



# **Visualização de Dados e *Data Storytelling*: Estudo de Caso**

Mestrado em Ciência de Dados

João Francisco Soares Paiva

Leiria, setembro de 2024



# **Visualização de Dados e *Data Storytelling*: Estudo de Caso**

Mestrado em Ciência de Dados

João Francisco Soares Paiva

Dissertação realizada sob a orientação da Professora Doutora Maria Beatriz Guerra da Piedade e da Professora Doutora Rosa Isabel Alves Cordeiro Matias

Leiria, setembro de 2024

# **Originalidade e Direitos de Autor**

A presente dissertação é original, elaborada unicamente para este fim, tendo sido devidamente citados todos os autores cujos estudos e publicações contribuíram para a 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 a mesma foi realizada, a saber, Curso de Mestrado em Ciência de Dados, no ano letivo 2023/2024, 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.

# Agradecimentos

Avizinhando-se o final desta jornada, e com um sentimento de dever cumprido por todas as etapas ultrapassadas, é hora de refletir e agradecer a todas as pessoas que de alguma forma fizeram parte deste projeto e deixaram uma parte delas nestas palavras que aqui escrevo.

Gostaria de deixar um profundo agradecimento a todos os que de forma direta ou indireta contribuíram para que este projeto fosse possível, tornando esta longa jornada num caminho menos sinuoso e mais alegre, e sem os quais certamente não teria alcançado o resultado a que me propus inicialmente.

Começo por agradecer às minhas orientadoras, Professora Doutora Rosa Isabel Alves Cordeiro Matias e Professora Doutora Maria Beatriz Guerra da Piedade, que sempre estiveram presentes em todas as fases deste estudo e possibilitaram que este fosse concluído com sucesso. A sua orientação marcada por um elevado rigor científico, as suas opiniões construtivas e assertivas sobre o desenrolar das diversas etapas e a sua disponibilidade tornaram este trabalho muito mais valioso e completo.

À minha família, que sempre me acompanha em todas as etapas da minha vida, e foi essencial nesta conquista dando-me todo o apoio que precisava e nunca deixando de acreditar em mim, independentemente das dificuldades e adversidades enfrentadas.

Agradeço à minha namorada, sempre presente e disponível ao longo de toda esta etapa, incansável no apoio demonstrado.

E agradeço também aos meus amigos, sempre prontos a dar uma palavra de ânimo e incentivo nos momentos cruciais desta jornada.

A todos vocês, o meu sentido agradecimento!

“O Homem é do tamanho do seu sonho.”

Fernando Pessoa

# Resumo

Desde os primórdios da civilização, a humanidade lida com dados de diferentes naturezas, porém os avanços computacionais provocaram tanto a explosão de informação, como o processo de a tornar interpretável num determinado contexto. Com o desenvolvimento de um volume cada vez maior de dados, surge a necessidade de desenvolver e aprimorar técnicas e ferramentas capazes de lidar com o armazenamento, processamento e apresentação da informação proveniente destas fontes. Dando resposta a estes desafios, as ferramentas gráficas evoluíram também para analisar estes dados massivos, tornando a sua leitura mais amigável. Estas ferramentas são capazes de facilitar a detecção de anomalias nos dados, relacionamentos, padrões ou tendências, tornando os pontos a avaliar mais evidentes nos gráficos selecionados.

Uma das problemáticas mais relevantes do estudo da visualização de informação recai sobre a escolha das representações gráficas adequadas para cada tipo de dados. Associar dados a elementos gráficos e a uma localização no espaço pode ser uma tarefa difícil que requer conhecimentos em tipos de dados, análise de dados e representações gráficas. Também o *storytelling* se apresenta como um tema bastante preponderante na questão da apresentação de informação proveniente dos dados recolhidos, entrando em ação, frequentemente, a par da temática da visualização de dados.

Como objetivo geral deste estudo, pretende-se o desenvolvimento de painéis sobre a forma de *dashboards* que sejam capazes de transmitir a informação presente numa base de dados considerada previamente (projeto UniBench) de forma que o utilizador final seja capaz de interpretar a informação apresentada e responder a um conjunto de questões definidas *a priori*. Durante este processo serão tidas em conta as técnicas de visualização adequadas para cada caso específico e, conseqüentemente, as técnicas de *storytelling* mais adequadas para contar a história desses dados.

No final do estudo foi possível inferir que os utilizadores dão primazia à simplicidade de utilização e navegação, sendo este um fator decisivo na sua escolha por determinada plataforma ou *dashboard*. É essencial que os tópicos mais relevantes sejam de fácil acesso e que se apresentem bem visíveis, devendo ser desenvolvido um esquema de navegação conciso, cujo acesso aos dados seja coerente e que ocorra mediante processos semelhantes,

de forma que o utilizador consiga identificar com facilidade a forma de resgatar determinados dados. Para finalizar, importa ressaltar que a história de dados deve sustentar os elementos visuais conferindo uma fonte adicional de suporte à navegação pelos diferentes painéis, complementando a informação apresentada com a interpretação do próprio utilizador.

**Palavras-chave:** Visualização de dados, *Data Storytelling*, *Dashboard*

# Abstract

Since the dawn of civilization, humanity has been dealing with data of different kinds, but advances in computing have triggered both the explosion of information and the process of making it interpretable in each context. With the development of large volumes of data comes the need to develop and improve techniques and tools capable of handling the storage, processing and presentation of information from these sources. In response to these challenges, graphic tools have also evolved to analyse this massive data, making it more user-friendly to read. These tools can facilitate the detection of anomalies in the data, relationships, patterns or trends, making the points to be evaluated more evident.

One of the most important issues in the study of information visualization is the choice of appropriate graphic representations for each type of data. Associating data with graphic elements and a location in space can be a difficult task that requires knowledge of data types, data analysis and graphic representations. Storytelling is also a very important topic when it comes to presenting information from the data collected, and often comes into play alongside data visualization.

The general aim of this study is to develop a dashboard capable of transmitting the information present in a previously considered database (UniBench project) in such a way that the end user is able to interpret the information presented and answer a set of questions defined previously. During this process, the appropriate visualization techniques will be considered for each specific case and, consequently, the most appropriate storytelling techniques to tell the story of the data.

At the end of the study, it was possible to infer that users give priority to simplicity of use and navigation, which is a decisive factor in their choice of a particular platform or dashboard. It is essential that the most relevant topics are easy to access and clearly visible. A concise navigation scheme should be developed, with consistent access to data and similar processes, so that the user can easily identify how to retrieve certain data. Finally, it should be noted that the data story should support the visual elements by providing an additional source of support for navigation through the different panels, complementing the information presented with the user's own interpretation.

**Keywords:** Data visualization, Storytelling, Dashboard

# Índice

<b>Originalidade e Direitos de Autor</b> .....	<b>iii</b>
<b>Agradecimentos</b> .....	<b>iv</b>
<b>Resumo</b> .....	<b>v</b>
<b>Abstract</b> .....	<b>vii</b>
<b>Lista de Figuras</b> .....	<b>x</b>
<b>Lista de tabelas</b> .....	<b>xi</b>
<b>Lista de siglas e acrónimos</b> .....	<b>xii</b>
<b>1. Introdução</b> .....	<b>1</b>
<b>1.1. Enquadramento</b> .....	<b>2</b>
<b>1.2. Objetivos gerais e específicos</b> .....	<b>4</b>
<b>1.3. Metodologia</b> .....	<b>5</b>
<b>1.4. Estrutura do trabalho</b> .....	<b>5</b>
<b>2. Estudo prévio</b> .....	<b>7</b>
<b>2.1. Dos dados à sabedoria</b> .....	<b>7</b>
<b>2.2. Visualização de dados e Introdução aos elementos visuais</b> .....	<b>10</b>
2.2.1. Fundamentos da Visualização de Informação .....	10
2.2.2. Elementos Visuais .....	13
2.2.3. Gráficos de Informação .....	24
<b>2.3. Desenho de <i>Dashboards</i></b> .....	<b>32</b>
<b>2.4. <i>Dashboards</i> e <i>Data storytelling</i></b> .....	<b>42</b>
<b>2.5. Perfis de utilizadores</b> .....	<b>50</b>
<b>2.6. Abordagens para projetos de <i>Data storytelling</i></b> .....	<b>51</b>
<b>2.7. <i>Data storytelling</i> na era do <i>Big Data</i></b> .....	<b>54</b>
<b>3. Planeamento e Conceção do caso de estudo</b> .....	<b>58</b>
<b>3.1. Caso de estudo: proveniência e análise</b> .....	<b>58</b>
3.1.1. Proveniência dos dados .....	58
3.1.2. Estrutura dos dados de suporte .....	58

3.2.	<b>Estrutura do sistema a desenvolver .....</b>	<b>59</b>
3.3.	<b>Enunciado das questões a responder .....</b>	<b>60</b>
3.4.	<b>Seleção de ferramentas.....</b>	<b>61</b>
3.5.	<b>Planeamento dos <i>Dashboards</i> .....</b>	<b>63</b>
3.5.1.	Definição de perfis de utilizadores .....	63
3.5.2.	Desenho de <i>mockups</i> .....	64
3.6.	<b>Planeamento do <i>Data Storytelling</i> .....</b>	<b>66</b>
4.	<b>Da conceção à narração: implementação do caso de estudo.....</b>	<b>69</b>
4.1.	<b>Implementação dos <i>dashboards</i>.....</b>	<b>69</b>
4.2.	<b>Conceção da história de dados.....</b>	<b>75</b>
5.	<b>Testes e Validação .....</b>	<b>78</b>
5.1.	<b>Questionário de usabilidade com um grupo de participantes.....</b>	<b>79</b>
5.2.	<b>Avaliação dos resultados .....</b>	<b>81</b>
6.	<b>Discussão de Resultados e Considerações Finais.....</b>	<b>87</b>
6.1.	<b>Discussão de Resultados .....</b>	<b>87</b>
6.2.	<b>Conclusão, Limitações do estudo e Futuras linhas de investigação.....</b>	<b>91</b>
	<b>Referências Bibliográficas .....</b>	<b>95</b>
	<b>Anexos .....</b>	<b>100</b>
	<b>Anexo A – <i>Mockups</i>.....</b>	<b>100</b>
	<b>Anexo B – <i>Dashboards</i> .....</b>	<b>103</b>
	<b>Anexo C – <i>Dashboards</i> tipo de Gestor Operacional.....</b>	<b>107</b>
	<b>Anexo D – <i>System Usability Scale</i>.....</b>	<b>110</b>
	<b>Anexo E – Questionário completo para teste com utilizadores .....</b>	<b>111</b>

# Lista de Figuras

Figura 1 - Pirâmide do Conhecimento (adaptado de Rowley, 2007).....	7
Figura 2 - Variáveis visuais de MacEachren (extraído de Roth, 2017) .....	22
Figura 3 - Modelo de classificação "Representações visuais - o que pretendo transmitir?" (extraído de Pereira, 2015).....	41
Figura 4 - Combinação entre narrativa, recursos visuais e dados (adaptado de Brent Dykes, 2016) .....	45
Figura 5- <i>Visual Data Storytelling Process</i> (extraído de B. Lee et al., 2015).....	46
Figura 6 - Exemplo da aplicação da <i>Data storytelling</i> (extraído de Echeverria et al., 2018).....	49
Figura 7 - Itens da <i>framework</i> de <i>data storytelling</i> (extraído de Zhang et al., 2022) .....	52
Figura 8 - Exemplo <i>dashboard</i> desenvolvido (extraído de Farias, 2020).....	53
Figura 9 - <i>The Analytic Hero's Journey</i> (extraído de Behera & Swain, 2019).....	56
Figura 10 - Modelo de Dados .....	59
Figura 11 - Esquema de transformação e visualização de dados (extraído de Wang, 2022).....	59
Figura 12 - 2024 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms (extraído de Schlegel, 2024).....	61
Figura 13 - <i>Mockup</i> Caraterização Cliente .....	64
Figura 14 - <i>Mockup</i> Análise de Vendas .....	65
Figura 15 - Modo de navegação pelos diferentes <i>dashboards</i> .....	66
Figura 16 - Planeamento do Data Storytelling (adaptado de Farias, 2020) .....	68
Figura 17 - <i>Dashboard</i> Caraterização Cliente .....	73
Figura 18 - <i>Dashboard</i> Análise de Vendas (sem filtros) .....	74
Figura 19 - Forma de navegação pelo <i>dashboard</i> .....	74
Figura 20 - Caraterização do cliente quanto à Localização geográfica (Europa).....	75
Figura 21 - Top 5 marcas com classificação mais baixa .....	76
Figura 22 - Evolução da Faturação comparando os meses de maio .....	77
Figura 23 - Classificação da pontuação SUS (extraído de Brooke, 2013).....	80
Figura 24 - Curva da descoberta de problemas em testes de usabilidade (adaptado de Nielsen, 2000) .....	81

# Lista de tabelas

Tabela 1 - Comparação entre visualização de dados e <i>data storytelling</i> (adaptado de Zhang et al., 2022).....	42
Tabela 2 - Tipos de dashboard (adaptado de Hayward, 2021) .....	51
Tabela 3 - Medidas descritivas de caracterização sociodemográfica.....	82
Tabela 4 - Resultado das respostas do questionário SUS e resultado ponderado .....	83
Tabela 5 - Tempo de realização de cada tarefa.....	84
Tabela 6 - Respostas ao questionário de <i>Storytelling</i> .....	85

# Lista de siglas e acrónimos

BI	<i>Business Intelligence</i>
ESTG	Escola Superior de Tecnologia e Gestão
IHC	Interação Humano-Computador
JSON	<i>JavaScript Object Notation</i>
TI	Tecnologias de Informação
VDSP	<i>Visual Data Storytelling Process</i>
VI	Visualização de informação
XML	<i>eXtensible Markup Language</i>

# 1. Introdução

O avanço tecnológico das comunicações e o crescente desenvolvimento dos sistemas digitais faz com que os utilizadores e organizações sejam inundados por uma grande quantidade de informação, surgindo uma grande dificuldade em perceber o que será e não será relevante. Numa sociedade em que cada vez mais os dados assumem um papel preponderante na cultura empresarial, e uma forma de maximizar o desempenho corporativo, as tecnologias digitais que permitem a recolha, produção e circulação tornam-se imprescindíveis para a exploração e compreensão dos próprios dados.

Este facto provocou nos indivíduos a necessidade de poder transformar os dados obtidos em informação e, posteriormente, em conhecimento. A informação obtém-se a partir do processamento, organização e representação dos dados, num determinado contexto, fazendo com que estes se tornem úteis e compreendidos. Após a transformação dos dados em informação surge o conhecimento, adquirido pela combinação das informações com experiências e *insights* que conferem benefícios aos utilizadores e às organizações (Pereira, 2015).

Uma forma de lidar com a interpretação do grande volume dados é o recurso a mecanismos visuais, onde as pessoas conseguem explorar, relacionar, interpretar e apresentar informação útil com rapidez e precisão. A visualização da informação consiste no processo de transformar dados em imagens ou representações gráficas com a finalidade de serem interpretadas e/ou apresentadas (Rodrigues & Dias, 2017). A utilização das técnicas de visualização de dados neste campo corresponde a uma estratégia emergente e de inovação que tem sido explorada em áreas distintas do conhecimento e recebido cada vez mais atenção em estudos e pesquisas, assim como por parte da administração das próprias empresas e organizações (Rodrigues & Dias, 2017).

Além da visualização de dados, também as estratégias de *data storytelling* se apresentam fundamentais para que os *insights* recolhidos a partir dos dados se tornem relevantes e sejam capazes de transparecer a informação realmente gerada através dos mesmos.

O termo *storytelling*, na sua tradução literal, significa "contar histórias", derivando dos termos “*story*”, em português, história e “*telling*” que significa contar ou narrar.

Habitualmente, o *storytelling* refere-se à arte de contar histórias de forma envolvente e cativante para transmitir informação, mensagens ou experiências a uma determinada audiência, sendo esta técnica utilizada em várias áreas, incluindo marketing, comunicação, literatura, cinema e negócios (Feigenbaun & Alamalhadaei, 2020).

O *storytelling* pressupõe a criação de uma narrativa que capta a atenção do público, gera empatia e transmite uma mensagem de maneira mais eficaz do que apenas apresentar factos e dados objetivos, permitindo que a informação seja contextualizada e envolva as pessoas emocionalmente, tornando a mensagem mais impactante e memorável (Feigenbaun & Alamalhadaei, 2020).

Ao refletir sobre a importância da aplicação destas estratégias, como referido por Cole Nussbaumer Knaflic (2015) “Em questões de linguagem, aprendemos como juntar palavras em frases e em histórias; com a matemática, aprendemos a entender os números. Mas é raro que esses dois temas se encontrem juntos, isto é, ninguém nos ensina a contar histórias com números.” (Knaflic, 2015). Com estas palavras, torna-se possível entender a lacuna que persiste na transmissão de informação de boa qualidade a partir de dados de boa qualidade, uma vez que sem a adoção de estratégias adequadas, o que outrora seriam dados com forte potencial de gerar informação de elevada qualidade acabam por se transformar em informações enviesadas, de duvidosa qualidade ou pouco perceptíveis para o recetor final.

O foco da presente dissertação está na análise da influência das estratégias de visualização de dados e *data storytelling*, e, de que forma as metáforas e elementos gráficos, em conjugação com as estratégias de *data storytelling*, poderão ou não promover um efeito decisivo junto dos indivíduos, quando a estes seja apresentado um conjunto de informações provenientes de dados de pequena, média ou grande dimensão.

## **1.1. Enquadramento**

Desde os tempos mais longínquos que a humanidade lida com dados, porém os avanços computacionais provocaram tanto a explosão de informação, como o processo de a tornar interpretável num determinado contexto. Com o desenvolvimento dos grandes volumes de dados (*Big Data*) surge a necessidade de desenvolver e aprimorar técnicas e ferramentas capazes de lidar com o armazenamento, processamento e apresentação da informação proveniente destas fontes. O surgimento do *Big Data*, apesar de carregar em si um enorme

valor intrínseco, acarreta também um conjunto de desafios em analisar informação de forma rápida e eficaz, de forma a trazer vantagem para quem a analisa.

Dando resposta a estes desafios, as ferramentas gráficas evoluíram também para analisar estes dados massivos, tornando a sua leitura mais amigável. Estas ferramentas são capazes de facilitar a deteção de anomalias nos dados, relacionamentos, padrões ou tendências, tornando os pontos a avaliar mais evidentes nos gráficos selecionados. Tornou-se claro que a exploração visual dos dados se apresenta como uma vantagem competitiva, melhorando os resultados obtidos durante este processo (Gina & Budree, 2020).

Uma representação eficaz dos dados ajuda o utilizador a direcionar a sua análise e perceber as evidências escondidas na heterogeneidade dos dados. Não só na exploração e análise de dados, as representações gráficas são úteis na apresentação de um negócio, conceito ou ideia.

Uma das problemáticas mais relevantes do estudo da visualização de informação recai sobre a escolha das representações gráficas adequadas para cada tipo de dados. Associar dados a elementos gráficos e a uma localização no espaço pode ser uma tarefa difícil que requer conhecimentos em tipos de dados, análise de dados e representações gráficas (Pereira, 2015). Também o *data storytelling* se apresenta como um tema bastante preponderante na questão da apresentação de informação proveniente dos dados recolhidos, entrando em ação, frequentemente, a par da temática da visualização da informação.

Este projeto surge com a motivação de perceber qual o tipo de gráfico e representação mais adequada para cada tipologia de dados, com o qual a sua interpretação se torne mais fácil, eficaz e adaptada num cenário concreto. Pretende-se desenvolver um estudo de caso expondo as melhores práticas de visualização de dados, que proporcionem a melhor experiência ao utilizador. Também neste seguimento, e na mesma linha de pensamento será aplicada à temática do *data storytelling*, no qual serão avaliadas e analisadas quais as técnicas mais adequadas e eficazes para cada tipo de dados e relacionamento entre dados.

Uma representação visual e o respetivo *storytelling* não só ajudam o utilizador final a interpretar e entender um determinado fenómeno, como podem ofuscar o seu significado, impedindo que o utilizador retire informação válida. Neste trabalho estuda-se, adicionalmente, para o cenário em causa, como os utilizadores abordam e resolvem os seus problemas, percebendo quais são as representações visuais úteis para adicioná-las às técnicas e ferramentas de visualização, e quais as respetivas técnicas de *data storytelling*. Este estudo

será realizado por intermédio do desenvolvimento de *dashboards* específicos nos quais serão adicionados e organizados diversos elementos de visualização de dados, e, posteriormente, um conjunto de utilizadores utilizará esses *dashboards* e será testado quanto à usabilidade dos mesmos. Também no que respeita ao *data storytelling* será realizado um processo semelhante, ou seja, será desenvolvida uma história de dados, que será apresentada ao utilizador, e, posteriormente, este terá a oportunidade de responder a um questionário no qual avalia a sua experiência e relevância da história de dados para a sua interpretação e análise dos dados.

## 1.2. Objetivos gerais e específicos

Como objetivo geral deste estudo, pretende-se desenvolver um conjunto de *dashboards* que seja capaz de transmitir a informação presente numa base de dados considerada previamente, de forma que o utilizador final seja capaz de interpretar a informação apresentada e responder a um conjunto de questões definidas *a priori*. Durante este processo são selecionadas técnicas de visualização consideradas as mais adequadas para cada caso específico e, conseqüentemente, as técnicas de *storytelling* para contar a história dos dados.

Quanto à base de dados utilizada, esta será baseada no projeto UniBench<sup>1</sup> desenvolvido pela Universidade de Helsínquia, que consiste no desenvolvimento de um conjunto de modelos de dados mistos que imita uma aplicação de comércio social, abrangendo modelos de dados que incluem *JavaScript Object Notation* (JSON), *eXtensible Markup Language* (XML), valor-chave, tabular, entre outros. Importa referir que a base de dados alvo de estudo, é previamente facultada antes de se iniciar a abordagem para este projeto. O projeto UniBench não tem, há data, um sistema visual composto por um conjunto de *dashboards* que permitam a visualização dos seus dados e o contar da sua história.

Como objetivos específicos serão definidos vários tipos de abordagens para efetuar um estudo sobre viabilidade e adequabilidade de *dashboards* desenvolvidos:

- **Estudo de usabilidade:** aplicação de um questionário de usabilidade de forma aferir qual foi a experiência do utilizador ao interagir com os *dashboards* desenvolvidos, e aferição do tempo requerido para desenvolver um conjunto de tarefas específicas;

---

<sup>1</sup> <https://www.helsinki.fi/en/researchgroups/unified-database-management-systems-udbms/research/unibench-towards-benchmarking-multi-model-dbms>

- **Estudo de *Data Storytelling* (ou estudo da narração/história dos dados):** aplicação de um questionário de forma a entender de que forma a história de dados influenciou a análise dos dados e a interação do utilizador com os *dashboards*.

### 1.3. Metodologia

Para dar início ao trabalho desenvolvido considerou-se uma base de dados relacional, a qual servirá de suporte a todo o trabalho e servirá também de comparação com os passos que serão apresentados em fases posteriores do desenvolvimento.

Após efetuado o devido levantamento teórico, optou-se, em primeiro lugar, por desenvolver um conjunto de *dashboards* tendo sido selecionado o Power BI, uma vez que é uma ferramenta completa e de acesso livre, que apresenta as características necessárias ao desenvolvimento dos *dashboards* pretendidos. Por outro lado, considerou-se que a familiaridade com a supramencionada ferramenta iria permitir um tempo de desenvolvimento mais eficaz. Na secção 3.4. é possível verificar outras ferramentas que foram consideradas antes de selecionar o Power BI. Após desenvolvidos os *dashboards*, segue-se a fase de criação do *storytelling*, tendo este de ser adequado à informação apresentada e às questões a que se pretende dar resposta, que serão devidamente identificadas *a priori*.

Após o desenvolvimento de todos os *dashboards* necessários será iniciada a fase de avaliação. Nesta fase serão avaliados diferentes aspetos relativos à apresentação, usabilidade, funcionamento dos *dashboards*, *storytelling* e demais aspetos relacionados com o projeto UniBench.

Nesta fase, a plataforma desenvolvida será testada por alguns utilizadores que irão contactar com as funcionalidades da mesma e tentar cumprir os objetivos propostos pela equipa de investigação. Após a sua utilização será realizado um questionário com o objetivo de avaliar a usabilidade da plataforma e a relevância percebida do *storytelling* apresentado.

### 1.4. Estrutura do trabalho

O trabalho apresentado é constituído por 6 secções distintas, correspondentes a diferentes fases e processos constituintes do estudo em causa.

A primeira secção corresponde à Introdução, e é nesta que se introduz a temática em estudo através de uma abordagem inicial onde são apresentados conceitos como o enquadramento, objetivos gerais e específicos, métodos utilizados e estrutura do trabalho.

A segunda secção corresponde ao Estudo Prévio, na qual é efetuada uma revisão de literatura e análise da melhor evidência disponível, dos temas que se pretendem retratar, por forma a dar suporte à aplicação prática da temática que se irá pôr em aplicação em fases posteriores deste trabalho.

A terceira secção apresenta o Planeamento e Conceção do caso de estudo desenvolvido. Nesta fase são descritos todos os processos adotados desde a proveniência dos dados a utilizar, estrutura a desenvolver, definição do tipo de estudo, identificação da população-alvo, definição das questões de investigação, seleção das ferramentas a utilizar e planeamento dos *dashboards*.

Na quarta secção será apresentada a Implementação dos *dashboards* e a Conceção da história dos dados. No que respeita à Implementação dos *dashboards* serão especificados e justificados os tipos de elementos visuais a utilizar, como estarão organizados, qual a forma de navegação entre os vários painéis e que tipos de painéis existirão. Relativamente à Conceção da história de dados será exposto o processo de construção da história, disponibilizando um resumo dos tópicos apresentados na mesma.

A quinta secção apresenta os Testes de Validação de resultados, no qual serão apresentados os testes realizados e os resultados obtidos com os testes de usabilidade e *storytelling* realizados com os utilizadores da plataforma.

Na sexta, e última secção, são apresentadas a discussão de resultados e as conclusões retiradas deste estudo, com o seu devido enquadramento na revisão de literatura teórica, sendo efetuada, também, a relação dos resultados obtidos com os objetivos propostos inicialmente, por forma a verificar se estes foram cumpridos integralmente, tal como proposto.

## 2. Estudo prévio

Neste capítulo, a seguir apresentado, é realizada a contextualização dos temas envolvidos, estabelecendo um ponto de partida inicial para o desenvolvimento do trabalho proposto. São abordadas temáticas como a pirâmide do conhecimento, a visualização de dados, os elementos visuais, os gráficos de informação, os *dashboards* e o *data storytelling*. Desta forma, é apresentado o processo que decorre desde o contacto com os dados em questão e o desenvolvimento dos *dashboards* e a apresentação do *data storytelling*.

### 2.1. Dos dados à sabedoria

Segundo (Rowley, 2007), os dados podem ser definidos "como sendo factos ou observações discretos e objetivos, que não estão organizados e não são processados e, desta forma, não têm significado ou valor devido à falta de contexto e interpretação". Assim, torna-se relevante entender de que forma é possível evoluir desde o momento em que apenas existem dados no seu estado primordial até à aquisição da sabedoria que estes nos permitem obter. Desta forma, diversos autores defendem a existência de uma hierarquia denominada pirâmide DIKW, cuja sigla significa *Data-Information-Knowledge-Wisdom*, também conhecida por Pirâmide do Conhecimento, que se refere a uma classe de modelos para representar supostas relações estruturais e funcionais entre dados, informação, conhecimento e sabedoria.

