 978-3-319-24558-4, pp. 233–240. DOI: [10.1007/978-3-319-24560-730](https://doi.org/10.1007/978-3-319-24560-730).
- Zhao, Zhao, S. Ali Etemad, et al. (2016). «Motivational Impacts and Sustainability Analysis of a Wearable-based Gamified Exercise and Fitness System». In: *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. CHI PLAY Companion '16. New York, NY, USA: Association for Computing Machinery, pp. 359–365. ISBN: 978-1-4503-4458-6. DOI: [10.1145/2968120.2987726](https://doi.org/10.1145/2968120.2987726). URL: <https://doi.org/10.1145/2968120.2987726> (visited on 12/14/2022).

## APPENDICES





## APPENDIX A - PROJECT GLOSSARY

---

This file contains a project deliverable aimed at providing additional context to section [4.1.2.4](#) of the main document, entitled "Terminology Dictionary". It contains a single main section composed by three subchapters documenting all relevant terminology that is interdisciplinary in this project, as well as a short list of terms exclusive to the business scenario, providing readers and project stakeholders alike with definitions for processes, technical terms or lingo that is unique to either discipline.

### A.1 GLOSSARY

Below, the reader can refer to either list to consult definitions for technical terminology used throughout the project, in alphanumerical order.

#### A.1.1 *Project-specific Terms*

- **CUDA:** Compute Unified Device Architecture, a programming platform developed by NVIDIA for general computing on graphical processing units, targeted at dramatically speeding up computing applications, such as neural network training cycles.
- **Gamevitae:** A hybrid gaming and healthcare platform currently in development by Techframe, heavily focused on harnessing the benefits of gamification and device integration for providing engaging alternative methods to therapy, especially muscular rehabilitation.
- **GripLand:** A proof-of-concept game for Gamevitae, in which the player can gradually develop a fictitious country with currency earned through correctly gripping the controller.
- **GripMusic:** A proof-of-concept game for Gamevitae, in which the player is required to grip the controller rhythmically, to the beat of a song.

- **GripRun:** A proof-of-concept game for Gamevitae based on an endless runner, where the player is required to grip the controlled at adequate intervals and strengths to jump between platforms.
- **Gripwise:** A controller device meant to be held in a grip position. Contains sensors capable of measuring the strength and duration of the user's grip. Is currently the only device integrated in the Gamevitae platform, and is used for all games currently available.
- **Techframe:** The business behind the development of Gamevitae, and the promoters of the data mining project included in this thesis.

#### A.1.2 *Healthcare Terms*

- **Exhaustion:** The state of feeling extremely tired.
- **Frailty Syndrome:** A clinical state of increased vulnerability that is recognized by progressive multi-systemic decline, reduced physiological reserve and ability to cope with acute stress, and increased adverse health outcomes (Ko and J. D. Walston, 2013).
- **FRIED Frailty Phenotype:** Assesses physical frailty through five criteria: unintentional weight loss, weakness or poor hand grip strength, self-reported exhaustion, slow walking speed, and low physical activity. Some variants may also include low balance ability.
- **Grip Strength:** A measure of how much force a person can generate with their fingers and thumb. In the context of this thesis, it is always measured in Kilograms of Force (Kgf).
- **GAD-7:** Generalised Anxiety Disorder assessment, a series of 7 questions aimed at providing an easy, self administered questionnaire for screening the severity of anxiety in an individual. Its scores range from 0 to 21.
- **PHQ-9:** Patient Health Questionnaire, a series of 9 questions that objectifies and assesses the degree of depression severity in a patient. Its scores range from 0 to 27.
- **Sarcopenia:** An age related, involuntary loss of skeletal muscle mass and strength (J. D. Walston, 2012).

### A.1.3 Data Mining Terms

- **Accuracy (Modelling):** A metric for assessing the success of a data mining task. When applied to a model, accuracy refers to the percentage of correct classifications in an arbitrary number of iterations.
- **Accuracy (Data Quality):** A measure of the extent to which data represents the true value of the attribute it is intended to measure.
- **API:** Application Program Interface, a service which provides a means to interact with a certain software system, such as data retrieval or on-demand predictions.
- **Availability:** A measure of how often the data is available to be accessed, consulted and used.
- **Centroid:** An imaginary or real observation representing the centermost point in a cluster.
- **Classification Problem:** A specific type of data mining problem aimed at predicting an observation's label by training a model on previously known predictor variables.
- **Cluster:** A collection of data based on similarities between the points.
- **Completeness:** The extent to which a dataset contains all the necessary elements and observations for a given purpose or analysis.
- **Collaborative Filtering:** A data filtering technique that can help identify outcomes that are preferable to a user, based on information from similar users.
- **CRISP-DM:** Cross-Industry Standard Process for Data Mining, an open standard process model that describes a general guideline for covering the most common aspects of a data mining project.
- **CSV:** Comma Separated Values, a type of file commonly used to store text data in a table structured format.
- **Inter-cluster Variance:** A measured used to quantify external separation between clusters. Defined as the sum of the squared distance between the global average point and each cluster centroid. Effective clustering aims to maximize this value.
- **Intra-cluster Variation:** A measure used to quantify internal cohesion in a cluster. Defined as the sum of the squared distance between the centroid and each point of the cluster. Effective clustering aims to minimize this value.
- **Missing Value:** An occurrence where no data value is stored for a given variable in an observation.

- **Outlier:** A data object that deviates significantly from the other data objects, usually showcasing a different behaviour as well.
- **Prediction Problem:** Data analysis aimed at predicting a future event, data or trend.
- **Predictor Variable:** Data attributes that are effectively used to predict other variables or outcomes, in a prediction problem.
- **Profiling:** The process of generating descriptions for clusters, in reference to the features used during cluster analysis.
- **SQL:** Structured Query Language, a programming language for storing and processing information in a relational database.
- **SQL Server:** A proprietary relational database management system developed by Microsoft.
- **Timeliness:** The expected time of availability and accessibility of the data.
- **Uniqueness:** The degree of singularity between data observations in a dataset. Uniqueness is effectively achieved when the information in a dataset appears once.
- **Validity:** The measure of the accuracy and reliability of information within a dataset, ensuring its conformity with standards, rules or other previously defined constraints.

## APPENDIX B - PROJECT PLAN

---

This file contains the deliverables and some additional considerations for section [4.1.4](#) of the main document, entitled "Project Planning". Each following section will further contextualize the elements deemed necessary for the successful planning of my project, namely:

- The project scheduling
- The resource allocation for each phase
- Tasks related to the post-deployment stage
- The result communication plan

### B.1 PROJECT SCHEDULE

Due to some internal difficulties with business definitions and scheduling issues with my superiors, my project suffered a significant delay prior to its start, thus falling out of reach of the intended original delivery date. Following this, the entire list of tasks had to move to later dates, and some processes' duration may have been affected during the adjustment. Regardless, I am confident in the revised project schedule and my ability to fulfill the proposed goals by the **31st of March of 2024**. The final project scheduling includes both the scientific research portion of the project, and all of the [CRISP-DM](#) related tasks, and can be consulted in [Fig. 13](#).

### B.2 RISK SCHEDULING

In addition to time scheduling, it is also important to document relevant risks attributions for the different phases of the project. Clearly, this project does not require an in-depth analysis of human personnel, since all tasks will be carried out by myself, but regardless, I consider the documentation of when different risks may happen an important step in avoiding them. In [table 20](#), a listing of the previewed project tasks and the associated risk

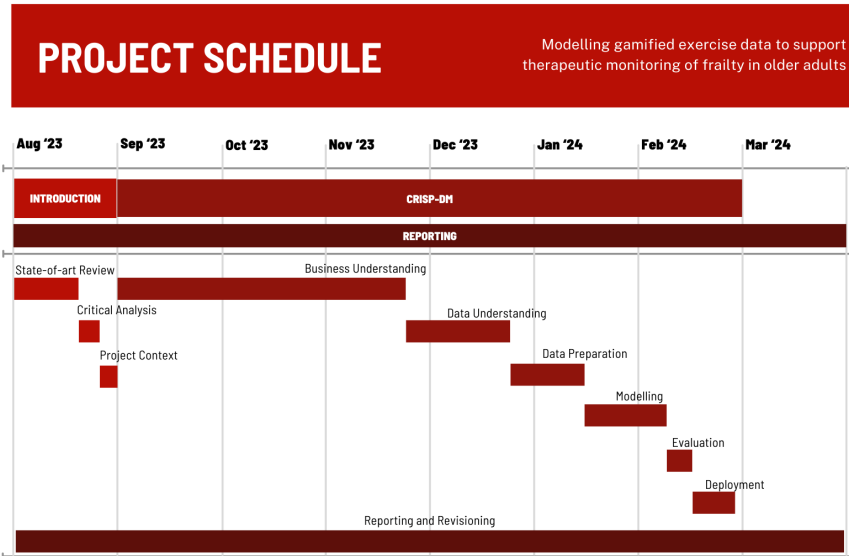


Figure 13: Project schedule

assessment is available. Information regarding the respective sections is also present, for easier navigation within the document.

Project Phase	Possible Risks	Document Location
State-of-art review	RC.01	Chapter 2
Critical Analysis	RC.01	Section 2.3
Project Context	RC.01	Chapter 3
Business Understanding	RC.01, RC.10	Section 4.1
Data Understanding	RC.01 to RC.07	Section 4.2
Data Preparation	RC.01, RC.07, RC.11	Section 4.3
Modelling	RC.01, RC.10, RC.11	Section 4.4
Evaluation	RC.01, RC.08, RC.10	Section 4.5
Deployment	RC.09, RC.10, RC.11	Section 4.6

Table 20: Project phases and associated risks

### B.3 COMMUNICATION PLAN

As discussed in the main document, the company superiors for this project have agreed to make its results publicly available via this thesis, which is the main document and reports on every project stage to great detail. However, I have devised a subsequent communication plan that includes the publication of a scientific article dedicated to supporting a particular model developed during the data mining project, establishing its relevance to the area.

#### B.3.1 *Publication Vehicle*

In-line with the expectations for papers reporting the development and evaluation of data models, the preferred vehicle of publication is an open-access journal, and the choice of publisher will be based on (a) the impact on the project financing when accounting the publication costs; (b) the impact factor of the specific journal issue selected for publication; (c) the expected time window between the first submission process and the subsequent answer, aiming to minimize the time-to-publish and (d) other external factors, such as previous experience with the publisher. Upon reaching an agreement with all interested parties and funding bodies, I have decided on the **Multidisciplinary Digital Publishing Institute (MDPI)** as the ideal publishing vehicle for this paper, due to having previously published my own works on their journals and being generally satisfied with the response times.

#### B.3.2 *Financing*

MDPI openly discloses their journal-specific Article Processing Charges (APC) ranging from 1000 CHF to 3000 CHF. At the time of writing this section, the Swiss franc to euro exchange rate is around 1 : 1.07, meaning that the writing and dissemination of results is a substantial expense which must be included in the financial breakdown of my project (see Tab. 6 of the main document). The financing required to support the publication of my article will be partly covered by Techframe, and partly covered by advisor teacher overseeing my project, for which I extend an additional thanks to both parties for providing the means for the project to develop as smoothly as possible.

### B.3.3 *Publication Time Window*

The review process at MDPI typically involves three reviewers weighing in on the viability and quality of the work proposed. The journal guidelines mention an average wait of two to three weeks for the first round of reviews, which falls in line with my previous experience with the institute. Additionally, there is added flexibility in case either the reviewers or the writers require more time to conclude their tasks. As such, I expect the publication of the paper to happen no later than the second half of May 2024, keeping in mind at least one round of reviewing and resubmitting.

## B.4 FUTURE EVENTS

This section is dedicated to documenting other tasks that would either fall out of the scope for the CRISP-DM process, or wouldn't logically fit within the time constraints for the project, despite their importance to its success. The tasks documented here justify the need for follow-up works and projects to close the gaps and solidify my contributions in the GameVitae platform.

### B.4.1 *Model Assessment Trials*

As stated in the model assessment phase (Section 4.4.4 of the main document), several of models faced shortcomings in the depth of their testing due to limitations with the time available for the completion of the project. Despite the impossibility of carrying out a complete experimentation process, the intention for this task remains as one of the main points of reference for future work.

The plan is to provide model access to a controlled set of patients and therapists for a determined period of time as a methodological way of assessing the subjective criteria referenced in the test design related to easy-of-use and interpretation. Based on the initial feedback collected through this trial, the model refinement process that follows might be complex enough to warrant an additional project. In specific, the model trials should aim to find gaps in quality, inconsistencies and inconvenient design approaches that may harm its adoption by GameVitae users. The following questions are to be used as guidelines:

- Are the existing groups enough to represent the majority (over 95%) of patient typologies?

- Is the patient classifier accurate enough as to not displace cluster density over time? (i.e.: Can the classifier comfortably fit newer patients into existing groups without significantly increasing intra-cluster distances?)
- Are the first recommended activities well received by the patients?
- How many iterations of feedback collection are needed to satisfy a patient's preferred activity layout?
- Are therapists satisfied with the speed and information of the analysis tool?
- Are the automatically provided interpretations about a patient's performance adequate and clinically relevant?
- Do therapists feel that some aspect of the tool is lacking or impertinent?

With these inquiries in mind, revised iterations of the models will be deployed, with adjustments that will hopefully reflect the wants and needs of the user groups. Even though the goals for the trial are clear, other aspects related to the duration and size of the trial are to be defined at a later date.

#### B.4.2 *Model Incubation*

Once models are up and running in the business, there will be a period of time during which the technical personnel of the business will need to adjust to the functioning of the platform. This includes learning how to properly use the provided endpoints, how to navigate the supporting files and database tables, and understanding the influx of information for each of the models. Naturally, most of this knowledge is achieved with time and hands-on practice, but given that it exceeds the timeline for the project, I have decided to include it in this section. One precaution to aid model incubation that was included in my project's scope is the writing and sharing of [API](#) documentation. While indispensable for maintenance purposes, the documentation file will also, at first, be the main point of contact between the engineers and the models. Usage examples and database schemas will also be provided, helping minimize lookup times and delays in my ability to provide personal support. The time frame for incubation was loosely discussed with business decision makers, who agree to target no more than 3 weeks after the deployment schedule for a comfortable integration of the models, putting an indicative deadline anytime between **1 July 2024** and **8 July 2024**.



## APPENDIX C - DATA UNDERSTANDING REPORT

---

This file is a project deliverable, providing context and documentation to the Data Understanding phase of the CRISP-DM methodology, present in section 4.2 of the main document. It will contain additional insights, visualizations, and guidance for all activities carried out in the Data Understanding Python notebook, namely:

- Data Loading
- Data Description
- Data Exploration
- Data Quality Assessment

For portability and code reproduction purposes, all tasks carried out during the Data Understanding phase were performed within a python notebook. Please refer to related notebook for source code and additional considerations that were better fitted to be read in-document.

### C.1 DATA LOADING

Many solutions exist to load CSV files into Python, and choice mostly comes down to the tools used for exploring the data. For my project, I am using *Pandas* as my preferred data analysis tool, which has built-in CSV file loading.

#### C.1.1 *First Look at the Data*

Immediately after loading the data, I printed out the first few rows of each set, allowing for some initial thoughts related to the available attributes.

Starting with patient information, a first look reveals that the anonymization process has kept the redacted columns in the data, which makes listing the removed information very easy. All information referring to **religion, race, sexual habits, names, national patient numbers and contact information** has been censored from the original data,

making these attributes irrelevant for work in future phases. Also, attribute names do not seem to follow a common naming process, and their order in the dataset appears random. The presence of categorical and continuous fields can also be averred.

The situation for activity data is different, as data is mostly numerical. At this point, it is not possible to identify any irrelevant fields, barring an unnamed column. The field reserved for storing a given user's score on the current activity has the particularity of allowing negative values, which seems to allow for a better understanding of deficiency / overshooting in certain cases.

### *C.1.2 Additional Considerations*

One important matter to discuss is whether the data can (and should) be merged as a part of the Data Preparation phase. As evidenced before, both sets share an attribute reserved for identifying patients. While this allows very easy merging, the issue of dealing with duplicate information arises. Additionally, the objectives set for data mining were created in a way that foresees working with separate sets of data, since **clustering** patients can be interesting from both a behavioural and activity standpoint. Therefore, data merging will not be considered a part of preparation, not excluding any situations where creating a unified view of the data is beneficial for exploration or modelling.

In the event that either of the datasets have missing values, it is decided that these situations should be analyzed and dealt with on a case-by-case basis, with proper documentation explaining the rationale for the inclusion/exclusion process of such fields.

## C.2 DATA DESCRIPTION

In this section, basic information regarding the data is recorded. In short, this step complements the dictionaries present in the main document by providing the shape of the data, the data types within it, and some central trends that classify the average patient. The following subsections refer to each of the sets, for ease of understanding.

