

**Levantamento, tratamento e seleção de dados  
para apoio à tomada de decisão -  
Schmidt Light Metal – Fundição Injetada, Lda**

Mestrado em Ciência de Dados

Carlos Pereira Ferreira, 2212829

Leiria, setembro de 2023

**Levantamento, tratamento e seleção de dados  
para apoio à tomada de decisão -  
Schmidt Light Metal – Fundição Injetada, Lda**

Mestrado em Ciência de Dados

Carlos Pereira Ferreira, 2212829

Relatório de estágio de Mestrado sob a orientação da Doutora Anabela Moreira Bernardino, Professora da Escola Superior de Tecnologia e Gestão do Politécnico de Leiria e da Doutora Eugénia Moreira Bernardino, Professora da Escola Superior de Tecnologia e Gestão do Politécnico de Leiria.

Leiria, setembro de 2023

## **Originalidade e Direitos de Autor**

O presente relatório de estágio é original, e foi elaborado unicamente para este fim, tendo sido devidamente citados todos os autores cujos estudos e publicações contribuíram para o elaborar.

Reproduções parciais deste documento serão autorizadas na condição de que seja mencionado o Autor e feita referência ao ciclo de estudos no âmbito do qual o mesmo foi realizado, a saber, Curso de mestrado em Ciência de Dados, no ano letivo 2022/2023 da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, Portugal, e, bem assim, à data das provas públicas que visaram a avaliação destes trabalhos.

Dedico este à minha família, à minha namorada Patrícia, e aos meus grandes amigos, por todo o apoio, motivação e força que sempre me deram, mas também pela paciência e por estarem lá sempre que precisei.

Sem vocês nada disto seria possível.

## Agradecimentos

Em primeiro lugar gostaria de agradecer a todos aqueles que, enquanto pessoas afetivas ou colegas de estudo, me acompanharam e contribuíram para a conclusão de mais uma etapa da minha formação e desenvolvimento pessoal.

Agradeço à empresa *Schmidt Light Metal* – Fundação Injetada, Lda que me abriu portas a este desafio e que me deu a oportunidade de estagiar nos seus departamentos de IT (*Information Technology*) e de Pessoas & Cultura, proporcionando-me todas as condições necessárias para a concretização deste estágio. Agradeço ao meu supervisor de estágio, Manuel António Valente, pela boa orientação e confiança que depositou em mim e no meu trabalho. À restante equipa do departamento (Sr. Mendes, Emmanuel, Pedro e Eduardo) deixo um agradecimento também, não só pela paciência que tiveram comigo, mas principalmente pelo apoio que me deram e pelo bom ambiente de trabalho que sempre procuraram proporcionar.

Um agradecimento também ao departamento de Pessoas & Cultura (com quem passei bastante tempo) representado pela Doutora Patrícia Villas-Boas, a quem muito devo, não só pela confiança que depositou em mim, mas, também, por tudo o que partilhou comigo e por ter sido Guia e Exemplo nesta caminhada. À restante equipa (Catarina, Daniela, José, Mariana e Ricardo), obrigado pelo excelente ambiente de trabalho que sempre me proporcionaram, pela forma como olhavam e valorizavam o meu trabalho e pelo enorme apoio e carinho que sempre demonstraram.

Quero também agradecer às minhas orientadoras de estágio, Professora Doutora Anabela Bernardino e Professora Doutora Eugénia Bernardino, que foram uma ajuda enorme para mim, ao longo deste período de estágio, não só pela orientação que me deram, mas também pela disponibilidade que sempre demonstraram, pela sua boa disposição e pela motivação que sempre me transmitiram.

Agradeço ao Instituto Politécnico de Leiria por me ter acolhido, enquanto estudante, e a todos os docentes de igual forma por todo o conhecimento e acima de tudo experiência que me foram transmitindo. Sem dúvida, são uma parte muito importante da minha formação.

Quanto aos grandes amigos, gostaria de deixar um agradecimento especial ao meu parceiro de mestrado, João António, por todo o trabalho que fizemos juntos, pelas

experiências, pela aprendizagem conjunta e, principalmente, por tudo aquilo que me ensinou e compartilhou comigo.

## Resumo

O presente relatório descreve o trabalho desenvolvido no âmbito da unidade curricular de Estágio do Mestrado em Ciência de Dados, da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria.

O principal objetivo deste relatório é descrever todo o trabalho desenvolvido ao longo do estágio curricular de 1080h (mil e oitenta horas). Este estágio foi realizado na empresa *Schmidt Light Metal* – Fundição Injetada, Lda., uma empresa multinacional, situada no parque industrial de Oliveira de Azeméis. Esta empresa dedica-se à fundição, produzindo peças em alumínio leve para o setor automóvel. Neste relatório farei uma descrição empresa, bem como das tecnologias e ferramentas utilizadas, a metodologia seguida, o trabalho desenvolvido e as respetivas conclusões.

Este estágio tinha como objetivo inicial o desenvolvimento de ferramentas de apoio à tomada de decisão (na produção), através da recolha, tratamento e seleção dos dados recolhidos pelos vários equipamentos e departamentos da empresa, desde os dados sensoriais de uma célula de injeção ao *headcount* feito pelo departamento de Pessoas & Cultura.

Durante este estágio foi desenvolvido e treinado um modelo de *Machine Learning* para facilitar a tomada de decisão no que respeita ao planeamento de produção, tendo, também, sido realizado bastante trabalho voltado para a “Análise de Dados”, através do lançamento de inquéritos internos, recolha de dados, tratamento e apresentação gráfica dos resultados obtidos a partir destes, sendo que este último trabalho foi realizado em cooperação com o departamento de Pessoas & Cultura.

O desenvolvimento do modelo de *Machine Learning* para apoio à tomada de decisão no processo de produção foi extremamente desafiante, uma vez que não existiam bases de informática ou código pré-mestrado e dada a complexidade do processo em si. A correção de falhas existentes que foram aparecendo e o constante “limar de arestas” também garantiu a satisfação de todos os *stakeholders*. Posto isto e como resultado deste estágio, foram adquiridas novas experiências, colocadas em prática novas técnicas e desenvolvidas inúmeras competências ao nível profissional, através dos vários projetos elaborados em ambiente empresarial.

**Palavras-chave:** Dados, *Machine Learning*, Processo produtivo, Tomada de Decisão.

## **Abstract**

This report describes the work developed within the scope of the curricular unit of Internship of the Master in Data Science, of the School of Technology and Management of the Polytechnic Institute of Leiria.

The main objective of this report is to describe the whole process involved throughout the curricular internship of 1080h (one thousand and eighty hours). This internship was carried out in the company Schmidt Light Metal - Fundição Injetada, Lda., a multinational company, located in the industrial park of Oliveira de Azeméis. This company is dedicated to casting, producing light aluminium parts for the automotive sector. Later on, I will give a full description of the company, as well as the technologies and tools used, the methodology followed, the work developed and the respective conclusions.

This internship had as its initial objective the development of tools to support decision making (in production), through the collection, processing and selection of data collected by the various equipment and departments of the company, from the sensory data of an injection cell to the headcount made by the People & Culture department.

During this internship it was developed and trained a Machine Learning model to facilitate decision making regarding production planning, and also a lot of work focused on "Data Analysis", through the launch of internal surveys, data collection, processing and graphical presentation of the results obtained from these, the latter work being done in cooperation with the People & Culture department.

The development of the Machine Learning model to support decision making in the production process was extremely challenging, since I have no computer science background or pre-mastered code and given the complexity of the process itself. The correction of existing flaws that appeared and the constant "edge trimming" also ensured the satisfaction of all stakeholders. Having said that and as a result of this internship, I have acquired new experiences, put into practice new techniques and developed numerous skills at the professional level, through the various projects in which I was involved in a business environment.

**Keywords:** Data, Decision Making, Machine Learning, Production Process.

## Índice

<b>Originalidade e Direitos de Autor .....</b>	<b>ii</b>
<b>Agradecimentos .....</b>	<b>iv</b>
<b>Resumo .....</b>	<b>vi</b>
<b>Abstract.....</b>	<b>vii</b>
<b>Índice.....</b>	<b>viii</b>
<b>Lista de Figuras.....</b>	<b>x</b>
<b>Lista de Tabelas.....</b>	<b>xii</b>
<b>Lista de Siglas e Acrónimos .....</b>	<b>xiii</b>
<b>1. Introdução .....</b>	<b>1</b>
<b>1.1. Motivação e objetivos.....</b>	<b>1</b>
<b>1.2. Cronograma do estágio .....</b>	<b>1</b>
<b>1.3. Descrição das atividades desenvolvidas .....</b>	<b>2</b>
<b>1.4. Metodologias utilizadas .....</b>	<b>3</b>
1.4.1. Algumas Metodologias de Trabalho.....	4
1.4.2. Metodologia <i>Scrum</i> .....	5
1.4.3. Principais princípios do <i>Scrum</i> .....	5
1.4.4. Utilização da metodologia <i>Scrum</i> no estágio .....	7
<b>1.5. Estrutura do documento .....</b>	<b>7</b>
<b>2. Enquadramento do estágio.....</b>	<b>9</b>
<b>2.1. Entidade de Acolhimento .....</b>	<b>9</b>
<b>2.2. Princípios corporativos .....</b>	<b>11</b>
<b>2.3. Ciência de Dados na tomada de decisão.....</b>	<b>12</b>
2.3.1. Os dados e a sua importância .....	12
2.3.2. <i>Business Intelligence</i> .....	13
2.3.3. A importância da tomada de decisão nas empresas.....	15
2.3.4. Recolha de dados em ambiente industrial .....	16
2.3.5. <i>Machine Learning</i> .....	17
2.3.6. <i>Machine Learning</i> em ambiente industrial.....	19
<b>3. Atividades desenvolvidas.....</b>	<b>21</b>
<b>3.1. Tecnologias utilizadas .....</b>	<b>22</b>
3.1.1. <i>Excel</i> .....	22
3.1.2. <i>PowerPoint</i> .....	23

3.1.3. <i>Google Colab em Python</i> .....	23
3.1.4. <i>Word</i> .....	24
3.1.5. <i>Power BI</i> .....	24
3.1.6. <i>Canva</i> .....	24
3.1.7. <i>Microsoft Forms</i> .....	25
3.1.8. <i>Mentimeter</i> .....	25
<b>3.2. Modelo preditivo de cumprimento de planeamentos produtivos</b> .....	<b>25</b>
3.2.1. Recolha e tratamento dos dados .....	27
3.2.2. 1º Teste com um modelo de regressão logística (“ <i>dataset</i> ”) .....	40
3.2.3. 2º Teste com um modelo de regressão logística (“ <i>dataset.w_hxt</i> ”) .....	42
3.2.4. 3º Teste com um modelo de regressão logística (“ <i>dataset.w_orders</i> ”) ..	44
3.2.5. 4º Teste com um modelo de regressão logística (“ <i>dataset.w_stocks</i> ”) ..	45
3.2.6. Escolha do modelo a utilizar .....	46
<b>3.3. Dashboards em Power BI</b> .....	<b>48</b>
3.3.1. Fotografia da população 2023 .....	49
3.3.2. ARTE .....	51
<b>3.4. Processos ETL e relatórios</b> .....	<b>54</b>
<b>3.5. Auditoria interna de segurança da informação TISAX</b> .....	<b>63</b>
<b>3.6. Apreciação crítica às atividades desenvolvidas</b> .....	<b>64</b>
<b>4. Conclusão</b> .....	<b>66</b>
<b>Bibliografia</b> .....	<b>68</b>
<b>Anexos</b> .....	<b>73</b>

## Lista de Figuras

Figura 1 - Cronograma do estágio .....	2
Figura 2 - Exemplo do dado "Dia".....	30
Figura 3 - Exemplo do dado "Ref".....	31
Figura 4 - Exemplo do dado "Obj_OPx" .....	32
Figura 5 - Exemplo do dado "sum.ddp".....	33
Figura 6 - Exemplo do dado "humanresources" .....	34
Figura 7 - Exemplo do dado "sum.hxt".....	35
Figura 8 - Parte 1 do <i>notebook</i> "Dataset_creator" .....	38
Figura 9- Parte 2 do <i>notebook</i> "Dataset_creator" .....	39
Figura 10 – Recorte do “dataset.csv” .....	40
Figura 11 - Parte 1 do <i>Notebook</i> de preparação do modelo e previsão.....	41
Figura 12 - Parte 2 do <i>Notebook</i> de preparação do modelo e previsão.....	42
Figura 13 - Alterações à coluna "Accomp" no 2º teste.....	43
Figura 14 - Alterações à coluna "Accomp" no 3º teste.....	44
Figura 15 - Alterações à coluna "Accomp" no 4º teste.....	46
Figura 16 - Página 1 do <i>Power BI</i> "Fotografia da população 2023".....	49
Figura 17 - Página 2 do <i>Power BI</i> “Fotografia da população 2023” .....	50
Figura 18 - Página 3 do <i>Power BI</i> "Fotografia da população 2023".....	51
Figura 19 - Análise de um excelente cumprimento do processo de avaliação de desempenho anual (ARTE) .....	53
Figura 20 - Análise de um bom cumprimento do processo de avaliação de desempenho anual (ARTE) .....	53
Figura 21 - "Inquérito com Sentido".....	55
Figura 22 - Apresentação de resultados obtidos do exemplo 1 .....	56
Figura 23 - Slide Modelo de apresentação de novas admissões.....	56
Figura 24 - Exemplo de Inquérito com uma implementação da medida mais demorada .....	57
Figura 25 - Apresentação de resultados obtidos do exemplo 1 .....	57
Figura 26 - Survey de consciencialização para a dádiva de sangue .....	58
Figura 27 - Inquérito de uma obrigação legal.....	59
Figura 28 - Exemplo de um <i>slide</i> de apresentação de dados tratados em <i>PowerPoint</i> (tema: Formação).....	60

Figura 29 - Slide exemplo de um anúncio de Vaga Interna..... 61

## **Lista de Tabelas**

Tabela 1 - Estrutura dos vários conjuntos de dados.....	28
Tabela 2 - Tipo de dado, por coluna .....	29
Tabela 3 - Resultados obtidos no 1º teste .....	42
Tabela 4 - Resultados obtidos no 2º teste .....	43
Tabela 5 - Resultados obtidos no 3º teste .....	45
Tabela 6 - Resultados obtidos no 4º teste .....	46
Tabela 7 - Comparação dos testes realizados .....	47
Tabela 8 - Inquéritos realizados .....	61

## **Lista de Siglas e Acrónimos**

- API – *Application Programming Interface*
- ARTE – Agilidade, Rigor, Tenacidade e Espírito de Equipa
- ATC – Autoconceptus - Projetos De Engenharia, Lda
- BI – *Business Intelligence*
- CNN – *Convolutional neural network*
- DM – *Data Mining*
- DMM – Desenvolvimento, Maquinagem E Montagem, Lda
- ESTG – Escola Superior de Tecnologia e Gestão
- ETL – *Extract, Transform and Load*
- FIFO – *First in, First out*
- IA – Inteligência Artificial
- IoT – *Internet of Things*
- IPL – Instituto Politécnico de Leiria
- IT – Information Technology
- ML – *Machine Learning*
- PCA – Análise de Componentes Principais
- RNN – *Recurrent neural network*
- SCADA – *Supervisory Control and Data Acquisition*
- SLM – Schmidt Light Metal, Fundação Injetada, Lda
- SVG – *Schmidt Light Metal Vertriebsgesellschaft, Mbh*
- SVM – Máquinas de Vetores de Suporte
- SWOT – *Strengths, Opportunities, Weaknesses and Threats*
- TISAX - *Trusted Information Security Assessment Exchange*

# 1. Introdução

O presente relatório descreve o trabalho desenvolvido ao longo do estágio de 1080 horas realizado no âmbito da unidade curricular de Estágio do Mestrado em Ciência de Dados, lecionado na Escola Superior de Tecnologia e Gestão (ESTG) do Instituto Politécnico de Leiria, e passará, mais concretamente, pelas motivações e objetivos a atingir, pela metodologia de desenvolvimento e tecnologias utilizadas, pela descrição das tarefas desenvolvidas no decorrer deste estágio e finalmente pelas conclusões e análises de resultados obtidos.

Neste capítulo é feita, na Secção 1.1, uma apresentação da motivação e dos objetivos deste estágio. É apresentado, na Secção 1.2, o cronograma de estágio, destacando as principais tarefas desenvolvidas. Na Secção 1.3 estas mesmas tarefas são descritas com maior pormenor e, na Secção 1.4 são apresentadas as metodologias adotadas no decorrer do estágio. Por fim, na Secção 1.5 deste capítulo, é feita uma breve descrição da estrutura deste documento.

## 1.1. Motivação e objetivos

A empresa *Schmidt Light Metal* – Fundição Injetada, Lda., por estar a atravessar um processo de transformação digital, e uma vez que era do seu interesse aproveitar os dados que já estavam a ser recolhidos para dar um passo rumo a uma fábrica mais inteligente e digital, criando ferramentas inteligentes de apoio à tomada de decisão, entendeu conveniente investir nesta área.

Da minha parte, interessava-me colaborar com uma empresa disposta a confiar e investir na ciência e poder dos dados para aplicar ferramentas de *Machine Learning* num ambiente industrial produtivo.

Mostrou-se assim benéfico, para ambas as partes, a minha entrada, enquanto estudante de Ciência de Dados, no departamento de IT desta mesma empresa.

## 1.2. Cronograma do estágio

O estágio curricular teve uma duração de 1080 (mil e oitenta) horas, tendo sido iniciado no dia 17 de outubro de 2022 e concluído no dia 17 de maio de 2023. Na Figura 1 é apresentado um cronograma com as tarefas realizadas ao longo do tempo.

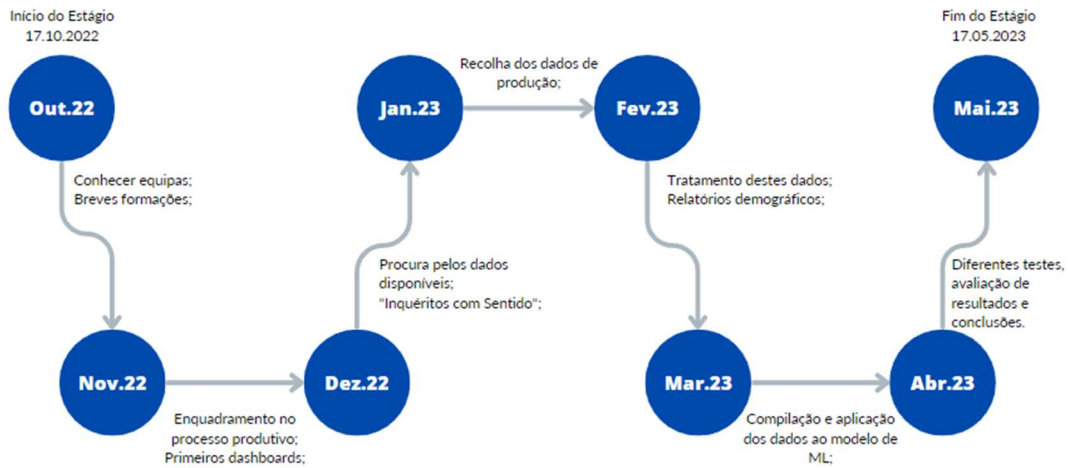


Figura 1 - Cronograma do estágio

Uma vez que este estágio decorreu em ambiente empresarial industrial, foi necessário um período de adaptação e introdução aos processos, que consistiu em momentos de explicação e entendimento dos vários passos produtivos e de planeamento industrial.

Enquanto isso, houve também um período de apresentação e formação relativa às ferramentas e dispositivos que viriam a ser utilizados no decorrer do estágio. Naturalmente, à medida que novos trabalhos eram desenvolvidos e novas ferramentas eram experimentadas, também a formação acompanhava estas experiências.

### 1.3. Descrição das atividades desenvolvidas

Durante o estágio, deu-se a oportunidade de estar por dentro de várias atividades que contribuem bastante para o desenvolvimento profissional e pessoal de cada um. Ao longo desse período, foram adquiridos experiência e conhecimentos em diferentes áreas, muito devido à criação de um modelo preditivo de ML (*Machine Learning*), ao desenvolvimento de *dashboards* no *Power BI (Business Intelligence)*, ao processamento de dados provenientes de questionários via *Forms* e ao acompanhamento de uma auditoria de segurança da informação *TISAX*.

Uma das atividades realizadas mais relevantes durante o estágio foi a criação de um modelo preditivo de ML para a previsão do cumprimento de planos de produção. Nessa tarefa, foram aplicadas técnicas e algoritmos de ML para analisar dados históricos e identificar padrões que pudessem ser usados para ajudar a prever o cumprimento (ou não) de planos de produção futuros. Esta atividade exigiu o processamento e tratamento dos

dados, a seleção de recursos relevantes e a construção do modelo preditivo, além da sua avaliação e ajuste, para garantir a melhor eficácia possível [5].

Além disso, também foram desenvolvidos e elaborados alguns *dashboards* no *Power BI*. Esta atividade envolveu a criação de painéis interativos e visualmente apelativos para apresentar informações e *insights* relevantes aos decisores da empresa. Utilizando o *Power BI*, foram aplicadas algumas transformações e filtros aos dados, assim como, foram criados gráficos e visualizações dinâmicas que ajudam a facilitar a compreensão dos dados e podem vir a auxiliar na tomada de decisões estratégicas.

Outra atividade desempenhada durante o estágio foi o processo de ETL (*Extract, Transform and Load*) de questionários via *Forms*. Nessa tarefa, trabalhou-se na extração dos dados das respostas dos formulários, no tratamento e limpeza destes mesmos dados, bem como na transformação dessas informações num formato adequado para a análise posterior.

Além disso, muitos destes resultados obtidos foram também apresentados através da criação de relatórios em formato PDF, apresentações em *PowerPoint* ou em *Canva*, com o objetivo de fornecer uma visualização clara e concisa dos *insights* extraídos das respostas dos questionários.

Por fim, também foi possível participar no acompanhamento de uma auditoria interna de segurança da informação *TISAX*. Durante esse processo, auxiliou-se na preparação dos documentos e procedimentos necessários, participou-se em reuniões com os auditores e também foi necessário apoiar a equipa na implementação das medidas corretivas recomendadas. Essa experiência proporcionou um entendimento mais amplo das práticas de segurança da informação e dos requisitos de conformidade, bem como a oportunidade de trabalhar num ambiente rigoroso e em conformidade com padrões internacionais.

## **1.4. Metodologias utilizadas**

Atualmente, o sucesso das empresas depende cada vez mais de processos eficazes e metodologias de trabalho que promovam a agilidade, colaboração e entrega de valor ao cliente. Diversas metodologias têm surgido para responder a estas necessidades, cada uma com as suas particularidades e abordagens específicas. Nesta secção são exploradas algumas das principais metodologias de trabalho utilizadas para gerir projetos e tarefas

(Secção 1.4.1), com especial destaque para metodologia *Scrum* (Secção 1.4.2), bem como para as suas principais características e princípios fundamentais (Secção 1.4.3). No final desta secção é explicado como a metodologia *Scrum* foi aplicada no estágio (Secção 1.4.4)

### **1.4.1. Algumas Metodologias de Trabalho**

Nesta secção são apresentadas algumas das metodologias utilizadas para gerir projetos e tarefas nas empresas. Neste contexto, destacam-se duas abordagens principais: a metodologia cascata (secção 1.4.1.1), caracterizada por um processo linear e sequencial, onde as etapas são seguidas em ordem, sem retrocessos; e o movimento ágil (secção 1.4.1.2), representado por metodologias como o *Scrum*, que enfatizam a flexibilidade, a colaboração e a entrega incremental de valor ao cliente. Ao explorar essas diferentes metodologias, torna-se possível compreender as diversas formas de organizar e executar projetos, identificando a abordagem mais adequada às necessidades específicas de cada projeto ou organização.