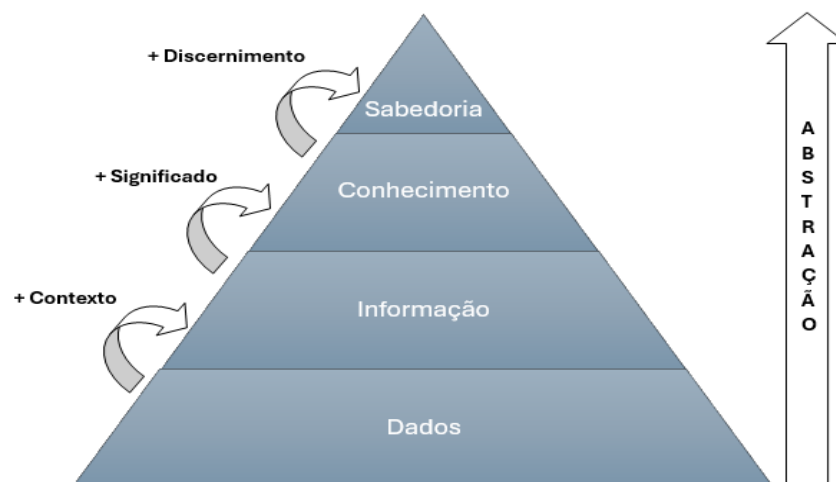


Figura 1 - Pirâmide do Conhecimento (adaptado de Rowley, 2007)

Como é possível observar na Figura 1, existem quatro conceitos distintos que constituem a Pirâmide do Conhecimento: Dados, Informação, Conhecimento e Sabedoria.

Os dados constituem o elemento primordial da Pirâmide do Conhecimento, e, como já referido anteriormente, quando apresentados de forma desorganizada e não processada não apresentam valor devido à sua falta de contexto. Assim, estes constituem o elemento com o menor nível de abstração que é apresentado na pirâmide.

A informação, no contexto da Pirâmide do Conhecimento, diferencia-se dos dados pelo facto de ser "útil". A informação é inferida a partir de dados, no processo de resposta a algumas perguntas (como por exemplo, "quem", "o quê", "onde", "quantos", "quando"), tornando os dados úteis para a tomada de decisão ou ação. Rowley (2007), descreve a informação como "dados organizados ou estruturados, que foram processados de tal forma que a informação tem agora relevância para um objetivo ou contexto específico e é, portanto, significativa, valiosa, útil e relevante" (Rowley, 2007).

A componente de conhecimento é geralmente aceite como um conceito abstrato e difícil de definir. Segundo o ponto de vista do DIKW, as definições do conhecimento podem referir-se ao facto de a informação ter sido processada, organizada ou estruturada de alguma forma, ou mesmo tendo sido aplicada ou posta em prática. Por outras palavras, o conhecimento pode ser definido como a compreensão humana de algum conceito com base na síntese da informação e experiência acumuladas (Van Meter, 2020). Tendo em conta esta definição, é importante ter em consideração que o conhecimento é um conceito coletivo, o que significa que se baseia na compreensão humana geral e não apenas na opinião ou crença de um indivíduo; este constitui uma síntese da informação recolhida e da experiência de vida, resultando da combinação destas duas vias, o que nos leva a inferir que a informação recolhida que não é confirmada pela experiência de vida pode não constituir conhecimento (Van Meter, 2020).

Embora seja comumente incluída como um nível da Pirâmide do Conhecimento, a referência à sabedoria é limitada nos debates sobre este modelo. Ackoff (1989) refere-se à compreensão como uma "apreciação do porquê", e a sabedoria como "compreensão avaliada", onde a compreensão é colocada como uma camada discreta entre o conhecimento e a sabedoria (Ackoff, 1989). Mais recentemente, Van Meter (2020) definiu a sabedoria como a aplicação de conhecimentos recolhidos para gerar uma compreensão da humanidade e da sociedade humana e do seu meio envolvente para orientar as ações e melhorar a vida

(Van Meter, 2020). A sabedoria envolve a utilização do conhecimento para um bem maior, sendo, por este motivo, a sabedoria mais profunda e exclusivamente humana. Requer um sentido do bom e do mau, do certo e do errado, do ético e do não ético.

Em suma, os dados não têm valor até serem processados numa forma utilizável, tendo em conta um contexto. A informação contém dados que foram objeto de algum tipo de organização e de análise sistemática. O conhecimento representa a informação que foi adquirida e utilizada, geralmente por um ser humano. A sabedoria é a posse de conhecimentos utilizados para estabelecer ligações inteligentes entre diferentes agentes e padrões necessários para compreender os princípios e os mecanismos subjacentes que regem o comportamento dos dados (Van Meter, 2020).

O nível mais baixo da pirâmide representa um menor nível de abstração, enquanto o nível superior representa o maior nível de abstração.

### ***Business Intelligence***

*Business intelligence* consiste no processo de recolha, análise e interpretação de grandes quantidades de dados para fornecer conhecimentos significativos e informações acionáveis que podem orientar a tomada de decisões estratégicas numa organização. Envolve a utilização de várias ferramentas, tecnologias e metodologias para recolher dados de fontes internas e externas, transformá-los em conhecimento valioso e apresentá-los num formato que seja facilmente compreensível e útil para os decisores (Bharadiya, 2023).

As ferramentas de *Business Intelligence* são definidas como o conjunto de *software* de aplicação utilizado para recolher, transformar e processar dados recolhidos a partir de fontes de dados organizacionais que permitem a realização de relatórios de negócio e melhorar o funcionamento da organização (Gina & Budree, 2020).

O objetivo do *Business Intelligence* é permitir que as organizações obtenham uma compreensão abrangente das suas operações, clientes, tendências de mercado e cenário competitivo. Este engloba uma série de atividades, incluindo a integração, extração e visualização de dados, a elaboração de relatórios e a monitorização do desempenho. Envolve frequentemente a utilização de *data warehouses* ou *data marts* para centralizar e armazenar dados estruturados e não estruturados, que podem depois ser analisados utilizando várias técnicas estatísticas e analíticas (Bharadiya, 2023). Os processos de *Business Intelligence* também envolvem cada vez mais uma efetiva representação gráfica dos elementos e a

comunicação de resultados de forma correta aos promotores dos projetos e a audiências alvo. Ressalva indicar que sem uma base sólida de dados devidamente integrados e limpos invalida os processos de transmissão de conhecimento obtidos.

## 2.2. Visualização de dados e Introdução aos elementos visuais

A visualização de dados está a tornar-se um componente cada vez mais importante na era da análise de dados. À medida que o volume e a variedade destes dados aumentam, a visualização dos mesmos torna-se ainda mais importante para instigar um diálogo colaborativo entre os grupos.

Lidar com dados pode conduzir, frequentemente, ao caos e a interpretações erróneas, sendo que, à medida que o volume destes dados se torna maior, os problemas de análise e interpretação podem tornar-se mais frequentes e de difícil resolução. O agrupamento visual de muitos pontos de dados melhora significativamente a qualidade da análise e proporciona uma forma muito conveniente para compreender as relações estabelecidas. A interpretação visual, por norma, ajuda a compreender a informação e a decidir rapidamente onde concentrar o esforço de investigação (Medeiros & Maçada, 2022).

### 2.2.1. Fundamentos da Visualização de Informação

A essência da visualização de informação encontra-se na relação entre o observador e a sua interpretação do conteúdo. Esta é uma forma de comunicação imediata e intuitiva, que se baseia na representação de símbolos.

No seu livro *“The Visual Display of Quantitative Information”*, Tufte explica os fundamentos da representação visual de informação. Desta forma, destacam-se os 7 princípios fundamentais que resumem esta temática (Tufte, 2007):

- **Remoção da desordem** – a informação é o mais importante, isto é, tudo aquilo que não esteja associado com a sua leitura deve ser eliminado;
- **Desconstrução de conceitos complexos** - comunicar ideias elaboradas com clareza pode ser moroso, pelo que a sua divisão em parcelas menores poderá ajudar na exposição da informação;
- **Transmissão dos conceitos de forma simples** - as ideias devem ser transmitidas da forma mais simples possível. Quantos mais recursos forem usados mais complexa a mensagem tende a tornar-se, pelo que a simplicidade ajuda não só a clarificar a informação, mas também a tornar a interpretação mais imediata;

- **Evitar distorção** - os meios devem adaptar-se à informação e não o oposto. A escolha do estilo de grafismos usados deve ser feita de forma a melhor representar o tipo de informação em questão. Não devemos negar os diferentes usos de cada género de representação nem devemos segregar informação ao género de representação;
- **Limitação do esquema de cores** – um esquema de cores reduzido e distinto ajuda não só na simplificação da comunicação, mas também serve para garantir que o leitor consegue identificar as diferentes cores usadas. Esquemas maiores tendem a ter combinações de cores muito semelhantes que podem criar confusão;
- **Garantir que o leitor se foca no conteúdo e não na forma** - a componente estética deve sempre servir a mensagem a transmitir, devendo guiar o leitor na direção da informação e não chamar a atenção em relação a si mesma;
- **Precisão** - a precisão é um parâmetro fundamental quando se trata de comunicar visualmente informações mensuráveis. Escalas, relações e proporções devem ser apresentadas com rigor para evitar falsas leituras.

Os dados apresentados de forma visual tornam menos exigente, para os indivíduos, a compreensão daquilo que um orador está a tentar transmitir.

Alguns dos princípios que são seguidos para criar uma visualização adequada são (Anuncia et al., 2020):

- **Ser direto e conciso** - a clareza é a parte mais simples, mas também a mais complexa da visualização. Por vezes, as informações adicionais acrescentam detalhe, mas podem tornar as tabelas ou gráficos mais difíceis de interpretar;
- **Utilização de ferramentas adequadas** - utilizar software de *Business Intelligence* adequado e utilizar software adequado para um público diversificado;
- **Narração da história** - deve utilizar-se texto apenas quando se torna relevante clarificar algum conceito que esteja a ser visualizado;
- **Recurso da ordenação** - os temas apresentados devem ser abordados de forma ordenada, consoante o seu carácter de relevância, devendo ainda ser apresentados de forma cronológica, caso isto se demonstre adequado à temática em questão;
- **Escolha dos melhores objetivos** - cada perceção ou tema tem o seu objetivo próprio, que se associa ao objetivo central;

- **Escolha das cores** - as cores mais vibrantes, como o vermelho, o amarelo, o laranja e o roxo são, por norma, as que sobressaem mais nos gráficos, podendo assim ser utilizadas para realçar elementos de destaque.

Ao abordar o conceito da palavra visualizar, esta representa o processo de tornar algo visual ou visível, enquanto o conceito de visualização pode ser definido como a transformação de conceitos abstratos em imagens reais ou mentalmente visíveis. No contexto da computação, visualizar é a conversão de números ou categorias para um formato gráfico que pode ser facilmente compreendido (Spence, 2009).

A partir de uma representação visual facilmente se identifica uma ideia geral a partir de um conjunto de dados, deduzida por meio de conhecimento do padrão visual, ideias estas que a partir de uma estrutura desorganizada de dados não são facilmente identificadas.

A percepção visual é considerada como sendo o produto final da visão, e consiste na capacidade de interpretação da informação transmitida por um estímulo luminoso. Esta relaciona-se com a visualização de informação uma vez que o intuito da mesma é criar representações que estimulem as propriedades de percepção do ser humano. A percepção pode ser influenciada em função das características físicas do estímulo.

Segundo a teoria de Gestalt (Chang et al., 2002) a percepção relaciona-se com um fator básico de forma e estabilidade e pode ser aplicada através dos seguintes princípios:

- **Pregnância** - considera-se que antes de se observar uma parte específica, analisa-se inicialmente o todo;
- **Semelhança** - todos os elementos que possuam características semelhantes (forma, cor, etc.) tendem a ser agrupados;
- **Proximidade** - elementos que se encontrem a uma proximidade reduzida dentro do mesmo espaço temporal tendem a ser perceptualmente agrupados;
- **Continuidade** - este princípio realça a importância da fluidez numa estrutura. Funciona como um instinto de seguir uma certa direção que é sugerida pelo campo visual aquando da observação, ou seja, um elemento que possuir um padrão contínuo, sem interrupções, possui uma boa compreensão e comunicação dos dados;
- **Closure** - o cérebro humano ao observar uma figura que possua formas inacabadas, tende a realizar o fecho dessas ligações através da análise à sua continuidade;

- **Unidade** - este princípio enuncia a capacidade humana de preencher os espaços vazios e visualizar o item como um todo;
- **Segregação** - realça a capacidade do cérebro humano em destacar as unidades de um todo.

A visualização da informação explora principalmente o sentido humano que possui maior afinidade para captação de informação temporal, isto é, a visão. Ainda assim, considerando a visualização como sendo um processo mental, outros fatores da percepção humana podem ser usados de forma a contribuir neste processo, auxiliando os procedimentos cognitivos humanos na recuperação da informação contida nas imagens criadas a partir dos dados (Jiang et al., 2023).

A visualização, para além de condensar os dados, permite gerar novas informação com maior facilidade, mesmo com dados relativamente elementares. De facto, a simples representação de uma tabela permite o levantamento de questões relativas aos mesmos. Para responder a estas questões com base apenas numa tabela, existe uma certa dificuldade, mas com gráficos circulares ou de linhas, pode resolver-se este problema de forma simples, para além da possibilidade de se realizarem novas inferências sobre os dados (Jiang et al., 2023).

Segundo Ware (2004), algumas das vantagens comumente associadas à visualização de informação são (Ware, 2019):

- Promove a capacidade de compreender grandes volumes de dados, disponibilizando de imediato as informações importantes;
- Permite a percepção de características que não são antecipadas apenas com os dados originais;
- Permite a identificação rápida de erros ou anomalias presentes nos dados;
- Facilita a compreensão de características dos dados quer em grande, quer em pequena escala;
- Facilita a formulação de hipóteses.

### **2.2.2. Elementos Visuais**

As representações visuais ou os modelos gráficos variam de acordo com a tipologia dos dados, que podem representar conceitos abstratos, valores ou relacionamentos do mundo físico em associação com um espaço temporal ou geográfico. A capacidade de visualizar na totalidade um conjunto de dados mapeado numa representação é essencial para descobrir

possíveis padrões e relações interessantes entre os dados. A visualização de dados tem evoluído como a abordagem que transforma conjuntos complexos de dados em informações úteis para o utilizador através de um meio rápido e apelativo. O método básico do utilizador é gerar representações visuais interativas de informação que exploram capacidades perceptuais do sistema visual humano e os recursos interativos da solução de problemas cognitivos (Anuncia et al., 2020). A conjugação entre o domínio da natureza de dados abstrata e o domínio das representações gráficas denomina-se mapeamento.

O mapeamento não é uma tarefa linear, uma vez que as representações visuais geradas nem sempre são claras. Ao longo do tempo, vários modelos têm sido propostos para definir as estratégias de mapeamento, podendo estes ser divididos nas seguintes categorias (Carvalho & Marcos, 2009):

- *Blind Matching Procedure* – os dados são separados em elementos. É executado o cruzamento de elementos de dados e dos atributos visuais. O número total de combinações possíveis de se obter entre os atributos é definido por:  $(n > m): n! / (n - m)!$ , onde  $n$  é o número de características existentes nos dados e  $m$  o número de atributos visuais;
- *Renaissance teams* – propõe um grupo formado por especialistas em visualização e na área destino da aplicação. Apresenta um custo elevado;
- *Bottom-up* – as imagens são geradas a partir das primitivas visuais ou uma linguagem gráfica. Depois as imagens são verificadas com um conjunto de condições (caraterísticas dos dados, ambiente computacional, utilizador, realidade, efetividade e expressividade);
- *Generate-and-test* – Consiste em testar as representações visuais que tenham sido geradas a partir do modelo anterior;
- *Top-down* – Permite somente técnicas complexas e coerentes. Escolhe a técnica que melhor se encaixe;
- *Code based reasoning* – é um procedimento que escolhe as representações visuais a partir de um conjunto de representações (consideradas) bem-sucedidas. É utilizado normalmente, após a aplicação do modelo anterior (Top-down).

As representações visuais contemplam um poderoso recurso, auxiliando o mapeamento entre os dados e um conjunto de propriedades gráficas com caraterísticas específicas definidas.

Através de variáveis visuais, substrato espacial e marcas visuais é possível criar uma imagem exploratória e representativa.

A visualização de dados permite aos indivíduos compreenderem os dados através de uma representação visual. A visualização corresponde à apresentação gráfica da informação, com o objetivo de proporcionar ao observador uma compreensão qualitativa do conteúdo da informação que se apresenta (Anuncia et al., 2020).

Para que esta compreensão efetivamente ocorra, durante o processo de visualização, devem ser consideradas as melhores técnicas e proceder à sua correta aplicação, uma vez que a sua utilização de forma inadequada gera resultados incorretos ou inconclusivos. De forma a evitar esses erros é necessário conhecer a técnica de visualização que irá conduzir a uma representação gráfica adequada para os dados.

Para que a seleção das melhores técnicas ocorra de forma adequada, torna-se necessário analisar previamente a tipologia de dados presentes, bem como as suas características, e assim, definir as diretrizes que levam à construção da representação visual. Para a definição destas mesmas diretrizes, segundo Daniel Keim (2002) é possível ter por base cinco classes de técnicas distintas (Keim, 2002):

- **Gráficos padrão 2D e 3D (*Standard 2D/3D display*)** - frequentemente utilizados no âmbito da estatística para alcançar uma estimativa, a distribuição ou a frequência de um atributo no conjunto de dados. Esta classe fornece gráficos para fazer comparações, classificações e determina a ligação entre atributos, sendo aplicados para a representação de dados quantitativos. Gráficos estatísticos são usados na análise de dados para descobrir padrões, estruturas nos dados e identificar *outliers*.
- **Técnicas de visualização geométrica (*Geometrically-transformed display*)** - têm por objetivo encontrar transformações “interessantes” em conjuntos de dados multidimensionais, mapeados num plano bidimensional. Os gráficos representados por estas técnicas são adotados para representar dados qualitativos e quantitativos. Gráficos que permitem uma vista pormenorizada ou global sobre todos os dados e que verificam a correlação entre atributos, padrões, comportamentos e regras são uma opção eficaz para grandes conjuntos de dados.
- **Técnicas iconográficas (*Iconic display*)** - a ideia é mapear os valores dos atributos de um item de dados multidimensional para as características de um determinado ícone. Aqui, os atributos são mapeados de acordo com propriedades que variam de

acordo com o valor dos próprios atributos. Gráficos de ícones representam cada dado individual o que permite perceber o comportamento de padrões dos dados. Os ícones utilizados podem ser definidos de forma arbitrária: podem ser pequenos rostos, ícones de estrela, *stick figure*, ícones coloridos, entre outros.

- **Técnica de pixéis densos (*Dense pixel display*)** - a ideia das técnicas de pixéis densos é mapear cada dimensão a um pixel colorido e agrupar os pixéis pertencentes a cada dimensão em áreas adjacentes. As técnicas de pixéis densos utilizam diferentes disposições para diferentes fins, pois ao dispor os pixéis de uma forma adequada, a visualização resultante fornece informação detalhada sobre correlações locais, dependências e *hot spots*. Dois dos exemplos mais conhecidos são:
  - Técnica do padrão recursivo: baseia-se num arranjo recursivo genérico dos pixéis e destina-se a representar conjuntos de dados com uma ordem natural de acordo com um atributo (dados de séries temporais, por exemplo). O utilizador pode especificar parâmetros para cada nível de recursão e controlar assim a disposição dos pixéis para formar subestruturas semanticamente significativas.
  - Técnica dos segmentos de círculo: representação dos dados num círculo dividido em segmentos, um para cada atributo. Dentro dos segmentos, cada valor de atributo é novamente visualizado por um único pixel colorido. A disposição dos pixéis começa no centro do círculo e continua para o exterior. A lógica desta abordagem é que, perto do centro, todos os atributos estão próximos uns dos outros, melhorando a comparação visual dos seus valores.
- **Técnica de visualização empilhada (*Stacked display*)** – técnica concebida para apresentar dados particionados de forma hierárquica, apresentando, tipicamente, uma estrutura de relação entre os dados. No caso de dados multidimensionais, as dimensões de dados a utilizar para particionar os dados e construir a hierarquia têm de ser selecionadas adequadamente. Uma estratégia recorrente diz respeito à escolha inicial das dimensões mais relevantes para o caso de estudo. Podem ser representadas por mapas de árvore, gráficos de mosaicos ou outros.

Para além das técnicas de visualização, para a exploração de dados, é necessário utilizar algumas técnicas de interação e distorção. As técnicas de interação permitem interagir e alterar dinamicamente as visualizações de acordo com os objetivos de exploração e permitem também relacionar e combinar múltiplas visualizações independentes. As técnicas

de distorção ajudam no processo de exploração de dados fornecendo meios de enfoque nos pormenores, preservando uma visão global dos dados (Keim, 2002). O objetivo destas técnicas é mostrar partes dos dados com um elevado nível de pormenor, enquanto outras são apresentadas com um nível inferior. Cada uma destas técnicas é apresentada, de seguida, com uma maior nível de detalhe (Keim, 2002):

- ***Dynamic Projections*** – consiste na alteração dinâmica das projeções de modo a explorar um *dataset* multidimensional. O número de projeções possíveis é exponencial no número de dimensões, ou seja, é intratável para uma dimensionalidade. A sequência de projeções apresentada pode ser aleatória, manual, pré-computada ou baseada em dados;
- ***Interactive Filtering*** – na exploração de grandes conjuntos de dados, é importante dividir o *dataset* em segmentos e incidir em subconjuntos interessantes. Isto pode ser feito através de uma seleção direta do subconjunto (navegação) ou por uma especificação das propriedades do mesmo (consulta). Porém, a navegação é muito difícil para grandes conjuntos de dados e a consulta muitas vezes não produz os resultados desejados. Por conseguinte, foram desenvolvidas várias técnicas de interação com vista melhorar a filtragem interativa, como por exemplo a *Magic Lenses*. Esta ferramenta usa uma lupa para apoiar a filtragem dos dados, diretamente na visualização, em que os dados sob a lupa são processados pelo filtro e o resultado é apresentado de forma diferente do que o restante conjunto de dados;
- ***Interactive Zooming*** – para efeitos de visualização, torna-se importante apresentar os dados de forma comprimida para fornecer uma visão geral dos dados, mas, ao mesmo tempo, permitir uma visualização variável dos dados em diferentes resoluções. O zoom não significa apenas ver os objetos de dados num tamanho superior, mas também que a representação dos dados mude automaticamente para apresentar mais detalhes em níveis de zoom mais elevados. A visualização inicial permite ao utilizador detetar padrões, correlações e valores atípicos no conjunto de dados, e, para explorar uma região de interesse, o utilizador pode aumentar o zoom, o que resulta que as linhas (ou colunas) afetadas sejam apresentadas com mais pormenor, possivelmente até em forma de texto;
- ***Interactive Distortion*** – estas técnicas apoiam o processo de exploração de dados, preservando uma visão geral dos mesmos durante as operações de pesquisa. O conceito prende-se com a apresentação de partes dos dados com um elevado nível

de pormenor, enquanto outras são apresentadas com um nível inferior de pormenor. As técnicas de distorção mais populares são as distorções hiperbólicas e esféricas, que são frequentemente utilizadas em hierarquias ou gráficos, mas também podem ser aplicadas a qualquer outra técnica de visualização;

- ***Interactive Linking and Brushing*** - o objetivo desta técnica é combinar diferentes métodos de visualização para ultrapassar as debilidades das técnicas individuais. Os gráficos de dispersão de diferentes projeções, por exemplo, podem ser combinados colorindo e ligando subconjuntos de pontos em todas as projeções. De forma semelhante, a *linking* e a *brushing* podem ser aplicadas a visualizações geradas por todas as técnicas descritas anteriormente. Como resultado, os pontos *brushing* são destacados em todas as visualizações, tornando possível detetar dependências e correlações. A conexão de várias visualizações através da *linking* e *brushing* interativas fornece mais informações do que apenas considerar as visualizações de componentes de forma independente.

As técnicas de representação apresentadas são classificadas com base nos critérios tipo de dados, técnica de visualização e a interação e técnica de distorção. Na Figura 2 é possível visualizar um referencial que correlaciona as grandezas utilizadas para a classificação das técnicas de visualização, bem como os constituintes de cada uma.

As variáveis visuais pretendem descrever as dimensões gráficas através das quais uma representação pode ser variada para codificar informação. Ao reduzir a representação aos seus elementos gráficos constituintes, os símbolos pouco claros podem ser identificados e redesenhados, melhorando a comunicação cartográfica. Esta desconstrução revela os juízos de valor e as relações de poder implícitas no processo de criação. As variáveis foram originalmente descritas pelo cartógrafo e professor francês Jacques Bertin no livro de 1967 “*Semiologie Graphique*”. A tradução inglesa, “*Semiology of Graphics*”, foi lançada em 1983 e até hoje é reconhecido como um trabalho de referência, tanto na área da cartografia como no domínio da visualização de informação (Bertin, 2011).

Originalmente, Bertin identificou 7 variáveis visuais (Bertin, 2011):

- **Localização** - descreve a posição do símbolo cartográfico em relação a um quadro de coordenadas. Esta é considerada uma variável visual "indispensável" e tem primazia visual sobre as outras, sendo usada principalmente para representar a componente espacial da informação no desenho cartográfico. No entanto, a

localização pode ser utilizada para representar informação de atributos, tal como a utilização da altura em perspetiva nos mapas de prisma. Em representações planimétricas, 2-D, os símbolos cartográficos perto do centro ótico tendem a aparecer, resultando numa interpretação visual mais imediata por parte do utilizador do mapa e, assim, colocá-los num lugar de maior importância na "hierarquia visual". Em contraste, os símbolos do mapa perto da periferia da página tendem a recuar para o fundo da hierarquia visual. Nas representações oblíquas ou tridimensionais, símbolos cartográficos que são percebidos como mais próximos ou mais altos tendem a sobressair;

- **Tamanho** - descreve a quantidade de espaço ocupado pelo símbolo do mapa, constituindo a principal variável visual manipulada em mapas de símbolos e cartogramas. Os símbolos maiores tendem a sobressair na figura;
- **Forma** – descreve a forma externa (isto é, o contorno) do símbolo. A forma é essencial para a conceção dos símbolos de pontos qualitativos normalmente utilizados na cartografia de referência. As formas destes símbolos cartográficos podem variar de muito abstratos, como círculos, quadrados ou triângulos, a altamente icónica, imitando diretamente o referente representado pelo símbolo do mapa. Os símbolos cartográficos que são mais complexos ou menos compactos tendem a sobressair na figura;
- **Orientação** – descreve a direção ou rotação do símbolo em relação à "normal". A orientação normal é relativa à linha de contorno do mapa, mas em alguns casos pode ser relativa ao sistema de coordenadas espaciais projetado (por exemplo, em relação à grelha) ou outra linha de base. Os símbolos de glifos multivariados utilizam a orientação para diferenciar os atributos representados. A orientação também é manipulada em mapas de fluxo para representar a direccionalidade do fluxo. Os grupos próximos de símbolos que têm a mesma orientação (alinhados ou desalinhados em relação ao normal) tendem a aparecer como um único grupo. De forma oposta, elementos individuais do mapa que estão desalinhados em relação à normal tendem a sobressair na figura;
- **Tonalidade (cor)** - descreve o comprimento de onda dominante do símbolo cartográfico visível no espectro eletromagnético (por exemplo, azul, verde, vermelho) e é uma das três variáveis visuais associadas à percepção da "cor". Um esquema de cores qualitativo ou espectral manipula a tonalidade da cor enquanto controla os seus

outros componentes; como se explica de seguida, estes esquemas de cores são apropriados para mapas coropléticos nominais. Os símbolos vermelhos do mapa tendem a sobressair, enquanto os símbolos azuis tendem na direção oposta;

- **Valor da cor** - descreve a quantidade relativa de energia emitida ou refletida pelo símbolo, também denominada de “luminosidade”. A variação do valor da cor resulta na percepção de sombreado, sendo importante para mapas coropléticos que apresentam informação ordinal ou numérica. Um esquema de cores sequencial ajusta o valor da cor numa direção, cruzando por vezes duas ou três diferentes para melhorar a discriminação. Um esquema de cores divergente ajusta o valor da cor em duas direções, afastando-se de um ponto médio, sendo cada direção denotada por alterações no valor da cor dentro de uma tonalidade de cor diferente. A relação figura-fundo para o valor da cor é relativa à quantidade de áreas claras e escuras no mapa. Os símbolos escuros tendem a sobressair em mapas que são maioritariamente claros, enquanto os símbolos claros tendem a aparecer em mapas que são maioritariamente escuros;
- **Textura** - descreve o grau de aspereza do padrão de preenchimento do símbolo. A textura pode ser descrita como uma variável visual de dimensão superior, composta por três componentes constituintes: a direccionalidade das unidades de textura (relacionada com a orientação da variável visual), o tamanho das unidades de textura (relacionada com o tamanho da variável visual) e a densidade das unidades de textura (que se aproxima do efeito percetivo de sombreado associado à variável visual do valor da cor). Relativamente a este último componente, os símbolos com uma textura mais densa tendem a sobressair na figura.