### *C.2.1 Patient Description*

This dataset is comprised by **874** rows and **31** columns, out which 9 columns are irrelevant as per an initial assessment. The following table (tab. 21) contains the *python dtype* of

each relevant attribute and the respective value ranges. The shaded column contains the average numerical or most common categorical value for each attribute.

Attribute	Data Type	Range	Average / Mode
fried	object	[non-frail; pre-frail; frail]	Pre-frail
gender	object	[F; M]	F
age	int64	[64, ..., 87]	74
hospitalization one year	int64	[0, ..., 6]	0
vision	object	[sees well; sees moderately; sees poorly]	Sees well
audition	object	[hears well; hears moderately; hears poorly]	Hears well
weight loss	object	[yes; no]	No
exhaustion score	int64	[1; 2]	1
balance single	object	[< 5 sec; > 5 sec; test non realizable]	> 5 sec
grip strength abnormal	object	[yes; no]	No
low physical activity	object	[yes; no]	No
bmi score	float64	[18.88, ..., 53.08]	30.67
sleep	object	[no sleep problem; occasional sleep problem; permanent sleep problem]	No sleep problem
depression total score	int64	[0, ..., 12]	2
anxiety perception	float64	[0, ..., 13.6]	4.71
pain perception	float64	[0, ..., 12.8]	5.07
activity regular	object	[no; < 2h per week; > 2h and < 5h per week; > 5h per week]	> 2h and < 5h per week
smoking	object	[never smoked; stopped (6 months); current smoker]	Never smoked
alcohol units	float64	[0, ..., 35]	1.65
medication count	int64	[0, ..., 18]	6
birthplace	object	[PT; ES; BR; AO]	PT

Table 21: Patient data technical description

### C.2.2 Activity Description

This dataset is comprised by **3525** rows and **15** columns, out which 1 column is irrelevant as per an initial assessment. The following table (tab. 22) contains the *python dtype* of each relevant attribute and the respective value ranges. The shaded column contains the average numerical or most common categorical value for each attribute.

Attribute	Data Type	Range	Average / Mode
game id	object	[GV001R; GV002M; GV003L]	GV001R
therapeutic difficulty	int64	[0, ..., 10]	6
target grip force	float64	[17.0, ..., 25.9]	20.4
target grip timing	float64	[35.1, ..., 89.0]	61.0
target instances	int64	[10, ..., 50]	28
grip force score	float64	[13.2, ..., 29.1]	21.7
grip timing score	float64	[20.0, ..., 91.98]	68.1
grip instances hit	int64	[20.0, ..., 91.98]	25
score	float64	[-38.6, ..., 47.4]	5.72
session target duration	int64	[45; 60; 90]	58
session real duration	int64	[23, ..., 90]	56
satisfaction query	float64	[1; 2; 3; 4; 5]	4.1

Table 22: Activity data technical description

## C.3 DATA EXPLORATION

This portion of appendix is dedicated to documenting the entire exploratory part of data understanding. In the notebook that complements this phase, I have carried out an extensive process for extracting as much information as possible from uni-variate and multi-variate data slices. There is also a portion of work exclusive for analysis that involve variables from both datasets, in an attempt to relate patient's health conditions to their performance.

The following subsections are complemented with tabular views and graphs also available in the notebook. To enrich the contents of this section, my personal insights and

interpretations will be present along the images, constituting what is essentially my learning process during exploration of the data.

### c.3.1 *Patient Data Exploration*

For a matter of convenience when exploring the patient dataset, the following columns were excluded from the whole process: **Religion; Race; Sexual Habits; E-mail; Address; NHS Number; First Name; Last Name.**

Firstly, an analysis on the age distribution (Fig. 14a) reveals that while the entire population is elderly (over 65 years old), most patients are between ages 70 and 74. There are also slightly more female patients (Fig. 14b), and the overwhelming majority of the study group was born in Portugal (Fig. 14d). The BMI distribution (Fig. 14c) suggests that only 490 patients (56.1%) have a healthy weight, while the remaining 384 patients (43.9%) are considered underweight or overweight. Note that for the elderly population, the proposed healthy BMI ranges differ slightly (Kıskaç et al., 2022).

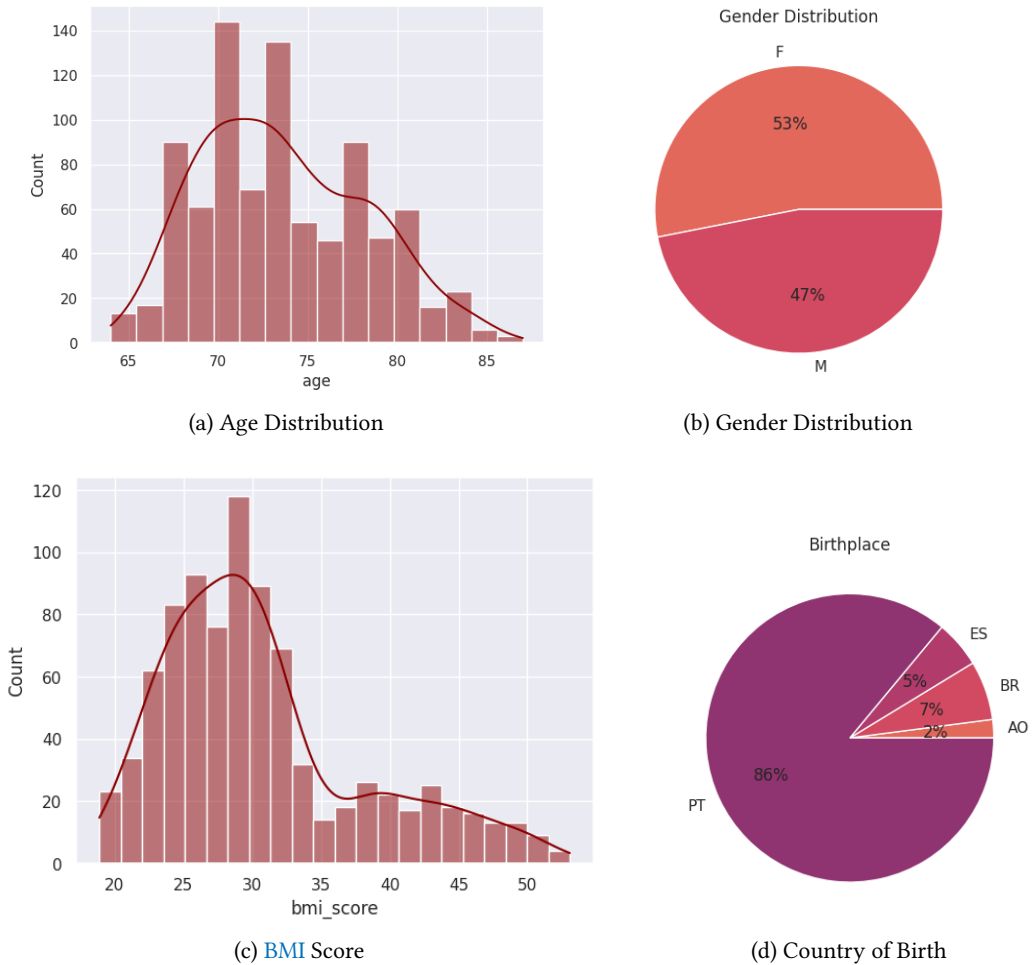


Figure 14: Patient demographic variable exploration

When exploring the incidence of self-reported symptoms, any conclusions must be drawn and interpreted with a degree of tolerance to error, since each individual’s perception is different and it is impossible to set a completely accurate scale. In relation to mental disorders, 709 patients (81.1%) scored ‘None-Minimal’ on the PHQ-9 depression scale, while 161 (18.4%) show a ‘Mild’ depressive state. Moreover, only 4 respondents scored ‘Moderate’ and this is the maximum severity recorded in the data (Fig. 15a). Anxiety is more pronounced among respondents, with only 391 (44.7%) scoring ‘None-Minimal’ on the GAD-7 ranges, 379 (43.4%) scoring ‘Mild’, and 104 (11.9%) classified as having moderate anxiety (Fig. 15b). Almost all (88.8%) of respondents claim to experience non-negligible pain levels constantly, with 276 (31.5%) considering their pain level moderate or worse (Fig. 15c).

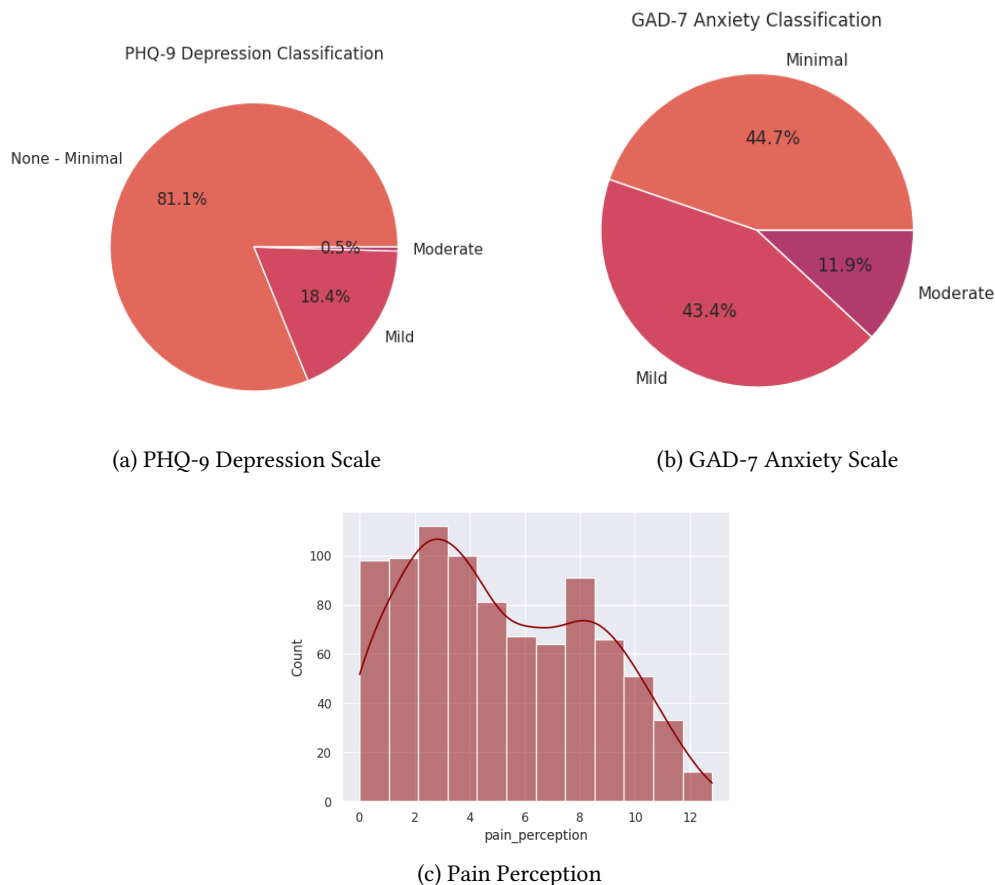


Figure 15: Self-reported variable exploration

The consumption of alcohol, tobacco and several concurrent prescription medications in the elderly may impact their overall energy levels and physical capacity, and could constitute a series of indirect factors for frailty aggravation. Understanding the presence of these habits in the data is important for making educated decisions about whether the inclusion/exclusion of the respective attributes is something to consider for future phases. Alcohol consumption is not necessarily prevalent in the data, with 860 (98.4%) respondents reporting a consumption of 14 or less units of alcohol per week (Fig. 16a), the threshold supported by the NHS Jani et al., 2021. Out of the 14 people who claim to exceed this value, the maximum amount of alcohol consumed in a week is 35 units. The study group is also majorly comprised of non-smokers, with 537 (61.4%) having never smoked, and 269 (30.7%) stopping over 6 months prior to data collection (Fig 16b). Contrarily, probably due to the comorbidity situations that elderly people usually face, over half of respondents (505) are taking 6 or more medications simultaneously, with 31 patients reporting taking over 15 administrations a day.

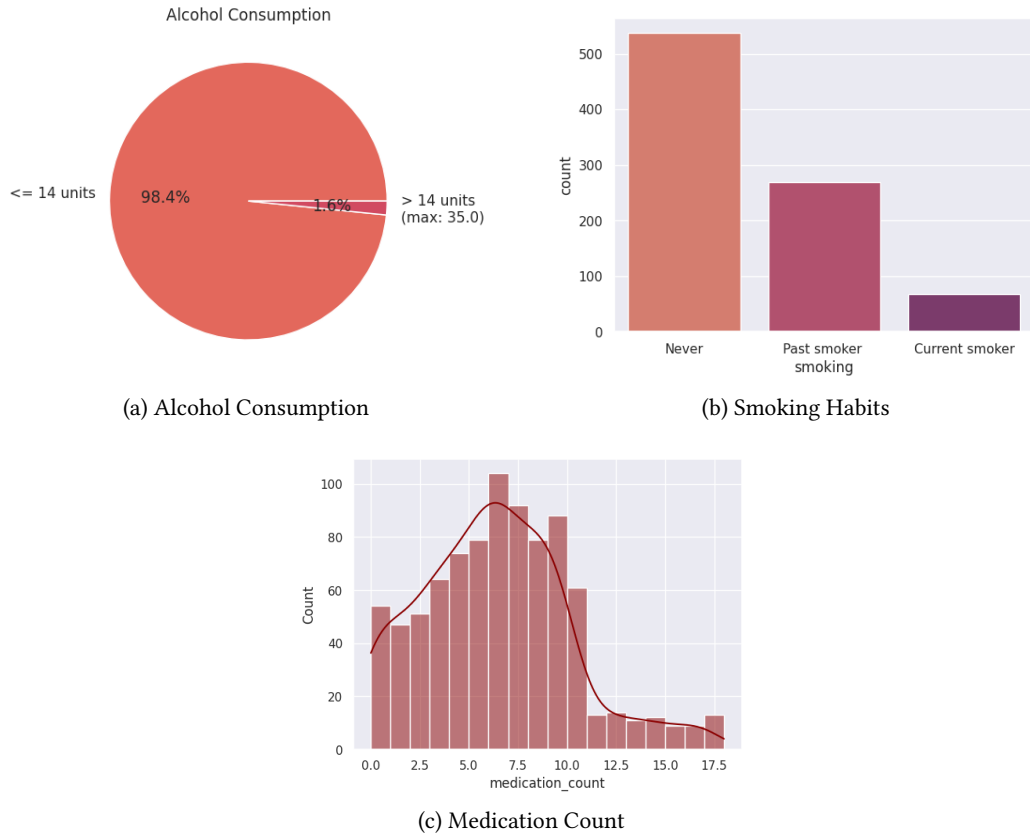


Figure 16: Substance consumption variable exploration

Improper rest and exercising habits are also important metrics that can directly influence feelings of sluggishness, exhaustion, a decrease in muscular strength, all of which are closely associated with frailty. In the evaluated data, 448 (51.2%) patients report having occasional or permanent sleeping disturbances, with 47 claiming the latter. When assessing weekly hours of physical activity, a staggering 216 patients (24.7%) report doing less than 2 hours of exercise per week, out of which 28 report absolutely no activity. The most common activity period is between 2 and 5 hours, with 375 (42.9%) patients fitting the metric (Fig. 17b).

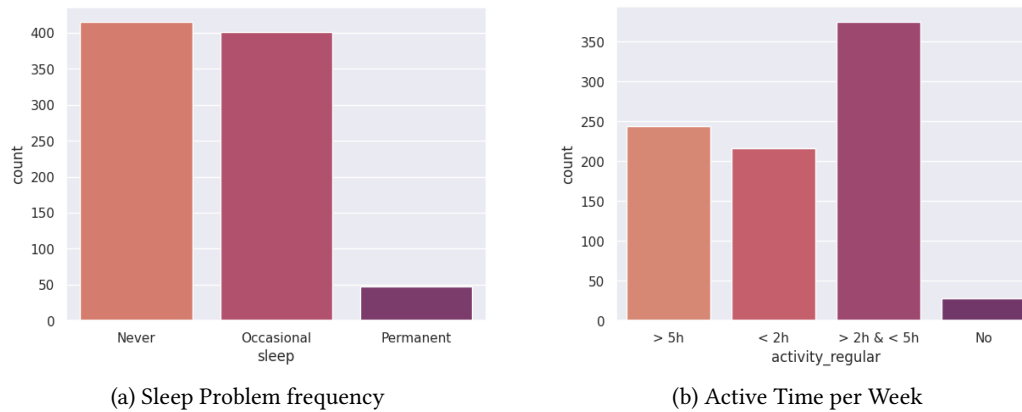


Figure 17: Rest and exercising variable exploration

Initially, the variables that track patients' eyesight and hearing quality were to be discarded, since, at first glance, they have insignificant relation to frailty and are simply assessments of well-being. At second-glance, these conditions may be interesting to relate to in-game performance, since poor eyesight and hearing may negatively impact score while not being necessarily measures of low muscle force (a patient with poor hearing might have trouble gripping the controller at the correct rhythm, when playing GripMusic). No patients seem to have significant hearing loss, however, 170 (19.4%) claim to have partial hearing (Fig. 18a). Eyesight problems are more pertinent in the study data, with 351 (40.1%) reporting altered vision quality, out of which 89 are considered legally blind without eye-wear (Fig. 18b).