#### **1.4.1.1. Cascata (*Waterfall*)**

A metodologia cascata é um modelo tradicional e linear, onde o desenvolvimento de um determinado projeto segue uma sequência de etapas bem definidas, como requisitos, *design*, implementação, testes e implantação, sem qualquer tipo de retornos ou ajustes significativos ao longo do processo. Embora seja uma abordagem estruturada, pode ser lenta para se adaptar a mudanças, o que pode ser um desafio em projetos mais dinâmicos [1]. Na fase dos inquéritos, foi adotada a metodologia de trabalho *Waterfall*, devido à sua natureza sequencial e estruturada, que se mostrou altamente adequada para o contexto específico dos projetos. O processo de recolha de dados e análise das informações necessárias seguiu uma abordagem sistemática, permitindo uma organização clara e precisa das etapas envolvidas. Com esta abordagem, foram definidas antecipadamente as fases do inquérito, estabelecidos prazos realistas e alocados recursos adequados para cada etapa. Além disso, a metodologia *Waterfall* proporcionou um maior controlo sobre o desenvolvimento do trabalho, facilitando a identificação de eventuais desvios e permitindo ajustes precisos, garantindo, assim, a qualidade dos resultados obtidos durante o processo investigativo.

#### **1.4.1.2. Ágil (*Agile*)**

O movimento ágil (*agile*) revolucionou a forma como as equipas desenvolvem produtos e serviços. Em vez de processos rígidos e inflexíveis, o Ágil baseia-se em

valores como indivíduos e interações mais do que em processos e ferramentas, *software* funcional mais do que documentação abrangente, colaboração com o cliente mais do que negociação de contratos, e resposta à mudança mais do que seguir um plano [2].

#### **1.4.2. Metodologia *Scrum***

A metodologia *Scrum* é uma das abordagens ágeis mais populares e amplamente utilizadas em projetos de desenvolvimento de *software* e outras áreas. Esta foca-se na divisão do trabalho em ciclos chamados *sprints*, que geralmente têm duração de duas a quatro semanas, durante os quais as equipas trabalham em incrementos funcionais do produto [3].

#### **1.4.3. Principais princípios do *Scrum***

Nesta secção são abordados os fundamentos essenciais que orientam esta metodologia ágil de trabalho. Entre os principais princípios, destacam-se a transparência (secção 1.4.3.1), que procura, essencialmente, uma comunicação aberta e honesta em todas as etapas do projeto; a inspeção e adaptação (secção 1.4.3.2), que incentivam a revisão contínua dos resultados e a procura por melhorias; a auto-organização das equipas (secção 1.4.3.3), que permite a tomada de decisões internas e a responsabilidade partilhada; e a entrega incremental (secção 1.4.3.4), que valoriza a entrega constante de funcionalidades ao cliente, possibilitando um *feedback* contínuo e ajustes ao longo do processo. Estes princípios são a base para o sucesso das equipas que adotam o *Scrum* como metodologia de trabalho ágil.

##### **1.4.3.1. Transparência**

O *Scrum* promove a transparência em todas as etapas do projeto. Isso significa que todas as informações relevantes sobre o progresso do trabalho, desafios enfrentados e resultados alcançados devem ser acessíveis a todas as partes interessadas. A transparência é essencial para uma comunicação aberta e colaborativa, facilitando a identificação precoce de problemas e a busca de soluções eficazes [3]. Durante a adoção da metodologia *Scrum* no departamento de IT, enfatizou-se a transparência como elemento essencial do processo de trabalho. As reuniões constantes permitiram a partilha de progressos e dificuldades entre os membros da equipa, nas *Sprint Reviews* apresentavam-se resultados e entregas, possibilitando *feedback* imediato e a cultura de abertura e respeito promoveu a livre expressão de ideias e preocupações, fortalecendo a

transparência e o espírito de equipa. O resultado foi uma abordagem colaborativa e projetos bem-sucedidos.

#### **1.4.3.2. Inspeção e Adaptação**

O *Scrum* incentiva a prática regular de inspeção e adaptação do processo. Isso é feito através de cerimónias específicas, como as reuniões de revisão de *sprint* e retrospectivas, onde a equipa analisa os resultados do *sprint* concluído e identifica oportunidades de melhoria para o próximo ciclo. A capacidade de se adaptar rapidamente às mudanças é uma das principais vantagens do *Scrum* [3].

#### **1.4.3.3. Equipas auto-organizadas**

As equipas no *Scrum* são auto-organizadas, o que significa que têm autonomia para tomar decisões relacionadas ao trabalho, planeamento e resolução de problemas. Essa abordagem permite que os membros da equipa se sintam mais responsáveis pelo projeto e que cada um traga as suas habilidades únicas para alcançar os objetivos comuns [3]. Durante a aplicação desta metodologia, no departamento, também se confere uma grande importância à criação de equipas auto-organizadas como elemento essencial do processo de trabalho. Através de um ambiente colaborativo, os diferentes membros da equipa foram encorajados a assumirem responsabilidade na tomada de decisões e na distribuição das tarefas. A cultura de confiança e respeito, uma vez mais, estimulou a cooperação e a partilha de conhecimentos, potenciando assim a auto-organização. O resultado foi uma abordagem mais ágil e eficiente, com a equipa a demonstrar maior capacidade de adaptação e superação de desafios ao longo do processo de trabalho.

#### **1.4.3.4. Entrega Incremental**

O *Scrum* enfatiza a entrega de incrementos de valor a cada *sprint*. Em vez de esperar pelo produto final, os clientes recebem funcionalidades e melhorias contínuas, possibilitando *feedback* precoce e ajustes conforme o projeto avança [3]. Como seria de esperar, esta entrega incremental está diretamente relacionada com o trabalho em *sprints*. A metodologia adotada, relacionada com a entrega incremental e a forma de trabalhar em *sprints*, na qual se desenvolvem e entregam partes funcionais do projeto em ciclos curtos de trabalho, resultou num desenvolvimento mais eficiente e alinhado com as necessidades dos vários *players*. Cada *sprint* resulta em entregas parciais que agregam valor ao projeto, permitindo a obtenção de *feedbacks* contínuos por parte dos *stakeholders*. Esta forma de

entregar proporcionou uma clara noção dos progressos e evidenciou a capacidade de adaptação rápida à mudança.

#### **1.4.4. Utilização da metodologia *Scrum* no estágio**

Como referido anteriormente, as metodologias de trabalho têm um papel fundamental na eficiência e eficácia das equipas e projetos. Durante o estágio, a metodologia *Scrum* demonstrou ser eficaz para gerir as tarefas de *Data Science*. Na fase de Recolha de Dados, foram estabelecidos objetivos de recolha, foram atribuídas responsabilidades e definidos prazos via *sprints*. No Tratamento e Seleção dos Dados, as reuniões diárias permitiram ajustar continuamente as estratégias de limpeza e preparação dos dados. Durante o desenvolvimento do modelo, as iterações regulares do *Scrum* possibilitaram adaptações com base no *feedback*. E finalmente, na fase de testes, a metodologia auxiliou na identificação e correção ágil de problemas, garantindo a qualidade do modelo. Resumindo, o *Scrum* proporcionou um ambiente de trabalho colaborativo e flexível que facilitou todas as etapas do processo de *Data Science* durante o estágio.

### **1.5. Estrutura do documento**

O presente relatório está estruturado e dividido em seis grandes capítulos.

O Capítulo 1 dá-se pelo nome de “Introdução”. Neste capítulo, como o próprio nome indica, é feita uma breve introdução à matéria deste relatório, apresentando aquelas que foram as motivações e objetivos (Secção 1.1) aquando da proposta deste estágio, qual o seu cronograma (Secção 1.2), todas as atividades desenvolvidas em ambiente de estágio (Secção 1.3), e por fim o tipo de metodologia utilizada no decorrer deste estágio (Secção 1.4). Apresenta-se ainda a estrutura do relatório em questão, nesta mesma secção.

O Capítulo 2, denominado “Enquadramento do estágio”, apresenta uma contextualização do estágio, descrevendo a empresa de acolhimento e as condições proporcionadas, de forma a garantir o bom desenrolar do mesmo (Secção 2.1). Ainda neste capítulo faz-se uma caracterização e enquadramento histórico da empresa em questão, apresentando os princípios corporativos (Secção 2.2) pelos quais esta se rege e, posteriormente, é feita também uma descrição de conceitos por forma a melhor entender a importância da Ciência de Dados na tomada de decisão (Secção 2.3).

No Capítulo 3, relativo às “Atividades desenvolvidas”, são apresentadas, tal como o nome indica, as atividades e tarefas desenvolvidas no decorrer deste estágio em Ciência de Dados. Primeiramente, é feita uma apresentação e descrição das mais variadas ferramentas utilizadas no decorrer deste mesmo (Secção 3.1). De seguida apresenta-se o desenvolvimento, fase de testes, avaliação e escolha de um Modelo preditivo de cumprimento de planeamentos produtivos (Secção 3.2). Posteriormente, é abordada a temática relacionada com as tarefas e projetos em *Power BI*, também estes para apoio à tomada de decisão (Secção 3.3). Para além destes, apresentam-se alguns outros processos ETL (*Extract, Transform and Load*) e relatórios realizados, também estes, e como seria de esperar, para apoio à tomada de decisão (Secção 3.4). De seguida, é feita uma apresentação das atividades acompanhadas no decorrer da auditoria interna *TISAX* (Secção 3.5) e por fim, como seria de esperar, é apresentada a apreciação crítica às atividades desenvolvidas (Secção 3.6).

O Capítulo 4 está dedicado à Conclusão, onde se faz um balanço geral do trabalho desenvolvido, analisa-se os resultados obtidos, bem como as dificuldades encontradas e, ainda, as soluções testadas para fazer face a tais dificuldades. São destacados ainda outros aspetos relevantes que surgiram ao longo da realização do estágio.

## 2. Enquadramento do estágio

Neste capítulo é apresentada a entidade de acolhimento (Secção 2.1), de uma forma pormenorizada, nomeadamente o setor em que o estágio decorreu, as equipas de trabalho e as condições disponíveis, assim como são descritas, brevemente, as diversas atividades desenvolvidas em estágio, bem como alguns dos princípios corporativos (Secção 2.2). Para além do supra mencionado, na Secção 2.3 é feita também uma descrição de conceitos relevantes para determinar a importância da Ciência de Dados na tomada de decisão.

### 2.1. Entidade de Acolhimento

A *Schmidt Light Metal* – Fundição Injetada, Lda. é uma empresa portuguesa fundada em 1989, que está diretamente ligada à indústria automóvel, através da produção de peças fundidas em alumínio [4].

Situada em Oliveira de Azeméis, dispõe de cerca de  $37.000m^2$ , sendo que destes, cerca de  $22.000m^2$  são de área coberta, estando esta distribuída por 6 naves, e contando com uma equipa de cerca de 430 profissionais especializados e extremamente comprometidos com o sucesso da empresa e com a satisfação do cliente final, o que permite que o trabalho que exercem seja feito com os mais elevados padrões de qualidade e eficácia.

Fazem parte deste grupo:

- A ATC (*Autoconceptus* – Projetos de Engenharia, Lda) que é a empresa de desenvolvimento de ferramentas, onde os moldes são desenhados/projetados e produzidos;
- A SLM (*Schmidt Light Metal* – Fundição Injetada, Lda) que é a unidade de fundição sob pressão de alumínio, onde 20 máquinas Bühler dão vida às peças;
- A DMM (Desenvolvimento Maquinagem e Montagem, Lda) onde, após a fundição sob pressão, se cuida do acabamento das peças com processos diferenciados, para atender às necessidades do cliente final;
- A SVG (*Schmidt Light Metal Vertriebsgesellschaft, Mbh*), uma empresa inaugurada no mês de julho/2021, com vista à expansão para novos mercados, sendo a Áustria um local estratégico de implementação.

É parte da missão do Grupo criar soluções integradas com agilidade e talento, fazendo com que, mais do que produzir peças fundidas em alumínio, se criem e desenvolvam soluções excepcionais, e desenhadas exclusivamente para cada cliente [4].

Cada empresa do Grupo *Schmidt Light Metal*, está inteiramente dedicada a fazer as melhores peças para automóveis, para toda a Europa. Em julho de 2021, a empresa expande para a Áustria, com a abertura da SVG- *Schmidt Light Metal Vertriebsgesellschaft, Mbh*. Esta unidade destina-se a maquinar componentes para automóveis, cujos materiais fundidos serão produzidos na unidade, SLM, em Portugal. Desta forma, torna-se possível a entrega das peças maquinadas ao cliente com uma periodicidade diária, de acordo com o caderno de encargos. Esta unidade austríaca permitiu também ter um *stock* avançado de produto acabado, com o objetivo de servir com maior rapidez os clientes da Europa central. A sua operação é completada ainda por um depósito em Salzgitter, na Alemanha.

A história da *Schmidt Light Metal* é feita de etapas conquistadas, que só foram possíveis alcançar com muito trabalho, total comprometimento com os seus clientes e uma equipa de pessoas dedicadas que se orgulham de representar o Grupo *Schmidt Light Metal*, desde o minuto em que ingressam na organização.

Foi em 1989 que foi lançada a primeira pedra do edifício que hoje alberga parte das suas instalações de fundição e, desde então, têm vindo a crescer, a melhorar e a investir.

Desde aí, o Conselho Administrativo vem trabalhando com firmeza para consolidar o Grupo *Schmidt Light Metal* como um importante *player* na Indústria Automóvel Portuguesa. Isso só é possível porque os seus líderes sempre tiveram uma visão clara do que desejavam perseguir e alcançar. São eles que moldam o futuro e o fazem acontecer, atuando como modelos dos seus valores e ética, inspirando confiança em todos os momentos. Esta é a visão partilhada, sendo o compromisso e o conhecimento partilhado também. Eles sabem que o sucesso vem do trabalho árduo, do desenvolvimento de uma estratégia focada nas partes interessadas, e que tem de ser colocada em ação em todas as decisões tomadas.

Com clientes extremamente conhecidos e conceituados, as peças que saem da SLM circulam pelas estradas de todo o mundo, e, atendendo à sua capacidade produtiva e à sua dimensão, esta desempenha um “*Major Role*” na indústria automóvel portuguesa. Como

principais clientes apresentam-se marcas como *Lamborghini, Maserati, Porsche, Bentley, Audi, Mercedes, Volkswagen*, entre outras.

## **2.2. Princípios corporativos**

Nesta secção são abordados os principais fundamentos que norteiam a identidade e o propósito desta organização, ou seja, são elencados os princípios éticos e morais que guiam o comportamento e a cultura da empresa. Estes valores servem como diretrizes para as ações dos demais colaboradores, e comportam-se como pilares, ajudando a manter uma conduta consistente com a identidade e a imagem da organização.

Ao definir estes princípios corporativos, a organização estabelece uma base sólida para todas e quaisquer operações bem como para as relações com todos os *stakeholders*. É assim que vincula a sua verdadeira identidade, os seus objetivos e os seus compromissos. Posto isto, seria pertinente abordar de uma forma simples, mas concisa, os valores organizacionais tão presentes.

Os valores são as crenças e convicções que guiam as ações dos colaboradores em todos os níveis da organização, refletindo a identidade e o propósito da empresa.

Esses valores servem como pilares fundamentais para a cultura organizacional, orientando grande parte das ações diárias, as relações internas e a forma como a empresa se relaciona com os seus clientes e parceiros. Ao viver esses valores, a empresa fortalece toda a sua identidade, constrói uma reputação sólida e alcança o sucesso sustentável no mercado em que atua. Tudo isto é muito nítido no *modus operandi* da maior parte das pessoas que ali trabalham. Observa-se resiliência, responsabilidade, dedicação e uma forte disposição para a melhoria contínua. Da lealdade, da determinação em ultrapassar obstáculos e do desejo de se superarem, vem a vontade e, principalmente, o orgulho em vestir a camisola do *SLM Group*. Centram-se no rigor, na organização e na qualidade, aspirando à excelência em todas as ações.

O compromisso integral com a comunidade local é adotado, com ênfase na prestação de serviço e cooperação, visando ao seu desenvolvimento e progresso, e para além disso, também a promoção da economia circular é uma prioridade. Acreditam que a determinação em cumprir os acordos com todos os parceiros contribui para a criação de

valor, mantendo um sólido compromisso com valores humanos e solidariedade para com todos os membros do Grupo.

O conhecimento especializado é fundamental para o bom desempenho, e os padrões de exigência são aplicados de forma consistente, tanto internamente quanto em relação a terceiros. Acreditam na sua capacidade de agir com rapidez, flexibilidade e disposição para inovar, reconhecendo a inevitabilidade da aprendizagem contínua. São evidenciados o respeito e a cortesia em todas as circunstâncias. Reconhecem que uma equipa bem estruturada, coesa, solidária, ágil e que promove uma cultura de feedback é capaz de atingir resultados significativos. Procuram colaborar harmoniosamente na procura de um objetivo comum e demonstram interesse no sucesso da organização como um todo, em detrimento de apenas focar no sucesso pessoal ou de apenas uma equipa.

### **2.3. Ciência de Dados na tomada de decisão**

Neste tópico são abordados temas cruciais relacionados à gestão de dados e à tomada de decisões em ambiente industrial. Dividida em seis secções, esta secção aborda questões fundamentais para a compreensão da relevância dos dados na era da informação (secção 2.3.1), o *Business Intelligence* e a sua importância (secção 2.3.2), bem como a importância da tomada de decisão no contexto empresarial (secção 2.3.3). Também são apresentadas as metodologias e tecnologias utilizadas para recolher dados em ambientes industriais (secção 2.3.4), juntamente com exemplos de como a aprendizagem de máquina pode ser aplicada na otimização de processos e na melhoria da eficiência operacional (secção 2.3.5) e finalmente uma abordagem acerca da utilização de *Machine Learning* em ambiente industrial (secção 2.3.6). Esta análise detalhada oferece uma visão abrangente e atualizada do panorama do estado atual da gestão de dados e do uso estratégico da informação no setor industrial, fornecendo uma base sólida para o aprofundamento nas secções subsequentes do relatório de estágio.

#### **2.3.1. Os dados e a sua importância**

A importância dos dados em ambiente industrial é cada vez mais evidente e relevante nos dias de hoje. Com o avanço da tecnologia e o surgimento da Indústria 4.0, as empresas industriais estão a transformar-se em ambientes altamente digitalizados, onde uma quantidade massiva de dados é gerada a cada instante.

Num contexto industrial, os dados desempenham um papel fundamental na tomada de decisões [5]. Fornecem *insights* valiosos sobre o desempenho das operações, qualidade dos produtos, eficiência dos processos e até mesmo para a saúde e segurança dos trabalhadores. Com base nesses dados, as empresas podem identificar problemas e oportunidades, implementar melhorias e antecipar exigências do mercado [5].

A análise de dados é especialmente relevante neste contexto, pois permite extrair valor e conhecimento a partir do grande volume de informações gerado pela indústria. Ao aplicar técnicas e algoritmos de análise, é possível identificar padrões, tendências e correlações nos dados, que muitas vezes passariam despercebidos aos olhos humanos. Isso proporciona uma compreensão mais profunda dos processos industriais e ajuda na identificação de possíveis melhorias e otimizações.

Além disso, esta análise, em ambiente industrial, também é essencial para a implementação de práticas de manutenção preventiva [6]. Através da monitorização constante dos equipamentos e análise dos dados recolhidos, é possível identificar padrões de falha e prever quando uma máquina ou componente pode apresentar problemas. Desta forma, a empresa pode planejar intervenções de manutenção de forma proativa, evitando paragens não programadas e reduzindo os custos de manutenção [6].

A análise de dados pode contribuir também para a segurança no ambiente industrial. Ao monitorizar e analisar dados relacionados à segurança, como incidentes, acidentes e comportamento dos trabalhadores, é possível identificar padrões de risco e implementar medidas preventivas mais eficazes. Isto pode levar a uma redução significativa de acidentes e incidentes, garantindo um ambiente de trabalho mais seguro para os funcionários [7].

Em suma, a importância dos dados em ambiente industrial está diretamente ligada à capacidade de uma empresa tomar decisões mais informadas, melhorar a eficiência operacional, otimizar processos, antecipar problemas e reduzir custos e desempenha assim um papel fundamental neste contexto, permitindo extrair conhecimentos valiosos dos dados e transformando-os em *insights* acionáveis.

### **2.3.2. Business Intelligence**

O *Business Intelligence* (BI) tem-se destacado como uma das principais vertentes da gestão empresarial contemporânea em Portugal [8]. Trata-se de uma abordagem

estratégica que visa a transformação de dados em informações valiosas, proporcionando uma visão analítica e aprofundada do desempenho organizacional [9]. Neste contexto, uma melhor definição do conceito “BI” revela-se essencial para compreender a sua evolução e o impacto nas organizações.

O BI fundamenta-se na utilização de ferramentas e tecnologias avançadas que possibilitam a recolha, análise e interpretação de dados provenientes de diversas fontes [10]. Ao integrar essas informações, é possível identificar padrões, tendências e *insights* relevantes para a tomada de decisões informadas e estratégicas. A capacidade de transformar dados brutos em conhecimento significativo torna o BI uma poderosa ferramenta para impulsionar o sucesso empresarial.

Uma das suas tendências notáveis é a crescente adoção de metodologias ágeis, como o *Scrum*, para o desenvolvimento de projetos de BI [11]. Essa abordagem favorece a flexibilidade, permite entregas contínuas e incrementais e possibilita aos utilizadores, uma análise em tempo real do ponto de situação e da capacidade de se adaptarem às necessidades que estão em constante evolução.

Além disso, também a visualização de dados é um elemento-chave do *Business Intelligence*. Através de gráficos, *dashboards* e relatórios interativos, o BI oferece uma representação visual apelativa das informações, facilitando a interpretação e a partilha do conhecimento em toda a organização [12]. Essa abordagem intuitiva contribui para uma comunicação eficaz e uma tomada de decisões mais ágil, potenciando o seu uso nas organizações.

Com a evolução da tecnologia, o BI incorpora, cada vez mais, soluções de BI baseadas em *cloud* [13], que permitem o acesso e a análise de dados em tempo real a partir de qualquer lugar, contribuindo para uma gestão mais descentralizada e facilitando o trabalho remoto, uma tendência que ganhou relevância durante o período de pandemia.

Em resumo, o *Business Intelligence* espelha toda a sua importância crescente na gestão empresarial atual por meio de ferramentas analíticas avançadas, metodologias ágeis e visualização de dados. O BI tem impulsionado a eficiência operacional e a tomada de decisões informadas. Todavia, é essencial reconhecer os desafios e trabalhar em prol de soluções que permitam potencializar ainda mais a sua aplicação e relevância no ambiente corporativo.

### 2.3.3. A importância da tomada de decisão nas empresas

A tomada de decisão é um processo fundamental em qualquer organização empresarial, influenciando diretamente o sucesso e o crescimento de uma empresa. Neste sentido, é importante perceber também a sua relevância e os principais desafios associados à tomada de decisão nas empresas.

Historicamente, a tomada de decisão nas empresas foi baseada em abordagens mais tradicionais, com as tão conhecidas técnicas de análise SWOT (*Strengths, Opportunities, Weaknesses and Threats*) e as análises custo-benefício [14]. Estas abordagens têm sido valiosas na identificação e avaliação de fatores internos e externos que impactam as operações empresariais, no entanto, e à medida que o ambiente empresarial se torna mais complexo, surgem desafios adicionais que requerem abordagens mais avançadas e é aqui que os dados ganham força [15].

Com o avanço tecnológico e a abundância de dados disponíveis, as empresas têm adotado cada vez mais abordagens baseadas em análise de dados para a tomada de decisão. A utilização de ferramentas de análise de dados, como *Data Mining* (DM) [16] e a Inteligência Artificial (IA) [17], permite que as empresas obtenham *insights* valiosos a partir de grandes volumes de informações e como será de esperar, esta abordagem auxilia na identificação de padrões, na previsão de tendências e no desenvolvimento de estratégias mais informadas.