O conjunto de variáveis visuais definido por Bertin foi alargado por Morrison (1974) para incluir duas variáveis adicionais (Morrison, 1974):

- **Saturação da cor** - descreve o pico espectral no espectro visível, e é a terceira de três variáveis visuais associadas à percepção da cor. As cores fortes ou saturadas emitem ou refletem energia concentrada do espectro visível, enquanto as cores pastel ou pouco saturadas emitem ou refletem energia de forma uniforme em todo o espectro visível. Do ponto de vista do design, o valor da cor pode ser conceptualizado como a quantidade de preto num símbolo, enquanto a saturação da cor pode ser conceptualizada como a quantidade de cinzento num símbolo. A saturação também pode ser designada como "intensidade" e "pureza" na teoria da cor. Símbolos

saturados e em negrito tendem a sobressair na figura, enquanto os símbolos pouco saturados tendem na direção oposta;

- **Organização** – descreve a disposição das marcas gráficas que constituem um símbolo. A disposição das variáveis visuais varia de regular (as marcas gráficas estão perfeitamente alinhadas numa estrutura semelhante a uma grelha) a irregular (as marcas gráficas são colocadas aleatoriamente ou agrupam-se em grupos). A disposição difere da variável visual textura na medida em que se assume que todas as texturas são dispostas regularmente, independentemente da direção inicial, tamanho e densidade da textura. Símbolos com disposições irregulares, e agrupados de forma específica, tendem a sobressair.

Em 1995, MacEachren em (MacEachren, 1995) identificou três variáveis visuais adicionais cuja manipulação é facilitada pelos métodos de produção digital. Inicialmente estas três variáveis visuais foram agrupadas numa única técnica chamada "foco", mas acabou por ser reconhecido que cada componente corresponde a uma variável visual, dada a sua potencial aplicação a outras formas de representação:

- **Nitidez** – descreve a nitidez/difusão do limite do símbolo. Em (MacEachren, 1992) é referido que a nitidez é a variável visual mais eficaz para representar a incerteza no contexto da simbolização de pontos. Os símbolos com um limite nítido tendem a sobressair na figura enquanto os símbolos com um limite difuso tendem na direção oposta;
- **Resolução** – descreve a precisão espacial com que um símbolo é apresentado, isto é, descreve a remoção significativa de pormenor no desenho do gráfico, uma vez que a complexidade do mundo real é omitida para se adaptar à escala reduzida do gráfico. A resolução como variável visual utiliza diferentes níveis de abstração para codificar informação, em vez do objetivo típico de generalização. Os símbolos que são representados com um nível de pormenor relativamente elevado tendem a sobressair;
- **Transparência** – descreve a quantidade de mistura gráfica entre um símbolo e o fundo ou os símbolos subjacentes. Em (MacEachren, 1992) refere-se originalmente a transparência como "nevoeiro" para sugerir uma barreira parcialmente opaca com impacto que afeta a clareza dos símbolos subjacentes. Símbolos opacos tendem a sobressair na figura.

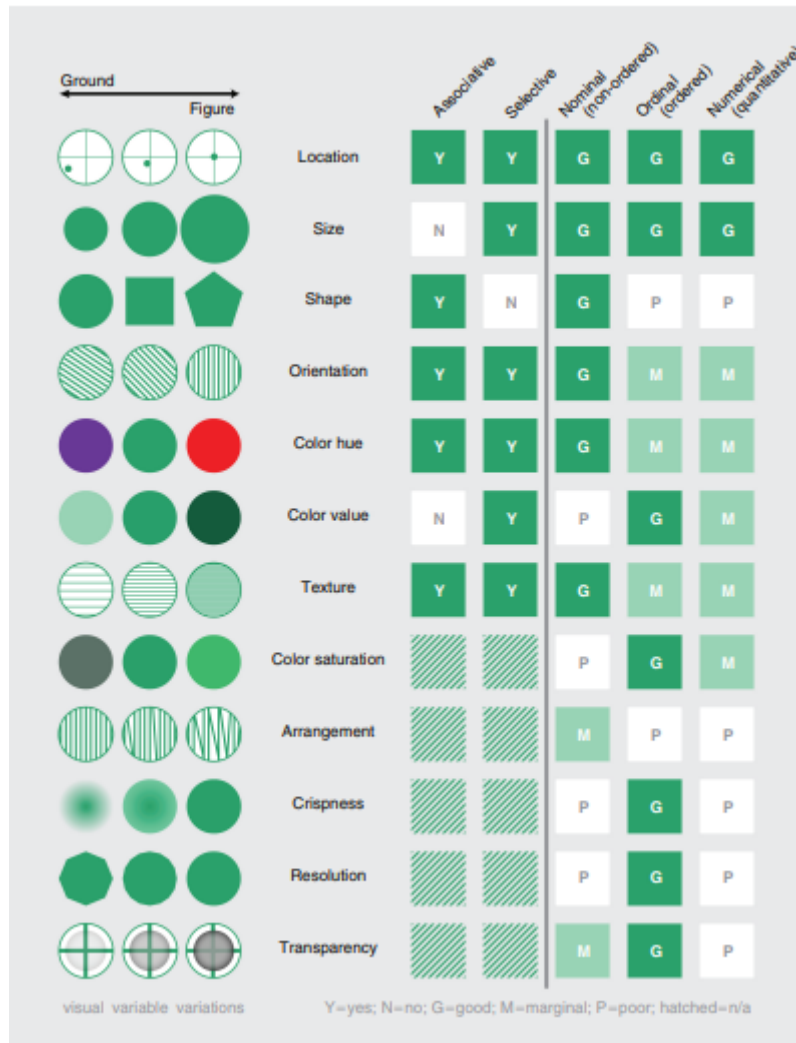


Figura 2 - Variáveis visuais de MacEachren (extraído de Roth, 2017)

Os princípios da psicologia da percepção preveem a forma como cada variável visual é processada pelos sistemas visual e cerebral e, por conseguinte, informam a utilização de uma variável visual, em detrimento de outras na representação da informação. Bertin (2011) apresentou quatro formas pelas quais a percepção visual pode informar a aplicação das variáveis visuais em mapas e outras visualizações, descrevendo estas quatro propriedades como "níveis de organização" (Bertin, 2011).

O primeiro nível de organização de Bertin é a percepção **associativa**. Com uma variável visual associativa, as variações da dimensão visual são percebidas com o mesmo peso, permitindo que o olho perceba todos os símbolos com a mesma variação, como sendo um grupo. Por exemplo, o olho não é atraído para uma tonalidade de cor sobre a outra ou para uma forma em detrimento de outra, permitindo a percepção de todos os símbolos com a mesma tonalidade de cor ou forma como um grupo associado. Como não existe nenhuma variação dominante, uma variável visual associativa permite ao olho atender visualmente a

outras variáveis visuais que também podem variar simultaneamente. Bertin defendia que a localização, forma, orientação, tonalidade da cor e textura como sendo variáveis visuais associativas. Com uma variável visual dissociativa, uma variação domina a percepção visual, sendo o olho atraído para essa variação, em detrimento das demais. Por exemplo, o olhar é atraído para valores de cor mais escuros, quando em contraste com um fundo branco, e para os tamanhos maiores. A variação numa variável visual dissociativa inibe a atenção para outras variáveis visuais. Desta forma, o tamanho e o valor da cor constituem variáveis visuais dissociativas.

O segundo nível de organização é a percepção **seletiva**. Com uma variável visual seletiva, o olho é capaz de se concentrar individualmente em cada variação da variável visual e ignorar as outras variações. Por outras palavras, é relativamente fácil isolar visualmente a distribuição de uma determinada categoria de símbolos quando caracterizada por uma variável visual seletiva. Por exemplo, é mais fácil ver a distribuição de símbolos vermelhos num conjunto de símbolos de várias cores do que a distribuição de símbolos triangulares num conjunto de símbolos com várias formas e a mesma cor, apesar de ambos os conjuntos codificarem a mesma informação. Bertin acreditava que a forma era a única variável visual que não era seletiva. Esta propriedade torna a forma útil quando cada símbolo deve ser interpretado individualmente, como no contexto de grandes conjuntos de símbolos de pontos qualitativos. No entanto, a forma não é útil quando se procuram padrões gerais, um objetivo importante da maioria das representações.

O terceiro nível de organização é a percepção **ordenada**, enquanto o seu quarto nível de organização é a percepção **quantitativa**. As variações nas variáveis visuais ordenadas são percebidas como ordenadas, com o olho a interpretar uma variação como "superior" ou "inferior" relativamente a outra variação. Para Bertin, a localização, o tamanho, a cor e a textura são consideradas variáveis visuais ordenadas; MacEachren argumentou mais tarde que a saturação da cor, a nitidez, a resolução e a transparência também são variáveis visuais fortemente ordenadas e que a textura é apenas marginalmente ordenada. A percepção quantitativa alarga a percepção ordenada, permitindo a estimativa de valores numéricos a partir de variações em variáveis visuais quantitativas (MacEachren, 1995). Bertin defendia que a percepção quantitativa se restringia apenas à localização e ao tamanho.

É importante notar que as variáveis visuais devem ser corretamente adequadas ao nível de medição da informação do atributo. Variáveis visuais não ordenadas - como a tonalidade da

cor, a orientação e a forma - são adequadas para codificar informação nominal. As variáveis visuais que são ordenadas, mas não quantitativas - como o valor da cor, saturação da cor, nitidez, resolução e transparência - são adequadas para codificar de forma ordinal. Variáveis visuais que são quantitativas - como a localização e o tamanho - são apropriadas para codificar informação numérica, mas também podem ser aplicadas à informação ordinal e nominal, dada a sua dominância visual. A Figura 2 apresenta as variáveis visuais de MacEachren, designando cada variável visual como boa, marginal ou má, adequando-a aos níveis nominal, ordinal, e numérico de medição da informação do atributo (Roth, 2017).

### 2.2.3. Gráficos de Informação

Com vista a explorar, analisar e comunicar dados, as representações visuais apresentam-se como um método eficaz na transmissão desta informação, uma vez que aumentam a capacidade de entendimento e absorção da informação. O avanço das tecnologias e ferramentas gráficas, impulsionaram a diversidade de modelos gráficos que o utilizador pode adotar como recurso. Através da exploração de propriedades gráficas, surgem vários tipos de representações visuais, desde tabelas, diagramas, gráficos e mapas (Bertin, 2011).

Desta forma, podem ser analisados vários tipos de representação, pelo que relacionar o propósito da visualização com as características gráficas surge como ponto de partida para a descrição dos tipos de representação gráfica que surgem de seguida (Pereira, 2015). Deste variado conjunto foram classificadas cinco principais classes de representações: gráficos, diagramas, tabelas, *charts* e mapas que se passam a descrever. Em cada classe são indicadas e descritas as metáforas visuais associadas.

**Tabelas** – Linhas ou colunas dispostas por uma grelha geométrica que contêm dados ou imagens. Habitualmente usadas quando se pretende analisar ou comunicar uma informação precisa e fácil de referenciar (Ribbecca, 2024);

- Gráfico de contagem: Categorias, valores, ou intervalos são colocados num eixo ou coluna (normalmente eixo vertical), e cada vez que um valor aparece, uma marca de registo acontece. Quando todos os dados estão representados, exhibe-se o total de ocorrências;
- Tabela de Tempo: Ferramenta usada para organizar e exibir eventos, atividades ou marcos em uma sequência temporal. Ajuda na visualização e gestão de atividades ao longo do tempo, fornecendo uma visão clara de quando cada evento ocorre e como

eles se relacionam entre si. Organizar dados numa tabela com uma ordem cronológica ou em ordem alfabética permite aos utilizadores uma consulta rápida (exibição de datas e horários de partidas/chegadas de transportes públicos, por ex.).

**Gráficos** – Apresentação de variáveis quantitativas e qualitativas como se fossem coordenadas, num sistema de referência bem definido, tendo por base um eixo vertical ou horizontal (Ribecca, 2024);

- **Gráfico de Linhas:** Indicados para exibir dados quantitativos ao longo de um período temporal, tornando-se adequados para revelar tendências e relações ao longo do tempo;
- **Gráfico de Área:** Semelhantes aos gráficos de linhas, mas com a área por baixo das linhas preenchida por uma cor ou textura. São utilizados para acompanhar o desenvolvimento de dados quantitativos ao longo de um período temporal, sendo úteis para exibir tendências e relações, uma vez que transmite valores específicos;
- **Gráfico de Barras:** Podem ser constituídos pelo uso de barras horizontais ou barras verticais. Adequados para exibir comparações entre categorias, uma vez que um eixo representa valores quantitativos e o outro representa dados com categorias específicas;
- **Caixa de Bigodes:** Representação gráfica que resume a distribuição de um conjunto de dados. Este exibe a mediana, quartis, e valores extremos de uma forma que facilita a visualização da dispersão e da simetria dos dados. Muito útil para identificar a centralidade, variabilidade, e a presença de *outliers* num conjunto de dados, além de permitir comparações entre diferentes distribuições;
- **Gráfico de Dispersão:** Exibe a relação entre duas variáveis numéricas. É composto por um conjunto de pontos, onde cada ponto no gráfico representa um par de valores, um de cada variável. Apresenta-se útil para identificar a correlação entre as duas variáveis analisadas. Se os pontos formam um padrão ascendente ou descendente, isso pode indicar uma correlação positiva ou negativa, respetivamente. Indicado para ajudar a visualizar a distribuição dos dados e a identificar possíveis *outliers*.
- **Gráfico de Bolhas:** Corresponde a uma extensão do gráfico de dispersão que permite visualizar três variáveis num único gráfico, no qual o tamanho de cada ponto (bolha) corresponde a uma terceira variável. Ideal para exibir dados que envolvem três dimensões, facilitando a análise comparativa de várias variáveis simultaneamente.

Útil para comparar e identificar padrões ou *outliers*, onde o tamanho das bolhas pode indicar a importância ou magnitude de uma observação específica;

- **Histograma:** Gráfico utilizado para representar a distribuição de um conjunto de dados contínuos ou discretos, mostrando a frequência com que diferentes intervalos de valores ocorrem num conjunto de dados. Útil para visualizar a distribuição dos dados, permitindo identificar padrões como simetria, assimetria, picos e dispersão, sendo ideal para visualizar a forma da distribuição dos dados em torno de certas classes de dados;
- **Gráfico de Mosaicos:** Ideal para exibir informação categórica através de várias variáveis. Ambos os eixos contêm um conjunto de variáveis, determinadas pela largura e altura de cada segmento. Ideal para revelar relações entre as categorias e subcategorias, revelando a proporção de cada categoria de dados;
- **Gráfico de Barras Agrupado:** Tipo de gráfico utilizado para comparar múltiplas categorias entre diferentes grupos. Organiza as barras em grupos, com cada grupo representando uma categoria e cada barra dentro do grupo representando uma subcategoria. Ideal para comparar os valores de subcategorias dentro de uma categoria principal, assim como entre diferentes categorias. Permite uma comparação clara e direta das diferenças entre grupos e subcategorias, ajudando a identificar tendências, diferenças e relações;
- **Pirâmide da População:** Gráfico de barras horizontal que ilustra a distribuição da população de um país ou região por idade e sexo. O gráfico é chamado de "pirâmide" pois, em populações em crescimento, ele frequentemente assume uma forma piramidal, com uma maior prevalência de jovens em detrimento de idosos. O eixo horizontal normalmente representa números da população, enquanto o eixo vertical exibe as características. Usado para visualizar a estrutura etária de uma população e entender sua dinâmica demográfica, como taxas de natalidade, mortalidade e expectativa de vida, podendo indicar tendências futuras, como envelhecimento da população ou crescimento populacional;
- **Gráfico de Radar:** Ferramenta de visualização que exibe múltiplas variáveis num formato bidimensional, onde cada variável é representada por um eixo que parte do centro do gráfico. Ideal para exibir as variáveis que marcam um alto ou baixo dentro de um conjunto de dados, tornando-se o principal gráfico quando se trata da

representação de um desempenho. Podem ser comparadas diferentes categorias de desempenhos;

- Gráfico de Barras Radial: Equiparável ao gráfico de barras, mas disposto num sistema de coordenadas polares, em vez de cartesiano. Usado principalmente para fins estéticos ou quando se deseja uma representação compacta que possa ser encaixada em *layouts* circulares. Permite comparar categorias de uma forma visualmente atraente, mas pode ser menos preciso na leitura dos valores exatos em comparação com gráficos de barras convencionais, uma vez que não existe proporção entre o tamanho da barra e o valor que ela representa;
- Gráfico de Barras empilhadas: Representam numa barra, segmentos de outras barras que correspondem a vários conjuntos de dados. Ideal para mostrar como diferentes subcategorias contribuem para um total. Útil para comparar a composição de várias categorias principais ao mesmo tempo, permitindo visualizar não apenas o total de cada categoria, mas também a proporção de cada subcategoria dentro desse total;
- Gráfico de Área agrupado: Variação do gráfico de áreas que combina elementos de gráficos de áreas empilhadas e gráficos de áreas agrupadas. Usado para visualizar a evolução de múltiplas séries de dados ao longo do tempo, com a vantagem adicional de mostrar a contribuição de cada série para o total num formato agrupado. Útil para comparar a evolução de várias séries de dados ao longo do tempo, mostrando como cada série contribui para o total em diferentes pontos no tempo;
- Gráfico de Intervalos: Tipo de gráfico utilizado para representar a variação ou o intervalo de valores de uma variável ao longo do tempo ou entre diferentes categorias. Ideal para visualizar a amplitude de variação dos dados, permitindo observar como os valores flutuam ao longo do tempo ou entre categorias. Útil para analisar a volatilidade ou a dispersão dos dados, permitindo a comparação entre diferentes intervalos, facilitando a identificação de padrões ou tendências de variação;
- Gráfico de Cascata - Tipo de visualização de dados que representa como um valor inicial é afetado por uma série de aumentos/diminuições sequenciais, resultando num valor final. Este é chamado de "cascata" porque os valores intermediários parecem "cair" em etapas, criando uma aparência semelhante a uma cascata. Amplamente utilizados em análise de dados para entender como diferentes componentes afetam uma métrica global ao longo do tempo ou num processo.

**Charts** – Exibição de dados quantitativos e qualitativos sem recurso a um sistema de referência. *Chart* é uma forma de representação que apresenta um meio visual para mostrar dados, com o objetivo de analisar, planejar, controlar e comunicar informação (Pereira, 2015);

- **Círculo de Pacotes:** Tipo de visualização de dados que organiza informações hierárquicas ou categorizadas em círculos de tamanhos variados. Estes círculos representam diferentes categorias ou subcategorias, e o seu tamanho (área) é proporcional à importância ou representatividade dessa mesma categoria. A cor também pode ser utilizada para exibir diferentes categorias ou variáveis. Ideal para representar dados que têm uma estrutura hierárquica ou agrupada, permitindo comparar o tamanho relativo das categorias e subcategorias de maneira visualmente intuitiva, destacando as proporções entre diferentes partes do todo;
- **Gráfico Circular:** Ferramenta de visualização que representa a composição de um todo dividido em partes, onde cada parte é proporcional à sua contribuição para o total. Ideal para mostrar a composição de um conjunto de dados, facilitando a comparação das partes em relação ao todo. Este tipo de gráfico é mais eficaz quando existem poucas categorias, pelo que é recomendado quando há no máximo 5-7 categorias a serem comparadas;
- **Gráfico de Donuts:** Variação do gráfico circular que apresenta um buraco no centro, criando uma forma de anel ou "donut". Assim como o gráfico circular, este é usado para mostrar a composição de um todo em partes, mas o espaço central vazio pode ser utilizado para adicionar informações adicionais ou simplesmente para melhorar a estética do gráfico;
- **Matriz de Pontos:** Tipo de gráfico que usa pontos individuais para representar dados quantitativos. Cada ponto na matriz corresponde a uma unidade de medida ou contagem, e os pontos são organizados numa estrutura regular, facilitando a comparação visual entre diferentes categorias ou grupos. São úteis para alcançar uma visão geral da proporção que cada categoria assume num conjunto de dados. Pontos com a mesma cor são exibidos em conjunto;
- **Gráfico em Caracol:** Ferramenta de visualização que representa informações ao longo do tempo ou outro eixo de progresso de forma contínua numa espiral. Em vez de seguir o *layout* linear ou circular típico de outros gráficos, os dados são dispostos numa espiral que se expande do centro para fora, criando um efeito visual semelhante

ao de um caracol. A área representa o valor da variável. Ideal para visualizar dados que têm uma natureza cíclica, como dados sazonais ou que se repetem em intervalos regulares. A forma espiral pode evidenciar padrões de repetição que não seriam tão óbvios em gráficos lineares ou circulares tradicionais;

- **Gráfico de Conjuntos Paralelos:** Utilizado para representar relações e comparações entre múltiplas categorias ou variáveis. Linhas horizontais ou curvas ligam pontos entre os eixos paralelos, indicando as relações ou fluxos entre as diferentes categorias em cada eixo. A espessura das linhas pode representar a frequência ou quantidade associada àquela combinação específica de categorias. Especialmente útil para visualizar dados categóricos complexos, permitindo a análise de como diferentes categorias interagem entre si. Cada caminho de fluxo deve ser colorido para comparar a distribuição entre as diferentes categorias;
- **Gráfico de Imagens:** Forma de visualização de dados que usa imagens ou ícones em vez de pontos, barras ou linhas tradicionais para representar informações quantitativas ou categóricas. Cada imagem ou ícone representa uma unidade de dados e é repetido várias vezes para indicar quantidade, frequência, ou outra métrica. Muito eficazes para comunicar dados de forma acessível e visualmente atraente, especialmente para públicos não técnicos. Podem ser usados para representar dados categóricos e quantitativos;
- **Gráfico de Área Proporcional:** Visualização de dados em que diferentes áreas são utilizadas para representar valores numéricos, sendo o tamanho de cada área diretamente proporcional ao valor que representa. Ideal para mostrar a magnitude relativa entre diferentes categorias, uma vez que a comparação direta entre áreas de diferentes tamanhos pode ser mais intuitiva do que comparar comprimentos ou alturas. Uma desvantagem é a dificuldade em estimar valores através da área, revelando-se mais adequados para a comunicação do que para análise;
- **Mapa de Árvores:** Ferramenta de visualização de dados que utiliza retângulos para representar informações hierárquicas, onde cada retângulo é subdividido em retângulos menores que representam as respectivas subcategorias, sendo o tamanho de cada retângulo proporcional ao valor numérico que este representa. Ideal para visualizar dados que apresentam uma estrutura hierárquica e para comparar tamanhos relativos de diferentes partes de um todo, permitindo ver rapidamente que categorias

ou subcategorias dominam em termos de valor ou quantidade, tornando-se útil quando existem muitas categorias para exibir em simultâneo;

- **Nuvem de Palavras:** Visualização gráfica que mostra palavras de um texto, onde o tamanho de cada palavra se apresenta proporcional à sua frequência. As palavras mais frequentes ou mais significativas aparecem em tamanhos maiores, enquanto as menos frequentes ou menos relevantes aparecem em tamanhos menores. Frequentemente usada para análise exploratória de textos, permitindo identificar rapidamente os temas e tópicos mais recorrentes num documento, podendo servir como um resumo visual de conteúdo, mostrando de forma visual que termos ou conceitos são mais importantes;

**Mapas** – Exibição de dados associados a um espaço físico, indicados quando se pretende comunicar dados com uma localização ou relação espacial (Hess, 2022);

- **Mapa de Bolhas:** Visualização de dados que combina elementos de gráficos de bolhas e mapas geoespaciais. Usado para representar dados quantitativos num mapa, onde as bolhas são colocadas em localizações geográficas específicas e o tamanho das bolhas reflete a magnitude ou a importância de uma variável. Ideais para comparar proporções em regiões geográficas;
- **Mapa de Cores:** Ferramenta de visualização geográfica que usa cores para representar dados em diferentes regiões de um mapa. Amplamente utilizado para exibir a distribuição de uma variável específica em áreas geográficas, facilitando a comparação e a análise visual de dados espaciais. São associadas cores específicas a valores ou intervalos de dados, sendo que, geralmente, a escala de cores é apresentada como uma barra ou gráfico que explica o que cada cor representa em termos de dados numéricos ou categóricos;
- **Mapa de Fluxo:** Ferramenta visual utilizada para representar o movimento ou a transferência de dados, recursos ou pessoas entre diferentes pontos ou regiões. Ilustra como algo se move através de um sistema, destacando as origens, destinos e volumes de fluxo. A quantidade é representada no tamanho de uma linha ou seta. Estas representações permitem perceber comportamentos de dados de distribuição geográfica.