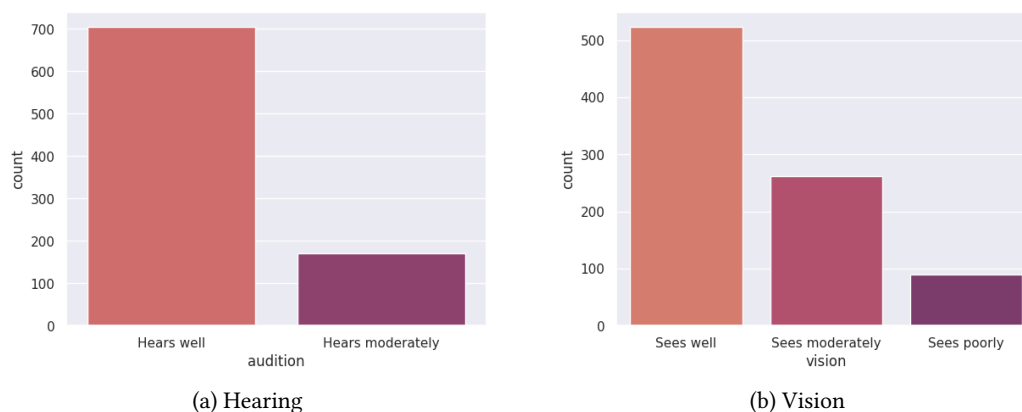


Figure 18: Audition and vision variable exploration

To close out uni-variate exploration, I studied five important factors for determining frailty present in the data. While not enough to define a state of frailty on their own, the presence of multiple factors constitutes a much higher risk of having muscular function

loss. In this dataset, the attributes that, together, compose a high risk of frailty are: involuntary weight loss; lack of balance; abnormal grip strength; sedentism; and exhaustion. Involuntary weight loss (of over 4.5kg in the past 6 months), while often associated with psychosocial factors, is a significant sign of frailty in the elderly, leading to more frequent falls and overall worse health outcomes (Gaddey and Holder, 2014). Out of all patients in the dataset, 300 (34.3%) have had significant unintentional weight loss (Fig. 19a). Balance is also closely associated with predicting frailty in older adults (Dayhoff et al., 1998). Out of the evaluated patients, 250 (28.6%) are not able to stand on one leg (Fig. 19e) for a period longer than 5 seconds, and a balance assessment was not possible on 42 (4.8%) patients, due to conditions of reduced mobility (bedridden, wheelchair, injury). As for grip strength (Fig. 19b), one of the most significant variables in my study, 256 (29.2%) patients show abnormally low grip force (under 30kgf in males, and 20kgf in females). Sedentism is also quite evident in the study group, with 292 (33.4%) respondents being classified as having low physical activity (Fig 19c), almost a quarter of respondents claim feeling exhausted to some degree (Fig 19d). I found it particularly interesting that across all risk factors, the percentages stay relatively close together, at around 30%, which could mean that the largest concentration of risk factors may be present within a specific group of individuals.

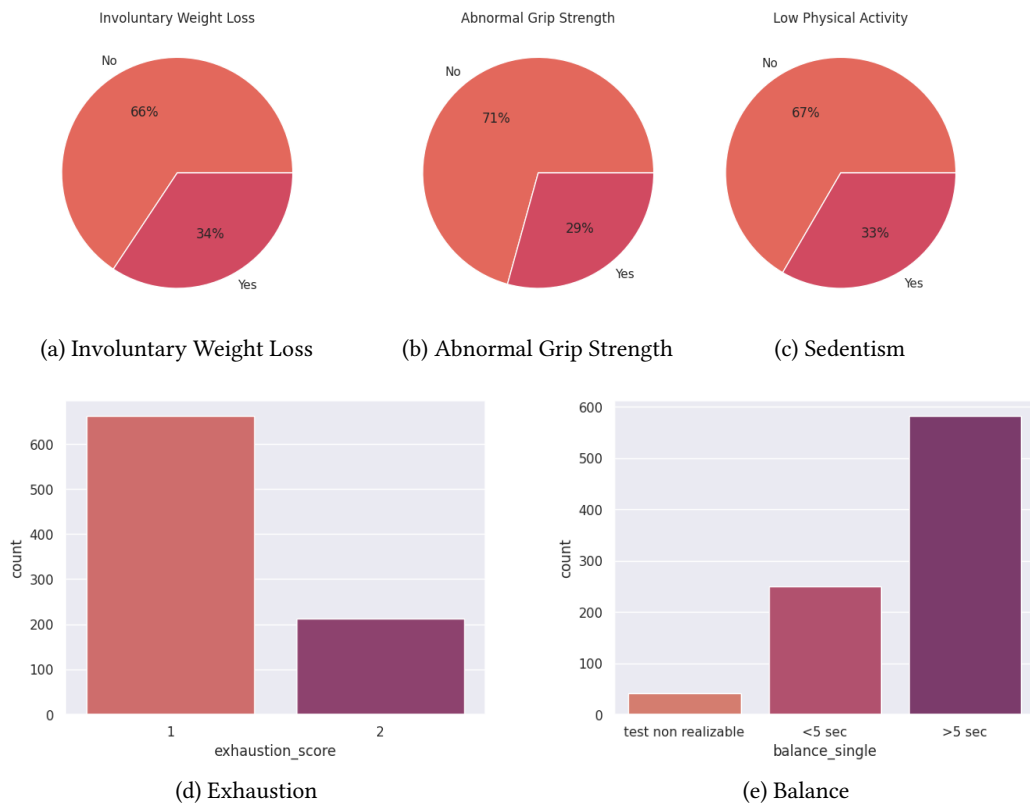


Figure 19: Key frailty factor variable exploration

The last remaining variable refers to the most recent FRIED classification of each patient, and is not a diagnostic variable based on the remaining factors. 396 (45.3%) patients were previously categorized as 'pre-frail', 272 (31.1%) as 'frail', and 206 were considered 'non-frail' (23.6%) prior to data collection (Fig. 20). This distribution is quite helpful for bi-variate analysis, since I have access to a fairly large sample of variable combinations to evaluate for having an impact in this particular population's frail assessment.

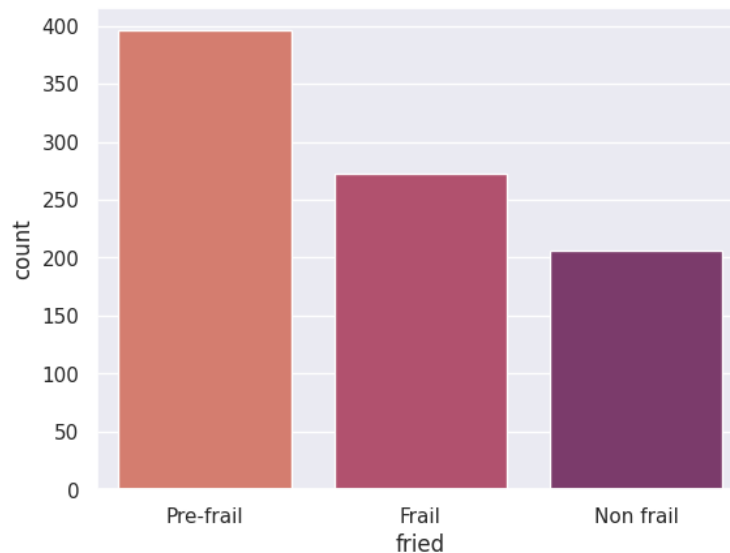


Figure 20: FRIED variable exploration

From this point onward, exploration was performed on sets of variables, in an attempt to find strong relationships between symptoms, demographic factors and the overall profile of a patient. I initiated bi-variate analysis by extracting a correlation matrix from numerical data, which revealed very weak relationships between the continuous variables. From this, I conclude that any pair formed between the variables: age; hospitalizations; exhaustion; bmi; depression; anxiety; pain; alcohol; and medication have no significant correlation or causality. Therefore, I focused on exploring relationships between categorical to categorical, and categorical to continuous relationships.

The first set of bi-variate analysis is focused on revealing potential links between demographic and anthropometric factors, (age, gender, BMI), and the severity of the FRIED classification assigned to the patient. The graphs presented in Figs. 21a, 21b and 21c do not indicate the presence of any relevant relationship between these factors, although males and people with unhealthy weights show ever so slightly (lower than 2% differential) more 'pre-frail' and 'frail' classifications.

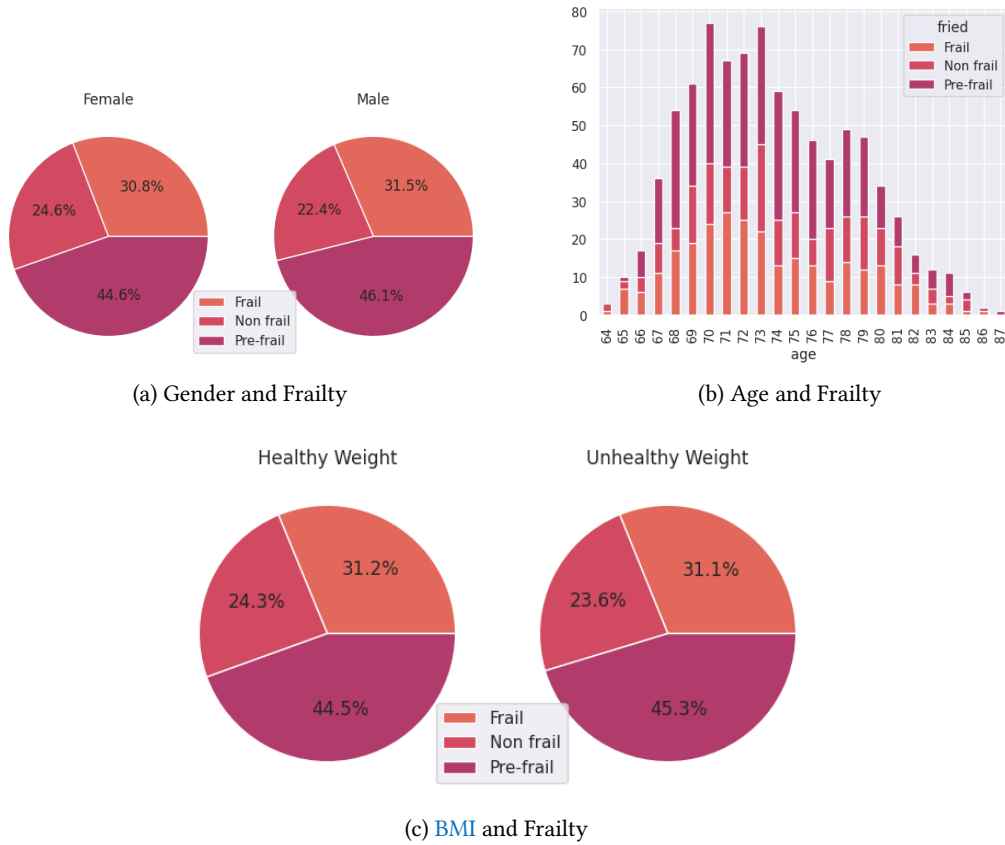


Figure 21: Demographic and anthropometric factors in FRIED assessment

Next, I evaluated the distribution of frailty scores between both categories of each risk factor, to determine those with the largest effect on the study group’s FRIED scale, the first of which being involuntary weight loss. The pie charts in Fig. 22a show an increase of 16.6% in patients classified as ‘pre-frail’ or ‘frail’ when they also report having unwillingly lost weight. As for patients with abnormally low grip strength, this increase was slightly lower, at 14.5% (Fig. 22b).

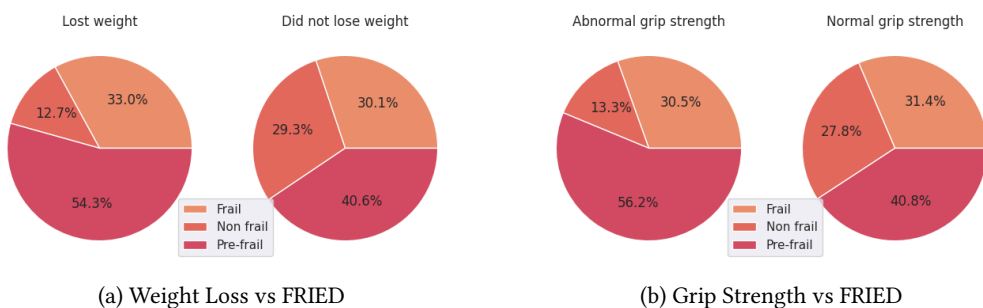


Figure 22: Effects of weight loss and low grip strength on frailty assessment

Sedentism is also clearly reflected on the frailty level of the population. Fig. 23a suggests that only 13.7% of patients with low physical activity were considered 'non frail', whereas this number climbs to 28.5% in the active population (increase of 14.8%). Contrarily, over half of patients (62.3%) who do not report feeling exhausted were actually classified as 'pre-frail', and a higher percentage (26.6%) of exhausted people are currently 'non-frail' (Fig. 23b).

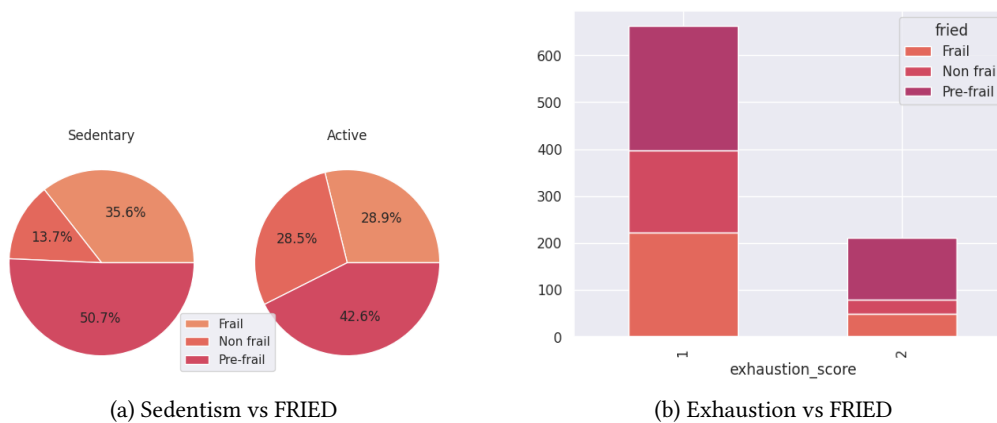


Figure 23: Effects of sedentary behaviour and exhaustion on frailty assessment

Lastly, the data also makes it clear that a person's single-leg balance influences their FRIED score (Fig. 24). Out of the 206 non-frail patients, 170 (82.5%) are able to stand on one leg for a period longer than 5 seconds. In comparison, this falls to 66.1% and 58.5% in frail and pre-frail patients, respectively.

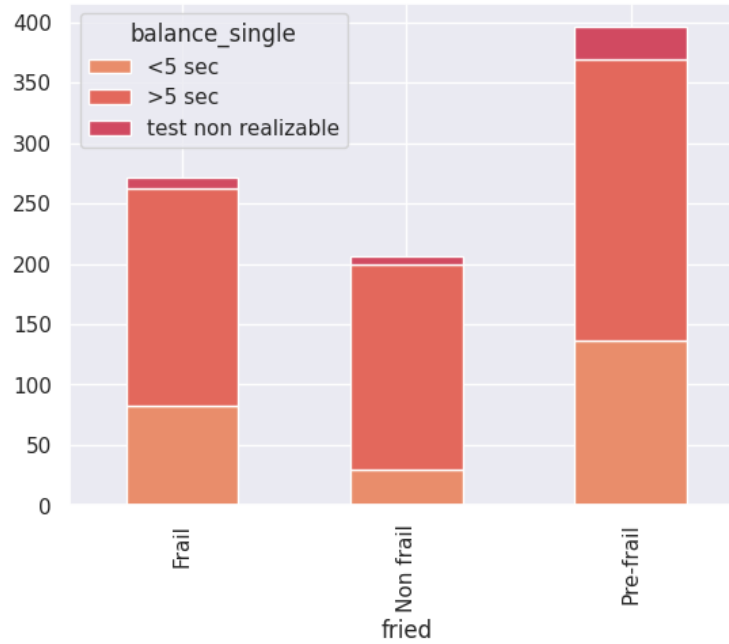


Figure 24: Effects of balance on frailty assessment

All in all, while each of the factors evaluated before is clearly linked to a patient's FRIED score, it is the presence of multiple factors that determine the severity of this assessment. I finished up patient specific data exploration by assessing the distribution of FRIED classes by patients with none to all five risk variables. In the plot presented in Fig. 25 shows that as the number of simultaneous factors increases, so does the relative number of pre-frail and frail patients. Most notably, the drop in non-frail patients between having one and two risk factors is quite evident, with non-frail patients making up 30.7% of the population with only one risk factor, and only 11.7% of patients with two risk factors. There is also only one patient that was still non frail despite having four risk factors, though they make up only 2% of that demographic.