Outra tendência emergente é a tomada de decisão colaborativa, na qual múltiplos *players* dentro da organização são envolvidos no processo decisório [18]. Através da partilha de conhecimentos e da participação de diferentes perspetivas, as empresas procuram evitar as tomadas de decisão unilaterais e, ao invés disso, procuram promover soluções mais abrangentes e eficazes. Esta abordagem colaborativa incentiva a diversidade de pensamento [18] e a criação de um ambiente mais inovador e inclusivo.

Num mundo empresarial caracterizado pela velocidade a que as coisas mudam, a tomada de decisão em tempo real tem-se mostrado cada vez mais relevante [19]. A capacidade de avaliar rapidamente as informações disponíveis e tomar decisões ágeis é essencial para responder a eventuais imprevistos e aproveitar oportunidades de curto prazo. A tecnologia desempenha um papel fundamental aqui também, sendo que é capaz de fornecer acesso quase instantâneo a dados e análises.

Assim, pode-se concluir, que a tomada de decisão nas empresas é uma atividade crítica que envolve a análise de informações, a avaliação de riscos e a consideração de inúmeros fatores [14]. À medida que o mundo empresarial evolui, é essencial que as empresas adotem abordagens mais avançadas e estratégicas. A análise de dados, a tomada de decisão colaborativa e a capacidade de tomar decisões em tempo real são tendências emergentes que estão a transformar a forma como as empresas operam.

#### **2.3.4. Recolha de dados em ambiente industrial**

A recolha de dados em ambientes industriais apresenta diversos desafios, sendo o volume, a variedade, a velocidade e a veracidade dos dados alguns dos principais obstáculos a serem superados [20].

Nestes ambientes, o acesso e a integração dos dados são particularmente difíceis, uma vez que é comum as empresas operarem em “ilhas”<sup>1</sup>, ou seja, de forma desligada entre os diferentes departamentos. Um dos desafios mais evidentes é o volume dos dados gerados nos ambientes industriais.

Com a adoção crescente de tecnologias de automação e monitorização, sensores, sistemas SCADA (*Supervisory Control and Data Acquisition*) e outros dispositivos é possível recolher uma quantidade massiva de dados em tempo real [21]. A recolha e tratamento de dados em tempo real desempenham um papel crucial na era da informação e das tecnologias avançadas. Esta prática permite que empresas e organizações obtenham informações imediatas e precisas sobre o seu desempenho operacional, comportamento do cliente e tendências de mercado [22].

Ao aceder a dados atualizados em tempo real, as empresas podem tomar decisões mais informadas e estratégicas, ajustando as suas operações de forma ágil e eficiente para otimizar resultados [19]. Além disso, a capacidade de processar dados em tempo real oferece a oportunidade de detetar e reagir a eventos críticos rapidamente, como ameaças de segurança cibernética ou problemas operacionais, minimizando assim potenciais danos. Em última análise, a recolha e tratamento de dados em tempo real impulsionam a inovação, melhoram a experiência do cliente e proporcionam uma vantagem competitiva no cenário empresarial moderno [20]. Esses dados incluem informações sobre a produção, a qualidade, a manutenção, os inventários, entre outros, e lidar com esse

---

<sup>1</sup> Os vários departamentos ou secções trabalham de forma independente.

volume de dados pode ser um desafio em termos de armazenamento, processamento e análise eficiente. Além disso, a variedade dos dados é outro desafio importante.

Nos ambientes industriais, os dados podem ser provenientes de diferentes fontes, como sensores de máquinas, sistemas de gestão de qualidade, sistemas de gestão de manutenção, entre outros. Essa diversidade de fontes resulta em formatos e estruturas de dados heterogêneos, o que dificulta a integração e a análise conjunta dessas informações.

A falta de padronização pode levar a problemas de compatibilidade e dificultar a obtenção de uma visão holística dos processos industriais [23]. A velocidade com que os dados são gerados também representa um desafio significativo. Em ambientes industriais, a produção é contínua e em tempo real, o que significa que os dados são gerados em alta velocidade. É essencial ter sistemas e infraestruturas adequadas a capturar, processar e armazenar esses dados em tempo útil [24]. A capacidade de lidar com a velocidade dos dados é crucial para garantir a tomada de decisões em tempo real e a resposta rápida a eventos ou problemas operacionais.

Por fim, a veracidade dos dados é um desafio crítico nos ambientes industriais. Devido à natureza descentralizada e fragmentada das operações, é comum encontrar inconsistências, incoerências e falta de qualidade nos dados [25]. Erros de medição, falhas nos sensores, dados manualmente alterados e problemas de integração podem levar a informações imprecisas ou incompletas. A garantia da qualidade dos dados é fundamental para evitar decisões erradas ou baseadas em informações incorretas.

### **2.3.5. *Machine Learning***

Na era dos grandes volumes de dados e das capacidades computacionais avançadas, a Ciência de Dados emergiu como um campo crucial para extrair informações valiosas a partir de vastas quantidades de dados. No cerne da Ciência de Dados encontra-se o *Machine Learning* (ML), um subconjunto da inteligência artificial que permite que os computadores aprendam com dados e melhorem o seu desempenho ao longo do tempo [26]. Esta secção tem como objetivo fornecer uma visão geral dos principais conceitos, metodologias e aplicações do ML no contexto da Ciência de Dados.

ML abrange uma variedade de algoritmos e técnicas que permitem que os sistemas aprendam padrões e façam previsões ou tomem decisões sem serem explicitamente

programados. Assim, os três tipos fundamentais de ML são a aprendizagem supervisionada, a não supervisionada e a de reforço [27].

- Aprendizagem Supervisionada

Nesta abordagem, os modelos são treinados utilizando dados rotulados, em que as entradas e as suas saídas correspondentes são conhecidas. O modelo aprende a relacionar as entradas com as saídas e pode posteriormente fazer previsões sobre novos dados não vistos. Algoritmos comuns de aprendizagem supervisionada incluem Regressão Linear, *Decision Trees*, *Random Forests*, Máquinas de Vetores de Suporte (SVM) e Redes Neurais [27].

- Aprendizagem Não Supervisionada

A aprendizagem não supervisionada lida com dados não rotulados, em que o modelo tenta encontrar padrões, estruturas ou agrupamentos dentro destes mesmos dados. É frequentemente usado para análise exploratória de dados e pré-processamento. Algoritmos comuns de aprendizagem não supervisionada são o Agrupamento *K-Means*, o Agrupamento Hierárquico, a Análise de Componentes Principais (PCA) e os *Autoencoders* [27].

- Aprendizagem por Reforço

Essa abordagem envolve o treino de um agente para interagir com um ambiente e aprender através do *feedback* na forma de recompensas ou penalidades. A aprendizagem por reforço é amplamente utilizada em robótica, jogos e problemas de otimização [27].

A integração de técnicas de ML na Ciência de Dados revolucionou várias indústrias e domínios. Algumas aplicações proeminentes incluem a análise preditiva, os processamentos de linguagem natural, a visão computacional e também os sistemas de recomendação.

Na Análise Preditiva, o ML permite que os cientistas de dados construam modelos preditivos que antecipam tendências futuras, comportamentos ou eventos [28]. Estes modelos são aplicados em previsões de vendas, previsões de procura, mercados financeiros, entre outros. Foi também este o tipo de aplicação utilizado em ambiente de estágio que será apresentado posteriormente.

O Processamento de Linguagem Natural (PLN) utiliza ML para permitir que os computadores entendam, interpretem e gerem, o máximo possível, linguagem idêntica à humana. Algumas aplicações incluem análise de sentimento, tradução de idiomas, *chatbots* e assistentes de voz [29].

Na Visão Computacional o ML desempenha um papel fundamental no avanço de tarefas de visão computacional, como detecção de objetos, segmentação de imagens, reconhecimento facial e condução autónoma em veículos mais recentes [30].

Os sistemas de recomendação são maioritariamente utilizados por plataformas de *e-commerce*, serviços de *streaming* de conteúdo e *sites* de redes sociais e utilizam algoritmos de ML para fornecer recomendações personalizadas aos seus utilizadores, aprimorando a sua experiência bem como a adesão aos seus conteúdos ou plataformas [31].

Apesar do amplo sucesso, o ML na Ciência de Dados enfrenta vários desafios. Alguns deles incluem a necessidade de modelos mais interpretáveis, o tratamento de dados tendenciosos ou manipulados, a garantia da privacidade e da segurança dos dados e o tratamento das implicações éticas da tomada de decisões impulsionada pela IA [32].

O ML é, sem sombra de dúvidas, um componente crucial da Ciência de Dados, principalmente por permitir que empresas e investigadores desvendem o potencial dos dados de maneiras sem precedentes. À medida que este campo continua a evoluir, a fusão da *expertise* humana com algoritmos inteligentes promete moldar um futuro cada vez mais orientado a dados, eficiente e inovador. No entanto, o uso responsável e ético do ML continua a ser essencial para que se possa aproveitar todo o potencial máximo em benefício da sociedade.

### **2.3.6. *Machine Learning* em ambiente industrial**

Os modelos de ML têm o potencial de ter um impacto significativo em ambientes industriais, oferecendo oportunidades para melhorar a eficiência operacional, otimizar processos, reduzir custos e aumentar a qualidade dos produtos finais [33]. Com o avanço da tecnologia e a disponibilidade de grandes conjuntos de dados, os modelos de ML estão a desempenhar um papel cada vez mais relevante na indústria [34].

Um dos principais impactos dos modelos de ML é a capacidade de previsão e prognóstico [35]. Por exemplo, um modelo de ML pode ser treinado para prever a procura futura de produtos com base em dados históricos de vendas, permitindo assim que as empresas ajustem a produção de acordo com a procura.

Além disso, os modelos de ML também podem ser usados para prever falhas em equipamentos e antecipar manutenção preditiva, reduzindo o tempo de inatividade não planeado e melhorando a eficiência dos mais variados processos [6].

Os modelos de ML podem analisar dados complexos e identificar padrões difíceis que podem conduzir a melhorias nos processos de produção. Por exemplo, eles podem otimizar os parâmetros de funcionamento de uma máquina, minimizando o desperdício de materiais [36] ou maximizando a eficiência energética. Isso resulta em economias de recursos e redução de custos para as empresas.

Para além disso, podem ser desenvolvidos modelos que analisam os dados de sensores e câmaras para detetar defeitos ou malformações em produtos, em tempo real [37]. Isso permite a identificação precoce de problemas e ações corretivas imediatas, melhorando a qualidade e reduzindo o desperdício.

No atual momento, os modelos de ML estão a ser aplicados em diversas áreas da indústria, como na área da energia, dos transportes e principalmente na logística [38]. Algoritmos avançados, como redes neuronais convolucionais (CNN) e redes neuronais recorrentes (RNN), têm sido utilizados com sucesso para resolver problemas complexos e melhorar o desempenho industrial. Além disso, a combinação de ML com outras tecnologias, como a “Internet das Coisas” (IoT) e a análise de *big data*, está a impulsionar ainda mais o impacto dos modelos de ML em ambientes industriais [39].

A disponibilidade de sensores e dispositivos conectados permite a recolha em tempo real de dados de várias fontes, fornecendo informações valiosas para os modelos de ML [22]. Ainda assim, é importante salientar que a implementação bem-sucedida de modelos de ML em ambientes industriais requer uma abordagem cuidadosa.

É necessário um planeamento adequado, seleção adequada de algoritmos, preparação e limpeza de dados, infraestrutura de tecnologias de informação robusta e colaboração interdisciplinar. Além disso, considerações éticas e de segurança também devem ser levadas em conta para garantir o uso responsável dos modelos de ML.

### 3. Atividades desenvolvidas

O presente capítulo apresenta uma descrição detalhada das principais ações e projetos realizados durante o período do estágio. Este capítulo está subdividido em quatro secções, cada uma delas abordando diferentes aspetos e tecnologias utilizadas para o desenvolvimento das atividades propostas.

Na secção denominada "Tecnologias Utilizadas" (Secção 3.1) são destacadas as ferramentas e plataformas tecnológicas empregues ao longo do estágio. São abordadas as funcionalidades e contribuições específicas do *Excel*, *PowerPoint*, *Word*, *Power BI*, *Google Colab*, *Python*, *Canva*, *Microsoft Forms* e *Mentimeter* para a manipulação de dados, criação de apresentações visuais, análises estatísticas, visualização interativa e muito mais.

Em seguida, na secção denominada "Modelo Preditivo de Cumprimento de Planeamentos Produtivos" (Secção 3.2) é detalhado o desenvolvimento de um modelo preditivo utilizando técnicas de ML. São explorados os métodos e algoritmos empregados para a criação desse modelo cujo principal objetivo eram o de melhorar a precisão na estimativa do cumprimento de planeamentos produtivos.

A secção denominada "Dashboards em *Power BI*" (Secção 3.3) foca-se na elaboração de *dashboards* interativos utilizando a ferramenta *Power BI*. São apresentados os indicadores e métricas-chave selecionados para compor cada um dos *dashboards*, permitindo uma visualização dinâmica e intuitiva dos dados para facilitar a tomada de decisão.

Na secção denominada "Processos ETL e Relatórios" (Secção 3.4) são descritas as etapas e técnicas utilizadas na extração, transformação e carregamento (ETL) de dados. Além disso, são apresentados alguns exemplos dos relatórios gerados a partir desses dados processados, fornecendo informações essenciais para a compreensão do desempenho e eficiência das operações realizadas no decorrer do estágio.

Na secção "Auditoria interna de segurança da informação TISAX" (Secção 3.5) foi feita uma breve apresentação do que se trata numa auditoria deste género, o que é, especificamente "TISAX", bem como as atividades acompanhadas.

Para terminar, na secção denominada “Apreciação crítica às atividades desenvolvidas” (Secção 3.6) foi elaborada uma análise reflexiva e ponderada das atividades realizadas em contexto de estágio. Neste parágrafo foi avaliado o desempenho, os resultados e os impactos das atividades de acordo com critérios específicos, incluindo elementos como eficácia, eficiência, conformidade com objetivos estabelecidos, ou até a qualidade do trabalho executado. Adicionalmente, esta apreciação crítica destacou o valor agregado das atividades em relação aos objetivos globais do estágio, fornecendo assim uma visão equilibrada e informada sobre o desempenho nestas mesmas atividades.

Ao explorar estas cinco secções, este capítulo pretende oferecer uma visão completa e aprofundada das atividades desenvolvidas durante o estágio, realçando as tecnologias e abordagens utilizadas para alcançar os objetivos propostos e os resultados obtidos ao longo do processo.

### **3.1. Tecnologias utilizadas**

A secção "Tecnologias Utilizadas" tem como objetivo apresentar as principais ferramentas tecnológicas empregues neste trabalho, que desempenharam um papel fundamental na recolha, análise e apresentação dos dados. As seguintes ferramentas são exploradas em detalhe: *Excel* (Secção 3.1.1), *PowerPoint* (Secção 3.1.2), *Google Colab* em *Python* (Secção 3.1.3), *Word* (Secção 3.1.4), *Power BI* (Secção 3.1.5), *Canva* (Secção 3.1.6), *Microsoft Forms* (Secção 3.1.7) e *Mentimeter* (Secção 3.1.8).

Com a exploração detalhada destas ferramentas tecnológicas, esta secção visa demonstrar a diversidade e a importância das soluções utilizadas para alcançar os objetivos propostos no desenvolvimento do trabalho. As funcionalidades e contribuições específicas de cada uma delas permitiram uma abordagem abrangente e aprofundada, enriquecendo o processo de análise e interpretação dos dados recolhidos.

#### **3.1.1. Excel**

O *Excel*, uma das aplicações mais conhecidas do *Microsoft Office*, é uma ferramenta de folha de cálculo eletrónica. No decorrer deste estágio, foram utilizadas, maioritariamente, funcionalidades para manipulação de dados, bem como para realizar cálculos e gráficos com os dados recolhidos dos inquéritos efetuados. O *Excel* facilitou na organização e na análise dos dados recolhidos, permitindo a criação de fórmulas e funções para realizar alguns cálculos mais complexos. Além disso, a ferramenta foi

essencial na criação de gráficos, tabelas, e outras infografias, proporcionando uma representação visual clara e concisa dos resultados obtidos. Com a sua capacidade de processar grandes volumes de informações de forma eficiente, o *Excel* agilizou o processo de análise e contribuiu para a obtenção de *insights* valiosos para o desenvolvimento deste trabalho.

### **3.1.2. PowerPoint**

O *PowerPoint*, uma das aplicações mais conhecidas do *Microsoft Office*, é uma ferramenta de criação de apresentações visuais. No decorrer deste estágio, foram utilizadas, maioritariamente, funcionalidades para a elaboração de *slides* envolventes e informativos. O *PowerPoint* permitiu comunicar os resultados e conclusões obtidos ao longo do trabalho de forma clara e impactante, enriquecendo a apresentação dos dados com recursos visuais, como imagens, gráficos e animações. Além disso, a aplicação possibilitou a organização e a formatação do conteúdo, garantindo uma apresentação coesa e profissional. Com o auxílio do *PowerPoint*, as informações relevantes foram destacadas de forma visualmente atrativa, contribuindo para uma melhor compreensão e assimilação do público-alvo.

### **3.1.3. Google Colab em Python**

O *Google Colab*, uma plataforma de desenvolvimento colaborativo baseada em *Python*, da *Google*, é uma excelente ferramenta para análise e manipulação de dados [40]. No decorrer deste estágio, foram utilizadas, maioritariamente, as funcionalidades do *Google Colab* em conjunto com a linguagem de programação *Python* para realizar tarefas avançadas de análise, tratamento e visualização de dados, principalmente aquando da criação do modelo de ML. Através do *Google Colab*, foi possível implementar algoritmos relativamente complexos, criar o modelo de ML e realizar cálculos estatísticos. A integração do *Python* no ambiente do *Google Colab* permite o acesso a bibliotecas especializadas, facilitando a manipulação e o processamento eficiente dos dados recolhidos. Com esta combinação poderosa, foram alcançados resultados mais abrangentes e aprofundados, contribuindo para a obtenção de *insights* significativos ao longo deste trabalho.

### **3.1.4. Word**

O *Word* é, também, um dos programas mais conhecidos do *Microsoft Office* e caracteriza-se por ser uma ferramenta de processamento de texto. No decorrer deste estágio, foram utilizadas, maioritariamente, as funcionalidades do *Word* para a elaboração de relatórios detalhados e para documentação dos métodos e resultados obtidos. Através do *Word*, foi possível organizar e apresentar as informações de forma clara e estruturada, garantindo uma apresentação coesa, simples e uniforme do trabalho. Além disso, o *Word* facilitou a formatação do conteúdo, tornando a apresentação mais profissional e esteticamente apelativa.

### **3.1.5. Power BI**

O *Power BI*, uma ferramenta de BI desenvolvida pela *Microsoft*, é uma ótima aplicação para visualização e análise de dados [41]. No decorrer deste estágio, foram utilizadas, maioritariamente, as funcionalidades do *Power BI* para a criação de painéis interativos e relatórios dinâmicos. Através do *Power BI*, foi possível transformar os dados recolhidos em visualizações significativas, como gráficos, *cards* informativos e mapas interativos. A ferramenta permitiu explorar os dados de forma mais aprofundada e compreender as relações existentes entre as diferentes variáveis. Com o auxílio do *Power BI*, a análise dos dados tornou-se mais intuitiva e eficiente, contribuindo para a obtenção de *insights* valiosos que enriqueceram o desenvolvimento deste trabalho.

### **3.1.6. Canva**

O *Canva* é uma aplicação *online* para a criação de materiais visuais atrativos. No decorrer deste estágio, foram utilizadas, maioritariamente, as funcionalidades do *Canva* para produzir elementos gráficos que enriqueceram a apresentação dos dados, bem como para realizar relatórios em *PDF* [42]. Com o *Canva*, foi possível criar infográficos, imagens ilustrativas e *layouts* personalizados para destacar as informações mais relevantes. Nele também foram construídos a maior parte dos *slides* para anunciar o lançamento de novos inquéritos. A facilidade de uso e a vasta coleção de recursos visuais disponíveis no *Canva* permitiram uma produção rápida e eficiente de materiais muito apelativos. Ao incorporar esses elementos gráficos no trabalho, a comunicação dos resultados e conclusões tornou-se mais envolvente e compreensível, proporcionando uma experiência mais agradável e impactante para o público-alvo, isto é, para os colaboradores da organização.

### **3.1.7. Microsoft Forms**

O *Microsoft Forms* é uma ferramenta de criação de formulários online da *Microsoft* e esta foi muito utilizada no decorrer deste estágio. As funcionalidades do *Forms* foram utilizadas, maioritariamente, para a recolha de dados através de questionários e pesquisas. Com o *Microsoft Forms*, foi possível criar formulários personalizados com perguntas e opções de resposta variadas, de forma rápida e fácil. A ferramenta permitiu uma recolha eficiente de dados relevantes, que posteriormente foram analisados no *Excel* e no *Power BI*, para obter *insights* importantes.

### **3.1.8. Mentimeter**

O *Mentimeter*, uma plataforma interativa de votação em tempo real, é uma aplicação que desempenhou um papel importante durante este estágio. As suas funcionalidades foram utilizadas, maioritariamente, para obter *feedback* e opiniões da população da organização durante inquéritos curtos e algumas formações [43]. Com esta plataforma foi possível criar questionários interativos com questões diversas, como inquéritos e pesquisas de opinião. A plataforma permitiu que os participantes respondessem às questões em tempo real através dos seus dispositivos móveis, proporcionando sempre uma interação dinâmica e envolvente, muito devido ao facto dos colaboradores poderem acompanhar o desenrolar das votações, em tempo real, nos seus equipamentos. As funcionalidades do *Mentimeter* enriqueceram as avaliações de formação e inquéritos, permitindo a captação imediata das demais opiniões e a participação ativa da população geral. Essas informações foram valiosas para avaliar o entendimento do conteúdo apresentado e para perceber as vontades das maiorias (quando falamos da sua utilização para os “*Inquéritos com Sentido*”). O *Mentimeter* foi uma ferramenta essencial para garantir uma experiência interativa e impactante durante as atividades realizadas durante o estágio.

## **3.2. Modelo preditivo de cumprimento de planeamentos produtivos**

Como seria de esperar, o propósito de um modelo preditivo de cumprimento de planeamentos produtivos, conforme o nome sugere, é prever e estimar a probabilidade de um planeamento produtivo ser cumprido com sucesso. Este tipo de modelos, geralmente, utiliza dados históricos e utiliza as suas variáveis mais relevantes em conjunto com algoritmos de ML para analisar padrões e tendências, por forma a permitir que as

empresas antecipem possíveis desvios ou problemas no cumprimento dos planos, ou até mesmo nos processos produtivos através do uso de sistemas computacionais. Com este auxílio, as empresas podem tomar medidas proativas e antecipadas para otimizar a produção, ajustar recursos (como mão de obra ou matéria-prima) e evitar atrasos ou interrupções, aumentando a eficiência e a produtividade geral.

Jean-Paul Tremblay e Paul G. Sorenson [44] identificam 4 problemas interrelacionados aquando da procura de soluções a partir de sistemas computacionais, sendo eles: (i) a compreensão profunda das relações existentes entre os diversos elementos de dados relevantes para a procura dessa solução; (ii) a decisão sobre as operações que devem incidir sobre esses elementos recolhidos; (iii) a conceção de métodos de representação desses elementos no sistema, para que sejam conservadas, da melhor forma possível, as relações lógicas entre elementos, bem como garantir que estas operações são efetuadas de forma fácil e eficiente; e (iv) a decisão sobre a linguagem de decisão de problemas mais adequada à solução do problema em causa.