**Diagramas** – Indicados quando se pretende ilustrar a relação entre diferentes elementos, conceitos ou dados, de forma a comunicar uma mensagem contida nas relações internas de um conjunto de dados (Ribecca, 2024);

- **Diagrama de Arcos:** Consiste num conjunto de pontos dispostos ao longo de uma linha horizontal, com arcos a conectar os pontos, de forma a indicar as relações entre os mesmos. Apresentam-se úteis para representar padrões de interconexão, especialmente quando se deseja evitar a complexidade visual que pode surgir noutros tipos de representações gráficas. A desvantagem desta representação pode surgir ao existirem muitas ligações desordenadas podendo tornar o diagrama difícil de interpretar;
- **Mapas Mentais:** Representação gráfica que organiza a informação de forma visual para facilitar a compreensão e o desenvolvimento de ideias. No centro do mapa mental, geralmente existe um conceito/tema principal, do qual se ramificam tópicos relacionados. São frequentemente usados em *brainstorming* e fases iniciais de planeamento de projetos, uma vez que a visualização em formato de mapa mental permite que sejam vistos o panorama geral de um assunto e as conexões entre suas partes, estimulando a criatividade e o pensamento crítico;
- **Diagrama de Cordas:** Representação visual usada para mostrar as relações entre diferentes elementos dentro de um conjunto de dados, destacando as conexões entre os mesmos. Composto por um círculo que representa um conjunto de dados, com secções ao longo do perímetro que correspondem a diferentes categorias/elementos e arcos/cordas que ligam estas secções, indicando as conexões ou fluxos entre os elementos. Estes diagramas são particularmente úteis para visualizar fluxos de informação ou interações destacando as relações internas de um sistema específico. A principal vantagem de um diagrama de cordas é a sua capacidade de representar múltiplas relações de forma clara e compacta, facilitando a identificação de padrões dentro dos dados apresentados;
- **Diagrama de Ilustração:** Representação gráfica que combina elementos visuais com um texto, de forma a explicar, esclarecer ou complementar essa mesma informação ou conceito. Este é usado principalmente para transmitir uma ideia de forma visual e intuitiva, auxiliando o observador a compreender rapidamente um conceito ou processo, tornando-se especialmente úteis quando o texto não é suficiente para descrever de forma clara e precisa a informação que se deseja transmitir;

- **Diagrama de Árvore:** Organiza a informação de forma hierárquica, semelhante à estrutura de uma árvore. Cada "nó" no diagrama representa um conceito, e as "ramificações" conectam esses nós, indicando a relação entre eles. O diagrama de árvore começa com um nó principal (a "raiz"), que representa o conceito central. A partir deste, ramificam-se nós secundários, que podem subdividir-se noutros níveis, formando uma estrutura de "ramos" que visualmente se assemelha a uma árvore. Esta estrutura facilita a visualização de como as informações se dividem e se conectam, tornando-se eficaz para demonstrar a estruturação de um tema nos seus componentes e para facilitar o entendimento das relações entre os diferentes elementos;
- **Diagrama de Venn:** Representação usada para mostrar as relações entre diferentes conjuntos de elementos. Este consiste na utilização de círculos que se sobrepõem parcial ou totalmente, onde cada círculo representa um conjunto distinto. As áreas de interseção entre os círculos indicam elementos que são comuns aos conjuntos representados, tornando-se adequado comparar e contrastar informações de forma visual e intuitiva;
- **Diagrama da Linha do Tempo:** Representação gráfica usada para mostrar eventos ou etapas em ordem cronológica ao longo de um eixo comum. Destaca a sequência e a duração dos acontecimentos ao longo do tempo, permitindo que se visualize claramente a progressão de eventos, marcos históricos, ou etapas de um processo, e facilitando a compreensão de como os eventos se relacionam ao longo do tempo, auxiliando na visualização de tendências e intervalos.

### **2.3. Desenho de *Dashboards***

Com o objetivo de desenvolver *dashboards* apelativos para o utilizador torna-se importante ter por base modelos específicos que permitam a criação de estruturas comuns, que sustentem o tipo de visualização que se pretende apresentar, surgindo desta forma os modelos do processo de visualização. Estes modelos servem de referência para o utilizador, apresentando um algoritmo específico que pretende facilitar a criação de uma representação visual.

Stuart Card em (Card et al., 1999), descreve a visualização como uma sequência de mapeamentos ajustáveis de dados para uma representação visual de modo a proporcionar interação do utilizador com o espaço de informação, concretizando o que foi denominado por cristalização do conhecimento. Desta forma, segundo o modelo de referência para

visualização apresentado por Card, no processo de visualização destacam-se três fases essenciais:

1. Transformação de dados;
2. Mapeamento visual;
3. Transformação visual.

A primeira etapa, denominada de **Transformação de Dados**, é onde ocorre a entrada de dados “em bruto”, que conseqüentemente são processados e organizados numa representação lógica mais estruturada, geralmente na forma de uma ou mais tabelas. O processamento pode envolver a eliminação de dados redundantes, errados ou incompletos, bem como a filtragem e o agrupamento dos dados relevantes. Além disso, pode ser feita a inclusão de novas informações, como, por exemplo, de resultados de análises estatísticas (média, soma total, desvio padrão e etc.) realizadas sobre os dados brutos.

Relativamente à etapa do **Mapeamento Visual**, ao recorrer ao uso de gráficos, pretende-se a obtenção de duas finalidades distintas: utilização como um meio para comunicar uma determinada informação; e utilização como processamento gráfico, em que cada utilizador pode manipular objetos gráficos, de forma a aumentar conhecimento sobre o que está a visualizar (Bertin, 2011). Este processamento inicia-se com dados não estruturados que existem nas mais variadas formas e que precisam de ser codificados na representação visual.

Existem duas formas fundamentais de dados (Steele & Iliinsky, 2011):

- **Entidades:** são geralmente o objeto de interesse, dados recolhidos de fontes aleatórias;
- **Relacionamentos:** baseiam-se nas estruturas que relacionam os atributos e criam um conhecimento, isto é, representam o propósito da visualização sob a forma de estruturas e padrões que se pretendem descobrir.

Nesta etapa é envolvida a construção de uma estrutura visual que pretende representar visualmente os dados da tabela. Toda estrutura visual pode ser decomposta em três partes distintas: substrato espacial, marcas visuais e propriedades gráficas das marcas (Nascimento & Ferreira, 2011).

O substrato visual caracteriza o espaço para a visualização, sendo normalmente representado por eixos, tais como os eixos X e Y do plano cartesiano. Existem quatro tipos elementares de eixos:

- U = *Unstructured Axis* (sem eixos);
- N = *Nominal Axis* (região dividida em sub-regiões);
- O = *Ordinal Axis* (região dividida em sub-regiões onde a ordem das mesmas tem importância);
- Q = *Quantitative Axis* (região que tem uma métrica, como valores inteiros ou reais).

As marcas visuais são símbolos gráficos utilizados para representar os itens de dados, sendo que os mais comuns podem compreender pontos (figuras geométricas simples com a finalidade de marcar uma posição), linhas, áreas, volumes e figuras complexas (cujos detalhes têm importância; também chamadas de ícones).

As propriedades gráficas são os atributos visíveis que caracterizam as marcas visuais. Algumas propriedades amplamente utilizadas são: a posição da marca dentro do substrato espacial; a forma, a cor, o tamanho, e a área ou o volume da marca; a orientação, o sentido e a inclinação de marcas do tipo de linha; e a textura de uma marca na forma de área. Certas propriedades gráficas dão um aspecto de dinamismo às marcas e servem para destacar atributos dos dados que requerem maior atenção. São exemplos dessas propriedades a possibilidade de animação ou de movimento da marca, a velocidade, a direção e sentido do seu deslocamento, e a frequência com que uma marca altera a sua forma ou posição na tela.

Nesta etapa, um substrato espacial deve ser escolhido e marcas visuais associadas aos itens de dados abstratos dentro do substrato. De seguida, cada atributo dos dados deve ser associado a uma ou mais propriedades gráficas das marcas.

A última etapa do processo de visualização é a de **Transformação Visual**, na qual é possível modificar e estender as estruturas visuais através de operações básicas como:

- testes de localização, que possibilitam obter informações adicionais sobre um item da tabela de dados;
- controlo de ponto de vista, os quais permitem ampliar, reduzir e deslocar a imagem com o objetivo de oferecer visões diferentes;

- distorções da imagem, visando criar ampliações de uma região específica em detrimento de outra.

Os mecanismos de interação implementados possibilitam a exploração de diferentes cenários para um melhor entendimento dos dados visualizados.

Segundo outros autores podem ser consideradas etapas adicionais, ou apresentadas de uma forma distinta. Segundo Veslava Osinska (2018), neste processo podem ser identificadas cinco fases distintas (Osinska, 2018):

- Recolha de dados: numa primeira fase os dados brutos são recolhidos de sistemas operacionais, e reunidos para a sua posterior utilização;
- Processamento de dados: nesta fase realiza-se a limpeza e tratamento dos dados recolhidos. É comum, nesta fase, que os dados se encontrem em diferentes formatos, pelo que se torna necessário uniformizar os mesmos, recorrendo-se habitualmente à sua conversão para tabelas;
- Preparação de dados: é realizada a preparação dos dados específicos a utilizar nas representações visuais desejadas;
- Redução: por norma, o conjunto de dados de entrada é representado por vetores de atributos, que são difíceis de visualizar em configurações 2D ou 3D. Uma das formas de resolver este problema prende-se com a redução das dimensões dos dados, através de técnicas específicas do âmbito da estatística e *machine learning*;
- Conceção da representação visual: esta fase, também denominada mapeamento, é onde toda a abordagem se centra na melhoria da apresentação visual através da correspondência de cores, construção de legendas e adição de funções de manipulação de dados como a ordenação ou a filtragem.

Além de todos os procedimentos realizados nas cinco fases anteriores, considera-se pertinente a inclusão de um estágio adicional, a **interação**, que pretende relacionar o utilizador com o processo de visualização. Este estágio ou mecanismo de vantagem poderá ser aplicado durante todo o processo de visualização:

- Acesso a um novo conjunto de dados para um novo pré-processamento, caso o conjunto de dados processados não seja o total ou completo;
- Modificação dos critérios de mapeamento visual, quer na escolha de outras representações quer na escolha de novas variáveis;

- Alteração dos critérios de filtragem aplicados ao conjunto de dados.

As etapas do processo de visualização dizem respeito a um conjunto de decisões que o utilizador deve tomar antes do resultado ocorrer: perceber qual o tipo de dados que o conjunto contém e defini-los (dados quantitativos ou qualitativos); conhecer os tipos de gráficos e seleccioná-los; saber quais as análises que podem ser adotadas para evidenciar as características dos dados; dominar a utilização de propriedades gráficas e construir uma história.

Ao longo do tempo foram sendo desenvolvidos diferentes princípios e teorias que nos direcionam no âmbito das melhores práticas no que respeita ao desenvolvimento de *dashboards*.

Tufte (2007) sugeriu cinco princípios da teoria dos gráficos de dados que produzem mudanças substanciais no design gráfico:

- Maximizar o rácio dados-tinta;
- Eliminar a “tinta” que não contém dados;
- Apagar a “tinta” de dados redundantes;
- Rever e editar.

O rácio de tinta de dados é definido por Tufte (2007) como a proporção da tinta de um gráfico dedicada à apresentação não redundante de informação. Maximizar a quota de tinta de dados significa que é preferível uma maior percentagem de tinta de um gráfico dedicada a dados. Toda a tinta num gráfico requer um objetivo, e quase sempre esse objetivo deve ser a apresentação de nova informação (Tufte, 2007).

Derivado do conceito de Tufte, Few recomendou dois princípios para destacar objetivos: "reduzir a tinta não associada a dados" e "melhorar a tinta dos dados". Para reduzir a tinta não associada aos dados, devemos primeiro remover a tinta e, em seguida, retirar a ênfase e regularizar a tinta não restante. A remoção da tinta não associada a dados desnecessária começa por colocar a seguinte pergunta em relação a cada componente visual: "Será que os dados sofreriam alguma perda de significado se isto fosse eliminado?" Se a resposta for negativa, então deve ser eliminada. Os elementos não informativos, de acordo com o seu papel de apoio, devem destacar-se do fundo suficientemente para servir o seu objetivo, mas não tanto que chamem a atenção para si próprios. Este objetivo pode ser alcançado através da utilização de linhas finas e cores suaves e neutras (Few, 2009).

A melhoria da tinta dos dados envolve duas etapas. O primeiro passo é remover a tinta de dados desnecessária; o segundo é enfatizar a tinta de dados mais importante. Não remover nada de crucial, mas assegurar que tudo o que é periférico aos interesses e objetivos dos leitores. Quanto ao objetivo de organização, a informação deve ser organizada de uma maneira que conte a história da forma mais clara possível. Quando organizamos a informação num *dashboard*, fazemo-lo conscientemente para contar uma história. O que é que devo dizer primeiro? O que é que devo guardar para o fim? O que é que devo realçar mais do que o resto? As respostas a estas perguntas assumem a forma de atributos visuais concebidos para realizar o seguinte:

- Agrupar - segmentar a informação em secções significativas;
- Priorizar - classificar a informação por importância;
- Sequenciar - dar indicações sobre a ordem pela qual a informação deve ser lida.

Shneiderman e Heer (2012) propuseram um ponto de partida útil para a conceção de uma interface gráfica, denominada "Mantra da procura de informação visual": primeiro a visão geral, zoom e filtro, depois pormenores personalizados". O autor propôs uma taxonomia de tarefas para a visualização de informações. As tarefas são ações de informação no domínio da tarefa que os utilizadores desejam realizar (Heer & Shneiderman, 2012). Os sete tipos de tarefas de abstração são as seguintes:

- Visão geral: obter uma visão geral de toda a coleção;
- Zoom: aumentar os itens de interesse;
- Filtro: filtrar os itens sem interesse;
- Detalhes personalizados: seleccionar um item ou grupo e obter pormenores quando necessário;
- Relacionar: visualizar relações entre itens;
- Histórico: manter um histórico de ações para permitir anular, repetição e refinamento progressivo;
- Extrair: permitir a extração de subcoleções e consultar parâmetros.

Munzner (2014) propôs um modelo de quatro níveis para a conceção e avaliação de visualizações. O nível superior destina-se a caracterizar os problemas e dados de um determinado domínio, o nível seguinte é para mapear esses problemas em operações abstratas e tipos de dados, o terceiro nível é para conceção da codificação visual e das

interações de apoio e suporte a essas operações, e o quarto nível, mais interno, destina-se a criar um algoritmo para executar essa conceção automaticamente e de forma eficiente. Munzner defende que muitos designers de visualização do passado efetuaram estas etapas, embora de forma implícita e não necessariamente nesta ordem (Munzner, 2014).

Munzner (2014) identificou ainda duas estratégias comuns para o design de *dashboards*: de cima para baixo ou de baixo para cima. Com o trabalho orientado para o problema, começa-se no nível superior do domínio e trabalha-se para baixo através de abstração, linguagem e decisões de algoritmo. No trabalho orientado para a técnica, trabalha-se num dos dois níveis inferiores, a conceção de expressões idiomáticas ou de algoritmos, em que o seu objetivo é inventar novas expressões idiomáticas que suportem expressões idiomáticas existentes (Munzner, 2014).

Segundo (Lin et al., 2018), tendo por base estudos efetuados por Tufte, Few, Shneiderman e Munzner, desenvolve o propõe um novo conjunto de princípios, que pretende guiar na direção das melhores práticas de desenvolvimento de interfaces de visualização de informação. Este princípio denomina-se “5S dashboard design principle framework” e tem por base 5 princípios fundamentais:

- **Seeing Both the Forest and Trees (Ver a Floresta e as árvores)** - Este princípio deriva do objetivo de organização de Few (agrupar, priorizar ou sequenciar) e do mantra de Shneiderman. O princípio de ver tanto a floresta como as árvores apresenta duas implicações. A primeira implicação é que o todo (floresta) deve ser classificado em diferentes regiões ou partes com base em certas características (tipos de árvores, habitats animais, ou características geográficas) para que os utilizadores possam identificar qual a parte da floresta que devem procurar para encontrar o que lhes interessa. A segunda implicação é a capacidade de ver a floresta, as árvores e as folhas das árvores num único *dashboard*, isto é, ser capaz de visualizar a totalidade de determinada informação e conseguir visualizar pormenores mais específicos;
- **Simplicity Through Self-Selection (Simplicidade através da autosseleção)** - Quase todos os guias e académicos salientam a importância da simplicidade na conceção do *dashboard*. Para atingir o objetivo de simplicidade, foram deduzidos três princípios operacionais: autosseleção, significância e síntese. No que diz respeito à autosseleção, foram utilizados na literatura vários termos e técnicas, como

seleção, filtragem, aprofundamento e detalhes personalizados. Uma característica crucial de um *dashboard* é que os utilizadores podem selecionar a dimensão ou categoria que lhes interessa através da função de filtragem, que é a característica mais essencial de qualquer ferramenta de BI;

- **Simplicity through Significance (Simplicidade através da Significância)** - Um dos principais objetivos de um *dashboard* (e, de facto, de toda a análise de dados) é fornecer informações acionáveis. Um *dashboard* eficaz facilita a identificação de informação vital, que pode ser utilizada pelos decisores para determinar se devem tomar medidas. Neste estudo, foram sugeridas duas formas de determinar a significância. A primeira foi o método tricotómico; as “árvores” foram agrupadas em três níveis (superior, médio e inferior com base nas suas características). O segundo método foi o teste estatístico. Considerando determinada característica, são efetuados testes estatísticos para verificar se determinado conjunto de “árvores” se destacam em relação à “floresta”;
- **Simplicity Through Synthesis (Simplicidade através da síntese)** - Uma característica crucial de um painel de controlo é a capacidade de revelar vários gráficos com indicadores relacionados, porém diferentes. No entanto, muitas vezes revelam-se dificuldades na tomada de decisão se diferentes indicadores apresentam resultados contraditórios. Para resolver este problema, foi sugerido um princípio de síntese para reduzir o stress da tomada de decisão. Nalgumas circunstâncias, pode ser utilizada uma abordagem sofisticada, como uma fórmula complexa ou uma modelação estatística para sintetizar todos os indicadores num índice composto. Ao apresentar um único índice reduz-se a complexidade da informação apresentada, facilitando a tomada de decisão e evitando a confusão;
- **Storytelling** - A literatura sublinha cada vez mais a importância da utilização do *storytelling* no desenvolvimento de *dashboards*. Desta forma, para o mesmo conjunto de *dashboards*, cada utilizador distinto pode direcionar a sua interpretação/análise para um âmbito específico, dependendo das questões a que pretenda dar resposta. Assim torna-se possível abranger uma multiplicidade de agentes e conteúdos, tendo por base o mesmo conjunto de dados.

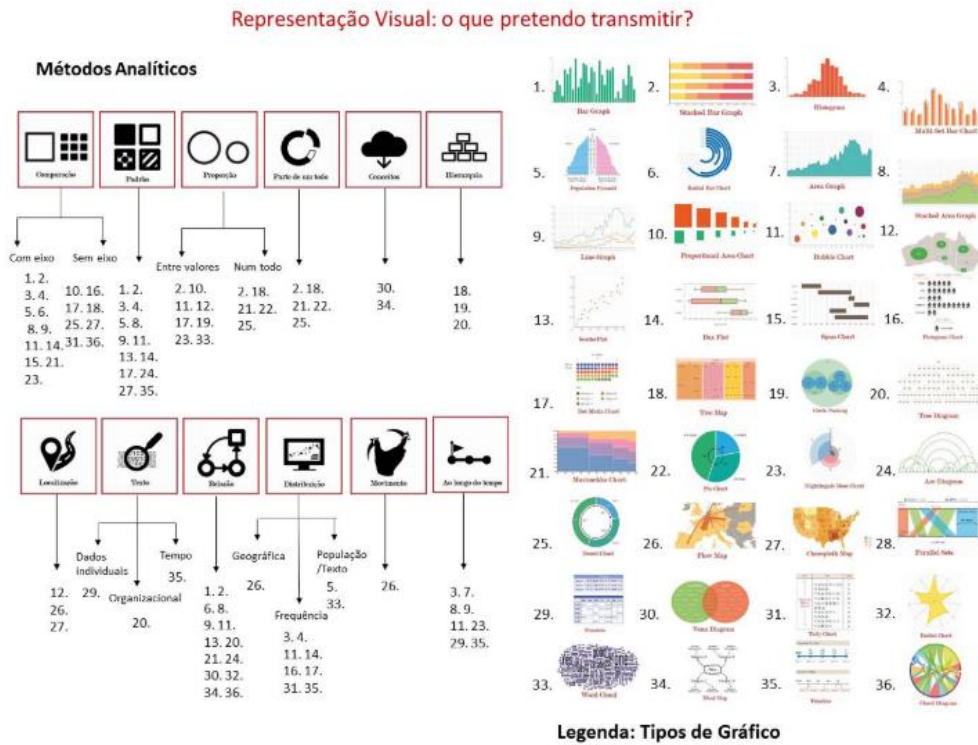
Antes do processo de visualização ser iniciado, o utilizador deve questionar-se cientificamente, de forma a despoletar o propósito da visualização. Devem ser levantadas questões temporais como “quando”, espaciais como “onde”, tópicos como “o quê” ou

relacionais como “com quem”, permitindo que o utilizador direcione e justifique as suas escolhas na construção de determinada visualização (S. Lee et al., 2016).

Perante o estudo das tipologias de visualização e dos métodos analíticos disponíveis, apresentam-se de seguida doze tipos de análises, que permitem estabelecer agrupamentos entre as várias classes de representação gráfica (Pereira, 2015):

- Comparação - Permite revelar semelhanças ou disparidades entre os dados apresentados;
- Proporção - Permite utilizar o tamanho da amostra para exibir semelhanças ou disparidades entre os valores de dados tendo em conta o todo;
- Relações – Permite relacionar os diferentes atributos através da comparação dos eixos vertical e horizontal;
- Hierarquia - Análise que exibe uma classificação ordenada de dados, numa estrutura ordenada definida sob um conjunto de critérios;
- Conceitos - Permite simplificar ideias e conceitos, de forma concreta;
- Localização - Permite visualizar a distribuição dos dados ao longo de uma região geográfica específica;
- Parte de um todo - Análise que permite visualizar uma porção específica ou várias porções de uma variável, tendo em conta o seu todo;
- Distribuição - Permite exibir a frequência de dados e como eles são agrupados ao longo de um intervalo e por diferentes categorias;
- Movimento - Permite exibir movimentos e alterações nos dados;
- Padrões - Processo de análise que examina e descobre padrões e formas para interpretar e compreender os dados;
- Ao longo do tempo - Exibe a evolução dos dados durante um período temporal específico, demonstrando as suas variações;
- Análise de Texto – Permite visualizar palavras de um texto, de forma mais concreta.

O objetivo da representação visual de dados é evidenciar e revelar as relações existentes, as características, diferenças e semelhanças. Poder efetuar uma comparação ou distribuição ao longo de uma variável e ainda organizar os dados em hierarquia é possível e existem gráficos adequados para revelar estes comportamentos e tendências.



**Figura 3 - Modelo de classificação "Representações visuais - o que pretendo transmitir?" (extraído de Pereira, 2015)**

Com o objetivo de aproximar a análise de dados e as representações gráficas, em (Pereira, 2015) é apresentado um modelo denominado “Representações Visuais – o que pretendo transmitir?”, evidenciando a relação entre os métodos analíticos e os vários tipos de gráficos.

No modelo apresentado, o grupo à esquerda da imagem, na Figura 3 representa os doze métodos analíticos, definidos anteriormente, que possibilitam a escolha de representações gráficas adequadas, uma vez que, a escolha de uma representação gráfica se encontra profundamente relacionada com a análise ou comportamento que se pretende revelar na visualização.

Após a fase em que o utilizador percebe qual o propósito da análise e o método analítico que melhor se relaciona com as principais características dos seus dados, torna-se necessário perceber qual a representação gráfica mais eficiente. O grupo à direita da imagem, na Figura 3, representa um conjunto de trinta e seis diferentes tipos de gráficos, que contemplam desde os três sistemas de coordenadas (cartesianas, polares e geográficas) às mais variadas marcas gráficas (linhas, pontos e áreas).

Desta forma, associado ao conjunto de métodos analíticos e representações gráficas surge a oportunidade de contextualizar a exploração ou representação através de um conjunto de ocorrências que desencadeiam o entendimento da representação visual final. A utilização das propriedades gráficas ao longo de um trabalho de visualização irá evidenciar comportamentos, padrões e tendências nos dados que após serem estudados e contextualizados devem ser comunicados através de uma narrativa visual, incorporando nesta fase o processo de *storytelling*.

## 2.4. Dashboards e Data storytelling

O termo *data storytelling* relaciona-se com o conceito de construir uma narrativa convincente, com base em dados e análises complexas que ajudam a contar a sua história e a influenciar e informar um determinado público-alvo (Zhang et al., 2022).

O *storytelling* aplicado à visualização de dados em *dashboards* é muito semelhante à narração humana, com a vantagem de oferecer os benefícios adicionais de conhecimentos mais profundos e provas de apoio através de gráficos, esquemas e quadros. Através do *storytelling*, a informação complicada é simplificada para que o público possa interagir com o seu conteúdo e tomar decisões críticas mais rapidamente e com mais confiança. Ao contrário da visualização tradicional de dados, que é fundamentalmente um processo orientado para a análise, a narração de dados visuais baseia-se num processo orientado para a narrativa (Zhang et al., 2022).

Tabela 1 - Comparação entre visualização de dados e *data storytelling* (adaptado de Zhang et al., 2022)

<b>Processo orientado para a análise de Visualização de dados</b>		<b>Processo orientado para a narrativa de Data storytelling</b>
Orientação para a análise	<b>Orientação</b>	Orientação narrativa
Profissionais	<b>Audiência</b>	Não-profissionais
Variáveis	<b>Unidade básica</b>	Unidades de informação
Mensagem simples	<b>Elementos de informação</b>	Composição da informação em forma de história
Baixo	<b>Entretenimento</b>	Elevado
Estático 2D	<b>Espaço visual</b>	Dinâmico 3D
Simple	<b>Complexidade visual</b>	Relativamente complexo

A integração de narrativas com gráficos oferece um meio mais rico para comunicar os resultados da análise do que a mera utilização de ferramentas gráficas normais para a análise de dados, do mesmo modo que a incorporação de elementos narrativos nas visualizações de dados convencionais pode contribuir para a comunicação da informação de forma eficaz e intuitiva. As visualizações melhoradas com elementos de narração de dados podem superar as técnicas de visualização convencionais quando a concepção se concentra exclusivamente nos dados essenciais e nos elementos visuais essenciais, reduzindo assim a complexidade e, conseqüentemente, encurtando o tempo necessário para compreender os dados (Shao et al., 2024).

Além da integração de narrativas com gráficos, as visualizações reforçadas com elementos de dados oferecem vantagens na medida em que tiram partido das capacidades perceptivas eficientes dos seres humanos para reconhecimento de padrões visuais (Knafllic, 2015).

Apesar das vantagens percebidas do *data storytelling*, as provas empíricas que sustentam a eficácia e eficiência relativas da narração de dados em relação às visualizações convencionais continuam a ser limitadas. Alguns trabalhos de investigação examinaram empiricamente outras vantagens potenciais; por exemplo, Zhao et al. concluíram que as visualizações enriquecidas com legendas são mais cativantes e fáceis de utilizar, em relação às demais. Do mesmo modo, a análise qualitativa do mesmo estudo sugeriu que as melhorias na narração de dados ajudam a concentração e a memória (Zhao et al., 2019). No entanto, um estudo efetuado por Zdanovic et al. não encontrou qualquer impacto significativo das visualizações melhoradas com elementos de narração de dados na memória a curto ou a longo prazo, em comparação com as visualizações convencionais (Zdanovic et al., 2022).