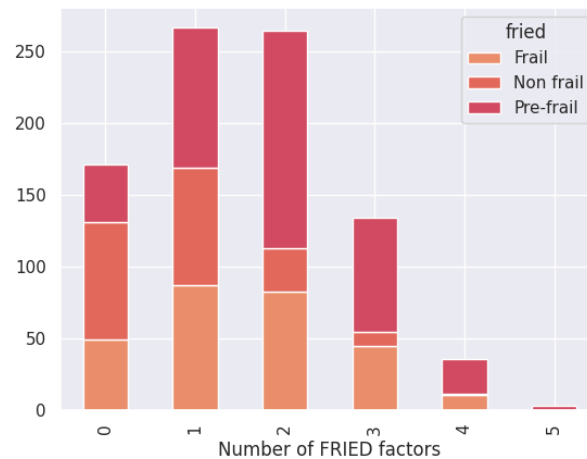


Figure 25: Presence of multiple risk factors on frailty assessment

The numbers for each amount of simultaneous factors are present in Tab. 23, alongside the relative percentage of total patients for each case. The majority (60.7%) of patients have between one and two concurrent factors, and only 3 people are afflicted by all of them.

Number of Factors	Non Frail	Pre-Frail	Frail	Percentage of Total Patients
0	82	40	49	19.5%
1	82	98	87	30.5%
2	31	151	82	30.2%
3	10	80	44	15.3%
4	1	24	10	4%
5	0	3	0	Under 1%

Table 23: FRIED category distribution by number of risk factors

### c.3.2 Activity Data Exploration

When performing exploration on the activity dataset, all columns were included in the process, since none were deemed redundant during the description phase. Moreover, this section will not be explicitly divided into uni-variate and multivariate analysis, and will instead include a more objective-based approach to exploration, containing a mix of visualizations that answer questions related to the adequability of activities and enjoyment/engagement measures.

The first goal was to understand the frequency of activities per game. As explained in the main document, three (albeit early versions) games are available to be played with the Gripwise controller. In this dataset, they are coded as such:

- GV001R: GripRun
- GV002M: GripMusic
- GV003L: GripLand

Out of the 3525 sessions that data has been collected from, 1942 (55.1%) were played on GripMusic, 1330 (37.7%) on GripRun, and only 253 (7.1%) on GripLand (Fig. 26a). The most common difficulty levels are 4, 5 and 9, with 556, 580 and 547 sessions played, respectively. These three difficulty levels account for almost half of all sessions played (47.7%) (Fig. 26b).

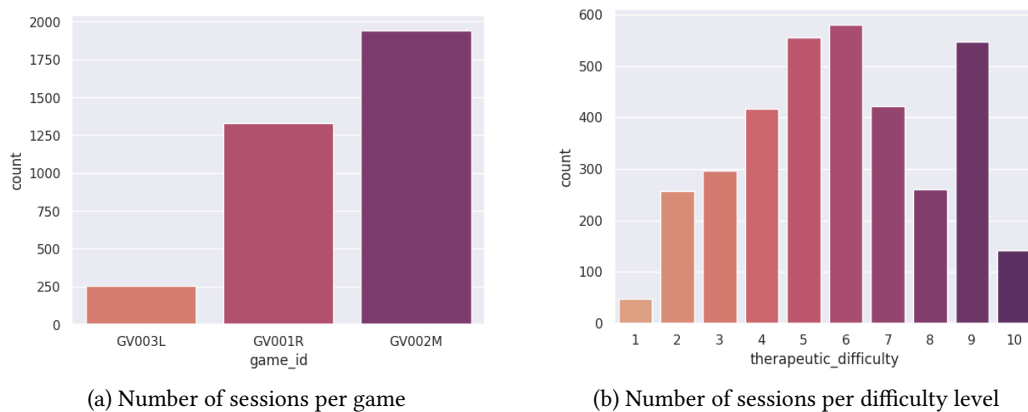


Figure 26: Distribution of sessions played per game characteristics

I also intended to understand whether the difficulty distribution changes significantly with the game played, and the data in Fig. 27 shows an evident difficulty spike in sessions played on GripMusic. Due to the difficulty of a session being dictated by the patient's physical state, this could mean that GripMusic is played mostly by less fragile users who have higher therapeutic thresholds to reach. In a similar way, GripLand has the lowest end of the difficulty range, being played primarily by patients with advanced frailty and several risk factors.

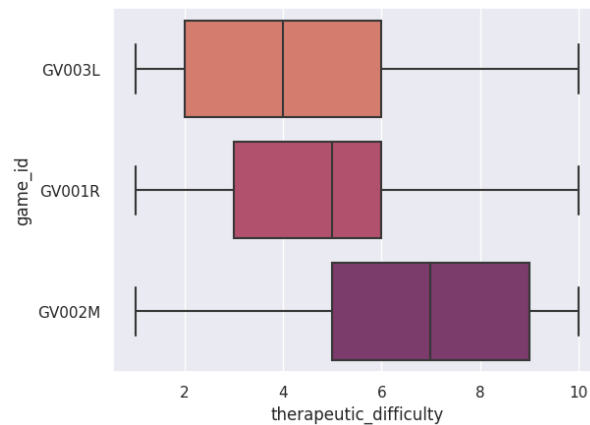


Figure 27: Difficulty range per game

The range of therapeutic objectives for each session, in theory, should change depending on whether the overall difficulty of the activity is higher or lower. Regardless, it may be useful to have an understanding on the central tendency of grip force, timing, and frequency scores in general. As a reminder, an explanation for the therapeutic measures is present in section 4.2.2.2 of the main document. The box charts in Fig. 28a shows that the majority of activities have a target grip force of 19.0Kgf to 21.5Kgf, whereas the proposed number of grip presses is concentrated between 20 and 35 (Fig. 28c). The larger percentage of sessions also targets between 52% and 71% 'on-time' grip presses (Fig. 28b).

Comparing the proposed values with the actual scores obtained in sessions, we can understand how well the activities match to the players. The data in Fig. 29a suggests that players' grip force is usually between 19.3Kgf and 24.2Kgf, though scores as low as 13.5Kgf are present. Curiously, the average player can comfortably outscore the proposed timing percentages, with around 59% to 80% accuracy, but several outliers with under 30% accuracy exist (Fig. 29b). The range of instances hit, due to having a fixed limit per session, is closer to the respective targets, with most falling between 16 and 34 presses (Fig. 29c).

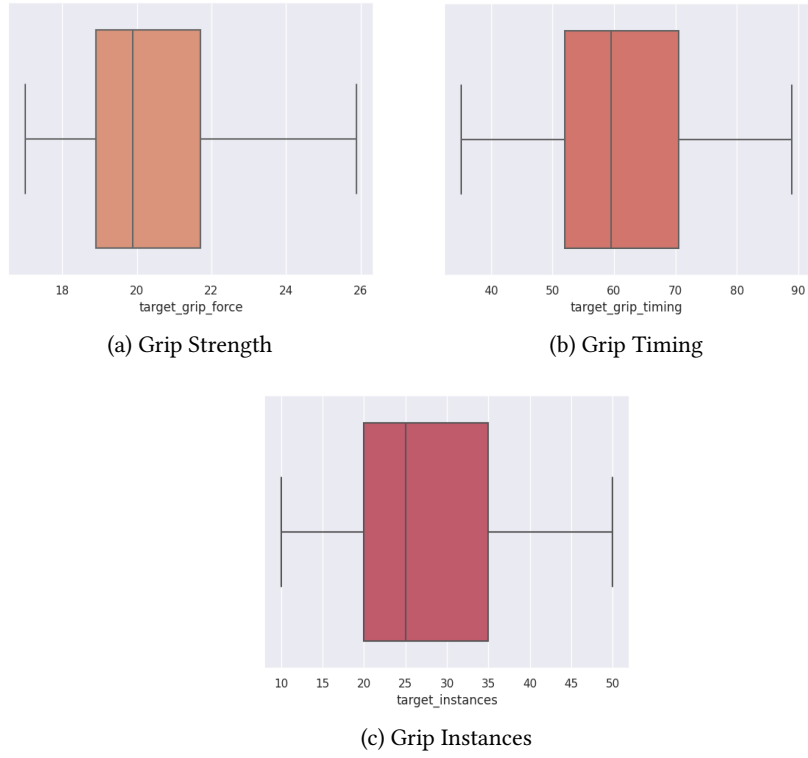


Figure 28: Range of therapeutic targets for activities

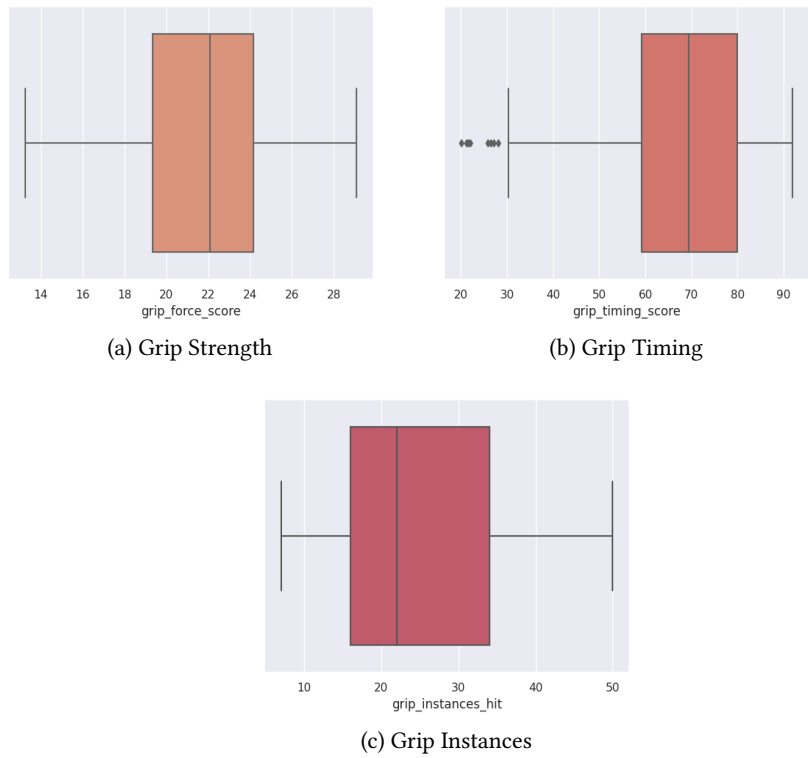


Figure 29: Range of therapeutic scores for activities

When scores are tabulated in accordance to passing and failing conditions, 62.7% obtained a positive score, though 1312 sessions ended with scores below zero. Out of all 'passes', only 13.9% managed to match or exceed every therapeutic target (Fig. 30a). The activity duration, while not considered in the final score, is a good indicator of patient engagement with the game. The underlying data in Fig. 30b shows that 60 second activities have the best engagement, with a 61.4% completion rate, while 45 second activities are only completed 40.8% of the time.

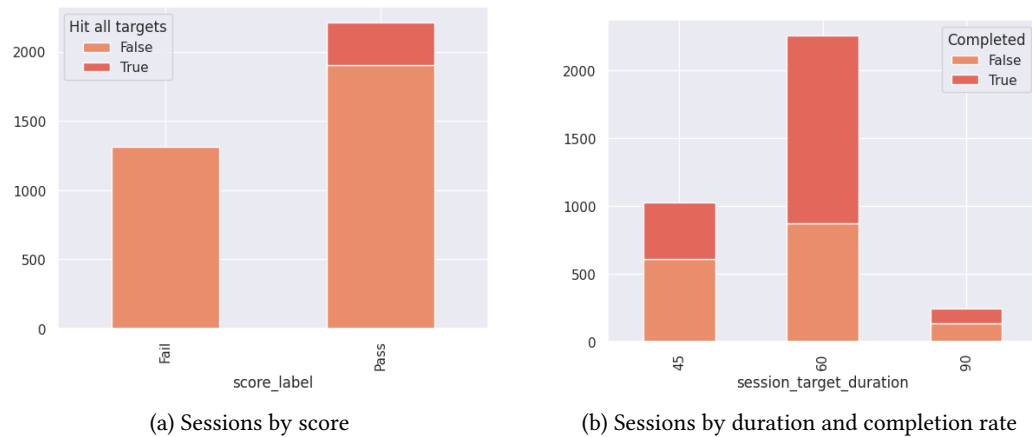


Figure 30: Success rate of activity sessions

As a side analysis, the distribution of scores per game was also explored, revealing that all games are quite balanced in terms of performance achieved by the players. GripMusic is the game in which the record score was achieved, at 47.4 points, and the lowest score of -38.6 was recorded on a GripRun session (Fig. 31). In general, lower scores are achieved in GripMusic, likely due to the fact that it is also the game with the highest difficulty spread, as shown before.

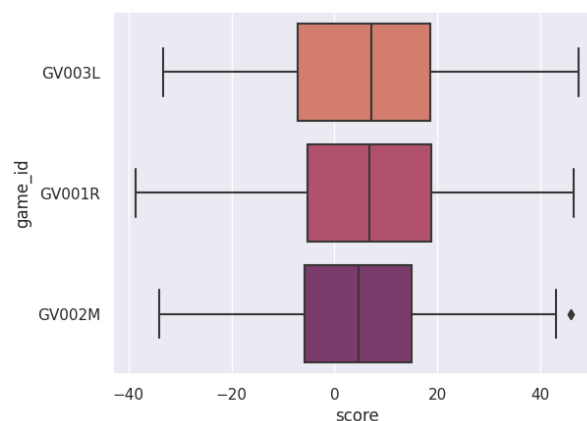


Figure 31: Score distribution per game

Finally, the feedback system built into the system provides very useful information regarding the overall enjoyment of the patients towards the sessions. In its current state, the system was well received by the majority of patients, with an average enjoyment level of 4.1 out of 5. In fact, 46% of all activities were marked with a satisfaction of 5 (Fig. 32a). When looking deeper into these evaluations per game, the data reveals that GripMusic is generally the most enjoyed game, with 59.6% of sessions being marked a 5 out of 5 for satisfaction.

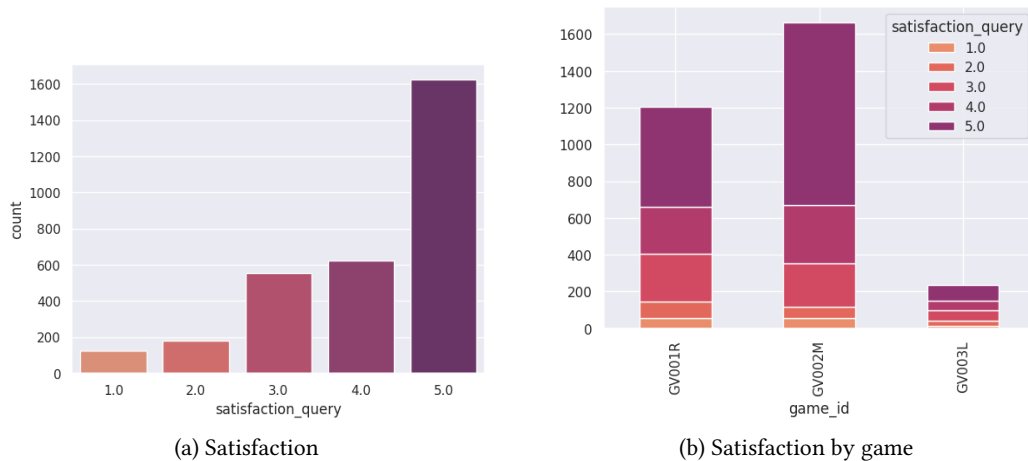


Figure 32: Patient satisfaction

### c.3.3 Merged Exploration

This is a shorter exploratory section dedicated to finding interesting distributions and relationships between variables of both datasets. As both datasets contain a '**patient\_id**' field, data merging was done through this key attribute, with duplicate information being kept.

To begin, I wanted to understand if patients with different levels of frailty assessments would perform differently, and found that the rate of activity success (positive score) is relatively constant across all FRIED levels (Fig. 33), at around 60-63%. This is a good indicator that the different levels of difficulty are doing their job in keeping activities challenging but not overly complicated at the upper ranges.