Estes autores [44] subdividem ainda a análise da resolução deste tipo de problemas em três fases, sendo estas: (i) a seleção de um modelo matemático adequado, tendo em conta a estrutura dos dados; (ii) a formulação de algoritmos com base na escolha da fase anterior; e (iii) a conceção de estruturas de armazenamento para os *datasets* já obtidos.

O modelo de ML trabalhado durante o estágio tinha, como objetivo final, prever o cumprimento (ou a falta dele) de planeamentos produtivos, e como tal, seria expectável que o resultado se traduzisse em “Cumpre” ou “Não cumpre”. Tendo isto em conta e olhando também aos conjuntos de dados disponíveis, depois de alguma pesquisa optamos por utilizar um modelo de regressão logística.

De acordo com Kevin Murphy, a regressão logística estima a probabilidade de um determinado exemplo pertencer a uma determinada categoria, classificando-a, utilizando uma função logística para modelar a relação entre as variáveis de entrada e a variável categórica. O autor afirma que “A isto chama-se regressão logística devido à sua semelhança com a regressão linear (embora seja uma forma de classificação e não de regressão!)” [45].

A regressão logística é especialmente adequada a problemas de classificação binária, em que a variável categórica possui apenas duas categorias (“0 ou 1”, “Sim ou Não”,

“Cumpre ou Não Cumpre”). Esta utiliza a *função logística* (também chamada de função sigmoide) para mapear a soma ponderada das variáveis independentes para um valor entre 0 e 1 [46], e é definida por “ $P(Y = 1 | X) = 1 / (1 + \exp(-z))$ ”, onde  $P(Y = 1 | X)$  é a probabilidade condicional de  $Y$  ser igual a 1 dado  $X$ ,  $\exp$  é a função exponencial e  $z$  é a soma ponderada das variáveis independentes. Essa soma ponderada é calculada multiplicando cada variável independente pelo seu coeficiente correspondente no modelo.

Posto isto, de seguida será explicado todo o processo de recolha de dados e o seu respetivo tratamento (secção 3.2.1), a preparação do modelo e o desenvolvimento do mesmo (secções 3.2.2, 3.2.3, 3.2.4 e 3.2.5). Para finalizar, são apresentados os resultados obtidos e é feita uma conclusão deste tópico (secção 3.2.6).

### **3.2.1. Recolha e tratamento dos dados**

A recolha de dados é uma etapa fundamental no processo de investigação. É neste momento que são recolhidas as informações necessárias para responder às perguntas de pesquisa.

Este momento é composto por um conjunto de etapas cuidadosamente planeadas e executadas, que têm como objetivo reunir informações relevantes e fiáveis de várias fontes e prepará-las para análise. É um processo fundamental para transformar dados brutos em *insights* acionáveis, capazes de impulsionar o desempenho das organizações em diferentes setores. Durante a recolha de dados são identificadas e selecionadas as fontes mais apropriadas para obter as informações necessárias. Isso pode incluir dados provenientes de sistemas internos, como bases de dados, registos de vendas ou informações de clientes, bem como fontes externas, como pesquisas de mercado ou dados públicos. A qualidade e a abrangência dos dados recolhidos são essenciais para garantir resultados precisos e significativos.

Uma vez recolhidos, os dados passam por um processo de tratamento, que envolve etapas como limpeza, padronização e transformação. Nesta fase, são aplicadas técnicas para corrigir erros, remover dados duplicados ou em falta, normalizar formatos e estruturas e combinar informações de diferentes fontes. [47] O objetivo é obter um conjunto de dados coerente, consistente e pronto para análise.

Neste caso, durante a etapa de recolha de dados, foram utilizadas fontes de dados em *Excel*, sendo todos eles relacionados por meio da chave primária "Dia". Essa abordagem

permitiu a integração e consolidação dos dados provenientes de diferentes fontes, proporcionando uma visão abrangente e consistente das informações.

O conjunto de dados recolhidos abrange um histórico de dois anos, desde o início de 2021, fornecendo uma perspetiva temporal ampla para análise e tomada de decisões fundamentadas.

Interessou então demonstrar que tipo de dados foram utilizados para o modelo. Na Tabela 1, abaixo, segue um breve resumo da estrutura de cada *dataset* utilizado.

Tabela 1 - Estrutura dos vários conjuntos de dados

Nome do <i>dataset</i>	Nº e Nome das colunas	Tipo de ficheiro	Forma de registo
obj_OP1	3 colunas – “Dia”, “Ref”, “Obj_OP1”	.xlsx	<i>Excel</i>
obj_OP2	3 colunas – “Dia”, “Ref”, “Obj_OP2”	.xlsx	<i>Excel</i>
obj_OP3	3 colunas – “Dia”, “Ref”, “Obj_OP3”	.xlsx	<i>Excel</i>
obj_OP4	3 colunas – “Dia”, “Ref”, “Obj_OP4”	.xlsx	<i>Excel</i>
ddp	3 colunas – “Dia”, “Ref”, “sum.ddp”	.xlsx	<i>Excel</i>
hr	2 colunas – “Dia”, “humanresources”	.xlsx	DB SQL
hxt	2 colunas – “Dia”, “sum.hxt”	.xlsx	<i>Excel</i>
orders	3 colunas – “Dia”, “Ref”, “Obj_order”	.xlsx	Excel
stocks	4 colunas – “Dia”, “Ref”, “available.stock”, “intransit.stock”	.xlsx	<i>Excel</i>

Para além desta, também se crê que seja relevante especificar, para cada coluna, o tipo de dados tratados. Assim, na Tabela 2 são apresentados os tipos de dados recolhidos e trabalhados, bem como, que tipo de informação é que estes tratam.

Tabela 2 - Tipo de dado, por coluna

Coluna de dados	Tipo de dado	O que se trata?
“Dia”	Data (dd/mm/aaaa)	Dia do registo
“Ref”	String	Referência de peça
“Obj_OP1”	Numérico Inteiro	Objetivo para a etapa 1 do processo
“Obj_OP2”	Numérico Inteiro	Objetivo para a etapa 2 do processo
“Obj_OP3”	Numérico Inteiro	Objetivo para a etapa 3 do processo
“Obj_OP4”	Numérico Inteiro	Objetivo para a etapa 4 do processo
“sum.ddp”	Numérico Inteiro	Somatório dos envios em trans. Especial
“humanresources”	Numérico Inteiro	Nº de pessoas a trabalhar
“sum.hxt”	Numérico	Nº de horas extras trabalhadas
“available.stock”	Numérico Inteiro	Stock em armazém
“intransit.stock”	Numérico Inteiro	Stock em trânsito

Detalhando pormenorizadamente cada um dos dados:

a) **“Dia”**- O dado “Dia”, como o próprio nome indica, serve para identificar o dia da ocorrência registada no *dataset*, conforme se pode ver na Figura 2. O uso deste tipo de dado (*Date*) na modelação de um modelo de dados é uma escolha vantajosa e geralmente recomendada, sujeita às necessidades específicas do sistema em questão, e, de entre as suas diversas vantagens destacam-se, por exemplo:

- **Precisão Temporal:** Este tipo de dado "Dia" permite o armazenamento preciso de datas, englobando informações como dia, mês e ano, e, caso necessário, horas.

- **Validação de Dados:** A utilização do tipo de dado "Dia" contribui para a validação rigorosa das datas inseridas no sistema, assegurando que estas cumpram um formato válido e obedeçam a um padrão consistente, reduzindo potenciais erros de entrada de dados.

- **Funcionalidades Temporais:** Através do tipo de dado "Dia", torna-se possível realizar cálculos temporais e efetuar consultas baseadas em datas, como, por exemplo, o cálculo da idade de uma pessoa com base na sua data de nascimento ou a filtragem de dados mediante intervalos temporais.

- **Ordenação e Filtragem:** A utilização do tipo de dado "Dia" facilita a ordenação e a filtragem de registos por data, uma funcionalidade essencial em inúmeras aplicações.

- **Histórico de Dados:** No caso de o modelo de dados englobar informações históricas, o tipo de dado "Dia" possibilita o rastreamento de eventos e alterações ao longo do tempo, mantendo um registo cronológico.

- **Compatibilidade:** É relevante salientar que o tipo de dado "Dia" goza de ampla compatibilidade com sistemas de bases de dados e linguagens de programação, o que o torna altamente interoperável e passível de integração sem dificuldades em sistemas já existentes.

	A	B	C	D	E
1	Dia	Ref	Obj_OF		
2	04/01/2021	148X	0		
3	04/01/2021	107AE	600		
4	04/01/2021	107AC	0		
5	04/01/2021	MF32	0		
6	04/01/2021	941	2200		
7	04/01/2021	R077	0		
8	04/01/2021	R047	0		
9	04/01/2021	238	400		
10	04/01/2021	239	600		

Figura 2 - Exemplo do dado "Dia"

b) **“Ref”**- O dado “Ref” diz respeito à referência de peça sobre a qual se faz um registo produtivo. A inclusão de um dado destes, do tipo *string*, é fundamental na modelação de um modelo de ML. Pode ver-se um exemplo deste mesmo dado na Figura 3. Esta informação, embora aparentemente simples, pode ser de extrema importância por várias razões:

- **Identificação Única:** O campo "Ref" permite identificar cada peça de forma única e inequívoca. Isso é essencial para garantir que o modelo de ML possa distinguir entre diferentes peças e suas características associadas.

- **Agrupamento de Dados:** Usar uma referência de peça como "Ref" possibilita agrupar os dados relacionados a uma peça específica. Isso é especialmente útil quando se lida com grandes volumes de informações, permitindo ao modelo analisar o desempenho ou características de cada peça individualmente.

- **Rastreabilidade:** A "Ref" permite rastrear e associar dados específicos a uma peça ao longo do tempo. Isso é crucial em cenários de manutenção preditiva, onde se pretende monitorizar o desgaste ou a degradação de uma peça ao longo de várias medições.

- **Personalização e Recomendações:** Em sistemas de ML, a "Ref" pode ser usada para personalizar recomendações ou previsões com base na história e no desempenho passado de uma peça específica. Por exemplo, ao prever a manutenção necessária para uma peça com base em seu histórico de "Ref," o modelo pode fornecer recomendações mais precisas.

- **Validação de Dados:** A "Ref" pode ser usada para verificar a consistência e a integridade dos dados. Certificando-se de que cada peça tenha uma "Ref" válida, o modelo pode garantir que os dados de entrada sejam confiáveis e relevantes.

- **Redução de Erros:** A utilização de "Ref" pode ajudar a minimizar erros de entrada de dados, uma vez que fornece uma referência clara e padronizada para identificar cada peça. Isso ajuda a evitar confusões ou duplicações de informações.

	A	B	C	D	E
1	Dia	Ref	Obj_OF		
2	04/01/2021	148X	0		
3	04/01/2021	107AE	600		
4	04/01/2021	107AC	0		
5	04/01/2021	MF32	0		
6	04/01/2021	941	2200		
7	04/01/2021	R077	0		
8	04/01/2021	R047	0		
9	04/01/2021	238	400		
10	04/01/2021	239	600		

Figura 3 - Exemplo do dado "Ref"

c) **“Obj\_OPx”**- Os dados registados nos campos “Obj\_OPx”, conforme se pode ver na Figura 4, são dados de histórico sobre as quantidades produzidas, de determinada peça, em determinado dia. A inclusão de um dado destes, do tipo numérico inteiro, desempenha um papel crítico na modelação deste tipo de modelos. Este tipo de dado é de suma importância por diversas razões:

- **Definição de Metas Claras:** O campo "Obj\_OPx" estabelece metas quantitativas específicas para cada ordem de produção. Isso é essencial para que o modelo de ML compreenda as expectativas de produção e seja capaz de avaliar o desempenho em relação a essas metas.

- **Monitorização de Desempenho:** Com base nos valores de "Obj\_OPx", o modelo de ML pode monitorizar o desempenho em tempo real e alertar quando a produção está aquém ou excede as metas estabelecidas. Isso permite a intervenção proativa para otimizar a produção.

- **Análise de Desvios:** O campo "Obj\_OPx" possibilita a análise de desvios entre a produção real e as metas estabelecidas. Essa análise ajuda a identificar causas de variações e a implementar ações corretivas.

- **Otimização da Produção:** Ao utilizar os valores de "Obj\_OPx", o modelo de ML pode sugerir ajustes em tempo real para otimizar a produção e atingir as metas de forma mais eficiente, como alterar a programação de máquinas ou priorizar determinadas tarefas.

- **Planeamento de Recursos:** Com dados consistentes de "Obj\_OPx", é possível realizar um planeamento mais preciso dos recursos necessários para atingir as metas de produção, como mão-de-obra, matérias-primas e tempo.

- **Tomada de Decisões Informadas:** Com base nos dados de "Obj\_OPx" o modelo pode fornecer insights para a tomada de decisões informadas, como a alocação eficiente de recursos, ajustes de produção ou priorização de ordens de produção.

	A	B	C	D	E
1	Dia	Ref	Obj_OF		
2	04/01/2021	148X	0		
3	04/01/2021	107AE	600		
4	04/01/2021	107AC	0		
5	04/01/2021	MF32	0		
6	04/01/2021	941	2200		
7	04/01/2021	R077	0		
8	04/01/2021	R047	0		
9	04/01/2021	238	400		
10	04/01/2021	239	600		

Figura 4 - Exemplo do dado "Obj\_OPx"

d) **“sum.ddp”**- Este dado, possível de verificar na Figura 5, regista a quantidade de peças enviadas em transporte especial para o cliente. A inclusão do dado "sum.ddp," do tipo numérico inteiro, é de extrema importância na modelação de um modelo de ML e apresenta as seguintes vantagens:

- **Deteção de Anomalias:** O campo "sum.ddp" serve como um indicador crucial para a deteção de anomalias na linha de produção. Se "sum.ddp" for maior que zero, isso indica que existiram remessas especiais fora do cronograma planeado, o que pode ser considerado uma anomalia ou uma falha no processo de produção.

- **Avaliação de Desempenho:** Utilizando "sum.ddp", o modelo de ML pode avaliar o desempenho da produção em relação ao cumprimento dos prazos de entrega acordados com o cliente. A identificação de remessas fora de data pode ajudar a identificar áreas problemáticas na produção.

- **Gestão de Qualidade:** O dado "sum.ddp" está intrinsecamente ligado à qualidade do processo de produção. Se as remessas especiais forem frequentes devido a prazos não cumpridos, isso pode ser um indicativo de problemas na qualidade ou eficiência da produção.

- **Tomada de Decisões em Tempo Real:** Com base nos valores de "sum.ddp," o modelo de ML pode tomar decisões em tempo real, como classificar uma ordem de produção como "falhada" quando as remessas especiais são frequentes. Isso permite uma resposta imediata para mitigar problemas e minimizar impactos negativos.

	A	B	C	D	E
1	Dia	Ref	sum.ddp		
2	08/01/2021	239	6		
3	01/04/2021	V839	10		
4	01/04/2021	H839	10		
5	06/04/2021	V839	30		
6	06/04/2021	H839	30		
7	08/04/2021	H839	6		
8	14/04/2021	055AC	5		
9	17/04/2021	055AE	5		
10	23/04/2021	V839	70		

Figura 5 - Exemplo do dado "sum.ddp"

e) **“humanresources”**- Este dado refere-se ao registo diário do número de pessoas presentes diariamente, nas instalações da organização. A inclusão destes dados, do tipo numérico inteiro, que representa a quantidade de pessoas diariamente presentes nas instalações da organização, é de grande relevância na modelação de um modelo de ML. Pode consultar-se na Figura 6.

Este tipo de dados desempenha um papel crucial por diversas razões:

- **Análise da Capacidade de Produção:** O campo "humanresources" fornece informações críticas sobre a mão-de-obra disponível no local. Isso é fundamental para entender a capacidade de produção real da organização em um determinado dia, uma vez que o número de pessoas presente pode afetar a eficiência operacional.

- **Impacto na Produção:** O modelo de ML pode analisar de que forma o número de pessoas presentes nas instalações afeta as quantidades produzidas. Essa relação entre recursos humanos e produção é essencial para identificar oportunidades de otimização e planeamento de produção mais preciso.

- **Gestão de Recursos:** O dado "humanresources" auxilia na gestão eficaz de recursos humanos, pois permite tomar decisões informadas sobre a alocação de pessoal, garantindo que haja pessoal suficiente para atender às necessidades de produção.

- **Adaptação a Mudanças:** O modelo de ML pode ajudar a organização a adaptar-se a mudanças nas circunstâncias, como variações sazonais na disponibilidade de pessoal, permitindo um planeamento de produção mais flexível.

	A	B	C	D
1	Dia	humanresources		
2	04/01/2021	306		
3	05/01/2021	318		
4	06/01/2021	304		
5	07/01/2021	308		
6	08/01/2021	312		
7	09/01/2021	140		
8	10/01/2021	5		
9	11/01/2021	296		
10	12/01/2021	305		

Figura 6 - Exemplo do dado "humanresources"

f) **“sum.hxt”**- O dado “sum.hxt” diz respeito à quantidade de horas extra realizadas ao fim de semana para suprimir as necessidades de cumprir entregas em atraso, por referência de peça, conforme se pode consultar na Figura 7. A inclusão do dado

"sum.hxt," demonstra-se fundamental na modelação deste tipo de modelos de ML. Este tipo de dado é de extrema importância por várias razões:

- **Indicador de Atrasos na Produção:** O campo "sum.hxt" é um indicador claro de que a produção atrasou, uma vez que as horas extras ao fim de semana foram necessárias para cumprir entregas em atraso. Isso é crítico para identificar problemas no processo de produção e no cumprimento de prazos.

- **Análise de Custos:** As horas extras representam custos adicionais para a empresa. O dado "sum.hxt" é essencial para a análise do impacto financeiro dessas horas extras e para a gestão eficaz dos recursos financeiros.

- **Gestão de Pessoal:** A informação de "sum.hxt" é fundamental para a gestão de recursos humanos, uma vez que indica quando e onde o pessoal adicional é necessário. Isto ajuda a empresa a garantir que possui pessoal suficiente para atender às exigências produtivas.

	A	B	C	D	E
1	Dia	Ref	sum.hxt		
2	16/01/2021	042B	18,50		
3	16/01/2021	093C	18,50		
4	16/01/2021	656	18,50		
5	16/01/2021	166H	18,50		
6	16/01/2021	308BM	18,50		
7	16/01/2021	941	18,50		
8	16/01/2021	188	18,50		
9	16/01/2021	107AE	18,50		
10	16/01/2021	262J	18,50		

Figura 7 - Exemplo do dado "sum.hxt"

g) **“available.stock”**- Este dado representa a quantidade de peças em armazém prontas para entrega (*stock*), por referência de peça, ao dia, e demonstra-se de suma importância na modelação de um modelo de ML pelas seguintes razões:

- **Gestão de Inventário:** O campo "available.stock" é fundamental para a gestão eficiente do inventário da organização. Ele fornece informações atualizadas sobre a quantidade de produtos disponíveis para entrega imediata, permitindo um planeamento preciso das operações de produção e logísticas.

- **Atendimento ao Cliente:** Ter acesso a dados precisos de *stock* é essencial para garantir que a organização pode atender às necessidades dos clientes de forma rápida e

eficiente, desde o momento da aceitação de pedidos de entrega. Isso contribui para a satisfação do cliente e a manutenção de relações comerciais sólidas.

- **Previsão de Procura:** O histórico de "available.stock" pode ser utilizado para prever a procura futura de uma determinada referência de peça, identificando tendências sazonais e padrões de consumo, ajudando a que a empresa se prepare adequadamente para picos de procura ou quedas de produção.

- **Redução de Custos:** Ao otimizar o inventário com base em dados de "available.stock," a organização pode minimizar os custos associados ao armazenamento de produtos em excesso e evitar custos de produção adicionais em caso de falta de *stock*.

- **Gestão de Riscos:** A gestão adequada do *stock* reduz o risco de falta de peças (produto acabado) em momentos críticos. Isso é especialmente importante em setores em que a disponibilidade de produtos é vital, como na indústria automóvel.

**h) "intransit.stock"**- Este dado, à semelhança do apresentado anteriormente, trata das quantidades de stock por referência de peça, mas, desta vez, stock em trânsito que poderá, eventualmente, influenciar planos produtivos e aceitação de pedidos de entrega. A inclusão deste dado é de extrema importância na modelação de um modelo de ML porque poderá, também ele, ajudar em:

- **Precisão de Stock:** O campo "intransit.stock" contribui para a precisão das informações de *stock* da organização. Permite que o modelo de ML leve em consideração não apenas o *stock* atual, mas também as peças que estão em trânsito, garantindo que as decisões sejam baseadas em dados completos e atualizados.

- **Planeamento de Produção:** As peças em trânsito têm um impacto direto no planeamento de produção. O modelo de ML pode utilizar os dados de "intransit.stock" para prever a disponibilidade futura de peças e ajustar o planeamento de produção de acordo com essa informação.

- **Aceitação de Pedidos:** O conhecimento da quantidade de peças em trânsito é fundamental para tomar decisões sobre a aceitação de pedidos de entrega. Isso evita a sobrecarga de pedidos quando a capacidade de entrega é limitada, devido às peças em trânsito.

- **Gestão de Prazos:** O dado "intransit.stock" permite a gestão eficaz dos prazos de entrega. O modelo de ML pode calcular prazos de entrega com base nas peças em trânsito, fornecendo aos clientes informações precisas sobre quando podem esperar a entrega.

- **Gestão de Custos:** As peças em trânsito representam custos associados ao transporte e armazenamento temporário. O modelo de ML pode ajudar a otimizar esses custos, garantindo que não haja excesso de peças em trânsito nem falta de *stock*.

Como seria de imaginar, para cada um destes dados recolhidos foram utilizadas técnicas e processos de limpeza e padronização diferentes.

Para os dados do tipo "Data", apenas foi feita uma normalização em todos os *datasets* para que a estrutura dos dados fosse a mesma em todos eles, no entanto, para os dados "sum.hxt", por exemplo, foi considerado que seria apenas relevante considerar como trabalho excecional as horas de trabalho-extra feitas aos sábados e domingos, e por esse mesmo motivo foi utilizado um processo de *data cleansing* para remover os dias úteis do *dataset* "hxt.xlsx".

Foi feita também uma etapa de *data cleansing* em todos os *datasets* .xlsx utilizados, para remover entradas de dados com valores relevantes que estavam omissos, pois, segundo Maria Helena Pestana e João Nuno Gageiro, "as não respostas podem resultar de erros de introdução ou de recolha dos dados, caso em que devem ser eliminadas, ou fazer parte da natureza intrínseca do fenómeno, caso em que devem ser retidas" [48].

As não respostas, atingindo valores iguais ou superiores a 20% dos dados, devem ser analisadas com cuidado, pois se não tiverem um comportamento aleatório poderão pôr em causa o estudo, enviesando-o, bem como os resultados obtidos, o que neste caso não se verificou.

Tendo os dados recolhidos e devidamente tratados, utilizou-se o *Google Colab* para, de uma forma muito simples e recorrendo a *Python*, poder-se compilar os diferentes *datasets*.

Para tal, foi criado um *notebook* no *Google Colab* (denominado "*Dataset\_creator*" e disponibilizado no Anexo A), *notebook* este, que permitiu compilar toda a informação recolhida.