Os resultados dos estudos que investigaram o potencial da narração de histórias com dados para promover a empatia ou influenciar as atitudes da audiência também foram díspares. Por exemplo, no estudo de Morais et al. (2021) não foram encontradas provas que apoiassem a ideia de que a adição de elementos antropomórficos às visualizações aumente a empatia (Morais et al., 2021). Do mesmo modo, Liem et al. não encontraram diferenças significativas nas respostas empáticas entre histórias de dados que utilizam narrativas pessoais e visualizações exploratórias convencionais. A literatura continua a revelar uma lacuna importante: não é claro se a adição de elementos de narração de dados às visualizações convencionais ajuda a identificar eficaz e eficientemente as ideias e os pontos de dados

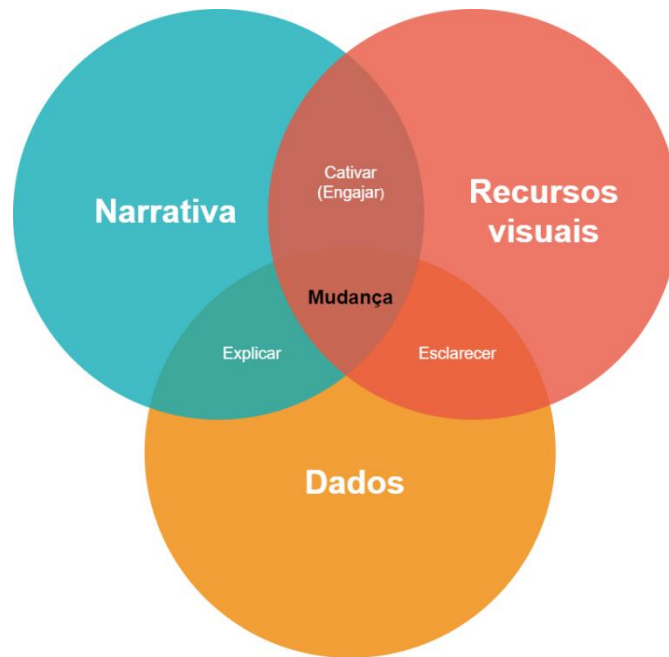
críticos correspondentes, em comparação com as visualizações convencionais (Liem et al., 2020).

Através de uma abordagem estruturada, o *storytelling* e a visualização de dados trabalham em conjunto para comunicar uma ideia através de três elementos essenciais: narrativa, recursos visuais e dados (Brent Dykes, 2016).

- **Construir a narrativa:** Ao contar uma história, os dados devem ser usados como pilares de apoio às ideias a apresentar. O público deve ser guiado de forma a compreender o ponto de vista apresentado, transformando informações complexas em pontos-chave chamativos. A narrativa e o contexto são o que irá conduzir a natureza linear da narrativa de dados;
- **Utilizar recursos visuais:** Os recursos visuais podem ajudar a guiar o público sobre a teoria apresentada. Ao combinar os recursos visuais (tabelas, gráficos, etc.) com a narrativa, o público torna-se mais envolvido com informações que, de outra forma, estariam ocultas e que fornecem os dados fundamentais para apoiar a teoria. Ao invés de ser apresentada uma única informação de dados para apoiar a teoria defendida, é útil mostrar vários dados distintos, tanto de baixo como de alto nível, para que o público possa compreender, verdadeiramente, o ponto de vista apresentado;
- **Mostrar dados:** Os seres humanos não são naturalmente atraídos pela análise, especialmente a análise que carece de contextualização utilizando a análise aumentada. O contexto e a crítica são essenciais para a interpretação completa da narrativa. A utilização de ferramentas analíticas pode ajudar a fornecer o contexto tão necessário em toda a história de dados.

Torna-se importante compreender como estes três diferentes elementos se combinam e funcionam em conjunto no desenvolvimento do *storytelling*. Quando a narrativa é associada aos dados, esta ajuda a explicar ao público o que está a acontecer nos dados e por que razão uma determinada informação é relevante. Quando os recursos visuais são aplicados aos dados, estes podem esclarecer o público sobre informações que não veriam sem tabelas ou gráficos. Muitos padrões interessantes e valores atípicos nos dados permaneceriam ocultos nas linhas e colunas das tabelas sem a ajuda de visualizações de dados. Por último, quando a narrativa e os elementos visuais são combinados, podem envolver e cativar (*engagement*) uma audiência.

Quando se combinam os elementos visuais e narrativas certas com os dados certos, obtém-se uma história de dados que pode influenciar e impulsionar a mudança (Brent Dykes, 2016).



**Figura 4 - Combinação entre narrativa, recursos visuais e dados (adaptado de Brent Dykes, 2016)**

Ao unir os conceitos de visualização de dados e *storytelling*, surge então um novo conceito denominado de histórias de dados visuais. Apesar de não existir um conceito perfeitamente definido e consensual, as histórias de dados visuais podem ser interpretadas através dos seguintes parâmetros (B. Lee et al., 2015):

- Uma história de dados visuais inclui um conjunto de elementos da história, isto é, factos específicos apoiados por dados (tal como a alteração do consumo de energia ao longo dos anos);
- A maior parte dos elementos da história são visualizados para apoiar uma ou mais mensagens pretendidas. A visualização inclui anotações (etiquetas, indicadores ou texto) ou narração para destacar e enfatizar claramente esta mensagem e evitar a ambiguidade (especialmente importante no caso de histórias assíncronas);
- Os elementos da história são apresentados numa ordem adequada ou com uma ligação visível entre eles, de forma a apoiar o objetivo de comunicação de alto nível (educar/entreter os espectadores ou convencê-los/persuadi-los com opiniões).

Na figura seguinte é apresentado o processo de narração de histórias de dados visuais, ou *visual data storytelling process* (VDSP) como modelo de trabalho, que deriva de outros modelos da literatura sobre jornalismo de dados.

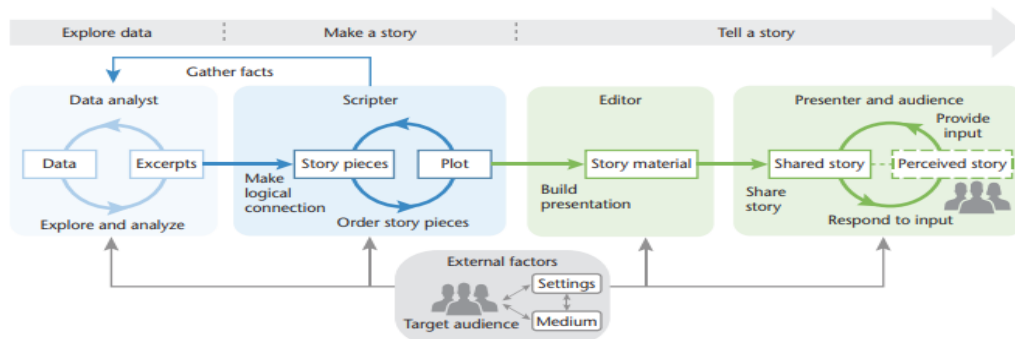


Figura 5- *Visual Data Storytelling Process* (extraído de B. Lee et al., 2015)

O VDSP resume as principais funções e atividades a que o orador se dedica quando transforma dados em bruto numa história visualmente partilhada, assim como os tipos de histórias que podem ser utilizadas. Embora o processo de contar histórias seja linear na figura apresentada, para efeitos de ilustração, este não tem de começar à esquerda e mover-se para a direita; pode conter muitos *loops* e múltiplas ocorrências em cada componente (B. Lee et al., 2015).

A fase da exploração de dados envolve o conjunto de atividades centradas na exploração e análise de dados, na qual os excertos pertinentes são recolhidos através desta análise. Estes excertos podem ser simples, como factos registados ou passos do processo de análise, ou podem ser mais complexos, tais como informações derivadas dos dados, conjuntos ou sequências interessantes, e/ou detalhes e variações do processo. Nesta fase do processo, esta coleção de excertos pode ou não estar ligada a qualquer representação visual específica (B. Lee et al., 2015).

Na fase da criação da história, os excertos de dados recolhidos na primeira etapa precisam de ser reunidos numa história que seja interessante, esclarecedora e convincente. A parte significativa da criação de uma história é o processo de construção do enredo ou trama. Neste processo, as atividades envolvidas são ordenar, estabelecer ligações lógicas, desenvolver o fluxo, formular uma mensagem e criar o desfecho. Estas atividades, que estão frequentemente entrelaçadas podem ser realizadas sequencialmente, simultaneamente, ou através de múltiplas iterações. Além disso, durante o desenvolvimento do enredo, pode ser necessário voltar à fase de exploração de dados para recolher mais excertos. O resultado

desta etapa do processo é um enredo que descreve como as peças da história estão relacionadas e o que significam num contexto geral (B. Lee et al., 2015).

A fase de contar uma história é o processo geral de materialização do enredo abstrato e de apresentação da história. Este processo consiste nas seguintes atividades: construir uma apresentação (criar o material da história), partilhar a história utilizando o material da história e, receber e tratar o *feedback* do público. Na fase de construção de uma história, um enredo e as peças da história são transformadas em material da história. O material da história é a materialização de cada peça deste conteúdo abstrato através do desenvolvimento de representações visuais, interações, animações, anotações ou narração. Este material da história transforma-se numa história partilhada quando é entregue a pelo menos uma pessoa. Em última análise, a história percebida é o que o público compreende através da experiência de contar histórias (B. Lee et al., 2015).

Para comparar de forma fiável as visualizações convencionais com as histórias de dados, torna-se necessário identificar claramente os elementos de narração de histórias de dados. Desta forma foi estabelecido um conjunto de ações de design que envolvem a adição ou exclusão de elementos de design visual de visualizações convencionais para as transformar numa história de dados. Este conjunto de ações tem por base o quadro para a narração de histórias sobre dados proposto por Zdanovic et al. (Zdanovic et al., 2022), que, por sua vez, se baseia nos princípios de narração de histórias de dados apresentados por Echeverria et al. (Echeverria et al., 2017), Ryan (Ryan, 2016) e Knaflic (Knaflic, 2015):

- **Princípio nº1 - Identificar um objetivo explícito:** Implica compreender o público, o seu nível de conhecimento tecnológico e a sua especialização no assunto, um tema comum na literatura sobre a narração de histórias com dados;
- **Princípio nº2 - Remoção de elementos redundantes:** Pressupõe a exclusão de elementos que não oferecem valor adicional, tais como etiquetas de dados, marcadores, grelhas, legendas, marcas de escala e etiquetas de eixos desnecessários. Este princípio é fundamental para uma visualização de dados eficaz e é realçado nos guias de narração de histórias de dados como um passo crucial para a organização;
- **Princípio nº3 - Utilizar elementos de narração com discernimento:** Utilizar anotações e pistas visuais criteriosamente para clarificar ou agrupar dados relacionados. No entanto, um excesso de anotações pode confundir a visualização e distrair a audiência, contrariando potencialmente o objetivo de dirigir a atenção;

- **Princípio nº4 - Captar o interesse:** Envolve a minimização de elementos visuais irrelevantes (ou seja, desorganização) enquanto se realça os componentes cruciais, por exemplo, utilizando cores ou elementos a negrito;
- **Princípio nº5 - Apelo à ação:** Criação de uma visualização coerente e sucinta que transmita a narrativa essencial, por exemplo, orientando os utilizadores para a ação pretendida através de um título explícito.

Zdanovic et al. descrevem um processo em três fases que inclui i) exploração inicial dos dados para encontrar as ideias a comunicar; ii) elaborar a visualização; iii) contar a história.

A criação de uma visualização convencional envolve, tipicamente, as fases i) e ii). Em contrapartida, a criação de uma história de dados implica considerar os seguintes elementos de narração de histórias de dados na ordem apresentada abaixo, particularmente nas fases ii) e iii).

1. **Utilizar a técnica de visualização correta de acordo com o objetivo do *storytelling*:** A seleção da técnica de visualização adequada, como um gráfico de barras ou um gráfico de dispersão, de acordo com o objetivo da comunicação de dados, é crucial para transmitir eficazmente a perceção relevante dos dados e preparar o terreno para uma história convincente. De acordo com as recomendações de Knaflic, uma narração de dados eficaz pode implicar a alteração dos tipos de gráficos para melhor transmitir a narrativa. Por exemplo, para realçar tendências temporais, a mudança de gráficos circulares para um gráfico de linhas pode ser mais ilustrativa;
2. **Eliminar a desordem:** A redução de elementos visuais estranhos pode tornar uma visualização mais facilmente interpretável, permitindo que os espectadores se concentrem apenas nos elementos-chave. Embora este seja um princípio geralmente aplicável a qualquer visualização de dados convencional, torna-se vital na narração de histórias sobre dados permitir que outros elementos visuais se destaquem e chamem efetivamente a atenção. É importante salientar que a organização também pode ser aplicada à utilização excessiva de anotações e outros elementos de narração de dados;
3. **Dirigir a atenção salientando pontos de dados e tendências:** No caso específico das histórias de dados, a contrapartida da organização é dar ênfase. Isto pode ser conseguido através de técnicas como a utilização de cores contrastantes, destacando pontos de dados específicos ou tornando certas linhas de tendência mais destacadas.

Estas técnicas ajudam a dirigir a atenção do espectador para pontos de dados e tendências cruciais associados às principais ideias identificadas na fase i), permitindo-lhe compreender a história que está a ser transmitida;

4. **Adicionar anotações:** As anotações (gráficas ou textuais) fornecem informações contextuais ou explicativas diretamente na visualização, ajudando a clarificar ou a realçar ainda mais os pontos-chave e as tendências.;
5. **Acrescentar um título explicativo:** Um título bem elaborado não só descreve a visualização como também orienta o observador para a conclusão pretendida ou para o apelo à ação, servindo de título para a história dos dados. Recomenda-se a adição de um título prescritivo que resuma as principais conclusões ou até sugira um curso de ação como um elemento poderoso na narrativa de dados.

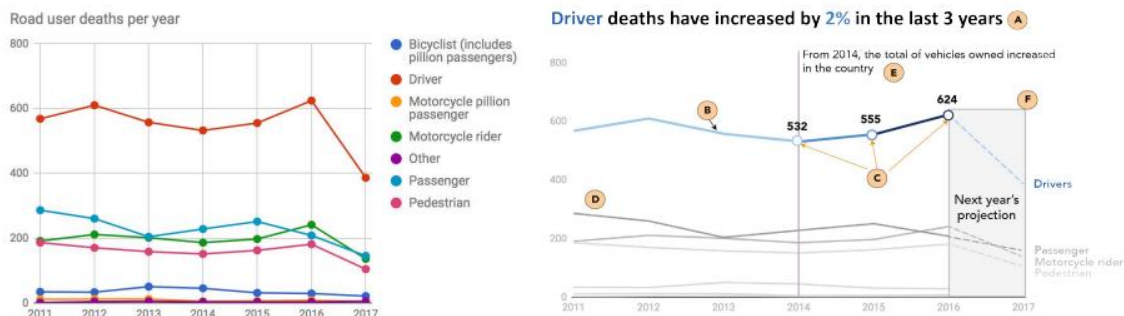


Figura 6 - Exemplo da aplicação da *Data storytelling* (extraído de Echeverria et al., 2018)

Através da Figura 6 é possível verificar um exemplo da aplicação de *storytelling* num gráfico de informação, relativo ao número de mortes, por ano, devido a acidentes rodoviários em estradas da Austrália. O exemplo à esquerda apresenta o resultado de uma ferramenta de visualização (folhas de cálculo do Google); o exemplo à direita apresenta a mesma visualização de dados com aplicação dos conceitos de *data storytelling*.

Se compararmos os dois gráficos, podemos ver como a mensagem é mais claramente percebida no exemplo à direita, sem nos perdermos a explorar a visualização. Por exemplo, pode ver-se como o título (A) dá uma mensagem sucinta que pretende comunicar explicitamente a ideia principal da visualização. Além disso, podemos ver como o destaque de uma linha chama a atenção visual para a tendência dos fatores (B). Destacar pontos de dados relevantes específicos e a remoção de pontos de dados irrelevantes torna explícito quais os dados que apoiam a afirmação no título (C). A remoção de grelhas, legendas e marcadores de pontos de dados elimina as distrações (D), enquanto a mensagem narrativa sobre o contexto, ajuda a obter uma melhor compreensão dos dados (E). Finalmente, a área

sombreada dá significado às secções do gráfico, dividindo-o em dados reais e projeções de dados (F).

## 2.5. Perfis de utilizadores

Tendo em conta o perfil dos utilizadores a quem se destinam os *dashboards* a desenvolver, torna-se relevante inquirir acerca dos diferentes tipos de *dashboards* que poderão ser desenvolvidos. Assim, podem ser consideradas três tipologias distintas (Hayward, 2021):

- *Dashboard* operacional: tem por objetivo a monitorização dos dados em tempo real ou transacionais em relação às principais métricas e KPIs, como o desempenho de vendas e os níveis de inventário. Estes foram concebidos para serem integrados no decurso do seu fluxo de trabalho diário, sendo que os trabalhadores, os supervisores e os chefes de equipa são os que mais beneficiam deste tipo de *dashboard*;
- *Dashboard* tático: utilizado para ligar o planeamento estratégico às atividades operacionais. Estas ferramentas oferecem dados em tempo real ou quase real para ajudar os utilizadores a tomar decisões informadas que afetam os objetivos e as estratégias a curto prazo. Enquanto os *dashboards* estratégicos se concentram em informações a longo prazo e os operacionais monitorizam as atividades diárias, os táticos oferecem informações detalhadas sobre o desempenho de um projeto ou departamento específico, desde o estado e os custos incorridos. Basicamente, acompanham o progresso das iniciativas em relação aos objetivos e metas estabelecidos;
- *Dashboard* estratégico: utilizado para monitorizar o estado dos KPIs relacionados com o desempenho geral da empresa, incluindo a sua posição no mercado e o seu desempenho financeiro. Ao contrário dos painéis de controlo operacionais, estes são atualizados com menos frequência, uma vez que se centram nas tendências e objetivos de longo prazo. Os executivos e os gestores de topo podem consultar estes *dashboards* de forma semanal, mensal, trimestral ou anual, permitindo ver o panorama geral e fazer previsões em conformidade, não existindo a necessidade/utilidade da sua consulta de forma tão frequente como nos *dashboards* operacionais.

Tabela 2 - Tipos de dashboard (adaptado de Hayward, 2021)

Pergunta que se pretende dar resposta	Dashboard Operacional	Dashboard Tático	Dashboard Estratégico
Que problema queremos resolver?	Maior conhecimento dos dados e acesso a dados sensíveis ao tempo	Acesso a tendências ou conhecimentos mais profundos	Visão direta dos KPIs organizacionais de topo
Público-alvo	Gestores intermédios e as suas equipas	Analistas e equipa executiva	Diretores e equipa executiva
Período de reporte	Desempenho diário	Desempenho semanal	Desempenho mensal, trimestral, anual
Quais os objetivos propostos?	Maior sensibilização dos funcionários e acompanhamento dos objetivos	Definição de objetivos analíticos e maior visibilidade dos principais processos	Definição de objetivos estratégicos, atingindo metas de KPI

## 2.6. Abordagens para projetos de *Data storytelling*

No presente trabalho pretende-se aplicar os conceitos de visualização de dados e *storytelling* a uma base de dados específica, podendo posteriormente averiguar se a aplicação destes conceitos se manifesta numa capacidade positiva de resposta às questões colocadas *a priori*. Neste sentido torna-se relevante explorar diferentes abordagens de diferentes projetos de aplicação e exploração de *data storytelling*.

Alguns investigadores, optam por tentar aplicar e desenvolver novas *frameworks* e métodos de aplicar os conceitos de *storytelling*. No estudo apresentado por Zhang et al, é proposta uma *framework* de aplicação dos conceitos teóricos estudados anteriormente. Neste caso, os investigadores dividem a *framework* em três perspetivas distintas: Conceito, Componente e Procedimento.

Do ponto de vista do conceito, é definida a intenção e as estratégias gerais da narrativa de dados; do ponto de vista da componente, são delineados o espaço de conceção alargado para a narração visual de dados e os elementos-chave da sua estrutura de informação; do ponto de vista do procedimento, é explicado o processo de criação de conteúdos de narração de histórias de dados visuais e a transformação da informação em aspetos visuais (Zhang et al., 2022).

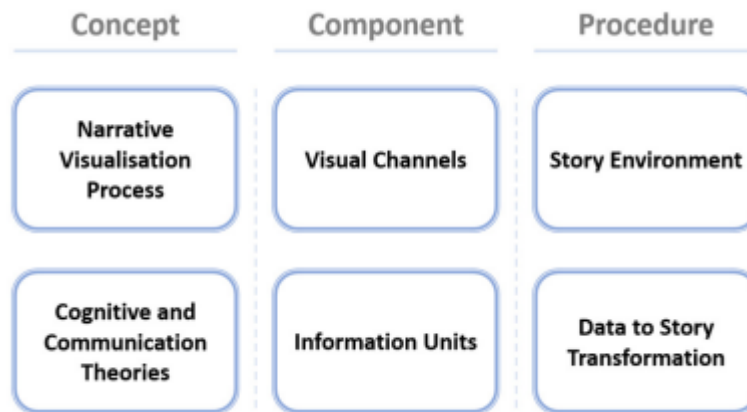


Figura 7 - Itens da *framework* de *data storytelling* (extraído de Zhang et al., 2022)

Outros investigadores, à semelhança deste presente trabalho, optam por aplicar os conceitos teóricos estudados a uma base de dados (real ou fictícia), de forma a perceber que práticas apresentam resultados mais promissores.

A título de exemplo, o estudo desenvolvido por Farias (2020) propõe-se a apresentar, através de um exemplo prático, técnicas que visam o aperfeiçoamento da visualização de dados (Farias, 2020). Através do desenvolvimento de *dashboards*, a fim de orientar os utilizadores de *business intelligence* a produzirem com foco em *storytelling* de dados, e considerando a escolha de um visual eficaz bem como a compreensão de contexto e público-alvo, propõe-se, dessa forma, alcançar a expectativa para o qual foi desenvolvido. No final propõe-se ainda apresentar um roteiro para a produção de um *dashboard* orientado a *storytelling* de dados. Neste caso de estudo específico, os dados utilizados dizem respeito a uma empresa de distribuição alimentar, cujo objetivo se prende com a apresentação de dados relacionados com KPI e métricas de análise de evolução da empresa, à sua equipa de gestão (Farias, 2020).

A figura abaixo apresenta um exemplo de um *dashboard* desenvolvido.

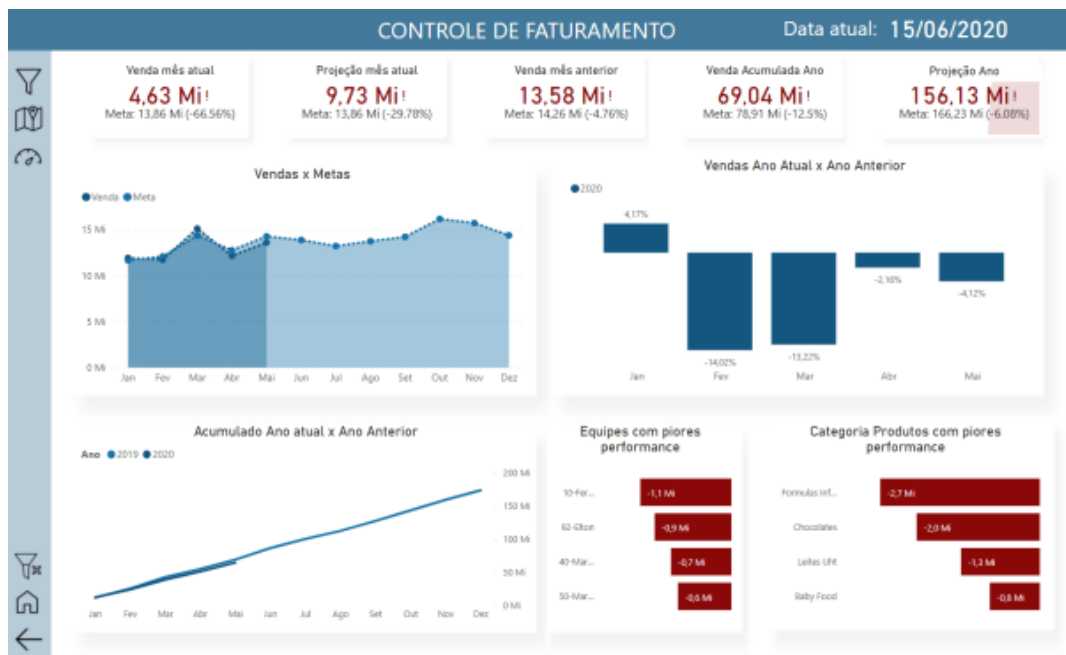


Figura 8 - Exemplo *dashboard* desenvolvido (extraído de Farias, 2020)

Em outros estudos, tem-se optado por verificar se a aplicação de técnicas de *data storytelling* se apresentam mais vantajosas, em comparação com as técnicas de visualização de dados convencionais, no que respeita a eficácia e eficiência na captação de ideias e compreensão de informação.

No estudo de Shao et al, são colocadas cinco questões de análise distintas (Shao et al., 2024):

- **Q1:** Em que medida é que o *data storytelling* permite discernir e compreender as informações críticas sobre os dados de forma mais **eficiente** do que as visualizações convencionais?
- **Q2:** Em que medida é que o *data storytelling* permite discernir e compreender os dados críticos de forma mais **eficaz** do que as visualizações convencionais?
- **Q3:** Em que medida as potenciais melhorias na eficiência e eficácia oferecidas pelos elementos de *data storytelling* variam entre as **tarefas de captação de ideias e compreensão de informação?**
- **Q4:** Quais são as percepções dos participantes sobre a utilidade dos **elementos de *data storytelling*** (por exemplo, anotações, destaques visuais e títulos)?
- **Q5:** Em que medida a relação entre a **literacia de visualização** dos indivíduos e a eficiência e eficácia das suas tarefas variam entre as condições de *data storytelling* e de visualização convencional?

Analisando os resultados obtidos, os investigadores formularam as seguintes indicações:

1. As técnicas de *data storytelling* podem ser mais eficazes do que a visualização convencional na comunicação de ideias (Q2), embora possam não ajudar necessariamente a encontrar as ideias mais rapidamente (eficiência) (Q1);
2. O *data storytelling* pode ser mais eficiente do que a visualização convencional para tarefas que envolvam a identificação de uma única ideia, mas não necessariamente para procurar pontos de dados específicos ou identificar ideias múltiplas/complexas a partir de uma visualização (Q3);
3. Para tarefas de captação de informação, especialmente para utilizadores com fortes conhecimentos de análise, pode ser preferível reconsiderar se realmente é necessário acrescentar elementos de narração de dados às visualizações convencionais (Q4);
4. A adição excessiva de elementos de narração de dados, em particular os que se baseiam em texto, pode dificultar tarefas simples ou distrair certos utilizadores (Q4);
5. O *data storytelling* ajuda na identificação eficaz de *insights* de forma relativamente igual para todos os utilizadores, independentemente dos seus níveis de literacia de visualização (Q5);
6. Da mesma forma, um nível mais elevado de literacia de visualização pode levar a uma identificação mais eficaz de ideias, quer utilizando *data storytelling* ou visualização convencional (Q5).

### **2.7. *Data storytelling* na era do *Big Data***

A terminologia *Big Data* tem sido utilizada em todo mundo para descrever o crescimento exponencial de dados, desde o início do presente século. O conceito foi utilizado pela primeira vez em 1997, sendo utilizado para nomear conjuntos de dados não ordenados e em rápido crescimento (Cox & Ellsworth, 1997). Atualmente, esta terminologia associa-se à área do conhecimento que se dedica ao tratamento, análise e extração de informações a partir de grandes volumes de dados. Para isto recorre-se a um conjunto de tecnologias, processos e práticas que permitem às organizações retirar proveito da análise destes dados com vista à tomada de decisão. Uma vez que o *Big Data* está associado a várias áreas e setores distintos, a definição deste conceito tem variado consoante as características de cada área.