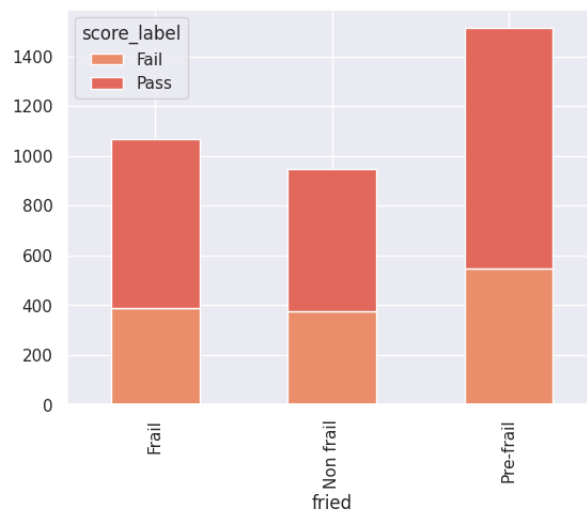


Figure 33: Activity success rate per FRIED level

Next, I studied the effects of specific risk factors in the corresponding therapeutic target scores. The data shows that, as expected, players with specific frailty factors have poorer performance in some aspects of gameplay, for example, exhausted patients have, on average, a worse timing score than non-exhausted patients (Fig. 34a). The same thing happens with patients with abnormal grip strengths having lower average scores in that therapeutic target (Fig. 34b). The remaining risk factors (sedentism, low balance and involuntary weight loss) are not traceable by the available activities, so I excluded them from this evaluation.

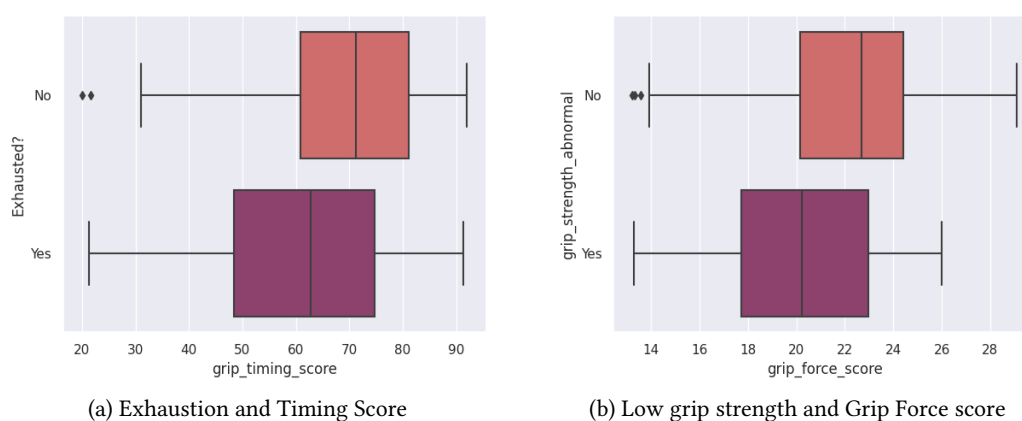


Figure 34: Risk factor impact on activity targets

The remaining high-risk factors are not easily measurable by a specific target in the activities, making its impact on patient performance harder to measure. For this, I analyzed the overall activity score distributions by discriminating patients with and without each

of these factors. From the analysis on Fig. 35a, it's not possible to discern any relevant impact on the scores of patients with incapacity to keep balance. Considering that all available games are played with a handheld controller, this is not unexpected. Contrarily, a dip in the scores is noticeable on patients who showcase sedentary behaviour (Fig. 35b), especially on the left side of the central limits depicted by the box. This means that low scores observed within sedentary patients are even lower than those of active patients. A similar phenomenon is observed in patients with reportedly unwanted weight loss, though the score range is larger on this subgroup (Fig. 35c).

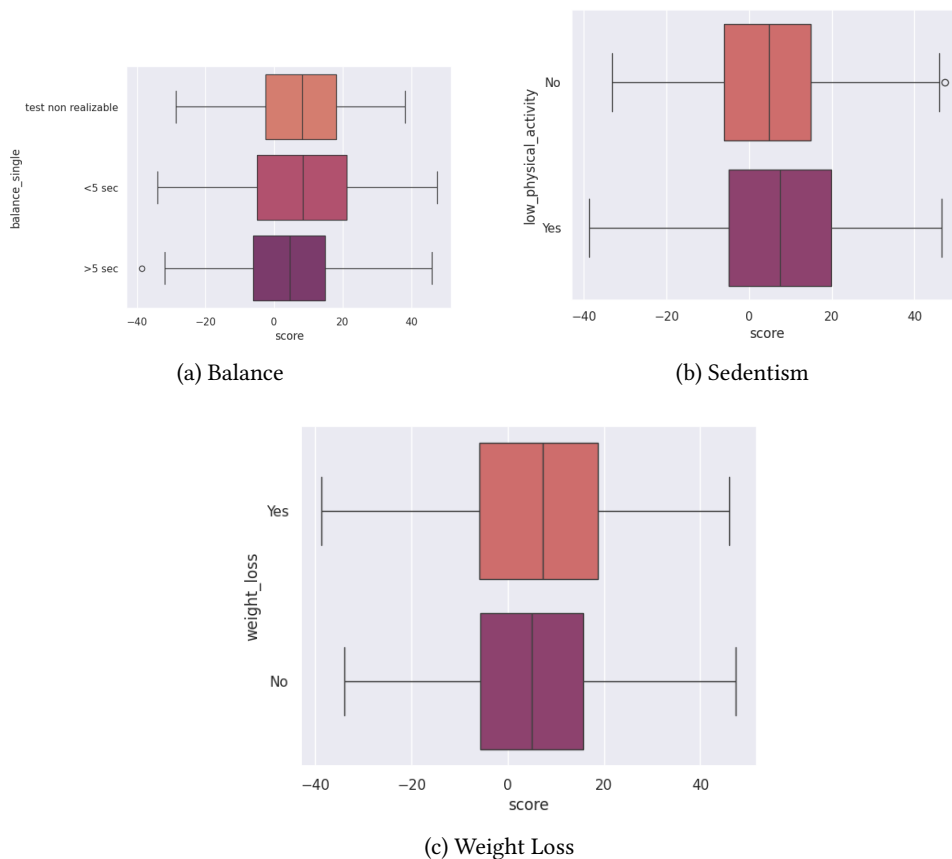


Figure 35: Remaining risk factors and overall activity scores

To close out exploration, I wanted to understand if patients with different FRIED assessments tend towards playing one specific game, and found that each game is a favorite amongst a specific frailty severity group (Fig. 35). Patients in a pre-frail state gravitate towards playing GripLand, with 60.1% of its sessions being played by people in this category. Non frail patients prefer the more challenging game, GripMusic, with a 33.6% share of sessions played, where as frail people concentrate more on GripRun, which, curiously, seems to be the middle ground in terms of difficulty, with 31.9% of sessions played.

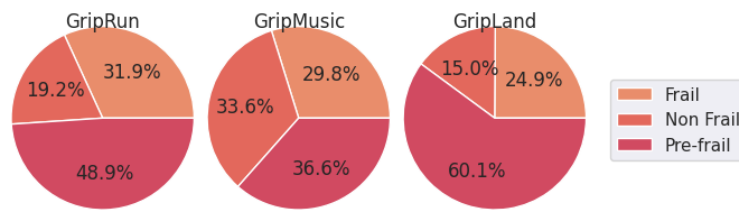


Figure 36: Share of sessions played by FRIED level

#### C.4 DATA QUALITY

The last section of this appendix documents the data quality assessment process. Due to having similar formats, I opted for an approach that is compact and consistent for both datasets, with goal of identifying:

- Measurement Errors: Usually data that is correct, but was measured in a way that is not adequate for data mining.
- Data Errors: With a focus on typos and inconsistent names for attributes.
- Coding Errors: Like multiple units of measurement present for a single attribute.
- Dictionary Inconsistencies: Columns that store values with a different meaning to what was expected.
- Missing Values: Which, if present, may lead to discussing replacement or elimination strategies.
- Deviations: Also known as outliers. Depending on the data mining goal, they may need to be removed to avoid working with noisy data.

I've split the process into three short tasks, explained below. The documentation present throughout this section is focused on recording all findings, whereas the main document section will be used to formulate reasons behind the quality gaps found, and make decisions on how to proceed in future phases.

##### C.4.1 Structural Analysis

A structural analysis deals with all issues that are not related to the data itself, but rather the meaning, formatting and measuring of each attribute.

Firstly, I went over the unique value counts for each dataset, and used the corresponding value to identify **redundant** (1 unique value), and **empty** (no unique values) variables. This process also reveals naming inconsistencies. I found that the patient dataset has 8 empty columns, as well as an two index columns that have no relevant information. The activity dataset has one index column with no relevant information. Studying each dataset for its data types and value ranges also revealed any potential measurement or coding-specific errors. All findings regarding structural problems in either dataset are documented in Tab.24. I refer to attributes in the patient dataset as *Patient.attribute*, and attributes in the activity dataset as *Activity.attribute*.

Attribute	Type	Issue	Solution
Patient.Unnamed: 0.1	int64	Index column with no significance	Remove from dataset
Patient.Unnamed: 0	int64	Index column with no significance	Remove from dataset
Patient.Sexual Habits	float64	Empty column	Remove from dataset
Patient.Religion	float64	Empty column	Remove from dataset
Patient.Race	float64	Empty column	Remove from dataset
Patient.E-Mail	float64	Empty column	Remove from dataset
Patient.Address	float64	Empty column	Remove from dataset
Patient.NHS Number	float64	Empty column	Remove from dataset
Patient.Last Name	float64	Empty column	Remove from dataset
Patient.First Name	float64	Empty column	Remove from dataset
Patient.age	int64	Measurement: Numerical is not an adequate way of measuring age for data mining problems	Replace age with a calculated year of birth
Patient.exhaustion_score	int64	Dictionary: Attribute stores a binary value, whereas its name suggests a scale	Codify values to an adequate binary field, and rename attribute for clarity
Patient.sleep	object	Coding: Field contains a 'nan' category	Consider as missing value
Patient.activity_regular	object	Coding: Field contains a 'nan' category	Consider as missing value
Activity.satisfaction_query	float64	Data: Values are in the integer range	Change data type

Table 24: Structural deficiencies in both datasets

### C.4.2 Missing Values

To allow an accurate count of missing values in both sets of data, I temporarily applied some rules to accommodate the issues present in the sleep and regular activity fields. As mentioned before, both of these variables have an extra category, 'nan', that needs to be removed. All rows with attributes belonging in this category will instead be mapped to a python **None**, which is essentially a null value and allows discovery by Pandas' missing value function. In the end, it was revealed that both Patient.activity\_regular and Patient.sleep have 11 missing values. In the activity dataset, there are 418 missing values present in the satisfaction query field, which makes sense considering that not all sessions prompt the user to rate their experience, or that patients sometimes don't feel like answering. A more detailed view of missing values per dataset is available in Tab. 25.

Attribute	Missing Values	% of Total	Action
Patient.sleep	11	1.2%	Replace with mode
Patient.activity_regular	11	1.2%	Replace with mode
Activity.satisfaction_query	418	11.8%	Categorize missing values with a zero, for 'did not answer'.

Table 25: Missing values in both datasets

### C.4.3 Outliers

Finally, I analyzed both datasets for the presence of outliers in some of the continuous attributes. It is important to note that for clustering tasks, some outliers can be useful for identifying groups of individuals with abnormal characteristics, whereas the presence of such rows can constitute noise and sway the results of classification models, highlighting the importance of dealing with deviations according to the proposed goal for the project.

The patient dataset has many variables that are either categorical or bound by a scale (BMI can be split into healthy / unhealthy, depression, pain and anxiety have their own scores), limiting the need to perform an exhaustive search. This leaves only the weekly alcohol and daily medication intake as the targets of deviation analysis for this set. The distributions can be seen in Figs. 37 and 37b, and reveal a spread of outliers throughout

the alcohol variable, and that only two patients deviate from the more common number of simultaneous medications.

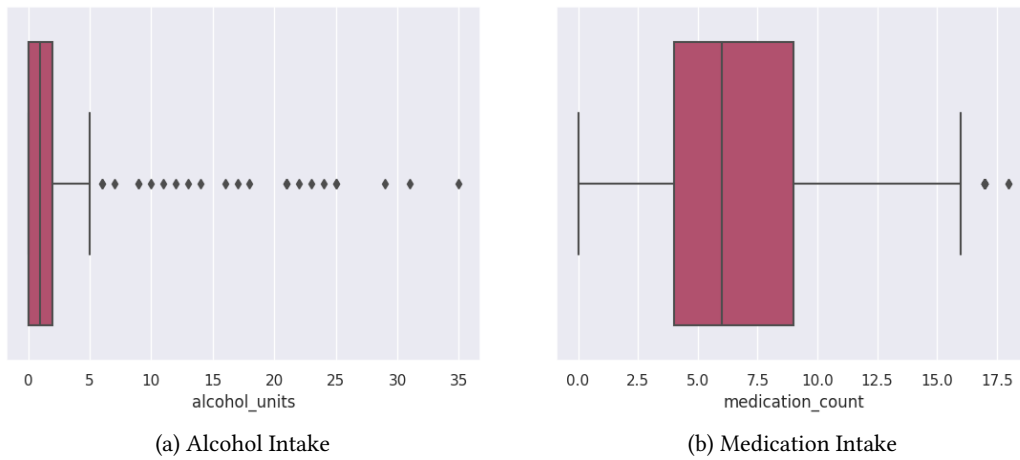


Figure 37: Risk factor impact on activity targets

Activity data is more interesting to explore, since the presence of outliers in scoring attributes may signify that some users can outmatch the remaining patients. Outlying sessions should be removed from any modelling task, since they are prone to causing under-fitting due to skewed relationships between variables. Thankfully, I already assessed that the activity difficulty is actually quite well fitted for each patient, and only detected the presence of 8 outliers in the component designed for measuring the timing accuracy. I still included every chart below, in Figs. [38a-38d](#).

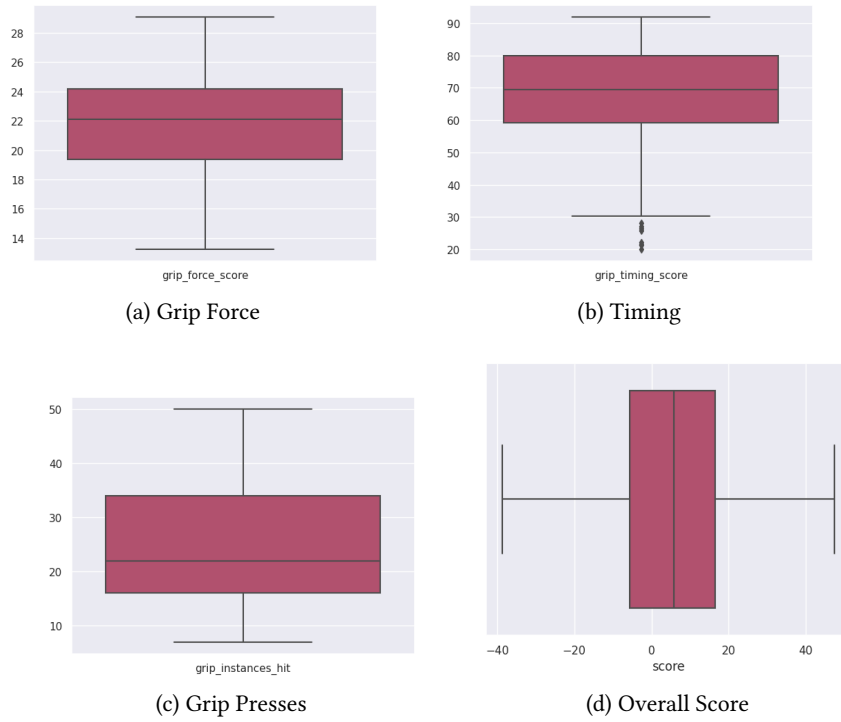


Figure 38: Outlier assessment for activity scores

## APPENDIX D - DATA CLEANING REPORT

---

This file is a project deliverable related to the Data Preparation task of the [CRISP-DM](#) methodology. It will be used to document the process of cleaning the data according to the findings raised during quality assessment. More specifically, the methods used for solving the following issues will be detailed:

- Structural Errors
- Missing Values
- Deviations

The purpose of this deliverable is enabling the tracking of modifications done to the original data, excluding any engineering processes that may also be associated with the Data Preparation phase. For portability, the tasks carried out for the entire Data Preparation phase, including data cleaning, were performed within a python notebook. Please refer to the related notebook for source code regarding the cleaning process.

### D.1 STRUCTURAL ERRORS

The first course of action in cleaning the data was to resolve any structural issues with either of the sets. As a reminder, these issues may be present in the form of unclear data, invalid categories, and incorrect data types. The identification of attribute engineering tasks may also pertain to structural deficiencies in the data, but these will be resolved in the appropriate section.

#### D.1.1 *Renaming Attributes*

Both datasets differ slightly in regards to the adequability of the names used for the attributes. Despite not necessarily jeopardizing the usability of the data, the presence of unclear attribute names makes data harder to use in the long run. For this reason, I decided to rename a few columns of the patient dataset. All modifications are identified in [Tab. 26](#).