Como é possível constatar na Figura 11, todos os *datasets* carregados para este mesmo *notebook* viram os seus dados “objetivo” e “dia” omissos, removidos. Para além disso, foram removidas também as entradas em que a formatação da coluna “Dia” não se encontrava conforme o padronizado.

```
import pandas as pd
import numpy as np

# Carregar os arquivos do Excel
df1 = pd.read_excel('obj_OP1.xlsx').dropna()
df2 = pd.read_excel('obj_OP2.xlsx').dropna()
df3 = pd.read_excel('obj_OP3.xlsx').dropna()
df4 = pd.read_excel('obj_OP4.xlsx').dropna()
df5 = pd.read_excel('ddp.xlsx').dropna()
df6 = pd.read_excel('hr.xlsx').dropna()

# Remover objetivos omissos
df1 = df1.dropna(subset=['Obj_OP1'])
df2 = df2.dropna(subset=['Obj_OP2'])
df3 = df3.dropna(subset=['Obj_OP3'])
df4 = df4.dropna(subset=['Obj_OP4'])

# Remover dias omissos
df1 = df1.dropna(subset=['Dia'])
df2 = df2.dropna(subset=['Dia'])
df3 = df3.dropna(subset=['Dia'])
df4 = df4.dropna(subset=['Dia'])
df5 = df5.dropna(subset=['Dia'])
df6 = df6.dropna(subset=['Dia'])

# Remover as linhas que contêm a string "(NULL)" da coluna Dia
df1 = df1.loc[~df1['Dia'].astype(str).str.contains('(NULL)')]
df2 = df2.loc[~df2['Dia'].astype(str).str.contains('(NULL)')]
df3 = df3.loc[~df3['Dia'].astype(str).str.contains('(NULL)')]
df4 = df4.loc[~df4['Dia'].astype(str).str.contains('(NULL)')]
df5 = df5.loc[~df5['Dia'].astype(str).str.contains('(NULL)')]
df6 = df6.loc[~df6['Dia'].astype(str).str.contains('(NULL)')]

# Passar a coluna dia de volta a "datetime"
df1['Dia'] = pd.to_datetime(df1['Dia'])
df2['Dia'] = pd.to_datetime(df2['Dia'])
df3['Dia'] = pd.to_datetime(df3['Dia'])
df4['Dia'] = pd.to_datetime(df4['Dia'])
df5['Dia'] = pd.to_datetime(df5['Dia'])
df6['Dia'] = pd.to_datetime(df6['Dia'])
```

Figura 8 - Parte 1 do *notebook* "Dataset\_creator"

Na Figura 12, conforme se pode constatar, foi feito um “merge” para compilar os diferentes ficheiros *Excel*, e exportou-se este ficheiro em *.csv* com o nome “*df.csv*”. Foi verificada a dimensão deste ficheiro *.csv* e, posteriormente, foi removida uma coluna *index* automaticamente gerada e foi adicionada uma “coluna condição”.

Feito isto, o *dataset* compilado foi extraído para ser utilizado no modelo preditivo.

```
# Criar dataset
from functools import reduce
dfs = [df1, df2, df3, df4, df5]
merge_df = reduce(lambda left,right: pd.merge(left, right, how='outer', left_on=['Dia','Ref'], right_on=['Dia','Ref']), dfs)
merge_df = merge_df.drop_duplicates()

merge_df = merge_df.merge(df6, how='outer', left_on=['Dia'], right_on=['Dia'])

# Exportar como csv
merge_df.to_csv('df.csv')
```

```
# Carregar o novo csv com o dataset gerado
dataset = pd.read_csv('df.csv')
```

```
# Estrutura do dataset
row_count = dataset.shape[0] # Conta o número de linhas
col_count = dataset.shape[1] # Conta o número de colunas

print('Número de linhas:', (row_count))
print('Número de colunas:', (col_count))
```

```
# Upload do Dataset "dataset" para tratamento
teste = dataset

teste.head()
```

```
# remover a coluna "Unnamed: 0"
teste = teste.drop(columns=['Unnamed: 0'])

# Converter a coluna "Dia" para o formato de date
teste["Dia"] = pd.to_datetime(teste["Dia"], errors="coerce")

# Remover as linhas com erro na conversão para o formato de data
teste = teste.dropna(subset=["Dia"])

# Adicionar uma nova coluna "Accomp"
teste["Accomp"] = np.where(teste["sum.ddp"] > 0, "0", "1")

teste.to_csv("dataset.csv", index=False)
```

Figura 9- Parte 2 do notebook "Dataset\_creator"

É importante referir de que não foi obtido, de forma nenhuma e para nenhum dos testes, um conjunto de dados que nos permitisse dizer se os objetivos de produção foram satisfeitos ou não, isto porque a organização não fazia qualquer tipo de controlo deste tipo e, por esse mesmo motivo, foi necessário criar uma coluna “condição” que o determinasse, mediante as condições dos diferentes cenários.

Segundo Jean-Paul Tremblay e Richard B. Bunt [44], este tipo de colunas condição são caracterizadas por poderem ser verdadeiras ou falsas. Na notação algorítmica, chamamos a esta condição uma expressão lógica ou booleana. Este tipo de expressão é calculado de forma análoga a uma expressão numérica, mas o seu valor é “verdadeiro” ou “falso” em vez de um número. Estes autores apresentam o seguinte exemplo: “Suponha que temos duas variáveis numéricas, chamadas A e B, que são garantidas para ter valores desiguais, e suponha que queremos imprimir o valor da maior. Se A for maior

do que B, queremos imprimir o valor de A. Por outro lado, se B for maior do que A, queremos imprimir o valor de B” [49].

### 3.2.2. 1º Teste com um modelo de regressão logística (“dataset”)

Na fase inicial de compilação, optou-se por tratar apenas seis conjuntos de dados ("obj\_OP1", "obj\_OP2", "obj\_OP3", "obj\_OP4", "ddp" e "hr"), uma vez que se considerava que seria suficiente para determinar a conclusão ou não dos planeamentos, ao lidar com os objetivos de produção e o número de pessoas a trabalhar.

No ambiente *Colab*, cada um dos seis ficheiros foi carregado e, seguindo uma abordagem semelhante ao que é observado na Figura 8 e Figura 9, procedeu-se à eliminação de entradas com dados relevantes em falta (como valores objetivo e dias), bem como à remoção de entradas com erros no campo Data. Deste modo, foi obtido o primeiro conjunto de dados compilado, denominado “dataset.csv”, o qual continha 46719 linhas e 9 colunas (Figura 10). Durante o processo de compilação dos dados, foi criada uma coluna denominada “unnamed”, que funcionava como coluna índice e que posteriormente foi removida.

	A	B	C	D	E	F	G	H	I	J
1	Dia	Ref	Obj_OP1	Obj_OP2	Obj_OP3	Obj_OP4	sum.ddp	humanresources	Accomp	
2	04/01/2021	042B	975.0	2000.0		1900.0		306.0		1
3	04/01/2021	042AC	1020.0	900.0		0.0		306.0		1
4	04/01/2021	M262C	412.5	0.0		0.0		306.0		1
5	04/01/2021	239	900.0	600.0		352.0		306.0		1

Figura 10 – Recorte do “dataset.csv”

Em seguida, procedeu-se à criação de uma nova coluna, com uma condição específica, destinada a obter a informação relativa ao cumprimento ou não dos objetivos de produção (coluna “Accomp” possível de consultar na Figura 10). Nesta abordagem, se o valor de “sum.ddp” fosse maior que zero, o objetivo era considerado como não atingido, e o novo conjunto de dados foi então exportado em formato *CSV* com o nome de “dataset”.

Posteriormente, importou-se o “dataset.csv” para outro *notebook* no *Google Colab*, designado por “ML\_IT\_testes”, e deu-se início à preparação do ambiente de treino e teste para a realização das etapas seguintes. Importaram-se as bibliotecas necessárias e começou-se a retificar o *dataset* para o modelo em questão.

Inicialmente, e como é possível verificar na Figura 11, começou-se por substituir todos os valores *NaN* por 0. Em seguida, selecionaram-se apenas as colunas numéricas, para não mexer com a coluna “Ref” (por exemplo), e foram separados os dados em “x” e “y”. Utilizou-se o *StandardScaler* para normalizar os dados e logo de seguida estes foram divididos em dois conjuntos: de treino e de teste. Esta divisão foi feita em 70/30, respetivamente.

```
[ ] # Ler o dataset para uma variável "df"
dataset = pd.read_csv("dataset.csv")

dataset[['Obj_OP1', 'Obj_OP2', 'Obj_OP3', 'Obj_OP4', 'sum.ddp', 'humanresources']] =
dataset[['Obj_OP1', 'Obj_OP2', 'Obj_OP3', 'Obj_OP4', 'sum.ddp', 'humanresources']].fillna(0)

dataset.to_csv('df.csv', index=False)

df = pd.read_csv("df.csv")

# Estrutura do dataset
row_count = df.shape[0] # Conta o número de linhas
col_count = df.shape[1] # Conta o número de colunas
print('Número de linhas:', (row_count))
print('Número de colunas:', (col_count))

[ ] # Selecionar apenas as colunas numéricas
df_num = df.select_dtypes(include=[np.number])

# Separar dados em "x" e "y"
x = df_num.drop('Accomp', axis=1)
y = df_num['Accomp']

# Normalizar os dados à mesma escala
sc = StandardScaler()
x = sc.fit_transform(x)

# Dividir os dados em conjuntos de treino e teste
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=1) # 70% training and 30% test

# Inserir os dados no modelo
lm = LogisticRegression()
lm.fit(x_train, y_train)

with open('modelo1.pkl', 'wb') as file:
    pickle.dump(lm, file)

# Fazer previsões no conjunto de teste
y_pred = lm.predict(x_test)
```

Figura 11 - Parte 1 do *Notebook* de preparação do modelo e previsão

Estes dados foram, assim, inseridos no modelo “*lm*” (*LogisticRegression*) pertencente ao *scikit learn*<sup>2</sup> e foram feitas as devidas previsões no conjunto de teste.

Por fim, e para avaliar o desempenho do modelo, foi extraída a quantidade de previsões 0 e 1 (sendo “0 = não cumpriu” e “1 = cumpriu”) com o código visível na Figura 12, e também foi calculada a *accuracy* de cada classe de previsão e a *accuracy* total. Os resultados obtidos podem ser observados na Tabela 3.

<sup>2</sup> Consultável em [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)

```
[ ] from sklearn.metrics import accuracy_score

# Calcular a quantidade de previsões de cada classe
quantidade_0 = sum(y_pred == 0)
quantidade_1 = sum(y_pred == 1)

# Calcular a accuracy de cada classe
accuracy_0 = accuracy_score(y_test[y_test == 0], y_pred[y_test == 0])
accuracy_1 = accuracy_score(y_test[y_test == 1], y_pred[y_test == 1])

# Calcular a accuracy total
accuracy_total = accuracy_score(y_test, y_pred)

# Imprimir os resultados
print("Quantidade de previsões 0: ", quantidade_0)
print("Quantidade de previsões 1: ", quantidade_1)
print("Accuracy de previsões 0: ", accuracy_0)
print("Accuracy de previsões 1: ", accuracy_1)
print("Accuracy total: ", accuracy_total)

Quantidade de previsões 0: 28
Quantidade de previsões 1: 13988
Accuracy de previsões 0: 0.8235294117647058
Accuracy de previsões 1: 1.0
Accuracy total: 0.9995719178082192
```

Figura 12 - Parte 2 do *Notebook* de preparação do modelo e previsão

Tabela 3 - Resultados obtidos no 1º teste

Nº Previsões 0	Nº Previsões 1	Accuracy de 0	Accuracy de 1	Accuracy total
28	13988	82%	100%	99,96%

### 3.2.3. 2º Teste com um modelo de regressão logística (“*dataset.w\_hxt*”)

No âmbito desta segunda etapa de testes, foram tratados exclusivamente os mesmos 6 conjuntos de dados ("obj\_OP1", "obj\_OP2", "obj\_OP3", "obj\_OP4", "ddp" e "hr") que haviam sido processados anteriormente, acrescentando-se ainda o conjunto de dados "hxt", no qual se encontrava compilada toda a informação relativa às horas extras efetuadas ao fim de semana. Esta modificação foi realizada com o objetivo de analisar as oscilações que os resultados apresentariam ao adicionar mais um critério ao modelo.

Novamente, no ambiente *Colab*, foi feito o carregamento de cada um dos 7 ficheiros *.xlsx* e foram executadas as etapas de tratamento previamente realizadas, tais como a exclusão de entradas com dados relevantes em falta (como valores objetivo e dias), bem como a remoção de entradas com erros no campo Data. Deste modo, obteve-se um novo *dataset* compilado, designado por "*dataset.w\_hxt*", o qual era constituído por 46719 linhas, tal como o primeiro conjunto de dados, porém com 10 colunas.

Para este segundo conjunto de dados, e em consonância com as razões expostas anteriormente, foi criada uma coluna com uma condição específica para a obtenção da informação relativa ao cumprimento ou não dos objetivos de produção. Nesta abordagem, quando o valor de "sum.hxt" fosse maior que zero e/ou o valor de "sum.ddp" fosse maior que zero, considerava-se o objetivo como não atingido. A linha de código utilizada para criar esta coluna condição pode ser consultada abaixo, na Figura 13.

Esta estratégia permitiu identificar de forma adequada e precisa a consecução dos objetivos de produção com base nas informações disponíveis.

```
# Adicionar uma nova coluna "Accomp"
teste["Accomp"] = np.where((teste["sum.ddp"] > 0) & (teste["sum.hxt"] > 0), 0,
np.where((teste["sum.ddp"] > 0) | (teste["sum.hxt"] > 0), 0, 1))
```

Figura 13 - Alterações à coluna "Accomp" no 2º teste

Após a extração do segundo conjunto de dados em formato *CSV*, denominado "dataset.w\_hxt", procedeu-se à sua importação para o *notebook* de teste. Em seguida, o ambiente de treino e teste foi novamente configurado, realizando alguns ajustes similares aos efetuados no primeiro teste. Os dados foram segregados em conjuntos "x" e "y". O processo de normalização dos dados foi realizado utilizando o *StandardScaler*. Posteriormente, os dados foram divididos em dois conjuntos, designadamente o conjunto de treino e o conjunto de teste, seguindo a mesma proporção utilizada no primeiro teste, ou seja, numa divisão de 70/30, respetivamente.

Os dados foram então inseridos no modelo e as previsões foram feitas de acordo com as configurações adequadas. Por fim, com o intuito de avaliar o desempenho do modelo, foram novamente extraídas as quantidades de previsões classificadas como "0" (não cumpriu) e "1" (cumpriu). Calculou-se a precisão para cada classe de previsão, assim como a precisão total. Os resultados obtidos podem ser visualizados na Tabela 4.

Tabela 4 - Resultados obtidos no 2º teste

Nº Previsões 0	Nº Previsões 1	Accuracy de 0	Accuracy de 1	Accuracy total
52	13964	85%	100%	99,93%

### 3.2.4. 3º Teste com um modelo de regressão logística (“*dataset.w\_orders*”)

No âmbito deste terceiro teste, foram tratados todos os oito conjuntos de dados disponíveis até então ("obj\_OP1", "obj\_OP2", "obj\_OP3", "obj\_OP4", "ddp", "hr", "hxt" e "orders"). É relevante mencionar que este último conjunto contém informações sobre os objetivos de produto acabado destinados a satisfazer os pedidos de encomendas por referência de peça. Esta alteração foi realizada com o intuito de examinar, mais uma vez, as alterações de comportamento do modelo ao receber mais uma fonte de "informação Objetivo" e verificar quais oscilações nos resultados ocorreriam ao incrementar mais um critério.

De forma análoga aos testes anteriores, no ambiente *Colab*, foram carregados os oito ficheiros *.xlsx* e procedeu-se às etapas de tratamento previamente realizadas, incluindo a remoção de entradas com dados relevantes em falta (como valores objetivo e dias) e a eliminação de entradas com erros no campo Data. Desta forma, obteve-se mais um conjunto de dados compilado, designado por "*dataset.w\_orders*". Este terceiro conjunto de dados era composto por um maior número de linhas (aumentando de 46719 linhas nos primeiros dois *datasets* para 50581) e continha 12 colunas.

Também neste caso, devido às razões apresentadas anteriormente, foi criada uma coluna com uma condição específica (parâmetros apresentados na Figura 14) para obter a informação relativa ao cumprimento ou não dos objetivos de produção. Nesta abordagem, se o valor de "sum.hxt" fosse maior que zero e/ou o valor de "sum.ddp" fosse maior que zero, o objetivo era considerado como não atingido. Esta estratégia foi adotada a fim de avaliar devidamente a consecução dos objetivos de produção com base nos dados adicionais disponibilizados.

```
# Adicionar uma nova coluna "Accomp"  
teste["Accomp"] = np.where((teste["sum.ddp"] > 0) & (teste["sum.hxt"] > 0), 0,  
np.where((teste["sum.ddp"] > 0) | (teste["sum.hxt"] > 0), 0, 1))
```

Figura 14 - Alterações à coluna "Accomp" no 3º teste

Mais uma vez, procedeu-se à extração do conjunto de dados em formato *CSV*, designado por "*dataset.w\_orders*". Foi importado para o *notebook* de teste e preparado novamente o ambiente de treino e teste. Foram realizados alguns ajustes ligeiros, seguindo uma abordagem semelhante à adotada no primeiro teste, e divididos os dados em conjuntos "x" e "y". Em seguida, foi efetuada a normalização dos dados utilizando o

*StandardScaler* e, logo após, foi realizada a divisão dos dados em dois conjuntos, nomeadamente o conjunto de treino e o conjunto de teste, utilizando a mesma proporção utilizada no primeiro teste, ou seja, numa divisão de 70/30, respetivamente.

Os dados foram inseridos no modelo apropriado, permitindo, assim, efetuar as devidas previsões.

Por fim, com o objetivo de avaliar o desempenho do modelo, procedeu-se, mais uma vez, à extração das quantidades de previsões classificadas como "0" (não cumprido) e "1" (cumprido). Adicionalmente, foi calculada a precisão para cada classe de previsão, bem como a precisão total. Os resultados obtidos foram registados na Tabela 5.

Tabela 5 - Resultados obtidos no 3º teste

Nº Previsões 0	Nº Previsões 1	Accuracy de 0	Accuracy de 1	Accuracy total
65	15110	88%	100%	99,94%

### 3.2.5. 4º Teste com um modelo de regressão logística (“*dataset.w\_stocks*”)

No âmbito deste último teste, após encontrar um novo conjunto de dados disponível, o "*stocks*", procedeu-se à sua incorporação, juntamente com os oito conjuntos de dados anteriores. Neste último conjunto de dados, encontrava-se a informação relativa à quantidade de peças em *stock*, tanto em armazém como em trânsito, por referência de peça. Esta alteração foi realizada pois considerou-se que essa informação seria relevante, uma vez que, no problema, lidamos com encomendas e ordens de produção que são afetadas pelo princípio FIFO (*first in, first out*) na hora da expedição.

Tal como nos testes anteriores, novamente no ambiente *Colab*, foi feito o carregamento de cada um dos nove ficheiros *.xlsx* e foram realizadas as etapas de tratamento de dados previamente estabelecidas. Dessa forma, foi obtido mais um conjunto de dados compilado, denominado "*dataset.w\_stocks*". Este último conjunto de dados era constituído por um maior número de linhas, naturalmente, passando de 50581 entradas no teste anterior para 52266, e também registou um aumento de duas colunas, passando assim a conter 14 colunas.

Também neste caso, devido às razões apresentadas anteriormente, foi criada uma coluna com uma condição específica para obter a informação relativa ao cumprimento ou

não dos objetivos de produção (parâmetros da coluna condição presentes na Figura 15). Nesta abordagem, se o valor de "sum.hxt" fosse maior que zero e/ou o valor de "sum.ddp" fosse maior que zero, o objetivo era considerado como não atingido. Essa estratégia foi adotada a fim de avaliar devidamente a consecução dos objetivos de produção levando em conta os dados adicionais sobre o stock disponível.

```
# Adicionar uma nova coluna "Accomp"
teste["Accomp"] = np.where((teste["sum.ddp"] > 0) & (teste["sum.hxt"] > 0), 0,
np.where((teste["sum.ddp"] > 0) | (teste["sum.hxt"] > 0), 0, 1))
```

Figura 15 - Alterações à coluna "Accomp" no 4º teste

Após o tratamento dos dados, procedeu-se à extração do conjunto de dados em formato *CSV*, intitulado "*dataset.w\_stocks*". Foi importado este conjunto para o *notebook* de teste e novamente executaram-se os mesmos passos preparatórios para o treino e teste do modelo, seguindo uma abordagem semelhante à utilizada nos testes e treinos anteriores. Foi realizado um ajuste ligeiro, utilizando o *StandardScaler* para normalizar os dados, e em seguida, os dados foram divididos em dois conjuntos, tal como, o que foi feito no primeiro teste. Essa divisão foi realizada de forma a utilizar 70% do conjunto de dados para o treino e 30% para o teste.

Os dados preparados foram inseridos no modelo e as devidas previsões foram realizadas.

Finalmente, para avaliar o desempenho do modelo, foi novamente extraída a quantidade de previsões classificadas como "0" (não cumpriu) e "1" (cumpriu). Além disso, calculou-se a precisão para cada classe de previsão, bem como a precisão total. Os resultados obtidos são apresentados na Tabela 6.

Tabela 6 - Resultados obtidos no 4º teste

Nº Previsões 0	Nº Previsões 1	Accuracy de 0	Accuracy de 1	Accuracy total
58	15622	84%	100%	99,93%

### 3.2.6. Escolha do modelo a utilizar

Uma vez realizados todos os testes e avaliações de resultados, foi compilada a informação numa tabela resumo (Tabela 7) onde é possível consultar os desempenhos de cada um dos conjuntos.

Tabela 7 - Comparação dos testes realizados

Tabela de comparação - Métricas de Avaliação						
	Nr. Previsões 0	Nr. Previsões 1	Previsões Totais	Accuracy dos 0	Accuracy dos 1	Accuracy Total
1° Teste (Obj, ddp e hr) 46719 entradas 9 colunas	28	13988	14016	82,35%	100%	99,96%
2° Teste (1° teste + HXT) 46719 entradas 10 colunas	52	13964	14016	85,25%	100%	99,94%
3° Teste (2° teste + orders) 50581 entradas 11 colunas	65	15110	15175	87,84%	100%	99,94%
4° Teste (3° teste + stocks) 52266 entradas 14 colunas	58	15622	15680	84,06%	100%	99,93%

Num primeiro teste, com menos dados, foi apresentada uma *accuracy* total mais alta, e por isso foi considerado pelos envolvidos que era o 3º conjunto de dados (segundo maior no que diz respeito ao número de entradas) que aparentava entregar melhores garantias de uso. Este foi o que apresentou um maior número de previsões de não cumprimento (previsões 0), foi também o que apresentou uma maior *accuracy* também para estes zeros e, tendo em conta a diferença de dimensão deste para o primeiro modelo, ter uma *accuracy* total inferior em apenas 0,02% mostrou ser um resultado bastante satisfatório.

Desta forma, e uma vez decidido que era este o modelo sobre o qual a empresa iria trabalhar no futuro, este foi extraído em ficheiro *.h5* para que pudesse ser utilizado, posteriormente, ao criar uma API (*Application Programming Interface*). Após a escolha do modelo que a empresa adotaria para as operações futuras, fez-se uma análise aos resultados obtidos e ao comportamento do modelo. Nessa análise, tornou-se evidente que os dados não estavam equilibrados, o que explica a presença de resultados enviesados. Neste contexto, torna-se perceptível a importância de considerar um processo futuro de balanceamento de dados. Em particular, destacamos a necessidade de equilibrar os casos de 'não cumprimento (ou 0)' em relação aos planos de produção. Sugere-se abordar esta questão utilizando dados históricos relacionados ao cumprimento ou não dos planos de produção, a fim de melhorar a qualidade e a fiabilidade das análises e previsões futuras.

### **3.3. Dashboards em Power BI**

No setor industrial, a obtenção e análise de dados precisos e relevantes são essenciais para garantir a eficiência operacional e a tomada de decisões, como já vimos anteriormente.