Quando se aborda o tema do *Big Data*, devem considerar-se diversas variáveis que fazem parte da sua composição. Além do volume significativo de dados surge também a sua variedade, uma vez que estes são recolhidos de diferentes fontes como sistemas de gestão

empresarial, comentários nas redes sociais ou até mesmo vídeos. Outra das variáveis é a velocidade, uma vez que cada vez mais é necessário realizar análises em tempo real. Além destas, também a veracidade e o valor que os dados devem transferir para o negócio, são duas variáveis bastante relevantes acerca deste tema. Este conceito pode assim ser definido por 5Vs do *Big Data*, representados pelas anteriores variáveis (Anuradha, 2015):

- Volume: refere-se à grande quantidade de dados gerados por pessoas e máquinas. Esta variável define o tamanho do conjunto de dados que se traduz num problema de armazenamento e análise para a tecnologia de base de dados;
- Velocidade: refere-se à rapidez com que dados são gerados e distribuídos, que se traduz num desafio quanto à análise em tempo real de dados;
- Variedade: refere-se aos diferentes tipos e formas de dados que derivam de diferentes fontes (dados estruturados e não estruturados);
- Veracidade: refere-se à integridade dos dados, que se pode traduzir em falhas de qualidade como incerteza, imprecisão, valores perdidos e distorção de dados;
- Valor: refere-se à descoberta de conhecimento que está relacionada com a qualidade da análise do conjunto de dados.

A partir deste conceito é possível identificar diversos desafios e problemas técnicos que foram classificados em três grupos: dados, processamento e gestão.

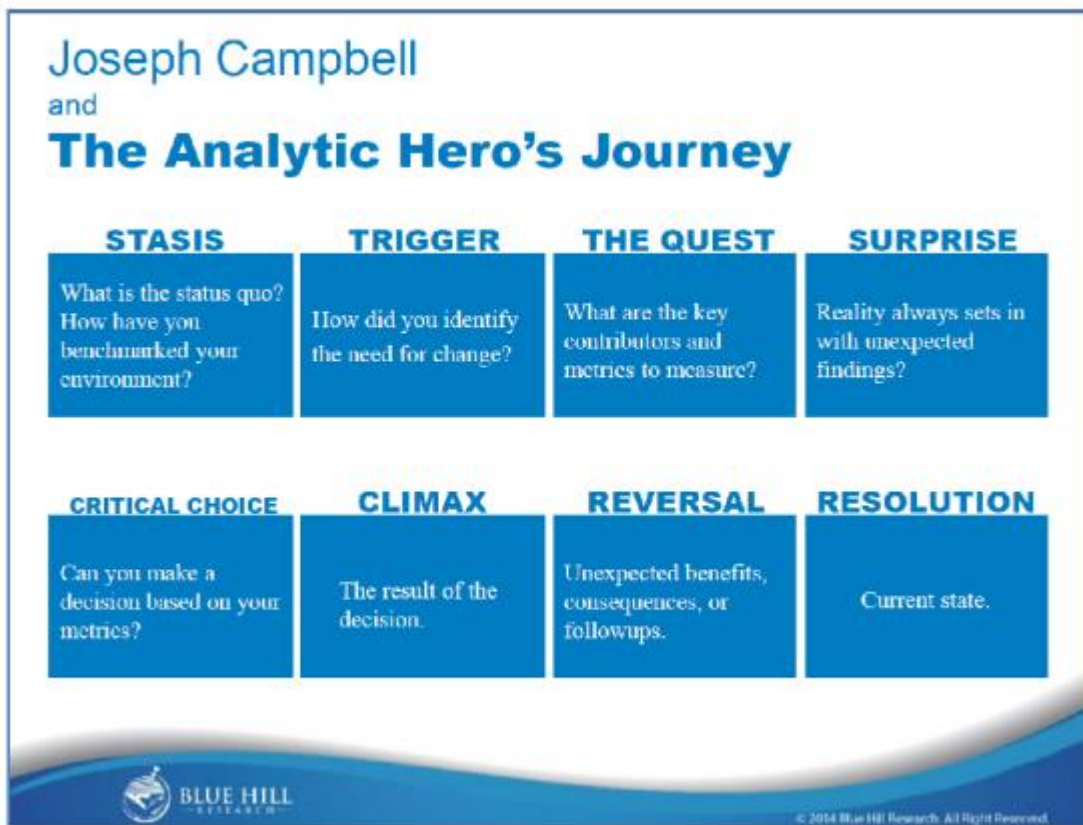
Os desafios de processamento relacionam-se com a recolha de dados, o cruzamento de dados de diversas fontes, a transformação e análise de dados e a representação visual final adequada ao processamento de informação humana. Os desafios de gestão estão relacionados com a segurança na recolha, no processamento e armazenamento de dados, que se traduz em problemas éticos e de privacidade controlados por políticas e regras dentro de uma organização ou noutro contexto que envolva armazenamento (Abdalla, 2022).

Uma vez que a definição de *Big Data* é um conjunto de conceitos teóricos que servem de base ao desenvolvimento de ferramentas para recolher, armazenar e processar dados, surgem desafios ao nível da recolha, armazenamento, análise, partilha, privacidade e visualização dos próprios dados. A visualização destes mesmos dados constitui uma técnica indispensável na transmissão de conhecimento, que está disponível na totalidade do volume de dados. A utilização de gráficos e esquemas constitui uma estratégia recorrentemente utilizada, desde há muito tempo, porém o complexo conjunto de dados torna a tarefa mais difícil e exigente. O desenvolvimento de técnicas de visualização para os dados heterogéneos que existem

atualmente serão úteis para simplificar e transformar a informação numa representação perceptível ao ser humano (Abdalla, 2022).

O objetivo da visualização é comunicar respostas a uma pergunta. Desta forma, alguns investigadores desenvolveram um *pipeline* generalizado de ciência de dados paralelizando-o com os elementos de uma história com princípio, meio e fim.

Em 2014, a Blue Hill Research adaptou a abordagem da Hero's Journey de Campbell à análise de dados, como mostra a Figura 9 (Behera & Swain, 2019).



**Figura 9 - *The Analytic Hero's Journey* (extraído de Behera & Swain, 2019)**

A análise de grandes volumes de dados requer a utilização das melhores tecnologias em todas as etapas, seja na recolha de dados, no processamento e na obtenção da conclusão final sob a forma de visualização (Behera & Swain, 2019). A melhor tecnologia ajuda a otimizar o poder de processamento do sistema global para a análise, que não só se pretende eficaz, mas preferencialmente em tempo real. A nível empresarial a análise de dados em tempo real tem-se mostrado uma arma poderosa no que respeita à compreensão dos seus dados, representando uma vantagem estratégica relevante. Além da análise de dados, também as ferramentas de visualização têm evoluído no sentido de representarem estes mesmos dados

em tempo real, ou com o mínimo de tempo possível, sendo alimentadas de forma automatizada pelos sistemas de recolha de dados.

Relativamente à recolha de dados uma das tecnologias mais utilizadas é o *Hadoop Distributed File System* (HDFS) que é um sistema de ficheiros distribuído baseado numa arquitetura mestre-escravo. No HDFS, um grande ficheiro único é dividido num número considerável de blocos, que são armazenados num conjunto de nós de dados. No que respeita ao processamento e armazenamento de dados a *Spark*, ferramenta da *Apache Software Foundation*, garante o aumento da velocidade de processamento, armazenando em cache os dados e os resultados no processamento em tempo real. Proporciona tolerância a falhas através do RDD (*Resilient Distributed Dataset*), que permite o armazenamento transparente de dados na memória e persiste no disco apenas quando necessário (Behera & Swain, 2019). Para aspetos de visualização, as ferramentas disponíveis serão apresentadas com maior detalhe nas secções posteriores.

## 3. Planeamento e Conceção do caso de estudo

No capítulo do Planeamento e Conceção do caso de estudo são apresentadas as diferentes etapas e processos que se seguiram de forma a desenvolver os *dashboards* finais, tendo por base a informação disponibilizada nos capítulos anteriores.

Desta forma, a literatura demonstra que têm sido desenvolvidos modelos e sistemas de classificação de maneira a otimizar a visualização de informação e adequar os gráficos e representações a cada tipologia de dado apresentado. Neste contexto, torna-se importante aferir que tipos de gráficos se apresentam mais representativos, para o espectador, ao comunicar diferentes tipos de informação, com diferentes objetivos de comunicação.

### 3.1. Caso de estudo: proveniência e análise

#### 3.1.1. Proveniência dos dados

Os dados que são tidos por base para suportar o trabalho apresentado são provenientes do projeto UniBench, já mencionado na secção 1.2, que apresenta o objetivo de desenvolver uma plataforma única de gestão de dados consolidada e eficiente que abranja uma variedade de coexistência de tipos de dados, a fim de identificar as mais valias e os problemas na integração de variedade em sistemas de bases de dados.

Neste caso, a base de dados utilizada refere-se a um sistema de *E-commerce*, em que os dados disponíveis se referem a características das encomendas, produtos vendidos, clientes, *feedback* dos clientes, entre outras características adicionais.

#### 3.1.2. Estrutura dos dados de suporte

Adicionalmente ao objetivo do projeto UniBench, que se prende com o desenvolvimento de um sistema de gestão de dados multimodelo, o objetivo deste trabalho foca-se na aplicação das melhores práticas ao nível da visualização de dados e *storytelling*.

Assim, os dados utilizados, apesar de inicialmente se encontrarem em diferentes formatos, foram transformados e transferidos para uma base de dados relacional, para, posteriormente, serem analisados com a ferramenta de BI. Neste caso específico, não é relevante a utilização de formatos de dados em vários tipos distintos, uma vez que o foco principal desta investigação não se centra na tipologia de dados a utilizar, mas sim no passo seguinte, ou seja, na forma como estes são apresentados ao utilizador final.

Assim, para este projeto foi desenvolvido um modelo de dados, constituído por seis tabelas que se relacionam tal como se pode observar na figura seguinte. O modelo de dados é dimensional, sendo composto por uma tabela de factos e tabelas de dimensões associadas à tabela de factos.

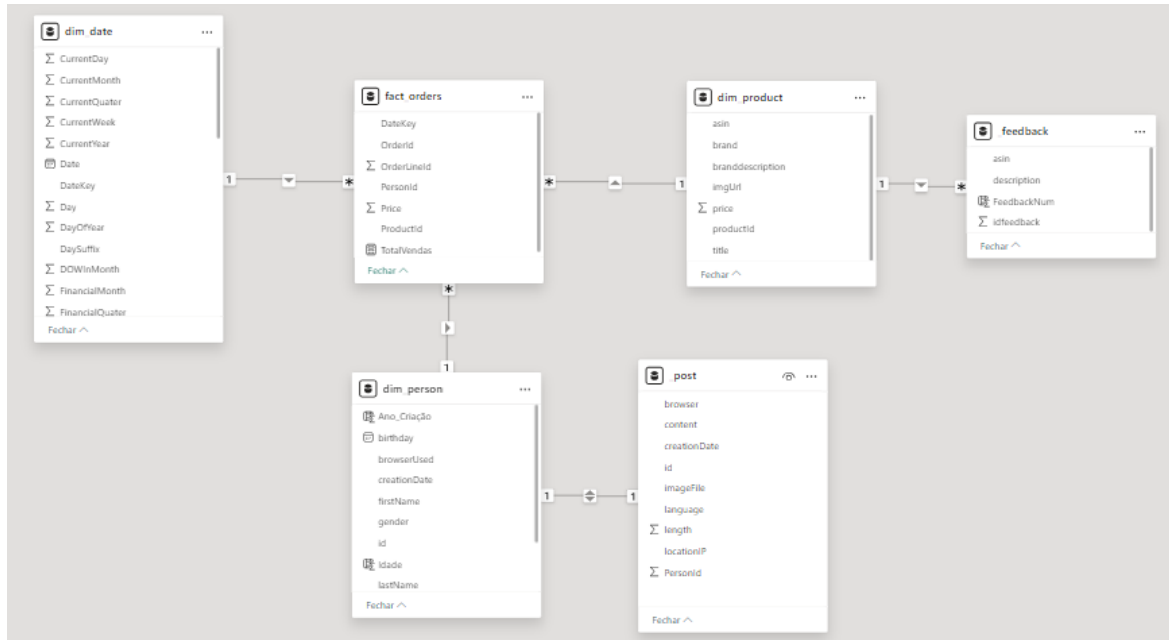


Figura 10 - Modelo de Dados

### 3.2. Estrutura do sistema a desenvolver

Numa primeira fase, após recolha dos dados desenvolvidos e agrupados no projeto UniBench, verificou-se que estes dados se encontram em diferentes formatos e tipos (XML, JSON, csv., entre outros). Para este trabalho, como referido anteriormente, não se manifestava necessária a diversidade de formatos, pelo que todos os dados recolhidos foram convertidos, de forma a criar uma base de dados relacional.

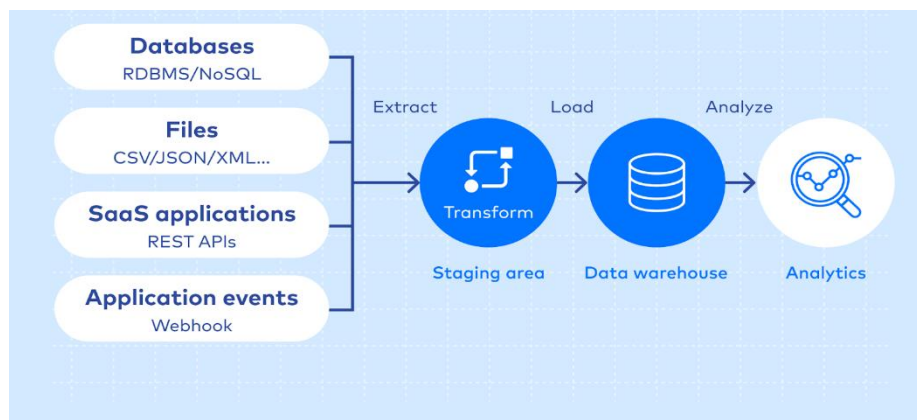


Figura 11 - Esquema de transformação e visualização de dados (extraído de Wang, 2022)

Assim, através da análise da Figura 11 é possível verificar que, numa fase inicial, os dados convertidos para formato relacional foram convertidos e agrupados num *data warehouse* (no caso foi utilizado o *SQL Server Management Studio*), no qual foram obtidos os dados. Após este processo, foi estabelecida uma conexão na ferramenta de *Business Intelligence*, e configurada de forma a permitir a atualização dos dados quando necessário. De seguida, já pela utilização da ferramenta de *Business Intelligence* são criados os *dashboards* para visualização dos dados disponibilizados de forma a dar resposta às diferentes questões colocadas *a priori*.

Por fim, a criar os *dashboards* é tido em conta o conceito de *storytelling*, de forma a criar a melhor experiência para o utilizador final, na qual este vai poder interagir com a ferramenta de *Business Intelligence*, de forma, a procurar dar resposta às questões colocadas.

### 3.3. Enunciado das questões a responder

As questões propostas para interação e resolução através da análise e visualização dos *dashboards* foram desenvolvidas tendo em consideração o seguinte método: em primeiro lugar foi analisada a base de dados de forma a perceber que tipologia de questões seria mais adequada a este caso de estudo; de seguida foram analisados alguns artigos relacionados com o projeto UniBench (Bimonte et al., 2020), que propõem algumas questões para análise da base de dados; a partir desta análise, foram selecionadas algumas questões consideradas pertinentes e que se adequassem à base de dados utilizada (uma vez que a base de dados utilizada não compreende a totalidade dos dados disponíveis pelo projeto UniBench, e desta forma algumas das questões não são passíveis de responder). As questões são na sua generalidade simples uma vez que se pretende obter valores iniciais para análise.

Q0 - Número de vendas por mês para determinados meses e anos

Q1 - Número de vendas por mês e género para determinados anos

Q2 - Número de vendas por mês e género para determinados produtos

Q3 - Valor total por ano para determinado género

Q4 - Valor total por idade para determinados navegadores

Q5 - Valor total por mês para determinados anos, géneros e marcas

Q6 – Ano com a maior percentagem de crescimento de faturação.

Q7 – Valor total das Top 5 marcas mais vendidas por ano

Q8 – Classificação atribuída para o Top 5 de marcas com maior preço médio

Q9 – Número de vendas por mês e idade do cliente

Q10 – Valor total por mês para determinados navegadores

### 3.4. Seleção de ferramentas

A visualização de dados encaixa na prática de *Business Intelligence*, e, por isso é vista como uma prática ou uma disciplina. Mais do que um conjunto de funções numa ferramenta de BI a visualização de dados transmite uma história ou um acontecimento, tornando o seu resultado preciso, atraente, oportuno e relevante.

A prática da visualização de dados apesar de melhorar a tomada de decisão ajuda a aumentar o poder da análise em toda a organização. O poder explicativo e exploratório da visualização de dados permite que o utilizador baseie as suas decisões em recursos visuais, uma vez que os meios visuais são mais eficazes do que os dados brutos. As ferramentas oferecem aos utilizadores recursos visuais para expor dados e funcionalidades de interação para uma análise rápida e vantajosa (Jääskeläinen, 2023). Com a evolução, cada vez mais acentuada, das tecnologias digitais, e em particular dos sistemas de análise e visualização de dados, surgiram ferramentas mais completas e capazes de transpor para o espectador a realidade exposta pelos dados.



**Figura 12 - 2024 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms (extraído de Schlegel, 2024)**

Na Figura 12 é apresentado o quadrante desenvolvido pela Gartner, no ano de 2024, que organiza as ferramentas de BI, tendo em conta a sua facilidade de utilização e a sua completude em termos de ferramenta, agrupando-as em quatro categorias distintas: Líderes, Visionários, Ferramentas Nicho e Desafiadores.

Tendo por base este quadrante apresentado, torna-se possível analisar com mais detalhe algumas ferramentas consideradas mais relevantes, nomeadamente alguns dos líderes de mercado (Jääskeläinen, 2023).

- Microsoft – apresenta um conjunto de ferramentas, conectores (personalizados e certificados) e serviços que permitem a visualização e análise de dados importados de diversas fontes desde ficheiros, servidores, serviços online e Azure. É composto pela aplicação Power BI que permite a transformação de dados para a visualização dos mesmos;
- Tableau – ferramenta que possibilita a criação de visualizações sendo elas visualizações da agregação de diversos dados, e a partilha destas no Tableau Server. O elemento principal da arquitetura do Tableau corresponde às várias fontes de dados a que se pode conectar sejam elas ficheiros ou servidores. O Tableau permite ainda a conexão a dados em tempo real;
- QlikView – plataforma de descoberta de dados para suporte à tomada de decisão. Os utilizadores podem interagir e visualizar os documentos e dados criados pela aplicação QV Developer que se conecta a diferentes fontes de dados (Data Warehouses, MS Excel, SAP, Salesforce, Oracle) que são autorizados a ver no QV Server e contém os documentos criados pelo QV Publisher no Back End;
- Looker (Google) – ferramenta de *business intelligence* considerada simples de utilizar e bem integrada com outros produtos Google, constituindo uma forma segura de aceder rapidamente aos dados brutos fornecidos. O Looker utiliza a sua própria marca de codificação de linguagem de pesquisa chamada LookerML.27. Pode conectar-se a diversas fontes de dados, como bases de dados SQL, serviços de *cloud*, e outras fontes de dados empresariais. Permite ainda a integração de dados em tempo real diretamente das fontes desejadas(Jääskeläinen, 2023).

Depois da análise de alguma das ferramentas existentes no mercado, optou-se por utilizar, para este trabalho, a ferramenta Power BI, da Microsoft, uma vez que apresenta uma versão gratuita com bastantes funcionalidades disponíveis e sem tempo máximo de utilização. Ao

contrário desta ferramenta, as restantes com opções gratuitas ou se apresentavam mais limitadas em termos de funcionalidades ou apresentavam um tempo experimental bastante reduzido, podendo comprometer a realização de todo o trabalho. Além destes fatores, pela análise da Figura 12, é possível verificar que esta é a ferramenta considerada como a mais simples de utilizar e, simultaneamente, a que se apresenta como uma ferramenta mais completa.

### **3.5. Planeamento dos *Dashboards***

#### **3.5.1. Definição de perfis de utilizadores**

Pela análise do tema apresentado na secção 2.5., é possível verificar que existem três tipos distintos de perfis de utilizadores: operacional, tático e estratégico.

Pelas diferenças apresentadas anteriormente, é possível perceber que o tipo operacional é aquele mais vocacionado para atividades diárias e relacionadas com processos operacionais e diários; o tipo tático encontra-se adequado para análises num espectro temporal mais alargado (semanal), analisando métricas simples do desempenho da operação da empresa; por fim, o perfil estratégico apresenta-se adequado para análise ao longo de espectros temporais mais longos (meses e anos), debruçando-se sobre a análise de KPI e métricas mais complexas com vista a servir de base para o delinear de aspetos estratégicos da organização.

Neste caso específico, o conjunto de *dashboards* desenvolvido corresponde a uma combinação entre os tipos tático e estratégico, tendo por base as definições apresentadas previamente. Isto ocorre, uma vez que o público-alvo pretendido corresponde à direção e equipa executiva da loja online, sendo que o objetivo se prende com a promoção de uma visão alargada sobre o desempenho da empresa e, conseqüente, suporte na definição de novos objetivos mediante as tendências verificadas através dos dados apresentados. Ainda assim, por ser um conjunto de *dashboards* que apresenta, maioritariamente, métricas simples, este não poderia ser considerado um *dashboard* do tipo estratégico. Desta forma, optou-se por unir os dois perfis criando um estado intermédio que engloba métricas simples, sendo capaz de promover uma visão de médio/longo prazo relativamente à atividade desenvolvida. É possível consultar o *dashboard* completo no Anexo B.

### 3.5.2. Desenho de *mockups*

Antes de ser iniciada a construção dos *dashboards*, foram desenvolvidos, previamente, os *mockups* relativos a cada um dos *dashboards* subsequentes. Nesta fase foram esboçados o que seriam os *dashboards* finais, tendo em conta todos os dados relevantes e a história de dados a apresentar.

Foram desenvolvidos dois tipos principais de *mockups*, que, por consequência, originaram, dois tipos distintos de *dashboards*:

- *Dashboards* de caracterização – têm por objetivo caracterizar clientes e produtos/marcas, não sendo estabelecida qualquer relação com número de vendas ou volume de faturação;
- *Dashboards* de Análise de Vendas – seguem-se aos *dashboards* de caracterização e é nestes que se apresentam os dados relativos a vendas e faturação, com a possibilidade de serem filtrados com as características disponíveis nos *dashboards* anteriores.

Na Figura 13 é possível verificar um exemplo de um *mockup* que deu origem a um *dashboard* de caracterização, neste caso a Caraterização de Clientes.

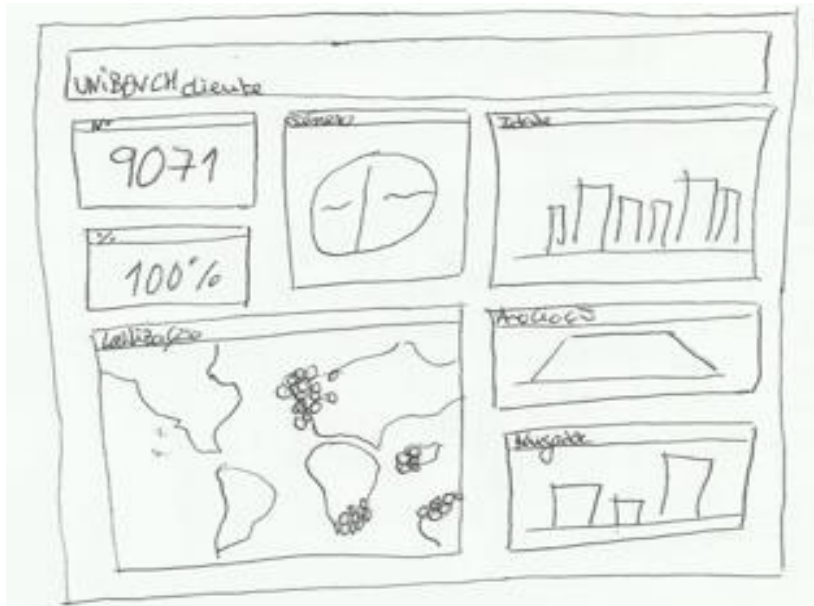


Figura 13 - *Mockup* Caraterização Cliente

O *mockup* apresentado demonstra as características essenciais dos *dashboards* a desenvolver, que serão aplicadas a todos os *dashboards* de caracterização.

- Barra superior - é apresentado o nome do projeto/empresa e aquilo que se visualiza em cada *dashboard* (Cliente, neste caso);
- Canto superior esquerdo – são apresentados dois cartões com informações relativas a número de clientes e percentagem do valor total (relevante quando aplicados filtros específicos);
- Corpo/centro do *dashboard* – são apresentados os gráficos/tabelas adequadas para cada tipo de dados apresentados.

Na Figura 14 é possível verificar um exemplo de um *mockup* que deu origem a um *dashboard* de análise de vendas.



Figura 14 - *Mockup* Análise de Vendas

O *mockup* apresentado demonstra as características essenciais dos *dashboards* a desenvolver, que serão aplicadas a todos os *dashboards* subsequentes.

- Barra superior - é apresentado o nome do projeto/empresa e aquilo que se visualiza em cada *dashboard* (Análise de Vendas neste caso);
- Eixo esquerdo - são apresentados dois tipos de informação distintos: dois cartões, na zona superior, com os valores relativos a número de vendas e volume de faturação;

e quatro cartões, nas zonas centro e inferior esquerda, correspondentes a filtros de informação (Género, Idade, Mês e Ano);

- Corpo/centro do *dashboard* – são apresentados os gráficos/ tabelas adequadas para cada tipo de dados apresentados.

Adicionalmente, na zona superior, podem ser adicionados filtros adicionais (Produto e Marca) que permitem filtrar os dados de forma mais específica.

No gráfico seguinte é possível verificar qual a forma mais adequada de se navegar pelos diferentes *dashboards*. Deve iniciar-se pela Página Inicial onde se selecionam os filtros adequados às visualizações que se pretendem; de seguida, segue-se para os *dashboards* de Caracterização de Cliente e Produto/Marca; por fim, segue-se para os *dashboards* de Análise de Vendas onde se irão visualizar questões relativas a vendas e volume de faturação. Nestes *dashboards* estão presentes alguns filtros que podem ser aplicados diretamente no *dashboard* em análise, porém, como nem todos os filtros estão presentes nestes *dashboards* pode ser necessário voltar à Página Inicial e redefinir os filtros a aplicar.