Attribute	Renamed To	Rationale
Patient.weight_loss	Patient.unwanted_weight_loss	Clarify the risk factor being assessed
Patient.exhaustion_score	Patient.is_exhausted	Correctly depict the attribute as binary
Patient.balance_single	Patient.balance	Remove unnecessary complexity
Patient.sleep	Patient.sleep_issues	Clarify the risk factor being assessed
Patient.depression_total_score	Patient.depression_phq9	Identify the scoring scale
Patient.anxiety_perception	Patient.anxiety_gad7	Identify the scoring scale
Patient.activity_regular	Patient.activity_frequency	Clarify the measurement
Patient.smoking	Patient.smoking_habits	Clarify the risk factor being assessed

Table 26: Attribute renaming process

No renaming operations were needed on the activity dataset.

#### D.1.2 *Categorical Correction*

During quality assessment, some structural issues pertaining to categorical attributes were identified. This is a more serious issue, since modelling invalid data will inevitably lead to errors or deficient results. Similar to before, the issues were present only on the patient dataset, and were solved as described in Tab. 27.

Attribute	Issue	Solution
Patient.sleep_issues	Invalid category	Mapped invalid values to <i>None</i>
Patient.activity_frequency	Invalid category	Mapped invalid values to <i>None</i>
Patient.is_exhausted	Inadequate value scale	Mapped numerical values to binary categories <i>Yes</i> and <i>No</i>

Table 27: Categorical correction process

### D.1.3 Data Type Correction

In the activity dataset, one attribute was measured with an inadequate data type. The correction of this issue was not as straightforward as the ones before, because a direct data type modification follows the assumption that the attribute is free of missing values. In this case, the correction outlined in Tab. 28 was performed after the replacement of missing values for that attribute.

Attribute	Original dtype	Correction	Rationale
Activity.satisfaction_query	float	integer	The scale for satisfaction queries does not allow decimals

Table 28: Data type correction process

## D.2 MISSING VALUES

Given the level of completeness of the data, dealing with the few existing missing values through replacement is the most adequate decision. Missing values exist in both datasets, and the replacement method for both cases varies. In patient data, the 1.2% NA presence in both Patient.sleep\_issues and Patient.activity\_frequency will be replaced with the mode for the corresponding attribute, leading to only a very minor skew of data towards the replacement value. In opposition, the decision to consider missing values a valid category in Activity.satisfaction\_query ensures that ratings will not be biased. The values used for replacement are present in Tab. 29.

Attribute	Missing Values	Replacement Strategy	Replacement Value
Patient.sleep_issues	11	Mode	No sleep problem
Patient.activity_frequency	11	Mode	> 2h & < 5h
Activity.satisfaction_query	418	New Category	0 (Zero)

Table 29: Missing value replacement strategy

### D.3 DEVIATIONS

To finalize data cleaning, the outliers present in activity data rows will be excluded, as to avoid misdirecting the predictive models. As a reminder, the decision to keep outliers found in patient data attributes is motivated by the clustering task to be performed, which may reveal more interesting distributions. The 8 outlying rows were found in Activity.grip\_timing\_score, and represent values in the lowest 1% quantile for values in that attribute (sessions with abnormally low timing scores). Upon the removal of deviations, the data is clean and exempt of errors, enabling the following preparation tasks to be performed with minimal issues.

## APPENDIX E - DATA ENGINEERING REPORT

---

This file is a project deliverable related to the Data Preparation task of the CRISP-DM methodology. It will be used to document the process of engineering the data to enable the application of the modelling techniques predicted for this project. Depending on each set of data, the following tasks may be applicable:

- Derived Attributes
- Record Creation
- Data Transformation
- Data Normalization

During this process, a significant amount of changes will be performed to each dataset. As such, the main objective of this appendix is logging the modifications done to either dataset, leading to the engineered versions saved in the end. For portability, all tasks carried out for this task, alongside all other Data Preparation tasks, were performed within a python notebook. Please refer to the related notebook for source code regarding the engineering process.

### E.1 ATTRIBUTE ENGINEERING

The first engineering task involved the generation of attributes that either solve outstanding issues with the data, or provide useful information for future processes. The following subsections provide further details on each of the attributes generated.

#### E.1.1 *Year of Birth*

In its first iterations, the patient dataset contains an attribute that records the age of each patient. During the quality assessment process, this attribute was considered a structural error due to the problems it may originate in the future, since anytime a birthday happens, the attribute needs to be updated to reflect the patient's new age. Instead, calculating the

patient's birth year is a permanent solution, since it is not mutable. For this, the reference year of 2023 was used in the calculations (year of data collection).

### E.1.2 *FRIED Count*

Knowing the sum of a patient's risk factors was very useful during Data Understanding, allowing me to easily study the impact that FRIED symptom parity has on patient's performance in activities. I have decided to make this attribute permanent for its convenience in assessing the results of some models, such as validating whether the proposed clusters can be easily discerned by the prevalence or lack of risk factors. Below, in Fig.39 the distribution of this new variable can be observed, though a similar graph is present in Appendix C.

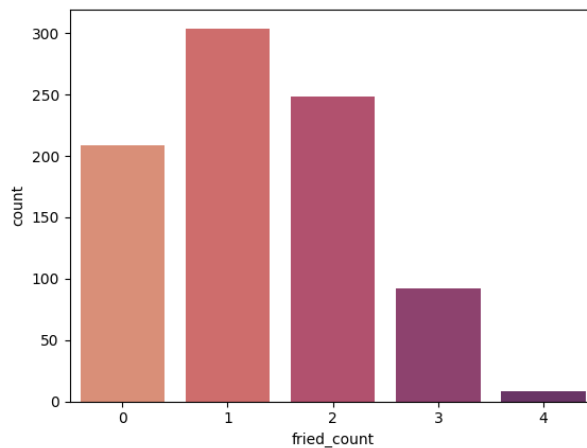


Figure 39: Count of risk factor coexistence

### E.1.3 *Engagement Score*

Measuring a patient's engagement with the activity enables the tracking of a target mentioned in the business objective list. According to the initial introduction to the data glossary, a patient is considered to have satisfactory engagement when they participate in the activity for at least 95% of its proposed duration. With this in mind, I derived an attribute that simply relates the proposed and real length of activities. Plotting this variable (see Fig. 40) leads to the discovery that approximately 30% of all patients fall under the proposed threshold, justifying the business objective related to the improvement of this metric.

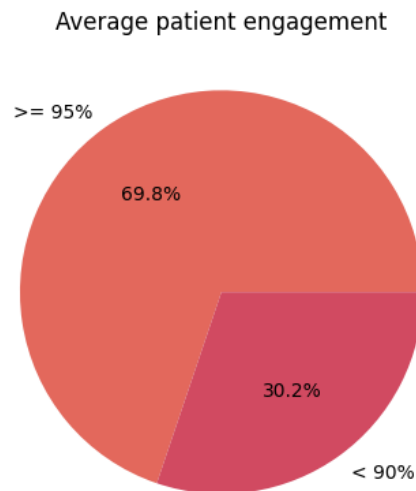


Figure 40: Split of patients with satisfactory engagement scores

## E.2 DATASET PREPROCESSING

Due to the multitude of modelling techniques I wish to employ to achieve the proposed goals, the data preparation phase included the construction of two additional data sources derived from the original sets, containing pre-applied formatting and merging operations to reduce the number of repeated actions during modelling. The data operations employed to generate these auxiliary data sources are detailed below.

### E.2.1 *Clustering Dataset*

During my initial tests in search of ideal combinations of features for clustering, I found myself constantly re-applying the same set of rules to the merged data. The process includes the standardization of all numerical variables and the removal of attributes of little relevance for the task. While this task isn't necessarily arduous and could be repeated, I have opted to save a preprocessed version of the dataset for reasons of consistency and reproducibility. When standardizing the scale of the data, I froze the weights of the scaling function to a permanent file which can be loaded on command. This means that regardless of the amount of new rows appended to the data, the consistency of the scaling function is guaranteed, which will prove particularly useful when a single row needs to be confronted against the entire data (KNN classifier). Circling back to the topic of attribute selection, I have decided to remove all attributes that satisfy any of the following conditions:

- The attribute did not represent a significant impact on any of the risk factors associated with frailty
- The attribute is, by itself, a method of classification
- The attribute relates to demographic data
- The attribute is engineered

The list of attributes removed is as follows: *fried, fried\_count, gender, hospitalizations\_one\_year, vision, audition, bmi\_score, sleep\_issues, smoking\_habits, alcohol\_units, medication\_count, patient\_id, birthplace, birth\_year*.

In total, 16 attributes were kept as relevant for clustering, out of which 10 are numerical and 6 are categorical. The resulting data resource has been saved locally.

### E.2.2 Recommendation System Dataset

The recommendation system is the most demanding modelling task for the project, and will derive its initial set of proposed activities based on data from existing patients. However, one patient can participate in many activities, and achieve a significantly wide range of scores, making a line-by-line analysis very difficult. To solve this issue, this aggregation task combines activity rows by two keys: patient and game, producing a much smaller dataset in which it is possible to observe central trends (average scores) for each patient, for each game they've played at least once. A major upside to this is the simplification of processes related to finding ideal candidates for the suggestion system, since data is prematurely averaged in a way that promotes easier identification of the most popular game and difficulty among a given set of individuals. The resulting file is considered an auxiliary data resource has also been saved locally. It is comprised of 1584 rows and 10 attributes.

## APPENDIX E - TEST DESIGN

---

This file is a project deliverable related to the Data Modelling task of the [CRISP-DM](#) methodology. It will contain the steps taken to evaluate the models produced, documenting questions regarding train/validation data splits, measures for evaluating supervised models, and, likewise, measures for evaluating unsupervised models. More specifically, the following sections will be included:

- Test Data
- Supervised Model Testing
- Unsupervised Model Testing
- Failing Criteria

The practical application of the test design to each of the developed models can be consulted in the Data Modelling notebook, and any resulting remarks and conclusions about model viability are available in section [4.4.4](#) of the main document.

### F.1 TEST DATA

Due to the nonexistence of dedicated data for the purpose of model testing, all models will be evaluated on a random 20% split of the original datasets, whenever needed. For supervised model testing, this confines the measurement of performance metrics to the application of the model to the entire test data. In unsupervised models, conclusions are to be drawn upon the observation of model results to a sample of the test data.

### F.2 SUPERVISED MODEL TESTING

The following test design is meant for the assessment of the supervised models mentioned below:

- Hybrid [KNN](#) Classifier

It includes specifications on the steps to be taken to ensure a proper and repeatable evaluation of the aforementioned processes.

### F.2.1 *Quality Metrics*

With the unsupervised task related to the clustering of patients, ground-truth labels were generated, enabling the measurement of specific error rate values of the **KNN** Classifier. In this case, the most accessible metric is accuracy, given by the ratio between the number of correct predictions and the total number of predictions (see Eq. 4 below).

$$A = \frac{\sum TP + TN}{\sum TP + FP + FN + TN} \quad (4)$$

Alongside accuracy, the performance of the method is also important to assess, with the perspective that the number of patients will increase steadily. For measuring the performance of the **KNN** Classifier, it makes sense to divide the process into two logical steps. First, the time it takes for preprocessing steps related to ensuring the data conforms with the specified formats. Secondly, the nearest-neighbor algorithm itself, using different numbers of neighbors to evaluate the sensitivity to this parameter.

### F.2.2 *Test Process*

The test design for this model foresees an exhaustive process to maximize all quality metrics in function of an ideal number of K neighbors for its calculations. To ensure reproducibility of results and minimize the impact of human perception, the following steps should be followed rigorously when assessing the model:

- **Stage 1 (Benchmark):** Evaluate all quality metrics for a 20% split of the data (172 rows). Use  $K = 14$  as a rule of thumb value based on  $K = \sqrt{x}/2$ , where  $x$  is the number of rows in the original data.
- **Stage 2 (Exhaustive):** Extensively test all K values from 5 to 86 (10% of the data size). Record all quality metrics for assessment.
- **Stage 3 (Optimize):** Analyze the previously collected data. Aim to optimize K in function of highest accuracy for least classification time.

### F.2.3 *Failing Criteria*

For the supervised methods, failure criterion can be defined as numerical thresholds on the previously evaluated quality metrics. The failure to meet or exceed these thresholds signifies that the model does not comply with the expected quality, and thus incurs in a risk specified in the risk analysis section of the main document (section 4.1.2.3). For the specific case of the **KNN** model, its approval is vital to the project, since most unsupervised techniques are based on the identification of patterns within a cluster or neighbor characteristics. To achieve satisfactory quality, the supervised method must satisfy **all** of the following criteria: At least 85% accuracy; Less than 0.5 seconds of preprocessing time; Less than 1 second of classification time.

In the event that no combination of parameters meets the expected quality, the K-Prototypes Clustering method is considered to be faulty (not the **KNN** algorithm) and should be re-assessed. The limit of reiterations with readjusted parameters is not given by a specific number, but rather by the calendar constraints of the modelling phase.

## F.3 UNSUPERVISED MODEL TESTING

The following test design is meant for the assessment of the unsupervised models and processes, namely:

- K-Prototype Clustering
- Therapeutic Difficulty Engineering (*process*)
- Recommendation Algorithm
- Performance Analysis Tool

It includes specifications on the steps to be taken to ensure a proper and repeatable evaluation of the aforementioned processes. Given the multitude of unsupervised mechanisms, some model-specific rules may be applied.

### F.3.1 *Quality Metrics*

As the name indicates, unsupervised methods do not rely on a set of ground-truth labels for the assessment of their quality. For this reason, defining objective quality measurements is not possible, and their evaluation will instead be based on subjective concepts that can not be easily quantified. Still, unsupervised approaches far outweigh supervised ones in my

project, and their importance is undisputed, making the definition of subjective metrics a task that should not be ignored. For all unsupervised methods, the most important measure of quality is the ease of deployment of a particular model. Since all of the systems foresee a modular use, the models should not contain design choices and formatting requirements that would hinder their usability in real scenarios. Similarly, the ease of interpretation of the results must take in consideration that target population (therapists and, by association, elderly patients), requiring a digestion of the data in favor of user-friendly results. Finally, the models' performance should be as close to real-time as possible, as severe downtime makes for a system frustrating to use for extended periods of time.

### F.3.2 *Test Process*

The test process for unsupervised models does not need to be standardized, since the only measurable metric is performance. The instructions detailed below are merely a formality and a reflection of the method I will use for assessing these models in the respective notebook section.

- **Performance Assessment:** Where performance measurement is applicable (Recommendation Algorithm, Performance Analysis Tool), perform an arbitrary number, N of sample runs and record the average elapsed time from start to finish. In the recommendation algorithm, exclude all time dedicated to recording user input.
- **Ease-of-Interpretation:** In models that lack direct user operation (Clustering, Reverse Engineering), perform a simplified data exploration process to validate the results and extract new insights. Additionally models that require human interaction must contain language fit for non-technical personnel to easily interpret.
- **Portability:** Evaluate all processes in regards to their data, structural and technical requirements. In models with common traffic (Recommendation Algorithm, Performance Analysis Tool), assess how viable the system is for deployment.

### F.3.3 *Failing Criteria*

This type of model varies greatly in goal and results, thus making the definition of failing criteria better suited for a case-by-case basis. Below, some requirements for the initial approval of each model are specified:

- **K-Prototype Clustering:** Each cluster must have one or more discernible features. If two or more clusters are indistinguishable, rerun the model with adjusted parameters (different K value, different set of attributes).
- **Therapeutic Difficulty Engineering:** The resulting auxiliary file must be easily understood by technical and non-technical personnel. Similarly, loading and lookup times for this file should be as short as possible. Different file formats can be tested to meet with performance expectations, as long as they do not incur in a steep difficulty increase for deploying the model.
- **Recommendation Algorithm:** The auxiliary structure for storing recommendations and supporting the feedback module must be easily understood by technical and non-technical personnel, and its access times as close to real-time as possible. Additionally, the model should be able to produce recommendations in less than 2 seconds, and must do so with the least function parameters as possible, as to minimize the possibility of human error.
- **Performance Analysis Tool:** Ease-of-use is absolutely necessary, since this model is targeted exclusively for non-technical personnel usage (therapists). The method should require the least number of user-submitted parameters as possible while still producing results in less than 2 seconds per iteration.

The non-compliance of any model with the proposed quality metrics is, by itself, also subjective. In situations where the ease-of-use of a specific model is a requirement, external validation in the form of clinical trials is needed to ensure the usability of the system.



## APPENDIX G - PROJECT EVALUATION

---

This file is an auxiliary file related to the Evaluation task of the [CRISP-DM](#) methodology. It will be used to document relevant auxiliary processes for the evaluation of the data mining project, and as a place to store additional context and reasoning behind some of the final choices present in the main document. This deliverable is not a standard process of the [CRISP-DM](#) methodology, and its creation stems from a personal necessity. Its content includes:

- Business Success Evaluation
- Data Mining Process Summary

References to this appendix will be present Section [4.5](#) of the main document. The following subsections are exclusively comprised of text documenting introspective commentaries, rationales, external validation processes and critical analyses, making this appendix a heavier read, but one necessary to contextualize the decisions made in the project evaluation.