Neste contexto, os *dashboards* do *Power BI* têm-se destacado como uma ferramenta poderosa para tratar, visualizar e analisar dados em tempo real. Estes *dashboards* oferecem uma interface visual intuitiva, permitindo que os profissionais tenham uma visão clara e organizada do estado atual dos processos produtivos. Com esta visualização, é possível identificar padrões, tendências e anomalias, facilitando a tomada de decisões rápidas e precisas [41].

Além disso, os *dashboards* do *Power BI* permitem uma monitorização em tempo real das operações, fornecendo uma visão atualizada e precisa do desempenho das máquinas, dos consumos de energia, do estado da produção, entre outros indicadores relevantes [41].

Neste caso em específico, os *dashboards* trabalhados destinaram-se apenas e só ao departamento de “Pessoas & Cultura”. Como sabemos, especialmente numa empresa que conta com mais de 400 colaboradores, os departamentos de recursos humanos desempenham um papel fundamental na gestão do capital humano.

Para obter *insights* relevantes sobre a composição demográfica dos colaboradores e tomar decisões estratégicas relacionadas a questões demográficas e culturais, a utilização de *dashboards Power BI* é de extrema utilidade. Estes permitiram ao departamento visualizar e analisar dados especificamente demográficos de forma clara e organizada. Através de gráficos, tabelas e indicadores interativos, é possível ter uma visão abrangente da estrutura etária, género, formação académica, experiência profissional, entre outros dados relevantes dos colaboradores.

Assim, são apresentados agora 2 *dashboards* elaborados no âmbito do estágio e que tinham como objetivo facilitar a consulta de dados relevantes e as consequentes tomadas de decisão. Na Secção 3.3.1 é apresentado o *dashboard* “Fotografia da população 2023” e na Secção 3.3.2 o *dashboard* “ARTE”.

### 3.3.1. Fotografia da população 2023

Nesta secção são apresentadas as 3 páginas do *Power BI* desenvolvidas no âmbito deste estágio e do *dashboard* “Fotografia da população 2023”.

Este *dashboard* foi elaborado porque, conforme referido acima, havia uma necessidade de conhecer melhor a população da organização e, para obter *insights* relevantes sobre a composição demográfica dos colaboradores e tomar decisões estratégicas relacionadas a questões demográficas e culturais, era necessário obter este tipo de representação.

Na primeira página (Figura 16), recorrendo apenas a 3 gráficos circulares e um cartão com um valor médio, foi tratada a distribuição de colaboradores pelas 4 empresas sediadas em Portugal, do Grupo, totalizando as 470 pessoas. Estas pessoas dividem-se, por género, em cerca de 18% do género feminino e 82% do género masculino.

A média de idade dos colaboradores do grupo fixa-se nos 42 anos, sendo que a empresa com a média de idade maior é a *Odibil* (ODB) e esta apresenta uma idade média de 54 anos. Por outro lado, a empresa com a média de idade menor é a *Autoconceptus* (ATC), apresentando um valor médio de 39 anos de idade.

Para além deste tipo de informação também é possível, nesta página, consultar a divisão das pessoas por Estado Civil. Este tipo de informação é relevante para auxiliar na decisão de campanhas de retenção, na escolha de benefícios e para pensar atividades de grupo a desenvolver.

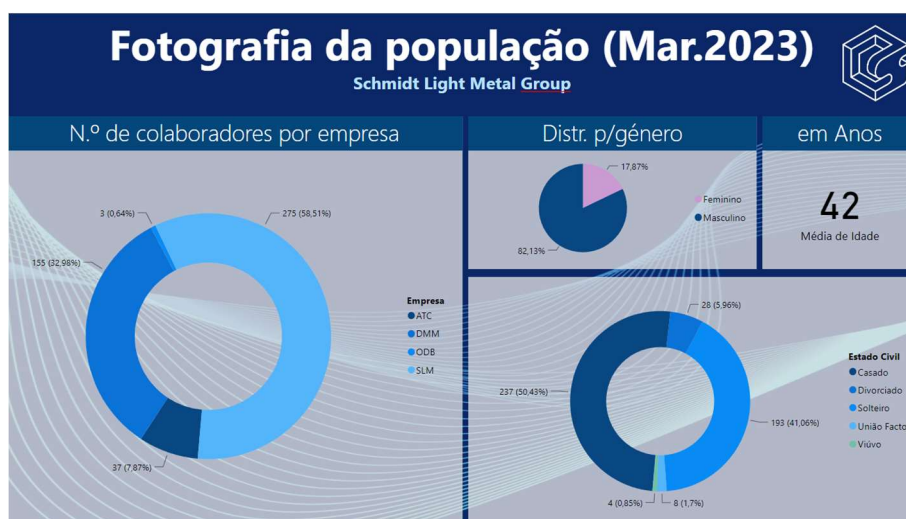


Figura 16 - Página 1 do *Power BI* "Fotografia da população 2023"

Na segunda página (ver Figura 17), utilizando dois gráficos de barras diferentes (o primeiro horizontal e o segundo vertical, da esquerda para a direita) foi possível representar a distribuição da população por código de habilitações ou nível de ensino e a distribuição de pessoas por departamento.

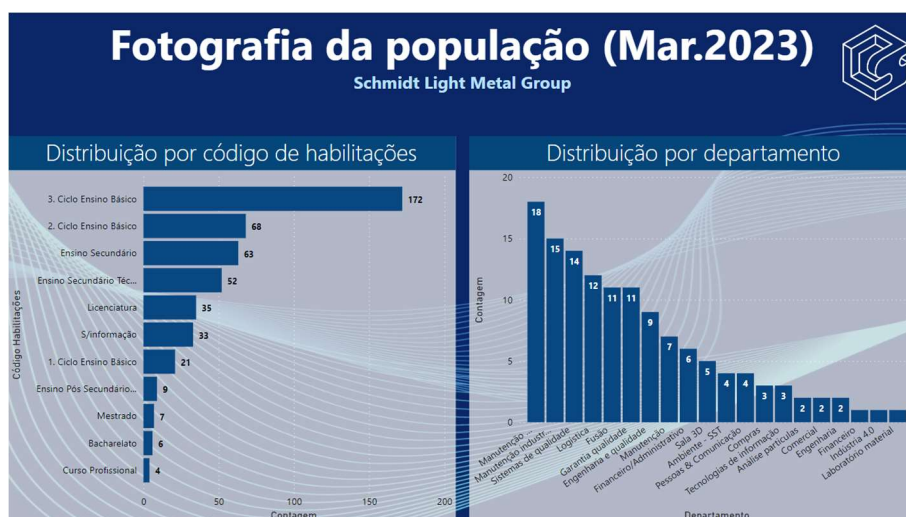


Figura 17 - Página 2 do Power BI “Fotografia da população 2023”

No primeiro destes, é perfeitamente perceptível que mais de metade da população se encontram com o 2º (segundo) ou 3º (terceiro) ciclo do Ensino Básico completo. Fazer este tipo de levantamentos e análises e compreender a distribuição por níveis de ensino da população de uma empresa é útil para o planeamento de recursos humanos, na identificação de lacunas de competências, no desenvolvimento de programas de formação, na análise de desempenho, e também na promoção da diversidade e inclusão. Essa compreensão permite que as empresas maximizem o potencial dos seus colaboradores, garantindo assim que se tomem medidas adequadas para o crescimento e sucesso da organização.

No segundo gráfico, foi feita apenas uma distribuição das pessoas não afetas ao departamento “Produção”, por departamentos, com um objetivo muito simples e concreto: utilizar a integração e o cruzamento dos dados do primeiro para o segundo gráfico e assim perceber qual o grau de ensino, por departamento. Conhecer a distribuição dos graus de ensino, por departamento, traz imensas vantagens, como por exemplo, a identificação de necessidades de competências específicas num determinado departamento, a otimização de contratações para fazer face às necessidades, a alocação eficiente de recursos humanos e possivelmente ajudará a trabalhar mobilidade interna, a

melhoria também da colaboração interna e, com tudo isto, ter um planeamento estratégico mais direcionado.

Nesta terceira e última página do *Power BI* “Fotografia da população 2023” (Figura 18) procurou-se fazer um gráfico de bolhas georreferenciado, onde as bolhas representam a quantidade de pessoas (da população da organização) em diferentes localidades geográficas, enquanto o mapa fornece o contexto espacial para esta distribuição. Esse tipo de visualização acaba por ser muito útil e prática para identificar padrões na distribuição geográfica das pessoas, mas também para identificar casos discrepantes na distribuição na população, como é o caso da pessoa que, trabalhando em Oliveira de Azeméis tem residência em Cantanhede (cerca de 72km de distância).

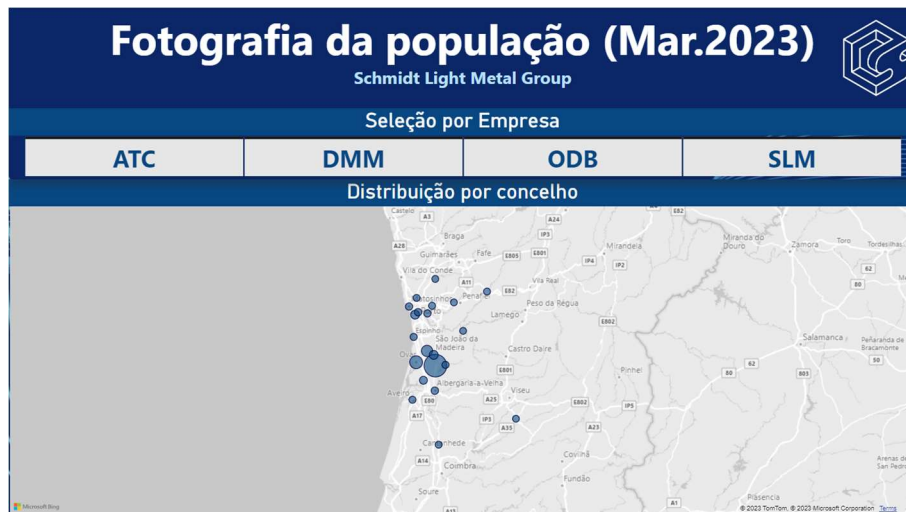


Figura 18 - Página 3 do *Power BI* "Fotografia da população 2023"

Perceber quais as principais zonas de residência da população de uma empresa, é útil não só para estratégias de recrutamento direcionadas, mas também para planear e adotar medidas de incentivo à deslocação e ajudas de custo, ou até para gestão de níveis de produtividade. Com este tipo de informação, a empresa pode ser mais assertiva na hora de oferecer benefícios, pois identifica, com maior precisão, as necessidades específicas de cada grupo de indivíduos.

### 3.3.2. ARTE

A Figura 19 apresenta uma página de controlo e análise da participação das várias chefias no modelo de competências interno. Para melhor contextualizar, de seguida é explicado o modelo.

No *Schmidt Light Metal Group* as pessoas não são avaliadas de uma forma quantitativa (recorrendo a escalas como as típicas de 1 a 5), mas sim de uma forma qualitativa. Este modelo de competências tem o nome ARTE e este nome é uma sigla que surge da junção das palavras “Agilidade”, “Rigor”, “Tenacidade” e “Espírito de Equipa”. Este modelo de competências geralmente tem a duração de um ano inteiro (de março a fevereiro) e divide-se em diferentes etapas.

Primeiramente, via *Forms*, cada pessoa do grupo responde ao seu próprio “balanço”, isto é, para cada uma das 4 competências acima mencionadas, um colaborador escolhe aqueles aspetos ou comportamentos em que mais se revê, destacando assim, de uma forma simples, os seus pontos mais positivos. Neste “balanço” existe também uma pergunta (por norma a última) onde pedimos que este mesmo colaborador identifique um aspeto em que gostaria de melhorar ou se desenvolver.

Posteriormente, as chefias vão fazer uma “avaliação” aos seus supervisionados identificando, de forma idêntica e para cada competência, os seus pontos fortes.

Feita esta “avaliação” das chefias ao grupo geral, são confrontadas as respostas de uns e de outros e os resultados são entregues a estes superiores hierárquicos.

De seguida, é expectável que estas mesmas chefias agendem conversas com os seus supervisionados, conversas essas designadas “Conversas com Sentido” e aí são discutidas (de uma forma saudável) as escolhas de uns e outros, é visto o aspeto que o colaborador menciona como “a melhorar” e é feito um Plano de Desenvolvimento de 1 ano onde a chefia se compromete a ajudar o colaborador a desenvolver, ou melhorar esta sua lacuna identificada.

Para fechar o ciclo, a 2 ou 3 meses do final, cada chefia volta a reunir com as suas pessoas a fim de averiguar qual a completude e satisfação destas mesmas em relação ao plano de desenvolvimento efetuado.

A página de *Power BI* abaixo (Figura 19) nada mais faz senão apresentar, ao departamento de Pessoas & Cultura (que é o que gere este processo de avaliação de desempenho) o nível de completude destes primeiros balanços. Nele, e como é possível ver no primeiro caso, do Sr. Manuel António Valente, é perceptível quantas pessoas este Diretor tem para avaliar (4), quantas avaliações já fez (4), qual a percentagem de completude até ao momento (100%), quantos balanços da sua equipa já tem feitos e

disponíveis para consulta (4), e qual a adesão total ao processo (100%). Ao olhar aos valores, neste caso, pode-se concluir que este Diretor e as suas pessoas estão prontas para passar à fase das Conversas com Sentido, uma vez que já todos cumpriram com a sua parte.



Figura 19 - Análise de um excelente cumprimento do processo de avaliação de desempenho anual (ARTE)

Na Figura 20 é ilustrado que a equipa do Sr. Paulo Silva está consideravelmente longe de poder avançar para a fase das Conversas, uma vez que 40% da sua equipa ainda não cumpriu com a sua função e não fez o seu próprio “balanço”



Figura 20 - Análise de um bom cumprimento do processo de avaliação de desempenho anual (ARTE)

. Constata-se ainda na Figura 20 que o Chefe já avaliou 19 das 20 pessoas ao seu encargo, e que desta forma já tem 95% das avaliações feitas. Neste tipo de casos, e olhando a este painel, o departamento de Pessoas & Cultura, muito provavelmente, sugeriria que o Chefe fizesse algum *pressing* junto da sua equipa, para que estes

cumprissem com a sua parte e para que se continuasse a dar seguimento ao processo. Com 12 dos 20 elementos que já fizeram o seu balanço, o Sr. Paulo Silva pode iniciar estas Conversas e conseqüentemente desenvolver os seus Planos de Desenvolvimento.

Escusado será dizer que também estes Chefes de Equipa, Sub-Diretores e Diretores tem o seu Plano de Desenvolvimento, ao se sentarem a conversar com os seus superiores hierárquicos, ou seja, com aqueles a quem reportam.

Este tipo de modelo e processo permite também, como seria fácil de prever, que o departamento de Pessoas & Cultura desenvolva e planeie ações de formação com base nestes Planos de Desenvolvimento.

Imaginemos, por exemplo, que um determinado número significativo de pessoas identifica como lacuna a sua falta de destreza (ou mesmo a falta de noções básicas) em *Microsoft Excel*. Faz-se um levantamento do número de pedidos deste tipo de formação e, normalmente, existem dois cenários possíveis: existe um número de pessoas suficiente para se formar uma turma internamente e contrata-se um formador para ensinar *Excel*, simultaneamente, a um grupo de pessoas do mesmo nível; existe um número muito reduzido (1 ou 2 pessoas) com uma necessidade de formação muito específica e para esses é paga uma formação fora de portas, por forma a permitir que também esses se desenvolvam.

Naturalmente, este tipo de lacunas e aspetos a desenvolver não se limita a ferramentas e *softwares* e pode passar por metodologias de trabalho (por exemplo, *Agile/ Scrum*), *coaching*, *mentoring*, entre outros. Interessa também salientar que este tipo de planos de formação/desenvolvimento deverá estar alinhado com aqueles que são os interesses da organização e com o enquadramento destas pessoas nesta mesma organização.

### **3.4. Processos ETL e relatórios**

Durante o estágio também foram realizados inquéritos para cumprir com as obrigações legais da organização, mas também para recolher *feedback* dos colaboradores da organização. Para realizar este trabalho foi utilizado, maioritariamente, o *Microsoft Forms*, que é talvez a ferramenta online mais conhecida para criar e distribuir questionários, mas para além desta, também foi utilizado o *Mentimeter*, uma ferramenta interativa de questionários e apresentação de dados em tempo real. Esta última foi sempre

mais utilizada nos nossos *Inquéritos com Sentido* (descritos mais à frente neste documento), mas também em ações de formação, onde no final era apresentado um teste aos formandos, ou até mesmo um inquérito de avaliação à ação de formação onde se procura medir a satisfação das expectativas dos formandos, a qualidade dos conteúdos apresentados e o desempenho do formador.

Explicar-se-á, agora, de que forma foi executado este trabalho, desde a criação do inquérito até à apresentação dos resultados finais, começando pelos tão famosos *Inquéritos com Sentido*.

O primeiro *slide* apresentado abaixo (Figura 21) remete, conforme foi referido anteriormente, para um dos vários *Inquéritos com Sentido*. Estes inquéritos, elaborados e partilhados via *Mentimeter* (ou Menti), tinham um propósito bem definido: aumentar o *engagement* junto da população da organização, perguntando-lhes algo que para eles teria um possível valor acrescentando, apresentando os resultados de uma forma relativamente fácil de interpretar, e procurando também que a implementação fosse muito simples e breve, o que permitiria, assim, tomar ações num curto espaço de tempo, fazendo com que os demais colaboradores percebessem que, efetivamente, as suas opiniões eram tidas em conta e que às suas opiniões e ideias seria dada a devida importância.



Figura 21 - "Inquérito com Sentido"

Esta medida foi tomada porque a organização se deparou com taxas de adesão a inquéritos internos muito baixas, fazendo mesmo com que inquéritos de carácter obrigatório não atingissem os 65% mínimos de respostas, o que fazia com que estes resultados não fossem considerados indexantes [50].

Neste primeiro caso em concreto, foi perguntado aos colaboradores do grupo se “gostariam de ver anunciadas as entradas de novos colaboradores para o Grupo”. O inquérito ficou “aberto” (a aceitar respostas) durante 5 dias úteis e ao 6º dia útil foi partilhado, com toda a população do grupo, os resultados obtidos num *slide* com representação gráfica simples e perceptível por todos (Figura 22).

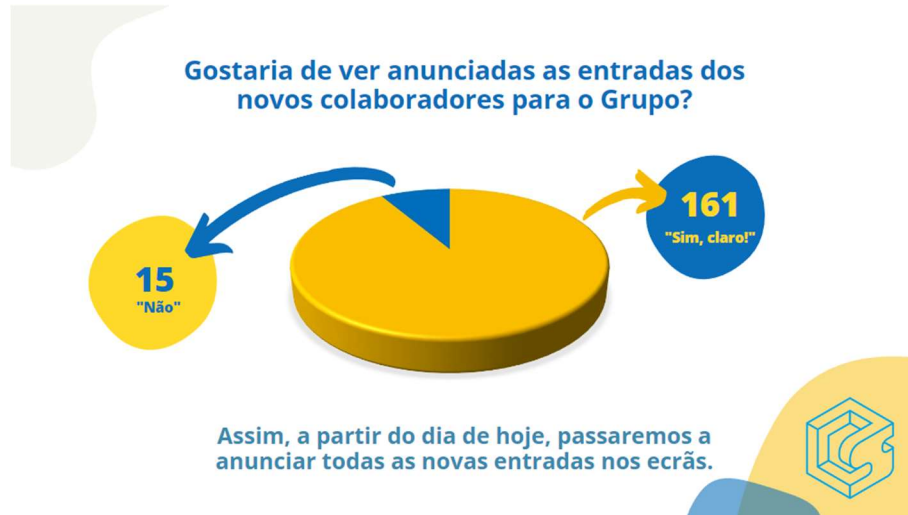


Figura 22 - Apresentação de resultados obtidos do exemplo 1

Uma vez tomada a decisão pela maioria, seria implementada a ação a que a empresa se propusera e assim, ao 7º dia útil começou-se a partilhar as novas admissões em *slides*, como o que está apresentado na Figura 23.



Figura 23 - Slide Modelo de apresentação de novas admissões

No caso de *Inquéritos com Sentido*, cuja implementação das novas ações que a organização se propunha realizar pudesse ser mais demorada, no momento da apresentação dos resultados obtidos, era deixada, sempre, a nota da data ou altura do ano a partir da qual essa mesma medida entraria em vigor. Segue o exemplo.

Num dos *Inquéritos com Sentido* realizados, a organização quis saber se as pessoas gostariam de receber, no seu posto de trabalho, e uma vez por semana, uma peça de fruta gratuita. Assim, foi lançado um inquérito com o *slide* apresentado abaixo (Figura 24):



Figura 24 - Exemplo de Inquérito com uma implementação da medida mais demorada

À semelhança de todos os outros inquéritos deste tipo, este foi deixado a aceitar respostas durante 5 dias úteis e no 6º dia útil foi partilhado, com todos os colaboradores, os resultados obtidos num *slide* à semelhança dos anteriores também, conforme se pode ver na Figura 25.



Figura 25 - Apresentação de resultados obtidos do exemplo 1

Conforme se constata, a esmagadora maioria decidiu que se avançasse com esta implementação, e como tal, tendo em conta a gestão a que obriga para que se entregue fruta todas as semanas a cerca de 400 pessoas, foi anunciado, neste mesmo *slide*, que a partir do mês seguinte ao do lançamento deste, se iria passar a entregar estas peças de fruta a todos os colaboradores (não apenas aos que votaram “Sim”).

Este tipo de *slides* e campanhas que envolviam processos ETL de dados serviam também para consciencializar ou apelar a outros momentos, também eles importantes, como o exemplo presente na Figura 26: uma colheita de sangue nas instalações da empresa.



Figura 26 - Survey de consciencialização para a dádiva de sangue

Abordando agora um pouco o assunto dos inquéritos de carácter obrigatório na organização, em primeiro lugar, era sempre criado o questionário desejado no *Microsoft Forms*, seleccionando e elaborando todas as questões relevantes, sendo que, em alguns casos, eram disponibilizadas opções de resposta múltipla com as opções mais adequadas, noutros a resposta seria aberta.

Era sempre personalizado o *design* de cada um destes e eram configuradas também as opções de privacidade, para garantir a confidencialidade das respostas. Em seguida, era distribuído o *link* de resposta ao questionário. Internamente, esta distribuição fazia-se via e-mail, *slide* com *QR Code* nos ecrãs internos, mas também através da afixação em quadros de comunicação, por forma a procurar sempre incentivar um maior número de pessoas a participar e a entregar *feedbacks* valiosos.

No *slide* apresentado na Figura 27, como é possível verificar, tratou-se de apelar à resposta ao inquérito “Consulta aos trabalhadores”. De uma forma muito simples, nos *slides* de lançamento destes inquéritos de carácter obrigatório, era dado o destaque ao nome do inquérito em questão e fazia-se um apelo à resposta, consciencializando sempre para a importância das respostas. A acrescer a essa informação, era também apresentado um *QR Code* para *scan*.

Uma vez mais, para além de partilhar o *slide* nos ecrãs internos e por email, também um *link* de resposta segue junto, por email, para que os colaboradores da organização possam optar pela forma como desejam aceder ao inquérito.

Por forma a proporcionar um melhor entendimento do conteúdo abordado neste tipo de inquéritos de carácter obrigatório, foi disponibilizada esta mesma “Consulta aos Trabalhadores”, na íntegra, no Anexo C.



Figura 27 - Inquérito de uma obrigação legal

As respostas vão chegando e quando se atinge o “deadline” para submeter resposta, utiliza-se a funcionalidade de exportação do *Microsoft Forms* para obter os dados recolhidos em formato *Excel*. Esta opção facilitou o processo de extração das respostas, pois já as organiza numa folha de cálculo, permitindo que se analise e interprete os resultados com maior facilidade.