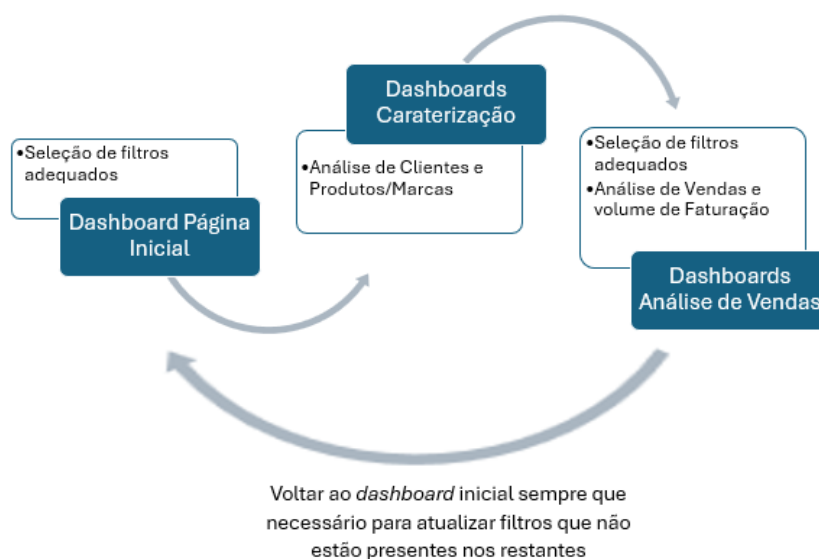


Figura 15 - Modo de navegação pelos diferentes *dashboards*

### 3.6. Planeamento do *Data Storytelling*

A primeira etapa de todo o processo de planeamento da história de dados, consistiu na definição do público-alvo desta mesma história, que no caso seria a direção e equipa

executiva da loja online. Após esta definição, o planeamento do *Data Storytelling* ocorreu pelo meio de três etapas distintas:

- Exploração dos dados: nesta fase foram agregados os dados provenientes da base de dados UniBench num Data Warehouse, através do qual se estabeleceu a comunicação com a ferramenta de BI. De seguida a base de dados foi editada, de forma a manter apenas os dados relevantes para a análise e adicionadas novas variáveis consideradas pertinentes. Após este processo, foi analisada a base de dados de forma a verificar qual a informação disponível e através da qual seria possível construir a história de dados a apresentar. No final desta fase foram definidas quais as medidas a utilizar para a apresentação, tendo em conta os dados disponíveis e o público-alvo.
- Construção da história: ao iniciar a construção da história de dados, foram tidos em conta as medidas disponíveis e que tipo de informação se pretende transmitir ao público. Uma vez que o público-alvo é a direção e equipa executiva da loja online, pretende-se transmitir informação relativa ao contexto atual do negócio, no que respeita a volumes de faturação e volumes de vendas. Optou-se assim por apresentar informação que permita estabelecer comparações entre diferentes períodos temporais, identificação de melhores casos de estudo e identificação de oportunidades de melhoria futura, passando ainda por alguns indicadores de desempenho considerados pertinentes na fase anterior. Durante esta fase foram definidos os elementos visuais que iriam constituir a história de dados, tendo a opção recaído sobre diferentes tipos de gráficos e cartões, suportados com algum texto explicativo.
- Apresentação da história: por fim, na última etapa do planeamento do *Data Storytelling* foi tido em conta o público-alvo da história, de forma a adaptar o conteúdo e forma de apresentação, assim como os objetivos pretendidos com a apresentação (maioritariamente relacionados com valores de faturação, comportamento das vendas ao longo do tempo e satisfação dos clientes com os produtos e marcas). Para finalizar, a apresentação ocorreu por meio de *dashboards*, nos quais foram aplicados filtros dinâmicos, de forma a criar uma maior envolvimento e detalhe, desenvolvidos com recurso à ferramenta Power BI, da Microsoft.

Através da análise da figura abaixo é possível identificar todas as etapas com maior detalhe.

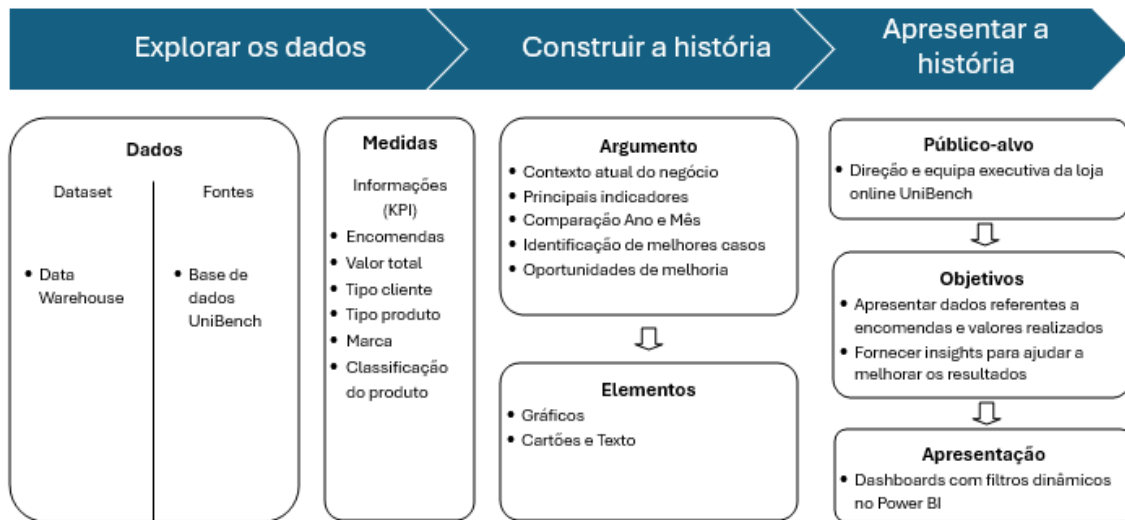


Figura 16 - Planeamento do Data Storytelling (adaptado de Farias, 2020)

## 4. Da conceção à narração: implementação do caso de estudo

Neste capítulo é onde se poderá verificar a aplicação prática de todos os conceitos estudados e planeados no capítulo anterior. Assim, nesta fase são implementados os *dashboards*, selecionando e definindo a organização dos elementos gráficos e a forma de navegação adequada pelos mesmos. De seguida, é também definida a história dos dados, tendo por base a organização definida anteriormente, de forma a criar um enredo com nexos e que cativa o público-alvo.

### 4.1. Implementação dos *dashboards*

#### Navegação pelos *dashboards*

Os *dashboards* desenvolvidos têm por objetivo corresponder às expectativas geradas e previstas para o perfil de utilizador idealizado. Este caso de estudo apresenta o objetivo de criar um conjunto de *dashboards* estratégicos/táticos, com o intuito de serem apresentados aos diretores e equipa executiva da loja de E-commerce UniBench. Desta forma, a construção dos *dashboards* tende a obedecer a uma ordenação lógica que conduz o utilizador a compreender a evolução das vendas da loja online analisada, permitindo verificar tendências de médio/longo prazo e tomar decisões informadas.

O conjunto de *dashboards* pode ser dividido em três secções distintas:

- Menu Inicial e Página inicial – servem de apresentação ao conjunto de *dashboards* que se seguem, englobando o título com a temática analisada, uma imagem ilustrativa, filtros a aplicar a todos os *dashboards* seguintes e as respetivas instruções de seleção dos filtros. A partir da página inicial é possível selecionar os filtros desejados de forma a serem apresentados apenas os resultados correspondentes a estes filtros, minimizando o tempo de análise; relativamente ao menu inicial é possível selecionar o *dashboard* a analisar a cada momento;
- *Dashboards* para caracterização de Cliente e Produto – nesta secção, constituída por dois *dashboards* distintos, é possível analisar a informação relativa, exclusivamente, a consumidores e produtos, não sendo apresentada qualquer informação relativa às vendas realizadas. Assim, torna-se possível caracterizar o

conjunto de clientes da loja online, destacando as suas características mais relevantes, assim como o conjunto de produtos disponíveis para venda e as suas respetivas características;

- *Dashboards* para análise de vendas – na terceira e última secção, é possível aceder à informação relativa às vendas e faturação da loja online, sendo possível cruzar esta informação com as características dos clientes e produtos. Nesta fase torna-se possível a verificação de tendências e padrões, bem como a evolução ocorrida ao longo do período em análise. É ainda nesta fase que se podem verificar se os objetivos traçados inicialmente estão a ser alcançados e de que forma se podem melhorar/atingir, uma vez que são utilizados elementos de inteligência temporal de forma a comparar períodos homólogos.

### **Seleção dos gráficos de suporte**

Tendo em consideração o tipo de dados disponíveis e o objetivo de análise para cada caso, foram selecionados diferentes tipos de gráficos com o intuito de disponibilizar a informação de forma organizada e intuitiva.

Neste caso de estudo, os elementos visuais mais utilizados foram:

- Gráfico de barras – utilizado para comparar elementos de diferentes categorias, no qual um dos eixos apresenta a categoria estudada e o outro eixo apresenta a escala de análise. Neste caso de estudo, uma das aplicações do gráfico de barras no *dashboard* de análise de vendas permite comparar diretamente o total de vendas realizadas por produto e marca, evidenciando aqueles que foram mais vendidos em determinado período temporal (encontrando-se igualmente o eixo horizontal ordenado pelo total de vendas, elucidando o utilizador relativamente aos produtos/marcas mais vendidos na loja online);
- Gráfico de barras agrupadas – semelhante ao gráfico de barras com a diferença de apresentar dois ou mais grupos de barras lado a lado. Possibilita a comparação entre os grupos para todas as categorias em análise. Neste caso de estudo pode verificar-se a sua aplicação num *dashboard* de análise de vendas onde é apresentado o total de vendas ao longo dos vários anos de atividade da loja online, sendo que as barras agrupadas apresentam as vendas do género feminino e do género masculino, permitindo a sua análise comparativa entre géneros;

- Gráfico circular – utilizado para analisar a proporção de determinada categoria dentro de um todo, possibilitando a identificação visual e intuitiva da prevalência de determinada categoria. Indicado para analisar categorias distribuídas em poucos grupos (5 a 6 grupos), sendo que neste caso de estudo apenas foi utilizado para analisar a proporção relativa ao gênero masculino e feminino, tendo em conta os clientes da loja online;
- Mapa de bolhas – utilizado para identificar elementos visuais associados a uma localização geográfica. Neste caso o maior tamanho das bolhas indica uma maior presença em determinada geografia. Foi utilizado para associar os clientes à sua respetiva geografia, indicando em que países/regiões se verifica um maior número de clientes;
- Gráfico de linhas – ideal para exibir dados quantitativos ao longo de um período, ou entre diferentes categorias. Neste caso de estudo, os gráficos de linhas foram associados a gráficos de barras de forma a estabelecer associações entre dois conjuntos de dados, quer de cariz temporal quer de cariz categórico. É exemplo deste caso um *dashboard* de análise de vendas em que o gráfico de linhas apresenta a evolução da faturação ao longo dos diferentes anos de atividade da loja online;
- Gráfico de área – semelhante ao gráfico de linhas, com um preenchimento colorido por baixo da linha. Neste caso, foi utilizado este tipo de gráfico porque é necessário observar qual é a tendência no registo na loja online de clientes ao longo do tempo, sendo que tanto este como os gráficos de linhas se apresentam adequados, por uma questão estética do *dashboard*, optou-se pelo gráfico de área;
- Gráfico de radar - ideal para comparar dados quantitativos com múltiplas variáveis, possibilitando o destaque das variáveis com valores discrepantes. Foi utilizado para verificar quais os meses com valores de encomendas superiores, sempre em associação com o seu ano específico;
- Gráfico de cascata - ajuda a ilustrar como um valor inicial é afetado por uma série de mudanças positivas ou negativas, sendo particularmente útil para entender a composição de um número e identificar as contribuições individuais de diferentes componentes. Neste caso, foi utilizado para analisar a evolução da faturação total, estabelecendo uma comparação entre os diferentes anos e a sua contribuição para o resultado;

- Medidor - utilizado para mostrar a performance de um único valor em comparação a um objetivo ou uma faixa de valores. Foi utilizado para identificar a média de preço de cada produto ou marca e a classificação atribuída a cada uma destas categorias. Permite identificar como é que estes valores se posicionam dentro do espectro total de valores possíveis;
- Tabela – neste caso de estudo, foram utilizadas, especificamente, para identificar de forma detalhada diversas categorias ou valores, por exemplo, no *dashboard* de caracterização do Produto, foi utilizada para apresentar características relativas às diferentes marcas. Tornam-se bastante úteis na medida em que permitem especificar conceitos e estruturar a informação desejada.

### **Organização dos elementos gráficos**

Tal como apresentado anteriormente, o conjunto de *dashboards* foi dividido em três secções distintas: Página inicial, Caracterização de Cliente e Produto e Análise de Vendas. Um dos elementos visuais comuns a todos os *dashboards* é o título apresentado, que remete e identifica a temática de cada *dashboard*.

Na página inicial e menu inicial, a nível de elementos gráficos apenas são apresentados títulos, imagens, textos explicativos e filtros a aplicar aos restantes *dashboards*. Nas duas secções seguintes é onde são apresentados os diversos gráficos, tabelas, cartões e filtros adicionais.

A nível de organização dos elementos dentro do próprio *dashboards*, optou-se por seguir um padrão comum para todos os *dashboards* apresentados. Na secção de Caracterização do Cliente e Produto são apresentados dois cartões, no canto superior esquerdo, que contém as informações que se pretendem destacar (nº de consumidores e % de consumidores – *dashboard* Cliente; nº de produtos e nº de marcas – *dashboard* Produto), seguindo-se de um conjunto de gráficos e tabelas que apresentam as categorias que caracterizam as duas temáticas.

Na Figura 17 é possível verificar o *dashboard* de caracterização do cliente, cujo *mockup* se encontra representado pela Figura 13.

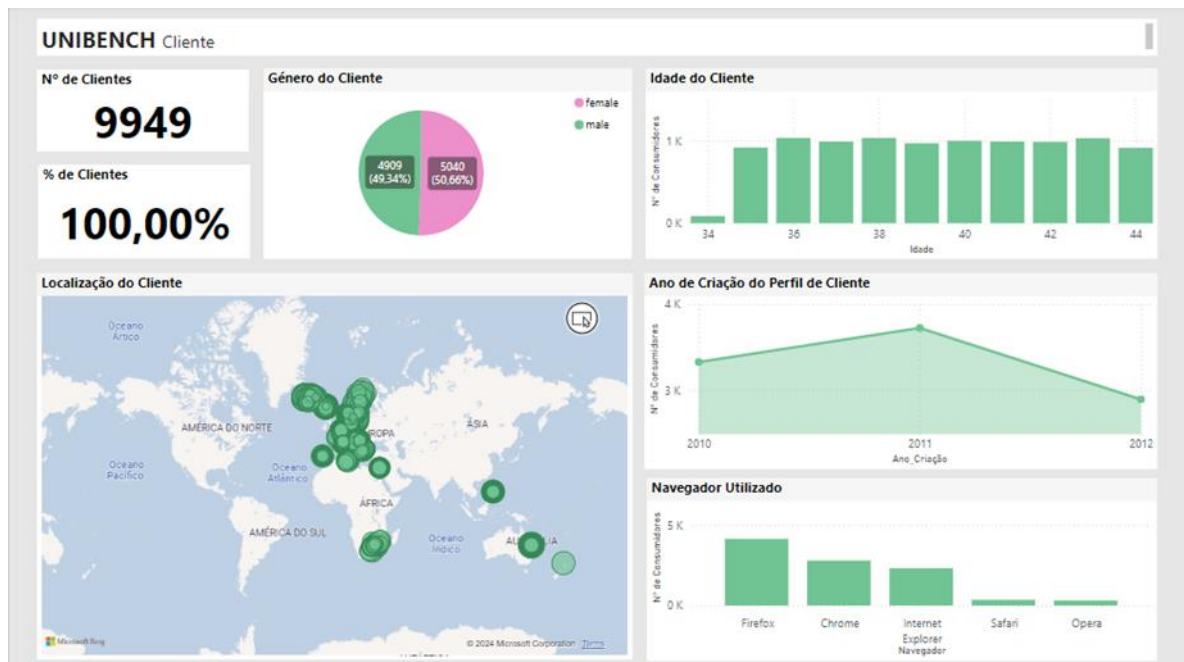


Figura 17 - *Dashboard* Caracterização Cliente

Na secção de Análise de Vendas são novamente apresentados dois cartões com as informações principais (nº de encomendas e valor de faturação), na mesma localização da secção anterior, e um conjunto de filtros localizados do lado esquerdo do *dashboard*. Adicionalmente, em alguns *dashboards*, são apresentados dois filtros na zona superior do mesmo. Além destes elementos são ainda apresentados o gráficos e tabelas que permitem fazer a análise das vendas desta loja online. Neste caso, optou-se por uma organização que se inicia com um tipo de análise mais abrangente (tendência ao longo do tempo e distribuição por mês e ano), seguindo para uma análise mais específica tendo em consideração as marcas e produtos vendidos, analisando o seu volume de vendas e de faturação, e termina-se com uma análise no âmbito da inteligência temporal, estabelecendo um comparativo entre períodos homólogos anteriores. Esta última permite verificar como tem ocorrido a evolução, quais as tendências verificadas e de que forma se pode atuar.



Figura 18 - *Dashboard Análise de Vendas (sem filtros)*

Na Figura 18 é possível verificar o *dashboard* de Análise de Vendas cujo *mockup* se encontra representado pela Figura 14.

De forma ilustrativa, na Figura 19 é representada a forma como deve ser feita a navegação/interação com os *dashboards*.

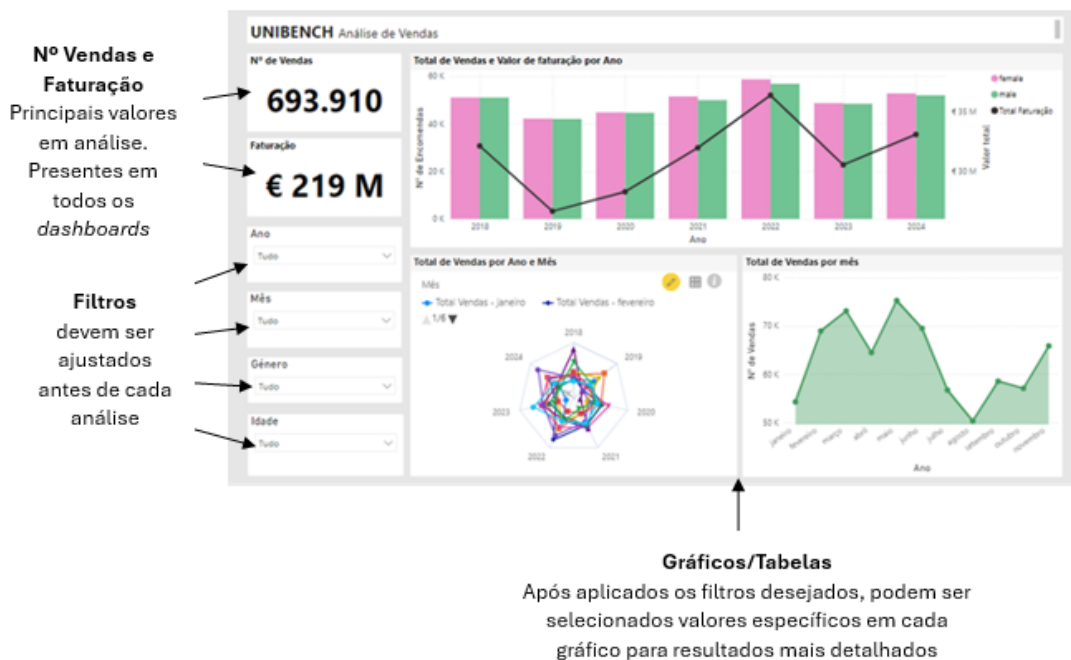


Figura 19 - *Forma de navegação pelo dashboard*

Os *dashboards* desenvolvidos e testados com utilizadores foram todos desenvolvidos para o tipo de utilizador tático/estratégico, porém, simultaneamente, foi desenvolvido um conjunto

de *dashboards* para um tipo de utilizador operacional, tendo em conta as características necessárias descritas no ponto 3.5.1. Este conjunto de *dashboards* pode ser consultado no Anexo C.

## 4.2. Conceção da história de dados

Como identificado anteriormente, o público-alvo deste estudo será a direção e equipa executiva do projeto UniBench, responsáveis por gerir e administrar o rumo da loja online analisada. Como resultado, pretende-se que através deste conjunto de *dashboards*, a equipa de direção seja capaz de entender a forma como a loja online tem evoluído, identificar pontos-chave para redirecionar a atuação da mesma e, em suma, otimizar os resultados obtidos. Assim, a história de dados a contar pretende fornecer esta contextualização, por forma a facilitar a tomada de decisão por parte da equipa.

De forma objetiva, e seguindo a sequência dos *dashboards* desenvolvidos, a história inicia-se pela análise dos clientes da loja online. Nesta fase é feita uma caracterização dos clientes, analisando como estes se distribuem quanto ao género, idade, localização geográfica (mundial), ano de criação do perfil de cliente e navegador utilizado. Aqui são destacadas as características que apresentam valores mais discrepantes e que possam indicar alguma tendência (localização e navegador, neste caso específico).

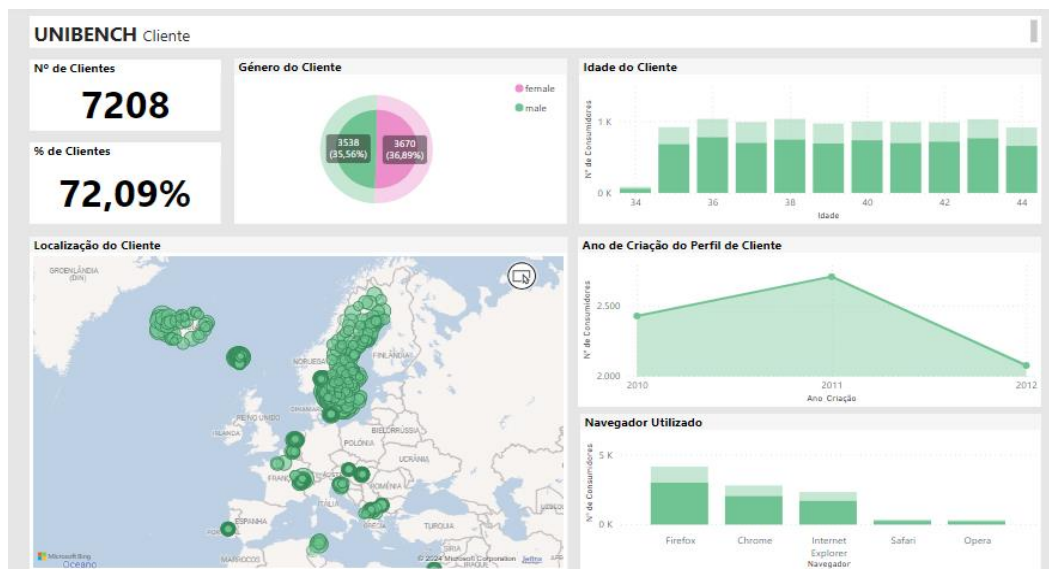
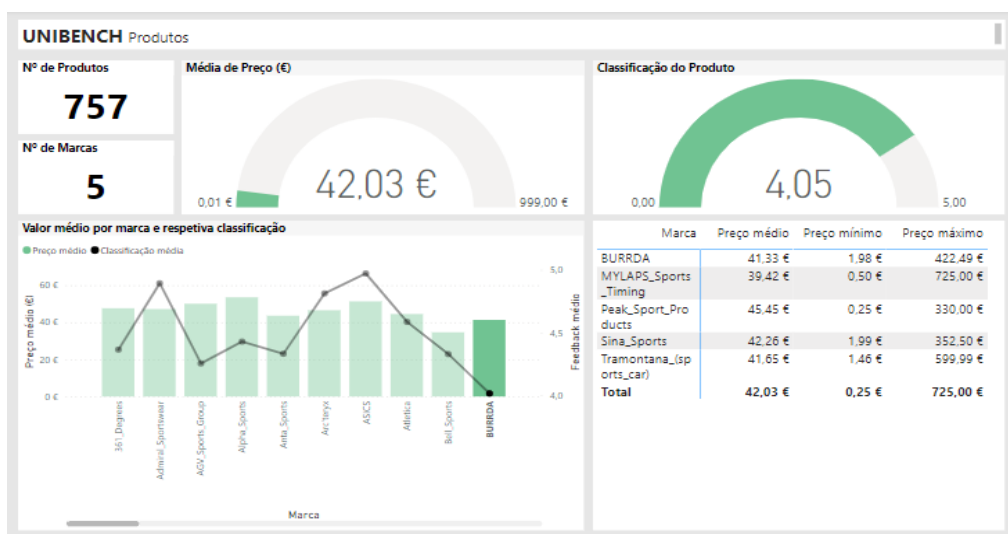


Figura 20 - Caraterização do cliente quanto à Localização geográfica (Europa)

Na Figura 20 é possível verificar o exemplo da aplicação da história para a localização geográfica dos clientes da loja online, na qual o mapa apresenta a sua localização específica

mais prevalente (Europa) e todos os gráficos ao redor caracterizam este conjunto de clientes selecionados.

Na segunda fase, segue-se a análise dos produtos e marcas disponíveis na loja online. Aqui são identificados o número de produtos e marcas distintas, sendo estes caracterizados, especialmente, pelo seu preço médio e classificação atribuída pelo cliente. Além destas características é ainda analisado o espectro de preços de cada marca, isto é, são analisados preços médios, mínimos e máximos por marca. Nesta fase são ainda realçadas as marcas com melhores classificações e com piores classificações, sendo estas comparadas com o seu preço médio, de forma a perceber se existe algum tipo de relação direta ou indireta.



**Figura 21 - Top 5 marcas com classificação mais baixa**

Na Figura 21 é possível visualizar um exemplo desta aplicação em que são apresentadas as 5 marcas com a classificação mais baixa atribuída pelo cliente. Pelos demais elementos gráficos torna-se possível relacionar a classificação com o preço médio da marca, permitindo uma retirar uma possível ilação entre estes dois fatores, para aplicação estratégica futura por parte da equipa de gestão.

Finalizando a análise das características de clientes e produtos, inicia-se a análise das vendas. Numa primeira fase, são verificadas as tendências gerais ao longo do tempo, ou seja, verifica-se a evolução da quantidade de encomendas e faturação ao longo dos sete anos em análise. É possível verificar que o ano com maior número de encomendas e maior faturação foi o ano de 2022, sendo que o mês de maio é aquele, que ao longo dos sete anos, obteve um maior volume de vendas. De seguida aprofunda-se um pouco mais a análise, correlacionando as marcas e produtos com o volume de encomendas e faturação. Aqui são analisados os

produtos e marcas mais vendidos, relacionando com o volume total de faturação gerado. Verifica-se que não existe uma relação aparente entre o volume de faturação e o volume de encomendas. Também nesta fase é possível verificar o crescimento/decrescimento anual da faturação, de forma percentual, sendo possível ainda analisar cada um destes pontos para cada produto e marca específica. Aproximando-se o final da história, analisa-se a evolução percentual da faturação comparando com períodos anteriores homólogos, de forma a entender em que anos se verificou um crescimento/decrescimento mais acentuado. Pelo gráfico é possível perceber que os anos de 2019 e 2023 foram anos de forte decrescimento, com quedas superiores a 16%. Já os anos de 2021 e 2022 foram anos de crescimento acentuado com taxas acima dos 13% ao ano. Para finalizar optou-se por verificar de que forma o melhor mês de vendas (maio), identificado anteriormente, se comportou ao longo dos vários anos, de modo a verificar se realmente a faturação se mantém constante ou foi fruto de algum acontecimento esporádico. Pelo gráfico verifica-se que houve um crescimento anormalmente acentuado no ano de 2020 (superior a 130%) face ao ano anterior, sendo que nos anos seguintes o valor mantém-se relativamente constante. Este exemplo concreto indica-nos que além da análise dos dados é necessária uma devida contextualização do negócio de forma a compreender de que forma o ambiente afeta a evolução dos resultados.