### G.1 BUSINESS SUCCESS EVALUATION

This section contains complementary information about the decisions made in Section [4.5.1](#) of the main document. First, the model evaluation meeting is summarized, where particular attention is given to its outcomes and business statements about the project results. Then, a discussion topic is introduced, where I compile this information with my own thoughts and visions on the project to provide logical reasoning behind the achievement (or non-achievement) of certain business goals.

#### G.1.1 *Result Presentation Meeting*

This meeting is an event that took place shortly after the model assessment phase (Section [4.4.4](#) of the main document). In it, myself and the project manager for GameVitae (Techframe CEO) have informally discussed the project thus far with the purpose of answering the following questions:

- Are the models useful from a business perspective?
- Do the project results answer to the business objectives?
- Are there any shortcomings in the project methodology?

The discussion that lead to the answers is an excellent piece of additional context for the evaluation phase of my project, hence why I have decided to summarize it in the following subsections.

#### G.1.1.1 *Models*

The first subject of the meeting included a short demonstration of the model's capabilities, and was done in two moments: (a) first, examples that were previously prepared by me were shown. These were controlled "runs" of the models whose outputs I have prior knowledge of, and were simply meant to demonstrate the basic functionality of the methods. (b) Then, an interactive demonstration took place, where the participant (project manager) could freely manipulate the inputs to observe the behaviour of the system, and ask questions about some of the intermediate processes and files. Throughout this demonstration, the models were very well received both in terms of their usability in the business and their perceived performance. The main takeaways from the project manager at this point in the meeting were:

- The cluster profiles make sense logically and match the expectations the business had in terms of quantity, size and predictor variables.
- The constraints set for activity target ranges help standardize activity setup process without over-complicating the process for the therapists involved. The idea of using previously available data for the definition of these metrics also helps in ensuring past activities are compliant with the new rule sets (backwards compatibility).
- Using the recently found information about patient neighbors as the basis for generating activity settings for completely new users is smart, and avoids overwhelming patients with questions about their preferred activities.
- The feedback system employed to gradually adjust recommendations is adequate at this stage of development, but might not capture the full spectrum of patient needs further down the production line.
- Caution must be used in abundance when allowing patients the control to optimize their experience. This is to avoid situations in which the patient can constantly request easier activities to make it seem like their scores are improving drastically.

- The performance analysis tool is much more efficient than the methods used previously, and the addition of cluster-wise comparisons will allow for a better monitoring of patients incurring in edge cases of improvement, degradation, or counter-therapy.

#### G.1.1.2 *Business Objectives*

Next, I presented an incomplete version of Table 16 (last two columns empty) to the project manager, and collaboratively filled the missing points while discussing the results of the project from a business point of view. This portion of the meeting was extremely fruitful because it allowed me to cover a significant amount of thinking work in a short time, since it is the project manager's responsibility to have a clear vision of where the project should go. During this discussion, a need to classify some goal achievements as "subjective" was felt. In one hand, we were able to come up with the necessary logical reasoning for considering some objectives as being met, while in the other, there was no tangible evidence that our reasoning would translate into the reality. Most of these situations of uncertainty were caused by a lack of practical testing (due to insufficient scheduling), and can only be definitively answered once the models are deployed in a future trial. Overall, the project was considered successful by the project manager, since compelling reasoning was found to support the complete or partial achievement of each objective.

#### G.1.1.3 *Methodological Shortcomings*

To close out the meeting, I offered a general rundown of the methods I used for the assessment of the project success, and gave particular attention to tasks, processes or steps I have skipped due to difficulties of a technical, calendar or scope nature. While I believe I was able to faithfully reproduce and follow the [CRISP-DM](#) methodology in this regard, the most evident issue is the lack of proper experimentation of some models by the personnel for which it is envisioned. This was the root cause of my difficulties with evaluating the quality of the models, and why some of the objectives were only achieved subjectively. As a future solution to this, the project manager proposed that a follow-up evaluation process be conceptualized and employed shortly after the first deployment phase, and its results be documented as part of a trial on the *usability of modern systems by the elderly and non-technical*. This would not only motivate future work in ensuring GameVitae's systems are compliant with user-friendliness necessities, but also contribute to the state-of-art in finding appropriate design choices for gamified platforms aimed this demographic.

### G.1.2 *Rationale for Objective Completion*

Below, each paragraph will be used to contextualize the objective and why I believe it has been achieved. The following reasoning was raised in collaboration with GameVita's project manager as a product of the result presentation meeting.

**BO.01 - IMPROVED RETENTION** The first objective relates directly to the 'engagement' metric (real and proposed activity time). As defined by the business and evaluated during data engineering, the patient can be categorized as "engaged" when they complete at least 95% of the activity. Prior to deployment, I measured the percentage of engaged patients to be at around 70%, which means this number should increase to roughly 85% (15% increase) for the success of this objective. The business believes this can be gradually achieved with deployment of intelligent activity suggestions based on patient feedback, which promotes a personalized experience and, with time, a better match for their interests and physical capabilities. With this in mind, the business agreed to mark this objective has **subjectively achieved**, pending confirmation from an actual trial for measuring its success.

**BO.02 - IMPROVED ENJOYMENT** The business wants to improve the average activity rating from 4.1 to 4.5 shortly after deployment. This success of this objective falls into the same reasoning as the previous one, where a belief that taking actions based on user feedback will eventually contribute to a much improved user experience and, in turn, allow for higher ratings, when patients feel like they are being listened to. Similarly, this objective is considered as **subjectively achieved** for the reasons stated before.

**BO.03 - LESS COUNTER-THERAPY** Counter therapeutic scenarios can happen in a variety of ways. A patient may be overexerting themselves and thus contributing negatively to their recovery, taking part in an excessive number of activities, or participating in tasks that are too easy for their physical level. All of the previous situations translate directly into therapeutic improvement not being achieved, and must be reduced to a minimum. The way that my project can contribute to solving this issue is by providing better methods for therapists to identify when patients are over/under-stimulated, and relates to the performance analysis tool I have developed. After the demonstration of this tool, this objective was marked as **subjectively achieved** due to its perceived benefits in processing time and depth of analysis, though it is currently impossible to assess whether the metrics studied are efficient in preventing counter-therapeutic situations from occurring.

**BO.04 - REDUCED ANALYSIS TIME** Related directly to the amount of manual labour needed to study a patient's performance across their activity records. The goal is to make this process as efficient as possible to allow on-demand reporting, replacing the bi-weekly strategy currently in place. Given the processing time assessment done to the performance analysis tool, I can comfortably consider this objective as **achieved**.

**BO.05 - REDUCED SETUP TIME** In line with the previous objective, here the goal is to greatly reduce the amount of time it takes for therapists to define activity parameters. The current approach involves manually assessing the patient's most recent score and making adjustments accordingly, and is simply non-scalable in the long run. The main contributor to the resolution of this project is the feedback collection mechanism's interaction with the activity recommendation model, which automates the task of performing parameter adjustments based on how patients felt about the previous activity. The business was very pleased with this alternative, and thus this objective was considered **achieved**.

**BO.06 - SCORE STABILIZATION** Having stable scores across different patient profiles means that regardless of a patient's physical ability, all activities are equally fair and challenging. Because the scoring function used in GameVitae is determined by negative and positive values (with zero being an exact match with the proposed targets), it is possible to explore results deeper than "success" / "failure" and determine exactly how much better or worse a particular patient performed. This objective relies heavily on every unsupervised model's success, and is by far the hardest to assess at this point. The business and I have agreed to mark it as **partly achieved**, since all of the related models have been proven to successfully carry out the intended tasks. However, determining the most efficient exercises is a criterion that is currently impossible to assess (requires experimentation phases), and the effect of the models on the score normalization can not be validated at this stage.

**BO.07 - STANDARDIZED DIFFICULTIES** This is a more straightforward objective, and relates to a problem that was introduced in the earlier stages of GameVitae with the introduction of a therapist-defined level of difficulty for each session. Due to a lack of normalization with the difficulty scale, the process became subjective and not bound by any specific constraints, making its meaning hard to understand. The process I have carried out to reverse engineer a set of rules has been well received by the business, and is the basis for the automatic generation of activities. This objective has been clearly **achieved**.

BO.08 - PERSONALIZED EXPERIENCE Finally, one of the main overarching business objectives for the project is enabling each user to have a personalized experience. Straying away from a "one-fits-all" mentality is one of GameVitae's goals as a platform, and providing patients with activities, insights and monitoring that is tailored to their individual needs helps building better relationships and keeps users engaged. I believe my project is a leap in this direction, with the feedback collection module contributing greatly towards a tailored experience. For this reason, this last object was considered **achieved**.

As a disclaimer, all of the previous rationale was agreed upon during the meeting with the project manager. The explanations provided are subject to interpretation, though I believe the arguments made are compelling and provide enough context for the decisions taken.

## G.2 DATA MINING PROCESS SUMMARY

This portion of the appendix contains a concise revision of [CRISP-DM](#) phases tackled from sections [4.1](#) to [4.4](#) of the main document, highlighting significant contributions, difficulties and alternate decisions that can lead to a better process in future projects. Each of the following subsections is with respect to one phase, and covers the mains activities and decisions for the completion of that phase.

### G.2.1 *Business Understanding Phase*

Out of all phases in my project, the business understanding phase was the most streamlined one. Some of the expected deliverables were generated from early business ideas, while others were formed as part of an interactive process between myself and the project manager. The fact that I am an active participant and constituent of the business has greatly facilitated my access to communication, which enabled this phase to proceed with only a few difficulties, despite its longer duration.

The development of this phase begun with a background check on Techframe as a business and how they are organized to work in GameVitae, the project to which I am contributing with my work. Immediately, the stage of maturity of GameVitae came to me as surprising due to how rudimentary its data and internal processes currently are. The open-ended nature of my project meant that the majority of insights are novel and interesting to the business, but the lack of predefined processes, workflows and scope

gave every business related task a degree of subjectivity, making it hard to define and harder still to measure.

This subjectivity played a significant part in the following stages, which included the definition of inventories, goals, criteria, budgets, risks, and perceived benefits. Arguably, this was the most thorough and time consuming process of this phase, partly due to a lack of a well-defined scope for GameVitae, which meant constantly reiterating on the previously mentioned deliverables based on feedback from the supervisors. On the bright side, I believe the level of completeness I have achieved in this phase has allow for a significantly easier modelling phase.

In future projects regarding GameVitae, a better approach to business understanding is warranted. The calendar for my project suffered from its longer duration, which although needed to overcome the previously mentioned issues, has slightly took away from the stages of the project that involved manipulating the data. This issue was unforeseen and unavoidable in my project, but this phase generated much needed knowledge and scope definition, which can be used to speed up business processes in the future (i.e.: follow-up data mining projects to the platform).

#### *G.2.2 Data Understanding Phase*

The tasks carried out in this phase were mostly disconnected from the previous one, and unveiled a massive amount of new information that would shape the data preparation phase. This time, subjectivity was not as big an issue, since a baseline set of data and the necessary entity-relationship were already defined prior to my intervention.

The phase began with documenting the data collection process, which in reality was an extremely simple process of analyzing the existing structure and requesting a collation of the attributes deemed relevant for analysis. Auxiliary steps performed to grant anonymity to the data were performed internally and were not included in my project's documentation.

After loading the data into the modelling tool (in this case, a Python notebook environment), I initially had some difficulties understanding the measurement units adopted by the business, likely due to my lack of medical education, which motivated an extra step of studying sources on the medical scaling used for the assessment of psychological conditions. This ended up adding unexpected delays to this phase but has ultimately lead me to a much better position to handle the data.

What ensues in the following week is an extensive exploration process meant to reveal as many insights about the data as possible. This was one of the most valuable tasks of the entire project, and warranted the creation of a large auxiliary file for the reporting of results (Appendix C). While the technical side of exploration itself was carried out with no issues (other than some hiccups in getting some visualizations to be properly displayed), I faced some difficulties in logically arranging the documentation around this step due to having an unbalance in content between the main document text, the notebook documentation, and the entire appendix. I settled on the decision of making the main document a lighter read, which I still endorse, however, for future projects, more planning needs to be assigned to the logical flow of data exploration, since the constant revisiting made it very difficult to maintain readability.

To close out this phase, the data was analyzed for inconsistencies, errors, and other mistakes that could compromise its understanding and viability. Thankfully, the few issues that were found had were of simple resolution, and this task proceeded with no major complications. The result of this phase was essentially a segue into preparing the data for modelling.

### G.2.3 *Data Preparation Phase*

In my project, the data preparation phase was simultaneously short and technically challenging. At this point, my knowledge of the data was much deeper and some of the first ideas for the models started to come into fruition.

The phase began with the correction most of the structural issues found in the data. This was quite straightforward as only a handful of naming and glossary issues were found. A few more serious inconsistencies in the data were also resolved with no difficulties, namely the correction of categorical problems and wrong data types. Finally, with the decision of replacing missing values and discarding only the outliers in activity records, the data was deemed clean and ready for the next stage.

During engineering, I tackled attribute generation first, closing out some issues that were still pending from the quality assessment. These changes to the data were permanent and independent of the models used. Before proceeding, an entire experimental process about modelling requirements was performed to identify repetitive data formatting processes. Afterwards, these processes were compiled into unique engineering tasks that gave origin to sub-versions of the dataset that were ready to model.

In fact, the way I approached data preparation in general included saving "snapshots" of the data after every major task, not only as a way of storing backups in case anything went wrong, but also because each model expects data with slightly different characteristics. This decision ended up being useful, but it also convoluted the process unnecessarily, and became a hindrance to progress due to the amount of file organization needed to ensure every piece of code and documentation was organized logically.

The last remaining task related to data preparation was the integration of both datasets via a merge of key attributes. In a way, this process was no different than the tasks done before to prepare specific versions of the data to work on particular models, but due to the fact that attributes from both data sources were used, I considered it an integration task instead.

Looking back at this process, no significant changes in methodology or major decisions come to mind, other than the possibility of having devised better file organization for this phase, since the large number of models meant an even larger amount of intermediate datasets, which ultimately lead to some issues in organizing and loading the correct data.

#### 6.2.4 *Data Modelling Phase*

The final stage abridged in this section is the modelling phase. I'm particularly satisfied with the way this phase was organized and timed, however, this was also the point in which the effects from having issues in previous phases became more apparent.

The initial assessment of modelling techniques ended up being nothing more than a technical introduction to each of the models. At this point in the documentation, I already had a very clear idea of what I wanted to build (even having tested most of the models' logic) which confined me to discussing the techniques I knew I wanted to incorporate. The documenting of these descriptions was iterative and the final texted suffered many modifications. I found it quite difficult to stay in topic and not repeat myself during the phase, as I felt the stages for this phase were the hardest to follow logically due to the cyclic nature of the [CRISP-DM](#) methodology. In fact, the previous phases were revisited multiple times during modelling for minor adjustments, and these changes quickly became difficult to track and integrate.

Upon deciding on the techniques I wanted to employ, I started conceptualizing the test design for assessing the quality of the models I would develop. During this task, I found the definition of quality metrics for unsupervised metrics to be quite challenging, an

issue that can be traced back to the first [CRISP-DM](#) phase, and one that has been present throughout the entire process.

Once a general idea of the model requirements was documented, I began converting the previously developed drafts into finalized versions of the models, simultaneously applying the test design to ensure quality throughout the entire process. Both the model building and the model assessment stages were carried out with minimal delays, although one major flaw was present in the latter task.

The issue was mentioned multiple times throughout the main document, and relates to the impossibility of testing some of the human operated models extensively to understand their matters of usability (ease-of-use, adequability of the methods, quality of results). The lack of a practical trial does not put the deployment of my models in question, but it raises additional concerns that must be addressed in follow-up approaches to GameVitae. In retrospective, I do not see an alternative set of choices that would have lead me to having been able to carry out this assessment, but perhaps an earlier assessment of model requirements would have revealed these needs earlier in the timeline, and allowed me to prepare better, as the issue is related to time constraints rather than difficulties in arranging the experiment itself.

Overall, this issue should not detract from the quality of models, which have all been very positively received by the business, and will be resolved once new initiatives for improving GameVitae open up. Regarding the documentation and dissemination of the projects results, the business is also supportive of the decision to use a technical paper as the means of communication, which may help position the platform in the state-of-art and bring in the attention of stakeholders.

#### G.2.5 *Critical Remarks*

Looking back at the way I carried out the different phases of the project, my inexperience with the business area, allied to the early development stage of GameVitae, was evidenced by some sub-optimal decisions that have costed me a non-negligible amount of time. At the same time, this unfamiliar environment has forced me to get out of my comfort zone and come up with solutions to the unusual problems that arose, which has undoubtedly allowed for a project of great academic interest. Despite the (now obvious) decisions that lead to issues with timing and organization, I believe all of the phases summarized in this section were successful and caused a positive impact on the final models. With that said, some pointers to provide better direction in future projects were evidenced through this revision process, making its value undisputed to myself and the business.