Uma vez com os dados no *Excel*, existe a necessidade de limpar e tratar as informações recolhidas. Este processo inclui a remoção de respostas duplicadas ou inválidas, padronização dos formatos de dados e a aplicação de fórmulas para realizar cálculos ou criar métricas relevantes. Como se sabe, ainda que não seja a tecnologia de ponta no que diz respeito a *Data Analysis*, o *Excel* oferece uma ampla gama de recursos para manipulação, tratamento e transformação dos dados, o que foi fundamental para garantir tanto a qualidade quanto a fiabilidade.

Com os dados tratados e analisados, o passo seguinte passa por criar apresentações em *PowerPoint* para comunicar os resultados aos colegas internamente ou noutros departamentos específicos (Figura 28). Durante as várias horas de estágio foram utilizados inúmeros gráficos e tabelas variadas, por forma a facilitar a visualização e

entendimento dos principais *insights* e tendências identificadas nestes inquéritos. Nas apresentações em *PowerPoint*, por norma, eram também adicionados alguns comentários e explicações para entregar um melhor contexto dos dados apresentados.

TENDO EM CONTA O MÓDULO “\_\_\_\_\_” AVALIE,  
DE 1 A 5 (SENDO 1- INÚTIL E 5- MUITO ÚTIL),  
O CONTEÚDO QUE LHE FOI APRESENTADO?

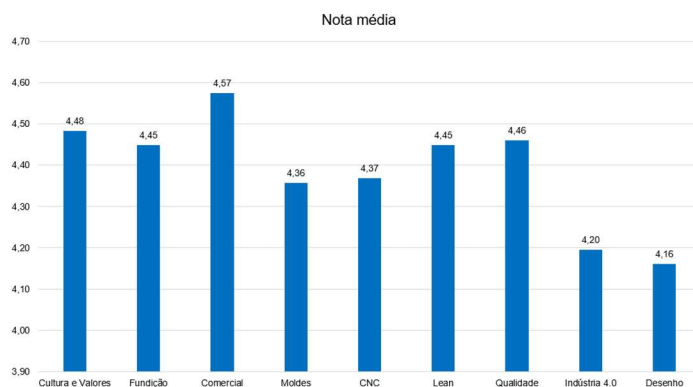


Figura 28 - Exemplo de um *slide* de apresentação de dados tratados em *PowerPoint* (tema: Formação)

Para além das apresentações em *PowerPoint*, também foram elaborados alguns relatórios mais detalhados em *Word* (posteriormente passados a *PDF*) para partilhar com os mais variados *stakeholders*. Nestes relatórios, eram destacados os principais resultados e conclusões a retirar do inquérito, fornecendo análises mais aprofundadas e quase sempre com recomendações/sugestões para melhoria.

Por vezes, alguns destes relatórios de inquérito eram apresentados em âmbito de reunião, utilizando, muitas vezes, os *slides* do *PowerPoint* ou as apresentações em *Canva* como suporte visual. Esta abordagem permitia uma comunicação eficaz dos dados recolhidos, e também facilitou a discussão e a tomada de decisões com base nas informações apresentadas.

No que diz respeito a processos ETL, também se aproveitavam estas ferramentas e capacidades de análise para promover vagas internas e analisar os perfis dos candidatos, aqui de uma forma simples, olhando apenas ao facto de cumprirem ou não com os requisitos mínimos para a função.

À semelhança dos exemplos anteriormente apresentados, também para estas vagas eram criados *slides* de apresentação da vaga, conforme é possível ver pelo exemplo

abaixo, na Figura 29, com *QR Code* para recolha de respostas via *Forms* e, quando seguia por email também se fazia acompanhar do *link* de resposta.



Figura 29 - Slide exemplo de um anúncio de Vaga Interna

Na Tabela 8 são apresentados os inquéritos mais importantes realizados ao longo do estágio utilizando a plataforma *Microsoft Forms*. Esta compilação abrange não apenas os inquéritos previamente discutidos, mas também uma variedade de outros relacionados com tópicos abordados ao longo do período de estágio. Esta abordagem permite ter uma visão completa das pesquisas conduzidas, refletindo a diversidade de temas e áreas de interesse exploradas durante esta experiência:

Tabela 8 - Inquéritos realizados

Nome do inquérito	Objetivo
Consulta aos trabalhadores	Inquérito de cariz obrigatório - serve para obter <i>feedbacks</i> e outras informações relevantes dos colaboradores de uma organização sobre diversos aspetos relacionados com o ambiente de trabalho, políticas, práticas, e satisfação no trabalho.
Inquérito do clima organizacional	Inquérito de cariz obrigatório - serve para avaliar a satisfação, o espírito, as atitudes e as perceções dos colaboradores em relação à empresa em que trabalham. O objetivo principal é identificar pontos fortes e áreas de melhoria na cultura da organização, para

	promover um ambiente de trabalho mais saudável, produtivo e motivador.
Campanhas internas e locais	Serve para registar e organizar a participação dos colaboradores em atividades internas da empresa, como dádivas de sangue, e em campanhas/eventos locais como caminhadas solidárias. Facilita a gestão logística, o planeamento de recursos e a comunicação com os participantes.
“Inquéritos com sentido”	Serve para aumentar o <i>engagement</i> junto dos colaboradores, perguntando-lhes algo que para eles teria um possível valor acrescentando, apresentando os resultados de uma forma relativamente fácil de interpretar, e procurando também que a implementação fosse muito simples e breve, o que permitiria, assim, tomar ações num curto espaço de tempo.
Vagas internas	Serve para identificar interessados dentro de uma organização em ocupar posições de emprego disponíveis antes de recorrer a contratações externas. Via inquérito recolhem-se as informações dos candidatos, verifica-se a sua compatibilidade com as competências necessárias e, assim, selecionar <i>short-lists</i> de candidatos.
Avaliações de formações	Serve para obter feedback dos participantes sobre a qualidade, eficácia e relevância de uma ação de formação, ajudando a melhorar o conteúdo das formações, bem como a medir o impacto das mesmas no desenvolvimento dos colaboradores.
Testes de competências	Serve para avaliar as capacidades e qualificações dos colaboradores que desejam candidatar-se a uma posição dentro da organização. Seja através de escolhas múltiplas ou “testes” mais complexos, este inquérito ajuda a identificar se os candidatos possuem as competências necessárias para desempenhar eficazmente as funções de uma determinada vaga em questão.

<p>“<i>Exit Discussions</i>” – Entrevistas de Saída</p>	<p>Serve para recolher feedback e informações dos colaboradores que estão a deixar a organização, sendo que o objetivo é compreender as razões da sua saída, avaliar a experiência que estes tiveram na empresa e identificar áreas que podem ser melhoradas.</p>
<p>Prendas de natal</p>	<p>Serve para recolher respostas dos colaboradores sobre que tipo de prendas os seus filhos gostariam de receber (da empresa), analisar e tratar para facilitar as compras e entregas.</p>

### 3.5. Auditoria interna de segurança da informação TISAX

TISAX, que significa "Trusted Information Security Assessment Exchange", é um padrão de segurança da informação desenvolvido pela indústria automóvel na Alemanha e, como tal, adequa-se perfeitamente, tanto à organização em si quanto aos seus clientes. Este padrão é projetado para avaliar e melhorar a segurança da informação em toda a cadeia de suprimentos automóvel. O TISAX é baseado em normas reconhecidas internacionalmente, como a ISO 27001, e estabelece critérios e procedimentos para avaliar a segurança da informação em organizações que trabalham, quase exclusivamente, com fabricantes de automóveis [51].

A certificação TISAX é frequentemente exigida por empresas deste setor para garantir que seus parceiros e fornecedores mantenham os mais altos padrões de segurança da informação ao lidar com dados confidenciais e informações sensíveis relacionadas à fabricação e *design* de veículos. Portanto, o acompanhamento de uma auditoria TISAX é um indicador importante de conformidade e segurança da informação em organizações que operam nesta indústria.

Ao acompanhar esta auditoria, ainda que de forma passiva, as atividades incluíram a observação das ações dos auditores, recolha de documentação e de evidências de ação, apoio à verificação da conformidade com normas e regulamentos, apoio administrativo, colaboração com a equipa de auditoria e contribuição na preparação de relatórios preliminares. Esta experiência ajuda a que haja uma compreensão mais profunda dos procedimentos de auditoria e da importância da segurança da informação na organização.

### 3.6. Apreciação crítica às atividades desenvolvidas

Ao longo do estágio, foi possível explorar e aplicar de forma prática os conhecimentos teóricos adquiridos durante o curso de mestrado em Ciência de Dados [52]. As atividades desenvolvidas proporcionaram uma oportunidade única para aperfeiçoar competências em áreas como modelagem preditiva, visualização de dados, processamento de informações e conformidade com padrões de segurança da informação. Esta secção tem como objetivo realizar uma apreciação crítica das atividades desenvolvidas, abordando os desafios enfrentados e os resultados alcançados.

Primeiramente, um dos principais desafios enfrentados durante o estágio foi a elaboração do modelo preditivo em ML. A qualidade e disponibilidade dos dados foram questões cruciais que impactaram significativamente o desenvolvimento do modelo. A obtenção de dados de alta qualidade, suficientes em quantidade e representativos para o problema em questão revelou-se uma tarefa extremamente árdua.

A dificuldade em encontrar dados adequados para treino e validação do modelo gerou impasses no processo de desenvolvimento. A escassez de informações relevantes e a falta de alguma melhor padronização entre as fontes disponíveis tornaram a tarefa ainda mais complexa. Além disso, a necessidade de lidar com dados incompletos e ruidosos obrigou a esforços adicionais para tratamento e pré-processamento. Para fazer face a estes “obstáculos” muito contribuíram as unidades curriculares do mestrado: *Big Data e Integração de Dados*.

Ainda assim, apesar dos obstáculos enfrentados, o compromisso com a qualidade dos resultados prevaleceu. Foram adotadas abordagens alternativas, como a utilização de técnicas de *data cleansing* e de integração dos dados, o que permitiu garantir a robustez do modelo preditivo final. A superação dessas dificuldades revelou-se fundamental para o sucesso do projeto e proporcionou valiosas lições sobre a importância da gestão e tratamento adequado dos dados em projetos de Ciência de Dados e ML.

É fundamental destacar que os resultados obtidos neste estudo podem estar sujeitos a possíveis inflações devido a diversos fatores. Primeiramente, a presença de uma coluna de “condição” no modelo, que determina o cumprimento ou não dos planos produtivos, pode influenciar significativamente os resultados, potencialmente levando a uma superestimação das previsões. Além disso, o desequilíbrio entre as classes, com poucos

exemplos de incumprimento em relação às situações de cumprimento, pode levar a uma tendência de viés em direção à classe majoritária, comprometendo a eficácia e a generalização do modelo. Outro ponto relevante é a ausência de imputação direta de registos históricos sobre o cumprimento prévio de encomendas passadas no *dataset*, o que pode gerar limitações na capacidade de capturar relações temporais e padrões de comportamento. É essencial estar ciente dessas possíveis influências e limitações ao interpretar os resultados, procurando sempre uma análise crítica e cautelosa das conclusões obtidas a partir do modelo preditivo. Certamente quem voltar a pegar no trabalho realizado deverá considerar o mencionado acima.

Quanto aos processos ETL, no geral, a utilização, tanto do *Microsoft Forms* quanto do *Mentimeter*, como plataformas para realizar inquéritos, combinado com as funcionalidades do *Excel*, *PowerPoint* e *Word*, proporcionaram um fluxo de trabalho bastante integrado e eficiente. Desde a recolha de dados até à comunicação dos resultados, este processo permitiu obter e entregar *insights* valiosos, e, quando conciliados com as competências consolidadas nas unidades curriculares de mestrado, nomeadamente “Análise Exploratória de Dados” e “*Business Intelligence*”, gerar impacto e fornecer informações acionáveis para a organização.

No que concerne às auditorias internas, durante o estágio deu-se a oportunidade de acompanhar uma de segurança da informação TISAX, o que me permitiu adquirir uma valiosa experiência na avaliação dos sistemas de informação de uma organização. Neste período, foi possível observar a importância crítica da garantia da segurança e da conformidade dos sistemas, assim como a complexidade das avaliações e procedimentos envolvidos. Esta oportunidade permitiu a aquisição de novos conhecimentos práticos sobre as melhores práticas em segurança cibernética, gestão de riscos e conformidade regulamentar, contribuindo também para a consolidação da compreensão das questões fundamentais relacionadas à proteção de dados e à integridade dos sistemas de informação num contexto cada vez mais digitalizado.

Em suma, e apesar dos desafios enfrentados no desenvolvimento do modelo preditivo, as atividades desenvolvidas ao longo do estágio proporcionaram uma experiência enriquecedora e contribuíram significativamente para o crescimento profissional.

## 4. Conclusão

Este documento descreve o trabalho desenvolvido no estágio realizado na empresa *Schmidt Light Metal*. O estágio consistiu na criação de um modelo preditivo de cumprimento de planeamentos produtivo que utiliza técnicas de *Machine Learning*, na criação de *dashboards* em *Power BI* e na criação de inquéritos e relatórios (que envolveram vários processos de ETL).

A experiência de desenvolver um modelo de *Machine Learning* no contexto de um ambiente fabril trouxe consigo desafios particulares, que exigiram capacidades e adaptações distintas para lidar com a complexidade dos dados. A aquisição e a consistência desses mesmos, frequentemente marcada por irregularidades e variações, impôs obstáculos claros, que foram superados com determinação, perseverança e trabalho conjunto. Embora nem sempre os resultados tenham sido alcançados de imediato, encaro essas dificuldades como oportunidades de aprendizagem valiosas que me ajudaram a aprimorar a minha capacidade de resolver problemas e lidar com situações imprevistas.

A cada teste e treino realizado ao longo desta caminhada que foi a elaboração de um modelo em *Machine Learning* deste género, pude aperfeiçoar as minhas técnicas de preparação dos dados, o que incluiu a remoção de valores omissos, o tratamento de erratas e a normalização dos mesmos. A minha capacidade de manusear dados provenientes de fontes diversas e heterogéneas foi também ampliada, e tal permitiu-me abordar de forma mais eficiente as fases de Extração, Transformação e Carregamento (ETL) necessárias para a análise de cada *dataset*.

Os inquéritos realizados através de ferramentas como o *Forms* e o *Mentimeter* foram uma componente essencial do estudo, permitindo-me recolher dados valiosos e diversificados. O tratamento destes, e a interpretação dos seus *insights* proporcionaram uma sólida base para a tomada de decisões informadas. A análise detalhada das respostas recolhidas permitiu também extrair conclusões significativas e destacar padrões e tendências relevantes.

Para além disso, a elaboração de relatórios de dados em diversos formatos representou uma etapa crucial para melhorar a minha forma de comunicar os resultados da pesquisa de forma eficaz. A diversidade de formatos, incluindo relatórios em *Word/PDF*, *designs* gráficos no *Canva* e visualizações interativas no *Power BI*, desafiou-

me a adaptar as minhas capacidades de *Data Visualization* às necessidades e preferências dos diversos *stakeholders*. Cada relatório exigiu uma abordagem única, garantindo a clareza e a eficácia na comunicação dos dados, desde os mais complexos aos mais simples, assim como, preservando sempre as informações relevantes.

Este documento representa o encerramento de um capítulo significativo na minha jornada académica e profissional. Com o encerramento deste estágio, é com grande satisfação que partilho a minha experiência ao integrar duas equipas tão distintas, uma de Tecnologias de Informação e outra de Recursos Humanos, durante a execução deste estudo. A colaboração entre estes dois departamentos revelou-se uma jornada verdadeiramente enriquecedora, onde a diversidade de conhecimentos e perspetivas proporcionou uma sinergia única e em muito contribuíram para que este estágio fosse tão rico. Ao unir esforços com tantos elementos diferentes, pude testemunhar como a complementaridade das competências resulta em soluções mais robustas e eficientes.

É com gratidão que reconheço o valor das oportunidades que este trabalho me proporcionou. A integração de equipas distintas, o conhecimento adquirido nas demais unidades curriculares do mestrado, o enfrentar de desafios na aquisição de dados, a realização de inquéritos tão variados e a elaboração de relatórios abrangentes foram pilares fundamentais para o crescimento do meu percurso profissional.

A integração entre as equipas de RH e IT demonstrou a importância da colaboração interdepartamental, evidenciando que a união de perspetivas diversas pode ser um motor de inovação e progresso. As capacidades desenvolvidas durante o processo deste estudo servirão como uma base sólida para enfrentar com sucesso os desafios futuros que surgirão, com toda a certeza, na minha carreira profissional.

Concluo assim, este relatório de estágio, com grande satisfação pelos resultados alcançados e com o sentimento de que esta jornada de aprendizagem, crescimento e descoberta foi uma experiência profissional extremamente enriquecedora. Sinto-me entusiasmado e preparado para encarar futuros projetos com determinação e motivação, confiante de que os conhecimentos adquiridos durante o mestrado e consolidados neste estágio contribuirão positivamente para o meu desenvolvimento contínuo. Com gratidão e otimismo, aguardo com entusiasmo as próximas oportunidades profissionais que a vida me reserva.

## Bibliografia

- [1] L. Hoory e C. Bottorff, “Forbes advisor - What Is Waterfall Methodology? Here’s How It Can Help Your Project Management Strategy,” *Forbes*, 25 Março 2022. [Online]. Available: [forbes.com/advisor/business/what-is-waterfall-methodology/](https://forbes.com/advisor/business/what-is-waterfall-methodology/). [Acedido em 10 Fevereiro 2023].
- [2] I. Sacolick, “InfoWorld - What is agile methodology? Modern software development explained,” 6 Abril 2022. [Online]. Available: <https://www.infoworld.com/article/3237508/what-is-agile-methodology-modern-software-development-explained.html>. [Acedido em 10 Fevereiro 2023].
- [3] Scrum, “scrum.org,” Scrum.org, [Online]. Available: <https://www.scrum.org/>. [Acedido em 10 Fevereiro 2023].
- [4] S. L. M. Group, “Schmidt Light Metal Group,” [Online]. Available: <http://performing.solutions/>. [Acedido em 31 10 2022].
- [5] S. C. Albright e W. L. Winston, *Data Analysis and Decision Making*, USA: Cengage Learning, 2010.
- [6] Z. Liu, N. Meyendorf e N. Mrad, “The Role of Data Fusion in Predictive Maintenance Using Digital Twin,” em *44th annual review of progress in quantitative nondestructive evaluation, volume 37*, Utah, USA, AIP Conference Proceedings, 2017, pp. 020023.1 - 020023.6.
- [7] C. Jacinto, F. P. Santos, C. G. Soares e S. A. Silva, “Assessing the coding reliability of work accidents statistical data: How coders make a difference,” *Journal of Safety Research*, vol. Volume 59, nº Safety, pp. 9-21, 2016.
- [8] M. J. I. d. Santos, “Business Intelligence na administração pública,” ISCTE Business School, Lisboa, 2019.
- [9] F. V. Primak, *Decisoes Com B.I. - Business Intelligence*, Brasil: Ciência Moderna, 2008.
- [10] E.-P. Lim, H. Chen e G. Chen, “Business Intelligence and Analytics: Research Directions,” *ACM Transactions on Management Information Systems*, vol. 3, nº 4, pp. 1-10, 2013.

- [11] D. Larson e V. Chang, “A review and future direction of agile, business intelligence, analytics and data science,” *International Journal of Information Management*, vol. 36, n° 5, pp. 700-710, 2016.
- [12] J. M. Munoz, *Global Business Intelligence*, New York: Routledge, 2018.
- [13] M. Nasr e S. Ouf, “Business Intelligence in the Cloud,” em *2011 IEEE 3rd International Conference on Communication Software and Networks*, Xi'an, China, 2011.
- [14] F. C. Lunenburg, “The decision making process,” *NATIONAL FORUM OF EDUCATIONAL ADMINISTRATION AND SUPERVISION JOURNAL*, vol. 27, n° 4, 2010.
- [15] F. Provost e T. Fawcett, “Data Science and its Relationship to Big Data and Data-Driven Decision Making,” em *BIG DATA*, Mary Ann Liebert, Inc, 2013, pp. 51-59.
- [16] A. Kusiak, J. Kern, K. Kernstine e B. Tseng, “Autonomous decision-making: a data mining approach,” em *IEEE Transactions on Information Technology in Biomedicine*, IEEE, 2000, pp. 274-284.
- [17] T. J. Loftus, P. J. Tighe, S. A. Brakenridge, P. A. Efron, A. C. Filiberto, A. M. Mohr, P. Rashidi, G. R. Upchurch e A. Bihorac, “Artificial Intelligence and Surgical Decision-making,” *JAMA Network*, 2019.
- [18] D. Owen, “Collaborative Decision Making,” em *Decision Analysis, Volume 12, Issue 1*, INFORMS, 2015, pp. 1-59.
- [19] B. Brehmer, “Strategies in real-time, dynamic decision making,” em *Insights in decision making: A tribute to Hillel J. Einhorn*, University of Chicago Press, 1990, pp. 262-279.
- [20] J. M. Sardroud, “Perceptions of automated data collection technology use in the construction industry,” em *Journal of Civil Engineering and Management, Volume 21, Issue 1*, Vilnius Gediminas Technical University, 2015, pp. 54-66.
- [21] M. W. Salter e A. Daneels, “WHAT IS SCADA?,” em *International Conference on Accelerator and Large Experimental Physics Control Systems*, Trieste, Italy, 1999.
- [22] T. Uhlemann, C. Schock, C. Lehmann, S. Freiburger e R. Steinhilper, “The Digital Twin: Demonstrating the Potential of Real Time Data Acquisition in Production Systems,” em *Procedia Manufacturing, Volume 9*, Elsevier, 2017, pp. 113-120.

- [23] M. Shanker, M. Y. Hu e M. S. Hung, “Effect of data standardization on neural network training,” em *Omega, Volume 24, Issue 4*, Elsevier, 1996, pp. 385-397.
- [24] P. Sarma, H. K. Singh e T. Bezboruah, “A Real-Time Data Acquisition System for Monitoring Sensor Data,” *International Journal of Computer Sciences and Engineering*, vol. 8, nº 6, 2018.
- [25] C. Taurion, *Big Data*, Brasil: BRASPORT, 2013.
- [26] Z.-H. Zhou, *Machine Learning*, Singapore: Springer, 2021.
- [27] Y. Zhang, *New Advances in Machine Learning*, Rijeka, Croatia: InTech, 2010.
- [28] J. H. Friedman, “Recent Advances in Predictive (Machine) Learning,” *Journal of Classification*, vol. 23, pp. 175-197, 2006.
- [29] P. M. Nadkarni, L. Machado-Ohno e W. W. Chapman, “Natural language processing: an introduction,” *Journal of the American Medical Informatics Association*, vol. 18, nº 5, pp. 544-551, 2011.
- [30] H. G. Barrow e J. M. Tenenbaum, “Computational vision,” *Proceedings of the IEEE*, vol. 69, nº 5, pp. 572-595, 1981.
- [31] A. Fanca, A. Puscasiu, D.-I. Gota e H. Valean, “Recommendation Systems with Machine Learning,” em *2020 21th International Carpathian Control Conference (ICCC)*, High Tatras, Slovakia, 2020.
- [32] G. Bonaccorso, *Machine Learning Algorithms: Popular algorithms for data science and machine learning - Second edition*, Birmingham, UK: Packt Publishing, 2018.
- [33] IT Convergence, “6 Benefits of Machine Learning in Manufacturing,” IT Convergence, [Online]. Available: <https://www.itconvergence.com/blog/6-benefits-of-machine-learning-in-manufacturing/>. [Acedido em 6 5 2023].
- [34] R. Rai, M. K. Tiwari, D. Ivanov e A. Dolgui, “Machine learning in manufacturing and industry,” *International Journal of Production Research*, p. 7, 2021.
- [35] T. Wuest, D. Weimer, C. Irgens e K.-D. Thoben, “Machine learning in manufacturing: advantages,” *Production & Manufacturing Research*, nº Production & Manufacturing Research, p. 24, 2016.
- [36] A. Zorpas e K. Lasaridi, “Measuring waste prevention,” em *Waste Management*, Elsevier, 2013, pp. 1047-1056.