## APPENDIX H - DEPLOYMENT PLAN

---

This file is a project deliverable related to the Deployment task of the [CRISP-DM](#) methodology. In it, the plan to integrate the project results in the business will be detailed, covering all manner of technical aspects, monitoring practices and contingency planning necessary for a successful adoption of the project. The following aspects will be discussed and accounted for in the deployment strategy:

- Deployment strategy for models and findings
- Technical details
- Planning for contingencies
- Monitoring and life cycles

### H.1 MODEL DEPLOYMENT

This first section details the strategy for deploying all of the **final models** of the project, with respect to each model's particularities, including deployment vehicles, usage habits and technical requirements.

Models are a type of result of the data mining project, and have the particularity of producing outcomes, usually in the form of classifications or predictions about the underlying data. In this project, models have been conceptualized, constructed and tested with a design philosophy that favours modularity, a decision that skews the deployment strategy towards using a common architecture for most models.

Table [30](#), available below, can be consulted as a overview sheet of the model deployment strategy, containing the essential details of each model's deployment.

To define each of the factors present in the table, a key is provided below.

- **Interactivity** Relates to the whether the model operates on a trigger that is (a) independent to the user, such as a frequency or an external condition being met, or (b) when the user performs a request specific to that model.

Model Alias	Interactivity	Deployment Vehicle	Trigger
P.01	Independent to user	API Endpoint	New game / difficulty level is added to GameVitae
S.01	On-Demand	API Endpoint	Every 12 weeks / Every new user / API Call
U.01	Independent to user	API Endpoint	Every 6 months
U.02	On-Demand	API Endpoint	Every day / API Call
U.03	On-Demand	API Endpoint	API Call

Table 30: Model deployment overview sheet

- **Deployment Vehicle** Is the technical method by which the model will be deployed or accessible. As mentioned before, in this project, all models will be part of an [API](#) that will be made accessible to the GameVitae platform, and thus all the models share this characteristic.
- **Trigger** Since not all models are exclusively operated by humans (independent), triggers are used to define the specific condition or action that needs to occur for a model to take effect.

#### H.1.1.1 *GameVitae API*

The deployment of models will be supported by an **Application Interface (API)** with several endpoints corresponding to the utilities developed in the context of this project. The existing hardware structure at Techframe favours this method of deployment, and the preexisting experience of the personnel in working with networking services will translate into lower training costs. Developing, deploying and managing an [API](#) is a technically demanding process with many aspects to discuss. This subsection will offer a detailed breakdown of all relevant technicalities of this strategy, separated by paragraphs. Similarly to before, the table below (Table 31) contains the essential information about the GameVitae [API](#).

**FRAMEWORK AND PROTOCOL** The application interface will be built on the REST architectural constraints, which fits the needs of client-server separation and individualization of the server components. The development of this service will be fully realized on the Python 3.11 programming language, using standard python functions for the definition of features, and the Flask micro web framework for the construction of the web application.

Technical Feature	Technology Used	Notes
<a href="#">API</a> Protocol	RESTful	N/A
Privacy	Hidden	Endpoints can only be accessible by authenticated users through the GameVitae platform
Language Version	Python 3.11	+ Compatible package versions
Framework	Flask	N/A
Servicing	Microsoft <a href="#">IIS</a>	On a business server
Versioning	Microsoft DevOps	N/A
Documentation	reStructuredText	Can be written in-code and parsed with <i>pdoc</i> for generating the <b>HTML</b> page
Testing	Postman	N/A

Table 31: GameVitae [API](#) technical information

Flask is favoured over other frameworks due to not requiring any additional tools or libraries, and containing only the vital necessities for web development, lacking database abstraction layers and other complexities that are not presently needed for this task. The entire functionality of application will consist on a single **.py** file, but supporting files and data references may be present in the [API](#) directory.

**AVAILABILITY** The [API](#) will be deployed using Microsoft's Internet Information Services ([IIS](#)) on a business-owned dedicated server. The application address will not be made available to the public at any point during deployment. Similarly, no web-based interface will be constructed to support the usage of the [API](#). Instead, all relevant endpoints will be accessible directly from the GameVitae platform when the business intends to conduct trials or make a public version of the platform available.

Data privacy is of great importance to the business, justifying the addition of several security layers to the [API](#), namely:

- A Network Flow Control module to prevent bot connections and distributed denial of service ([DDoS](#)) attacks.
- A Web Application Firewall to deny [SQL](#) Injection attacks.
- Access Control Rules to ensure only authenticated users can perform requests, and only authorized users can access specific endpoints.

**MAINTENANCE** To aid in making the [API](#) easy to use and extend its life cycle, documentation will be written in **reStructuredText** format directly in the IDE and exported

in HTML format for internal use. Additionally, versioning through Microsoft **DevOps** will be considered throughout the entire deployment process to keep track of hotfixes and changes to the functionality (more details in Section [H.1.1.1](#)). For testing and response time monitoring, mock requests will be frequently set up with **Postman** and logged for analysis.

**STRUCTURE** The following structure is planned for the GameVitae [API](#). It includes expected endpoints, auxiliary files and package requirement specification files needed for the correct functioning of the service.

```

gv_api.py
├── /ping
├── /cluster
├── /knn
├── /feedback*
├── /diff
├── /recommend
├── /perf
├── auxfiles
│   ├── db
│   │   ├── diff_lookup.csv
│   │   ├── merged_data.csv
│   │   └── rec_lookup.csv
│   └── scaler.pkl
└── requirements.txt

```

There are some important aspects that have been purposefully left out from the previous tree or marked with an asterisk (\*) for added context:

- **Logging File** Logging will be performed on a new [SQL](#) table in the existing GameVitae database for easier access within the business. The logging messages will all be constructed by error handlers in the [API](#), and directly written to the aforementioned table, using a secure Python to [SQL](#) connection.
- **\*Feedback Collection** This functionality is almost entirely dependent on a future front-end interface for displaying the questions and collecting the answers, making it impractical for interacting with an endpoint. Given that this is a crucial part of the recommendation algorithm, the current solution expects a string of numerical digits (equating to the answer options) provided by the user. For example, the following input:

2 3 1 13

would tell the system that the targets for grip force and grip instances were too easy in the previously played activity. The spacing denotes different questions, whereas sequential numbers indicate multiple answers to one question.

- **User Weights** Both the recommendation system and the feedback collection module perform constant accesses and modifications to an auxiliary file that stores the most recent recommendation for each user. To avoid excessive I/O operations in the machine and problems caused by concurrent accesses to physical files, this functionality will be supported by an [SQL](#) table in the existing GameVitae database. All *CRUD* operations to this table will be performed and managed by the [API](#).

#### H.1.1.1.1 *API Monitoring*

As soon as the [API](#) is deployed internally, an extremely important monitoring process will begin. Monitoring includes the collection and analysis of several metrics for the identification of usage patterns, processing flaws, peaks/dips in performance or other aspects that may affect the validity of the results. Monitoring is a model dependent task in the sense that each one requires the tracking of a particular set of factors or influences, hence the documentation of this task being of great importance to the business. In some cases, a significant change in the data (such as rapid growth) can completely invalidate a model, incurring in an situation of expiration which could have measurable repercussions in the business if not dealt with in a timely manner. It is also the duty of adequate monitoring to prepare for these changes and define the actions to be taken upon model expiration.

Table 32 contains the monitoring details by model. It's worth noting that at this point in the GameVitae development cycle, model expiration does not translate into retiring that particular model in favour of another, but rather employ alternative modelling strategies during a reduced period of time. The idea behind "refining over substituting" is to potentially find scalable solutions to the problems the business might face with the least amount of structural change as possible.

Model Alias	Measured Metrics / Influences	Monitoring Plan	Expiration Criteria	Actions
P.01	Number of games in GV Number of therapeutic objectives Query time	When new game or objective is added, rerun P.01 and test file query times automatically.	File access times > 1.0s	Two week trial with alternative file format (Feather).
S.01	User growth Classification time Classification accuracy	Every week, run internal process for measuring accuracy, performance, optimizing K-Value.	Accuracy falls below 65% for two consecutive weeks.	Three week trial with centroid-based classification. Alternative set of classification attributes.
U.01	User growth	Every 2 weeks, generate KElbow plot and distribution plots for current K-Value.	Optimal K-Value is different than 4 or cluster profiles no longer applicable to groups.	Re-evaluation of cluster profiles Alternative set of clustering attributes with reflected changes on S.01.
U.02	Activity rating Engagement Indirect user feedback	Generate feedback reports every week based on collection mechanism answers and recent activity ratings.	Average engagement lowers in three consecutive weeks. Activity rating falls below 3.0 average. Answer patterns reveal consistently inadequate recommendations.	Two week complimentary R&D stage for applying model updates based on user rating and engagement patterns. Alternative rule set for generating recommendations
U.03	Monitored user performance Usage frequency Therapist feedback	Every week, collect usage metrics to identify decreases in model reliance. Conduct biweekly feedback collection with select therapists	Number of endpoint requests lowers by 50% for three consecutive weeks. Other unexpected therapist-provided feedback.	Two week complimentary R&D stage for developing and trialing additional performance metrics and visualizations, based on therapist feedback.

Table 32: Model monitoring overview

In the future, should the business face a similar issue to those solved with this project, the models can be repurposed and redeployed, given that proper documentation regarding the use-cases and technicalities exists. For this reason, the business has also expressed that they wish for [API](#) documentation to be considered a critical aspect of monitoring, during which, small immediate changes to the models' parameters and functionalities may be applied (hotfixing) and must be adequately reflected in the documentation and in the versioning tool. Access to [API](#) versions will first be limited to main developer (myself) and the project manager only. When deployment succeeds, additional developers tied to the business may be invited to contribute with optimizations. After a model has expired, it's original code will be made available in a public repository.

#### H.1.1.2 *Contingency Planning*

This subsection deals with documenting and preparing for any foreseeable problems during the deployment of the [API](#). Given my past experience in deployment models in the business, familiarity with the process has prepared me to expect several hurdles, most of them related to issues with file access permissions and connection to [SQL](#) databases. The table present below (Table 33) contains a list of possible technical problems and a reference procedure to solving them based on previous experience, aiming to minimize delays in having the models deployed in the business.

## H.2 DEPLOYING FINDINGS

Despite not being as technical and tangible as models, findings are nonetheless important results which may drive the business to make smarter decisions about the product they are developing. Due to the freshness of the data when it was first provided to me, I found myself in a privileged position in which the majority of the exploratory process was novel to the business, and thus relevant to include in the deployment plan. This section will detail the processes that will ensure all findings are adequately communicated with the business. Note that the findings specified in Table 19 of the main document are referred to as **Conclusive Findings** in this section, to reduce ambiguity with the overwhelming number of smaller discoveries throughout the project.

Issue	Contingencies
Host Name Resolution errors	<ol style="list-style-type: none"> <li>1. Redo site bindings with unused IP-Port combinations</li> <li>2. Verify state of business VPN</li> <li>3. Ping the address for response. Restart IIS and clear web browser caches</li> </ol>
FastCGI issues	<ol style="list-style-type: none"> <li>1. Disable and re-enable component with administrator permissions</li> <li>2. Modify website inactivity timer</li> <li>3. Verify web.config file</li> </ol>
Pandas / Numpy package errors	<ol style="list-style-type: none"> <li>1. Verify and rerun requirements.txt installation</li> <li>2. Upgrade packages and dependencies</li> <li>3. Downgrade malfunctioning packages and dependencies</li> </ol>
Pyodbc connection errors	<ol style="list-style-type: none"> <li>1. Manually specific a different ODBC driver in the connection statement</li> <li>2. Verify authentication account permissions</li> </ol>
Pyodbc transaction errors	<ol style="list-style-type: none"> <li>1. Use auto-commit on Stored Procedure transactions</li> <li>2. Manually commit on UPDATE transactions</li> </ol>
Local file access errors	<ol style="list-style-type: none"> <li>1. Verify IUSRS account permissions on files on the same folder as the code</li> <li>2. Add API ports to firewall exceptions</li> </ol>
Failover logic issues	<ol style="list-style-type: none"> <li>1. Deploy hotfix for uncaught exception and attempt to reproduce issue for documentation purposes</li> </ol>
Load underestimation	<ol style="list-style-type: none"> <li>1. Analyze legitimacy of overflowing requests and implement timeouts if necessary</li> <li>2. Perform API profiling to identify choke points or bottlenecks in performance</li> <li>3. Implement recurrent pinging to keep services "primed"</li> </ol>
Test case insufficiency	<ol style="list-style-type: none"> <li>1. Implement more robust error reporting functionality</li> <li>2. Request feedback directly</li> </ol>

Table 33: Technical difficulty contingency planning

### H.2.1 *Conclusive Findings*

The most important set of data findings will be introduced to the business in the form of an informal presentation, where all of the human resources identified for this project are required to be present. The meeting will be comprised of three distinct phases, with the goal of promoting an engaging discussion and critical thinking about what the data means to the future of GameVitae.

- **Result Presentation:** The presentation will consist on a curated approach to the exploratory phases of the project, focused on highlighting the most impactful findings through graphs and other visualizations. A short introduction to the developed models is also planned to be included here.
- **Discussion:** After presenting the results, all members are invited to share their opinions and thoughts, prompting a productive and interactive conversation in which feedback will be taken in consideration.

- **Future Directions:** To finalize the meeting, the topic will shift towards the meaning of these findings on a broader business scope. The goal is to find logical directions and new possibly new ways to leverage the findings.

### H.2.2 *Other Findings*

Through frequent interactions with the business throughout the initial stages of the data mining project, I've come to realize that very little was known about the data prior to my engagement with it. Due to this, I have decided to extend the finding dissemination plan to also include the entire data understanding report (Appendix C). The business' technical personnel and decision makers will be provided with a refined version of the appendix (with added covers and context) and asked to analyze it and draw their own conclusions. There currently isn't a predetermined plan to act upon this data, hence why the goal for this sharing is for the eventuality that the analysis gives rise to particularly interesting motivations for future projects.

### H.3 DEPLOYMENT SCHEDULE

Due to requiring additional development and the preparation of auxiliary deliverables, deployment and monitoring will last for a significant amount of time (exceeding the data mining project's schedule). To keep each individual task clear and reasonably constrained in time, a deployment schedule has been devised and is presented below, in Fig. 41. This schedule is not as strict as the data mining project schedule, since delays and unforeseen problems are a common occurrence in software development, but rather is to be understood as the estimated timing for each individual task. Note that in this figure, the 'lenience period' towards the end is intended to allow myself a margin of time for acting upon collected feedback, if need be.

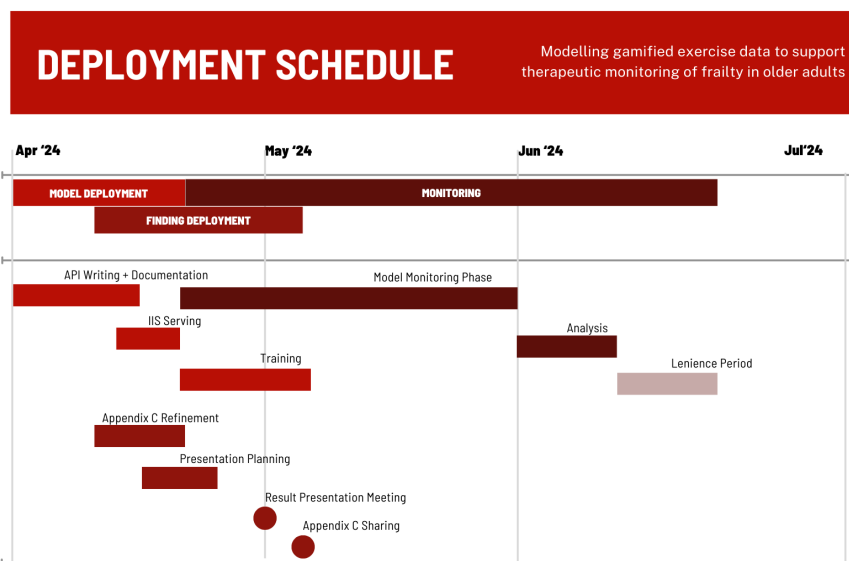


Figure 41: Result deployment schedule



## STATEMENT

---

Declaro, sob compromisso de honra, que o trabalho apresentado nesta dissertação, com o título “*Modelling gamified exercise data to support therapeutic monitoring of frailty in older adults*”, é original e foi realizado por Estudante João Miguel António (2210786) sob orientação de Professor Doutor Ricardo Malheiro ([ricardo.malheiro@ipleiria.pt](mailto:ricardo.malheiro@ipleiria.pt)).

*Leiria, Março de 2024*

---

Estudante João Miguel António