- [37] C. Liu, Y. Zhang e S. Mao, “Image Classification Method Based on Multi-Agent Reinforcement Learning for Defects Detection for Casting,” *Sensors*, n° Deep Reinforcement Learning and IoT in Intelligent System, 2022.
- [38] A. Gonfalonieri, “Towards Data Science,” 7 Novembro 2019. [Online]. Available: <https://towardsdatascience.com/how-to-implement-machine-learning-for-predictive-maintenance-4633cdbc4860>. [Acedido em 8 06 2023].
- [39] V. Anneketh, V. Singh, J. Abhishek, B. Shivam, B. Aashima e S. Aarushi, “IoT and Machine Learning Approaches for Automation of Farm Irrigation System,” *Procedia Computer Science*, vol. 167, n° Computer Science, pp. 1250-1257, 2020.
- [40] Google, “Damos-lhe as boas-vindas ao Colaboratory,” Google, [Online]. Available: [https://colab.research.google.com/?utm\\_source=scs-index](https://colab.research.google.com/?utm_source=scs-index). [Acedido em 05 2023].
- [41] Microsoft, “Transforme os seus dados em impacto imediato,” Microsoft, [Online]. Available: <https://powerbi.microsoft.com/pt-pt/>. [Acedido em 05 2023].
- [42] “Canva,” Wikipédia, [Online]. Available: <https://pt.wikipedia.org/wiki/Canva>. [Acedido em 05 2023].
- [43] “Mentimeter,” Wikipédia, [Online]. Available: <https://en.wikipedia.org/wiki/Mentimeter>. [Acedido em 05 2023].
- [44] J.-P. Tremblay e P. G. Sorenson, An introduction to data structures with applications, Singapore: McGraw-Hill, Inc., 1987.
- [45] K. P. Murphy, Machine Learning: A probabilistic perspective, Massachusetts, United States of America, 2012.
- [46] s. learn, “scikit learn,” [Online]. Available: [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html). [Acedido em 14 Novembro 2022].
- [47] W. Hillier, “What is Data Cleaning and Why Does It Matter?,” Career Foundry, 14 Setembro 2023. [Online]. Available: <https://careerfoundry.com/en/blog/data-analytics/what-is-data-cleaning/>. [Acedido em 3 04 2023].
- [48] M. H. Pestana e J. N. Gageiro, Análise de dados para ciências sociais - A complementaridade do SPSS, Lisboa: Sílabo, 1998.

- [49] L. Zatolokina, “Mobidev - Machine Learning in Manufacturing: Industrial Use Cases,” 22 Março 2022. [Online]. Available: <https://mobidev.biz/blog/machine-learning-application-use-cases-manufacturing-industry>. [Acedido em 08 06 2023].
- [50] D. C. Hoaglin, F. Mosteller e J. W. Tukey, *Análise exploratória de dados: técnicas robustas - Um Guia*, Lisboa: Salamandra, 1983.
- [51] D. Global, “DQS - Certificação TISAX,” DQS, [Online]. Available: <https://www.dqsglobal.com/pt-pt/certifique/certificacao-tisax>. [Acedido em 17 Maio 2023].
- [52] T. Escovedo e A. Koshiyama, *Introdução a Data Science - Algoritmos de Machine Learning e métodos de análise*, Casa do Código, 2020.
- [53] J.-P. Tremblay e B. R. Bunt, *Introduction to computer Science*, Singapore: McGraw-Hill, Inc., 1989.

## **Anexos**

[A] Google Colab “Dataset\_creator” disponível para consulta em <https://colab.research.google.com/drive/1NbkEn1WUfNwJKvLsWhNuMYwTA5bCd1Zm?usp=sharing>

[B] Google Colab “ML\_IT\_testes” disponível para consulta em [https://colab.research.google.com/drive/125xlKrF7ML26Gw\\_DjwkMg2T-N9vIXGEr?usp=sharing](https://colab.research.google.com/drive/125xlKrF7ML26Gw_DjwkMg2T-N9vIXGEr?usp=sharing)

[C] FORMS – Consultável em [https://forms.office.com/Pages/ResponsePage.aspx?id=DQSIkWdsW0yxEjajBLZtrQAAAAAAN\\_\\_iZadaVUMFkzQ0dJMIMxQ01UTloxM1RCQ0swMjdRSC4u](https://forms.office.com/Pages/ResponsePage.aspx?id=DQSIkWdsW0yxEjajBLZtrQAAAAAAN__iZadaVUMFkzQ0dJMIMxQ01UTloxM1RCQ0swMjdRSC4u)

[D] PDF – “Manual de recolha de dados | Schmidt Light Metal Group

Schmidt  
Light  
Metal  
Group

---

# RECOLHA DE DADOS

---

Carlos Ferreira  
Instituto Politécnico de Leiria

2023

# INTRODUÇÃO

---

Uma descrição clara e completa da recolha e tratamento de dados é essencial para garantir a integridade e a qualidade dos resultados obtidos.

Este documento tem como objetivo apresentar uma visão geral do processo de recolha e tratamento de dados utilizados neste projeto, bem como os métodos e técnicas utilizados para garantir a precisão e a fiabilidade dos dados.

A recolha de dados é uma etapa fundamental no processo de investigação. É neste momento que são recolhidas as informações necessárias para responder às perguntas de pesquisa.

Após a recolha, os dados foram tratados para estarem aptos a serem utilizados no modelo. Este processo incluiu a validação dos dados para garantir que não havia erros ou inconsistências, bem como técnicas de limpeza.

Este documento descreve com detalhe o processo de recolha e tratamento de dados.

O objetivo deste documento é fornecer uma descrição clara e completa do processo de recolha e tratamento de dados, garantindo a transparência e a objetividade dos resultados obtidos.

# INDÍCE

---

1. hr.xlsx
2. obj\_OP1.xlsx
3. obj\_OP2.xlsx
4. obj\_OP3.xlsx
5. obj\_OP4.xlsx
6. hxt.xlsx
7. ddp.xlsx
8. orders.xlsx
9. Dataset Creator (Colab)

# HR.XLSX

1. Aceder aos registos de ponto do ELO;
2. Criar uma consulta do total de picagens (por dia) na portaria;
3. Dividir os valores de picagem por 2 (porque a maioria das pessoas tem registo de entrada e de saída);
4. Recolher essa consulta em .xlsx com duas colunas ("Dia" e "humanresources") e guardar com o nome "hr".

Preview da estrutura:

	A	B	C	D	E
1	Dia	humanresources			
2	04/01/2021	306			
3	05/01/2021	318			
4	06/01/2021	304			
5	07/01/2021	308			
6	08/01/2021	312			
7	09/01/2021	140			
8	10/01/2021	5			
9	11/01/2021	296			
10	12/01/2021	305			

# OBJ\_OP1.XLSX

1. Através de um ficheiro partilhado na rede foi consultado o número de injeções esperadas de cada Referência (de peça), por dia e ano, e a partir daí, olhando à cadência de injeção, foi calculado o Obj OP1 (Objetivo em Operação de Produção 1);
2. Para este cálculo, **na célula AA2**, utiliza-se a seguinte fórmula SE.ERRO(((G2\*H2)\*7,5;"")) ;
3. Interessa também criar uma coluna o mais à direita possível, nomeada "Ref", onde, com base na coluna "Molde", se retira a Referência de peça associada. Para tal, utiliza-se, **na célula AB2**, a fórmula SE.ERRO(ESQUERDA(F2;(PROCURAR(" ";F2)));F2) ;
4. Finalmente, e para um ficheiro diferente a que chamamos "obj\_OP1" extraem-se (copiando) as colunas "Data", "Ref" Obj\_OP1", ordenando conforme apresentado abaixo.

Preview da estrutura:

	A	B	C	D	E
1	Dia	Ref	Obj_OP1		
2	04/01/2021	0			
3	04/01/2021	0			
4	04/01/2021	042B	975		
5	04/01/2021	042AC	1020		
6	04/01/2021	0			
7	04/01/2021	0			
8	04/01/2021	M262C	412,5		
9	04/01/2021	0			
10	04/01/2021	239	900		

# OBJ\_OP2.XLSX

1. Através de um ficheiro partilhado na rede foi consultado o planeamento por Referência (de peça), por dia e ano, bem como o planeado, definido por Obj OP2 (Objetivo em Operação de Produção 2);
2. Depois de consultado o ficheiro com o nome "Produção Diaria Acabamento - SEM 01- 2021" (onde "SEM 01" diz respeito ao nº da semana e "2021" diz respeito ao ano da semana que se consulta), foi seleccionada a informação relevante e foi criado o ficheiro de extração dessa informação, denominado "Ficheiro.extracao.op2".
3. Utilizando a macro CTRL+r é recolhida toda a informação relevante (colunas "Data", "Ref" e "Obj\_OP2") e depois interessa copiá-la e colá-la num ficheiro Excel diferente nomeado "obj\_OP2" com a seguinte organização:

Preview da estrutura:

	A	B	C	D	E
1	Dia	Ref	Obj_OP		
2	04/01/2021	148X	0		
3	04/01/2021	107AE	600		
4	04/01/2021	107AC	0		
5	04/01/2021	MF32	0		
6	04/01/2021	941	2200		
7	04/01/2021	R077	0		
8	04/01/2021	R047	0		
9	04/01/2021	238	400		
10	04/01/2021	239	600		

# OBJ\_OP3.XLSX

1. Através de um ficheiro partilhado na rede foi consultado o planeamento por Referência (de peça), por dia e ano, bem como o planeado, definido por Obj OP3 (Objetivo em Operação de Produção 3);
2. Depois de consultado o ficheiro com o nome "apa-S01-2021" (onde "S01" diz respeito ao nº da semana e "2021" diz respeito ao ano da semana que se consulta), foi seleccionada a informação relevante e foi criado o ficheiro de extração dessa informação, denominado "Ficheiro.extracao.op3".
3. Utilizando a macro *CTRL+r* é recolhida toda a informação relevante (colunas "Data", "Ref", "Obj\_OP3") e depois interessa copiá-la e colá-la num ficheiro Excel diferente nomeado "obj\_OP3" com a seguinte organização:

Preview da estrutura:

	A	B	C	D	E
1	Dia	Ref	Obj_OP3		
2	08/01/2021	148X	240		
3	08/01/2021	189	165		
4	08/01/2021	MF32	0		
5	08/01/2021	941	14190		
6	08/01/2021	R077	108		
7	08/01/2021	R047	0		
8	08/01/2021	238	2700		
9	08/01/2021	239	2200		
10	08/01/2021	188	5472		

# OBJ\_OP4.XLSX

1. Através de um ficheiro partilhado na rede foi consultado o planeamento por Referência (de peça), por dia e ano, bem como o planeado, definido por Obj OP4 (Objetivo em Operação de Produção 4);
2. Depois de consultado o ficheiro com o nome "obj.sem.pr.acab\_mod", foi selecionada a informação relevante e foi criado o ficheiro de extração dessa informação, denominado "Ficheiro.extracao.logist".
3. Utilizando a macro CTRL+r é recolhida toda a informação relevante (colunas "Data", "Ref", "Obj\_OP4") e depois interessa copiá-la e colá-la num ficheiro Excel diferente nomeado "obj\_OP4" com a seguinte organização:

Preview da estrutura:

	A	B	C	D	E
1	Dia	Ref	Obj_OP4		
2	04/01/2021	148X	0		
3	04/01/2021	189	0		
4	04/01/2021	107AE	800		
5	04/01/2021	107AC	0		
6	04/01/2021	MF32	0		
7	04/01/2021	941	2475		
8	04/01/2021	R077	0		
9	04/01/2021	R047	0		
10	04/01/2021	238	450		

# HXT.XLSX

1. Primeiramente é extraído do Sage uma query com o código "HXT\_SUM" que contem a quantidade de lançamentos de horas extra (HXT3, HXT7, HXT9,...) por dia do ano, desde inicio de 2021.
2. Depois é consultado o ficheiro extraído com o nome "HXT\_SUM", é selecionada a informação relevante e é criado o ficheiro de extração dessa informação, denominado "Ficheiro.extracao.hxt".
3. Neste ficheiro, começa por ser feito um agrupar de *hxt* por dia do ano, para que a cada dia seja alocado apenas um "*sum.hxt*" (um somatório das horas extra por dia do ano).
4. Decidimos internamente que a existência de horas extras à semana não funcionavam como trabalho excecional mas mais como complemento. Desta forma filtramos apenas os dois dias de descanso (sábado e domingo).

Preview da estrutura:

	A	B	C	D	E
1	Dia	Ref	sum.hxt		
2	16/01/2021	042B	18,50		
3	16/01/2021	093C	18,50		
4	16/01/2021	656	18,50		
5	16/01/2021	166H	18,50		
6	16/01/2021	308BM	18,50		
7	16/01/2021	941	18,50		
8	16/01/2021	188	18,50		
9	16/01/2021	107AE	18,50		
10	16/01/2021	262J	18,50		

# DDP.XLSX

1. Através de uma consulta SQL foram retirados todos os registros de "ddp" (transporte especial de mercadorias). Este registro é consultado entrega, por dia e referência de peça, a quantidade de envios especiais.
2. Há que formatar o ficheiro extraído no seu campo "Dia" para o formato DD/MM/AAAA, e o campo "Ref" para o tipo "Texto".

Preview da estrutura:

	A	B	C	D	E
1	Dia	Ref	sum.ddp		
2	08/01/2021	239	6		
3	01/04/2021	V839	10		
4	01/04/2021	H839	10		
5	06/04/2021	V839	30		
6	06/04/2021	H839	30		
7	08/04/2021	H839	6		
8	14/04/2021	055AC	5		
9	17/04/2021	055AE	5		
10	23/04/2021	V839	70		

# ORDERS.XLSX

1. Nesta consulta aos pedidos de encomenda é feita uma extração em SQL do nº da encomenda ("Order\_nr"), das quantidades encomendadas ("Obj\_order"), da referência de peça pedida ("Ref") e do dia de registo da encomenda ("Dia").
2. À semelhança de outros registos extraídos em excel, também aqui foi importante proceder a algumas formatações, nomeadamente
  - a. Coluna "Dia" para o formato DD/MM/AAAA;
  - b. Coluna "Ref" para o tipo "Texto";
  - c. Coluna "Order\_nr" para o tipo "Texto";
3. Findada esta formatação, o dataset deverá estar com um aspeto similar ao apresentado abaixo, pronto a ser utilizado.

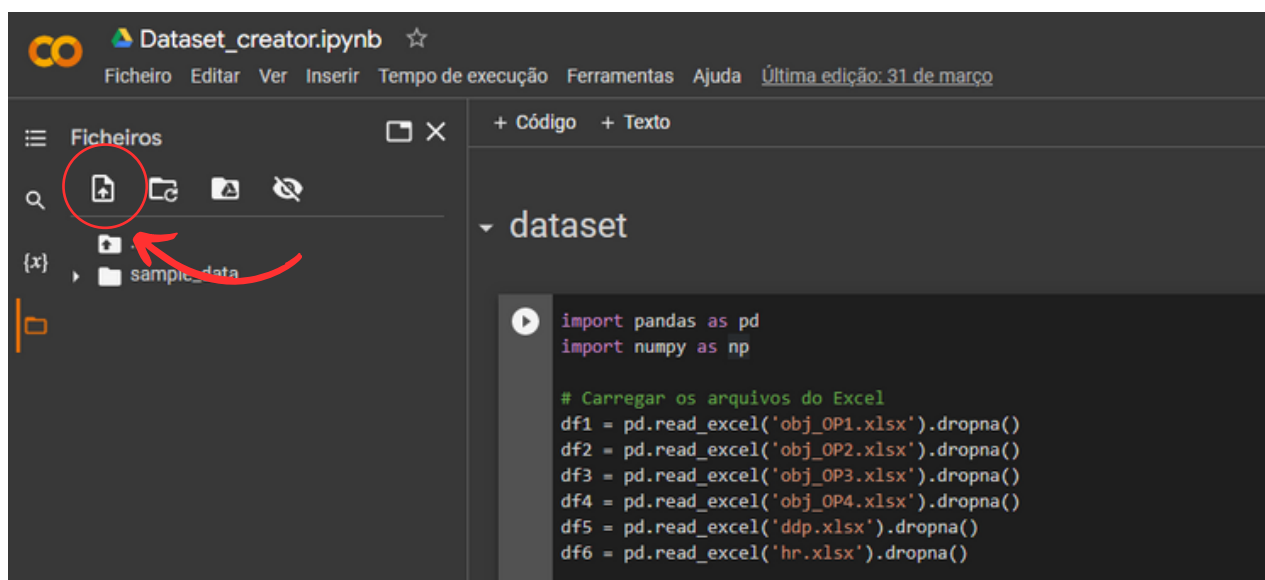
Preview da estrutura:

	A	B	C	D	E
1	Dia	Ref	Order_nr	Obj_order	
2	04/01/2021	4M7AF	108300506	180,00	
3	04/01/2021	307BN	108300653	120,00	
4	04/01/2021	307BN	108300653	480,00	
5	04/01/2021	308BL	108300654	240,00	
6	04/01/2021	308BM	108300714	330,00	
7	04/01/2021	262J	108300791	88,00	
8	04/01/2021	262J	108300791	704,00	
9	04/01/2021	K262R	108300893	180,00	
10	04/01/2021	263N	108300925	42,00	

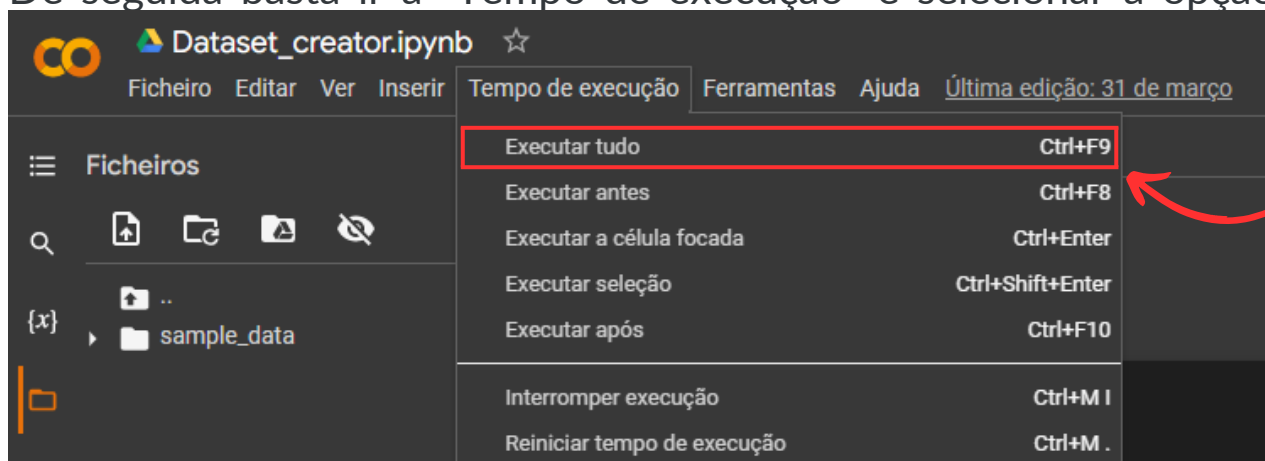
# DATASET CREATOR (COLAB)

Tendo os diferentes ficheiros Excel tratados podemos agora passar para o Google Colab onde deverá colar o código que está pronto (e pode ser consultado no ficheiro "notebook\_dataset.creator.txt") para se avançar para a criação do dataset final, compilando toda a informação relevante.

Primeiramente interessa fazer upload de todos os ficheiros .xlsx. Para isso basta clicar no botão para "upload", seleccionar os ficheiros e carregar em "Abrir":



De seguida basta ir a "Tempo de execução" e seleccionar a opção

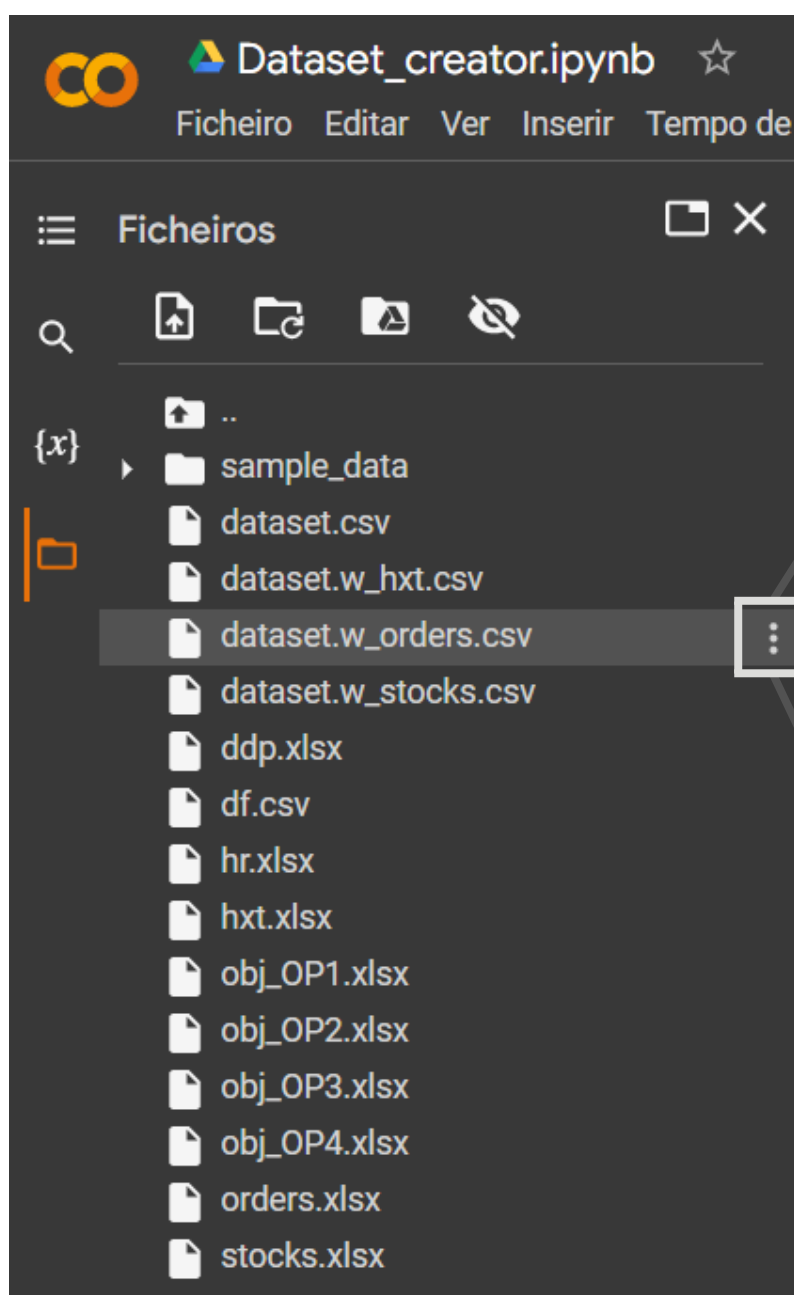


# DATASET CREATOR (COLAB)

Uma vez executado o Notebook, é expectável que um ficheiro .csv seja criado (do lado esquerdo) e estará assim sujeito a transferência para se utilizar no modelo treinado.

Tendo em conta os resultados obtidos e a comparação entre testes, o .csv que queremos é aquele cujo nome é "dataset.w\_orders.csv".

Segue o exemplo abaixo de como transferir:



- 1º. No .csv correto carregar nos 3 pontos;
- 2º. Selecionar a opção "Transferir".
- 3º. O .csv em questão deverá ir para a sua pasta de *Transferências*.