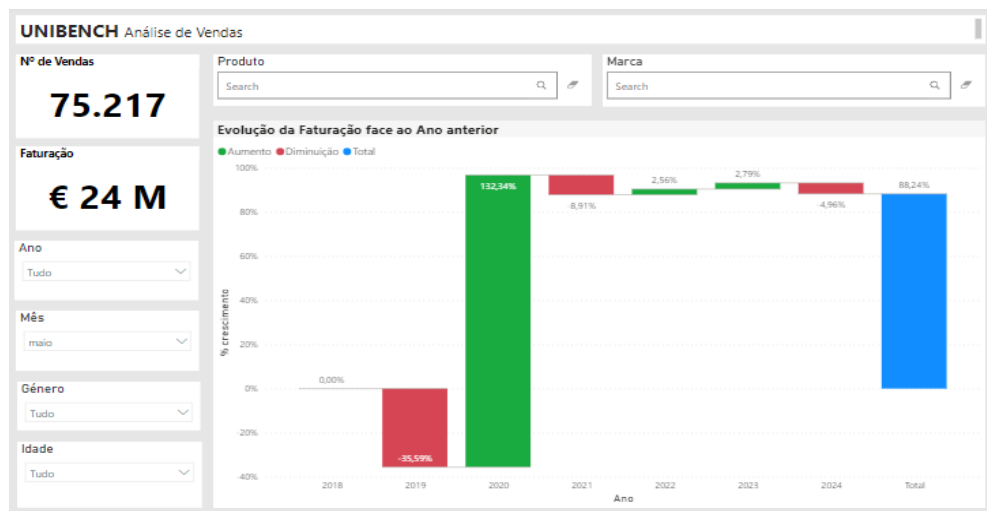


Figura 22 - Evolução da Faturação comparando os meses de maio

Durante a conceção da história dos dados, optou-se por não apresentar muitas anotações nem textos explicativos, uma vez que sendo os *dashboards* dinâmicos, poderia criar algum ruído adicional. Porém foram adicionadas algumas descrições de forma a facilitar a interpretação de dados mais complexos. Um exemplo pode ser verificado na tabela que apresenta a taxa de crescimento face ao ano anterior, em que ao passar com o cursor sobre a tabela é apresentado um texto explicativo de como interpretar os ícones presentes na tabela.

## 5. Testes e Validação

Após o desenvolvimento dos *dashboards*, são realizados testes de usabilidade e de relevância do *storytelling* a uma amostra da população. Nesta fase pretende-se identificar utilizadores que desempenhem funções na área da gestão, cujas decisões a tomar e contexto laboral sejam de carácter tático/estratégico. Assim, o objetivo será testar a usabilidade dos *dashboards* e a relevância da apresentação do *storytelling* para as suas tomadas de decisão.

### **Critérios de inclusão de participantes:**

- Idade igual ou superior a 18 anos;
- Capacidade de compreensão e expressão em língua portuguesa;
- Desempenhar funções relacionadas com a área de gestão, em contexto profissional, nomeadamente tomada de decisões de carácter tático;
- Utilização anterior de *dashboards* e gráficos dinâmicos e interativos.

Os testes são constituídos por diferentes fases distintas:

- Explicação da organização/funcionamento dos *dashboards* - de forma a contextualizar o utilizador com os dados apresentados;
- Apresentação do *storytelling* – nesta fase conta-se a história dos dados, apresentando as características principais e mais relevantes dos dados;
- Utilizador interage com os *dashboards* – nesta fase o utilizador tem acesso ao conjunto de *dashboards* desenvolvidos, na qual executa uma série de processos de forma a dar resposta a onze questões relacionadas com a análise de dados. Durante a realização de cada tarefa será registado o tempo que cada tarefa demora a ser executada;
- Resposta/preenchimento do questionário final – na última fase o utilizador irá preencher um questionário que será constituído por três secções distintas: análise sociodemográfica; questionário de usabilidade; e questionário de opinião/perceção do *storytelling* apresentado.

## 5.1. Questionário de usabilidade com um grupo de participantes

Após a concepção dos *dashboards* e a sua utilização pela população-alvo, torna-se relevante testar a usabilidade do sistema, de forma a aferir se as técnicas aplicadas se mostram adequadas e apreciadas pelos utilizadores.

Assim, dado que a usabilidade num dado momento é definida pelo contexto da utilização de um sistema tem-se que, em geral, o modo como se mede a usabilidade também será necessariamente definido por esse contexto. A norma ISO 9241-11 (*ISO 9241-11:2018(En) Ergonomics of Human-System Interaction*, 2018) divide a medição da usabilidade em três componentes distintos que devem ser definidos em relação ao contexto de utilização:

- Eficácia (se o utilizador consegue efetivamente concluir as suas tarefas e atingir os seus objetivos);
- Eficiência (a medida em que gasta recursos para atingir os seus objetivos);
- Satisfação (o nível de conforto que sente ao atingir esses objetivos).

Um dos testes mais reconhecidos para avaliar a usabilidade de sistemas é o SUS (*System Usability Scale*), cujos objetivos se prendem com o fornecimento de uma medida das perceções subjetivas dos utilizadores sobre a usabilidade de um sistema; e ser realizado num curto espaço de tempo durante uma sessão de avaliação (Brooke, 2013).

### Como é constituído o questionário e como é efetuada a pontuação?

O questionário SUS é constituído por 10 itens, que correspondem a afirmações relativas à utilização do sistema, sendo estas distribuídas igualmente entre afirmações de cariz positivo e afirmações de cariz negativo. Cada item apresenta uma escala de pontuação que varia entre 1 e 5, na qual 1 corresponde a “Discordo totalmente” da afirmação e 5 corresponde a “Concordo totalmente” com a afirmação.

A contribuição da pontuação de cada item varia de 0 a 4. Para os itens 1, 3, 5, 7 e 9 (os itens com afirmações positivas), a contribuição da pontuação é a posição na escala menos 1. Para os itens 2, 4, 6, 8 e 10 (os itens com afirmações negativas), a contribuição é 5 menos a posição na escala. Em seguida, multiplica-se o somatório das pontuações, dos 10 itens, por 2,5 para obter o valor global do SUS.

A razão de se multiplicar o somatório por 2,5 apenas se justifica pela maior probabilidade, por parte dos utilizadores, de compreender uma escala que varia entre 0 e 100 em relação a

uma escala que varia entre 10 e 50, sendo que o importante é conseguir captar a atenção num curto espaço de tempo.

Naturalmente, a utilização de itens positivos e negativos leva a um nível de complexidade; as respostas às perguntas indicam a força de concordância ou discordância, pelo que discordar fortemente de uma afirmação negativa é equivalente a concordar fortemente com uma afirmação positiva. A desvantagem desta abordagem é, obviamente, a tendência para que as pontuações entre 0 e 100 sejam entendidas como percentagens.

Na Figura 23 é possível verificar as designações atribuídas a cada pontuação final do teste de usabilidade. Além da informação fornecida o valor médio da pontuação atribuída pelo SUS situa-se nos 68 pontos (Brooke, 2013).

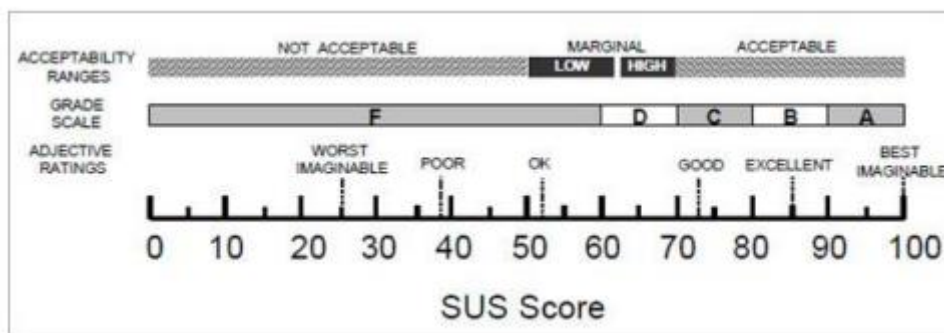


Figura 23 - Classificação da pontuação SUS (extraído de Brooke, 2013)

Relativamente ao número de participantes necessários para a correta realização de testes de usabilidade, vários artigos indicam números semelhantes. Segundo Nielsen o tamanho ideal da amostra situa-se nos cinco participantes, sendo que a sua explicação para este número é simplesmente matemática (Nielsen, 2000). No seu artigo, Nielsen expõe que o número de problemas encontrados num teste de usabilidade com  $n$  utilizadores seria:  $N(1-(1-L)^n)$ , onde  $N$  é o número total de problemas de usabilidade e  $L$  é a proporção de problemas descobertos ao testar um único utilizador. O valor típico de  $L$  é de 31%, calculado com base na média de um grande número de projetos estudados. Traçando a curva para  $L=31\%$  obtém-se o seguinte resultado:

Esta fórmula matemática demonstra que as maiores diferenças nos testes surgem nos primeiros três utilizadores e, depois, a curva estabiliza, sendo as novas descobertas insignificantes em comparação com as três primeiras. Assim, estatisticamente, é preferível fazer os testes de cinco em cinco do que fazer um grande grupo de utilizadores para testar (Nielsen, 2000).

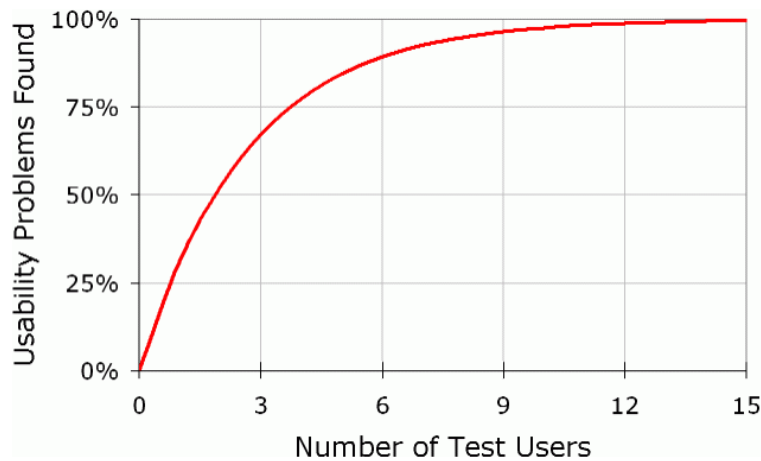


Figura 24 - Curva da descoberta de problemas em testes de usabilidade (adaptado de Nielsen, 2000)

Mais recentemente, outros estudos têm sugerido um tamanho ligeiramente superior da amostra, entre oito e doze participantes (Sauro, 2010).

Além do questionário de usabilidade, foi aplicado um questionário com o intuito de averiguar qual a relação, na perspectiva do utilizador, entre a aplicação da história de dados e a interação com os *dashboards* apresentados.

Este questionário é constituído por 6 itens, que correspondem a afirmações relativas ao impacto do *storytelling* na análise e interação com os dados, sendo estas distribuídas igualmente entre afirmações de cariz positivo e afirmações de cariz negativo. Cada item apresenta uma escala de pontuação que varia entre 1 e 5, na qual 1 corresponde a “Discordo totalmente” da afirmação e 5 corresponde a “Concordo totalmente” com a afirmação.

## 5.2. Avaliação dos resultados

Tendo por base o tamanho de amostra sugerido por diferentes estudos, foram identificados seis utilizadores, com o perfil indicado para realizar os testes referidos. Todos os utilizadores seguiram o mesmo protocolo de teste. Iniciou-se o processo por apresentar os *dashboards* e o tema em análise; de seguida, foi apresentada a história de dados; após a apresentação da história de dados seguiu-se um momento de interação com os *dashboards* de forma a que o utilizador se familiarize com a sua utilização, e, logo após, a aplicação das questões/tarefas ao utilizador; no final de todo o processo foi aplicado o questionário final que contém a componente sociodemográfica, questionário SUS e questionário relativo a *storytelling*.

## Caraterização sociodemográfica

Tabela 3 - Medidas descritivas de caraterização sociodemográfica

Variáveis		Medidas descritivas (n,%)
Género	Feminino	4 (66,7%)
	Masculino	2 (33,3%)
Idade	18-24	1 (16,7%)
	25-29	4 (66,7%)
	40-49	1 (16,7%)
Habilitações académicas	Licenciatura	2 (33,3%)
	Mestrado	4 (66,7%)
Área de estudo	Gestão	3 (50%)
	Contabilidade e gestão	3 (50%)
Atividade profissional	Consultor/Técnico de gestão	4 (66,7%)
	Gestor de projetos	1 (16,7%)
	Diretor geral	1 (16,7%)

Como apresentado na tabela anterior, a amostra é constituída por seis utilizadores, todos com áreas de estudo relacionadas com a gestão e com atividade profissional compreendida nas áreas de consultoria de gestão, gestão de projetos e direção geral de empresas. Todos os participantes apresentam idades compreendidas entre os 18 e os 49 anos (sendo que a maioria apresenta entre 25 e 29 anos), prevalência do género feminino (66,7%) e com habilitações académicas no mínimo ao nível da licenciatura.

### Questionário SUS

Como referido anteriormente, cada participante após utilizar e testar os *dashboards* desenvolvidos respondeu a um questionário de avaliação da usabilidade do sistema que utilizou previamente.

Na tabela seguinte é possível verificar as respostas oferecidas por cada um dos participantes, sendo apresentado o resultado ponderado para cada questionário.

Tabela 4 - Resultado das respostas do questionário SUS e resultado ponderado

Questão	U1	U2	U3	U4	U5	U6
<b>Acho que gostaria de utilizar este produto com frequência.</b>	5	4	5	4	5	4
<b>Considereei o produto mais complexo do que necessário.</b>	1	2	1	2	2	4
<b>Achei o produto fácil de utilizar.</b>	5	4	4	4	4	5
<b>Acho que necessitaria de ajuda de um técnico para conseguir utilizar este produto.</b>	1	2	2	3	1	1
<b>Considereei que as várias funcionalidades deste produto estavam bem integradas.</b>	4	4	5	5	4	5
<b>Achei que este produto tinha muitas inconsistências.</b>	1	2	1	1	2	1
<b>Suponho que a maioria das pessoas aprenderia a utilizar rapidamente este produto.</b>	5	4	5	5	2	5
<b>Considereei o produto muito complicado de utilizar.</b>	1	2	1	1	1	1
<b>Senti-me muito confiante a utilizar este produto.</b>	5	4	4	3	4	5
<b>Tive que aprender muito antes de conseguir lidar com este produto.</b>	1	2	1	1	1	1
<b>Total</b>	<b>97,5</b>	<b>75</b>	<b>92,5</b>	<b>82,5</b>	<b>80</b>	<b>90</b>

Através da tabela anterior, é possível verificar que os resultados ponderados se situam entre os 75 e os 97,5 pontos. Segundo a caracterização do questionário, apresentada no ponto 5.1, é possível afirmar que a partir da pontuação 73 a classificação é considerada com um nível “Bom”, pelo que, neste caso, todos os utilizadores consideraram a usabilidade do sistema no mínimo num nível “Bom”. De forma mais específica, as respostas dos utilizadores permitem enquadrar os resultados em duas categorias, no que respeita à sua usabilidade: “Bom” (3 utilizadores), que compreende as classificações entre 73 e 89,5; “Excelente” (3 utilizadores), que compreende as classificações entre 90 e 99,5.

### Questões/tarefas realizadas

Além do questionário SUS, uma outra forma de analisar a usabilidade e coerência dos *dashboards* é através da realização de algumas questões/tarefas, às quais o utilizador terá de dar resposta. Como apresentado no ponto 3.3, foram realizadas onze questões/tarefas a cada utilizador, e foi contabilizado o tempo que cada um demorou a realizar cada umas destas mesmas tarefas.

Na tabela seguinte é possível verificar o tempo que cada utilizador demorou a realizar cada tarefa, assim como a média ponderada para cada tarefa.

Tabela 5 - Tempo de realização de cada tarefa

Questão/Tarefa	U1	U2	U3	U4	U5	U6	Média
<b>Q0 - Número de encomendas por mês para determinados meses e anos</b>	50s	60s	57s	151s	21s	173s	<b>85s</b>
<b>Q1 - Número de encomendas por mês e género para determinados anos</b>	20s	25s	15s	56s	9s	10s	<b>23s</b>
<b>Q2 - Número de encomendas por mês e género para determinados produtos</b>	60s	108s	16s	61s	51s	101s	<b>66s</b>
<b>Q3 - Valor total por ano para determinado género</b>	30s	19s	19s	31s	23s	40s	<b>27s</b>
<b>Q4 - Valor total por idade para determinados navegadores</b>	31s	94s	92s	80s	75s	80s	<b>75s</b>
<b>Q5 - Valor total por mês para determinados anos, géneros e marcas</b>	37s	37s	46s	50s	47s	33s	<b>42s</b>
<b>Q6 - Ano com a maior percentagem de crescimento de faturação.</b>	22s	50s	79s	15s	41s	52s	<b>43s</b>
<b>Q7 - Valor total das Top 5 marcas mais vendidas por ano</b>	19s	77s	64s	51s	23s	57s	<b>49s</b>
<b>Q8 - Classificação atribuída para o Top 5 de marcas com maior preço médio</b>	20s	52s	30s	80s	73s	65s	<b>53s</b>
<b>Q9 - Número de encomendas por mês e idade do consumidor</b>	27s	29s	51s	44s	23s	30s	<b>34s</b>
<b>Q10 - Valor total por mês para determinados navegadores</b>	31s	36s	37s	53s	96s	75s	<b>55s</b>

Analisando a tabela anterior, é possível verificar que três tarefas/questões (Q0, Q2 e Q4) se destacam por apresentarem uma média de tempo, para a sua execução, visivelmente superior às demais, sendo as únicas com médias de execução superior a um minuto (60 segundos). Abordando especificamente cada uma destas tarefas/questões, é possível tentar encontrar algumas justificações para estas diferenças verificadas.

Relativamente à questão Q0, é possível assumir, que por ser a primeira questão, o utilizador ainda se encontra na fase de exploração inicial em que necessita de encontrar os elementos corretos, pelo que poderá constituir um fator relevante. Ainda assim, antes de iniciar qualquer questão foi dado um pequeno período de exploração dos *dashboards*. Na questão Q2, é possível verificar que a filtragem por produto ocorre de uma forma um pouco distinta dos demais filtros aplicados. Aqui, ao invés do utilizador selecionar o filtro que pretende através de uma lista pendente restrita, a seleção/filtro é realizada por uma de duas formas:

ou através da procura pelo produto através da escrita do seu nome, num filtro próprio, ou através da sua seleção no gráfico que apresenta todos os produtos vendidos, ordenado por número de vendas. Por este motivo, pode assumir-se que por ser uma tarefa cuja execução ocorre de forma distinta das restantes, surge alguma confusão e, naturalmente, a resposta ocorre de forma mais tardia. Relativamente à tarefa Q4, mais uma vez a razão da sua demora pode dever-se à forma como é executado o filtro necessário. Neste caso, ao contrário dos restantes filtros, o filtro do “Navegador” apenas é passível de ser aplicado na página inicial (sendo assim aplicado a todo o conjunto de *dashboards*), enquanto os restantes filtros podem ser aplicados na página inicial, mas também em cada *dashboard* específico. Isto promove alguma confusão uma vez que o utilizador, ao estar familiarizado com a filtragem dos dados em cada *dashboard*, vê-se obrigado a recuar até à página inicial para executar este filtro.

### Questionário *Data Storytelling*

De forma a averiguar a opinião dos utilizadores relativamente à implicação do *data storytelling* na interpretação, análise e interação com os *dashboards*, foi desenvolvido um questionário aplicado no final de todo o teste. Neste sentido, pretende-se aferir se existe alguma relação, percecionada pelo utilizador, entre o *data storytelling* desenvolvido e a forma como a interação se realiza com os *dashboards*.

Na tabela seguinte é possível verificar as respostas conferidas por cada utilizador, a cada uma das questões aplicadas.

Tabela 6 - Respostas ao questionário de *Storytelling*

Questão/Tarefa	U1	U2	U3	U4	U5	U6
<b>1 - Através da história de dados apresentada fui capaz de identificar as principais questões em análise.</b>	5	4	5	5	5	5
<b>2 - A história de dados apresentada criou mais confusão na minha interpretação dos dados.</b>	1	2	1	1	1	1
<b>3 - A história de dados apresentada facilitou a minha interpretação dos dados apresentados.</b>	4	4	5	5	5	4
<b>4 - A história de dados apresentada deu ênfase a tópicos que considerei pouco relevantes.</b>	1	4	1	1	4	3
<b>5 - Se não fosse apresentada a história de dados a minha navegação pelos diferentes dashboards seria mais difícil.</b>	4	4	4	5	5	1
<b>6 - A apresentação da história de dados apresentada não influenciou a minha interpretação dos dados.</b>	2	3	2	1	3	1

Através dos dados apresentados, é possível verificar que das seis questões aplicadas, três delas são de caráter positivo (1, 3 e 5) e as restantes três são de caráter negativo (2, 4 e 6). Nas questões de caráter positivo, praticamente todas as respostas, de todos os utilizadores, apresentaram respostas entre 4 (concordo) e 5 (concordo totalmente), o que permite aferir que, tendencialmente, os utilizadores em causa consideram que a história de dados permitiu identificar as principais questões em análise, apresentadas nos *dashboards*, facilitou a interpretação subsequente dos dados apresentados e facilitou a navegação pelos diferentes *dashboards*.

Relativamente às questões de caráter negativo, a interpretação dos dados não se apresenta tão linear como nas questões de caráter positivo. Na questão 2, as respostas conferidas pelos utilizadores apresentam-se relativamente homogêneas, com cinco dos utilizadores a discordar totalmente da afirmação e um utilizador a discordar da mesma, indicando que todos consideram que a história de dados apresentada não constitui um fator de confusão na interpretação dos dados apresentados. Na questão 4, as respostas dos utilizadores apresentam-se muito pouco homogêneas, com uma grande dispersão entre valores. Neste caso, três utilizadores discordaram totalmente da afirmação, dois destes concordaram e um utilizador nem concorda nem discorda da mesma. Desta forma, é possível entender que a história apresentada deu ênfase a um conjunto de tópicos cuja relevância não é unânime entre os utilizadores. Na questão 6, as respostas conferidas pelos utilizadores concentram-se entre os valores 1 (discordo totalmente), 2 (discordo) e 3 (nem concordo nem discordo), com uma distribuição igual entre os três valores. Neste caso, à exceção de dois utilizadores, que nem concordam nem discordam da afirmação, todos os restantes quatro utilizadores discordam, de forma mais ou menos veemente, da ideia de que a história de dados não influenciou a sua interpretação dos dados.

## 6. Discussão de Resultados e Considerações Finais

### 6.1. Discussão de Resultados

O projeto desenvolvido assume como principal objetivo o estudo de técnicas de visualização de informação e técnicas de *data storytelling* quando aplicadas ao desenvolvimento de *dashboards*.

Como exposto anteriormente, os *dashboards* são desenvolvidos a partir de uma base de dados elaborada no contexto do projeto UniBench, que recria uma loja online de produtos desportivos (Lu et al., 2023). Nesse projeto são gerados dados de forma aleatória, mas controlada, com o objetivo de avaliar a exequibilidade de coexistência no mesmo sistema de diferentes formas de armazenamento e representação. Contudo verifica-se que à data o sistema carece de mecanismos interativos para explorar e apresentar esses dados. Assim, verifica-se a necessidade de identificar quais são as tendências nesses dados disponíveis e aleatoriamente gerados sobre aspetos como a evolução ao longo do tempo e do espaço geográfico e de tópicos como as vendas e o volume de faturação. Por outro lado, o sistema desenvolvido é posto à consideração de utilizadores finais para avaliação de interação e perceção de usabilidade.

Pelos resultados apurados através de questionários SUS, é possível afirmar que as respostas conferidas pelos utilizadores permitem enquadrar os *dashboards* desenvolvidos em duas classificações: “Bom” (3 utilizadores), que compreende as classificações entre 73 e 89,5; “Excelente” (3 utilizadores), que compreende as classificações entre 90 e 99,5.

No desenvolvimento dos *dashboards*, são utilizados gráficos e outros elementos visuais selecionados e configurados de acordo o estudo bibliográfico realizado procurando facilitar a compreensão dos dados e a legibilidade. Neste contexto recorreu-se à utilização de cartões para apresentar os indicadores para a análise, de gráficos de barras para efetuar comparação entre valores, gráficos de linhas para visualização da evolução dos dados a longo do tempo e uma combinação entre estes dois anteriores para visualização de possíveis associações entre variáveis. Igualmente são utilizados mapas para caracterizar a distribuição geográfica dos clientes, tabelas para apresentar pequenas quantidades de valores e gráficos circulares na presença de um número limitado de variáveis.

Como descrito anteriormente, os *dashboards* foram desenvolvidos tendo por base as orientações na literatura estudada. O jogo de cores utilizado pretende a criação de um aspeto homogéneo e neutro, promovendo, quando necessário, o destaque. Por estes motivos anteriormente descritos, a classificação da usabilidade do sistema desenvolvido apresentou como classificação “Bom”.

Com descrito no protocolo de realização dos testes, a usabilidade do sistema foi avaliada por intermédio do questionário SUS e pela realização de um conjunto de tarefas relacionadas com a interação com os *dashboards*, nas quais foi contabilizado o tempo de execução dessas tarefas. No total cada utilizador realizou onze tarefas distintas, abrangendo todos os *dashboards* desenvolvidos.

Das onze questões em análise é possível verificar que três destas tarefas apresentam tempos de execução superiores às demais, sendo as únicas cujo tempo médio de execução ultrapassa os 60 segundos. Para cada uma destas tarefas são identificados motivos que podem justificar o elevado tempo de realização.

A primeira tarefa/questão que apresenta um tempo de execução superior às demais é a Q0. Uma das justificações prende-se com o facto de ser a primeira tarefa realizada e o utilizador ainda se encontrar na fase de exploração. A tarefa proposta apresenta-se semelhante (a nível de processos) a outras questões apresentadas posteriormente, mas verifica-se um tempo de execução mais prolongado. Além disto, esta tarefa proposta não se apresenta particularmente exigente a nível de processos de busca, uma vez que os filtros requeridos e o resultado esperado podem ser executados e visualizados em diversos *dashboards*, recorrendo a elementos visuais.

A segunda tarefa/questão que apresenta um tempo de execução elevado é a Q2. Neste caso, é necessário aplicar um filtro usando o campo designação do produto que, como referido na secção de Avaliação dos Resultados, apresenta um mecanismo de interação diferente da maioria dos outros filtros. Ao invés do utilizador aplicar o filtro através de uma lista pendente restrita, aplica o filtro de uma de duas formas: ou através da escrita da sua designação/expressão contida ou através da sua seleção num gráfico de barras que ordena os produtos por ordem decrescente do número de vendas. Constata-se que a aplicação deste filtro não se apresenta tão intuitiva como os demais contribuindo para tempos maiores de realização. Uma vez que nesta fase inicial a filtragem se processa de forma diferente das demais filtragens, é gerada, inicialmente, alguma confusão pois o utilizador vê-se forçado a

escrever o nome de um produto específico/expressão contida no nome ou a selecionar diretamente num gráfico de barras. Esta forma de pesquisa e filtro por produto é considerada devido à quantidade elevada de produtos (9600).

A terceira e última tarefa/questão que apresenta um tempo de execução superior às demais é a Q4, que, mais uma vez a forma como é executado o filtro necessário (“Navegador”) pode condicionar a realização da tarefa. Neste caso específico, não é possível apresentar todos os filtros em todos os *dashboards*, pelo que se optou por colocar todos os filtros gerais na página inicial e, subsequentemente, adicionar filtros mais específicos em cada *dashboard*. Naturalmente, o utilizador tenderá a recorrer aos filtros que dispõe em cada *dashboard*, evitando regressar à página inicial. Porém este método de distribuição de filtros pode tornar-se uma ferramenta poderosa, pois, após a devida familiarização com o sistema, permite uma navegação mais eficaz evitando a aplicação de filtros específicos em cada um dos *dashboards*.

No que concerne à avaliação da relevância do *data storytelling* para a interação com os *dashboards* e interpretação dos dados, como referido anteriormente, foi aplicado um questionário de forma a aferir a opinião dos utilizadores acerca desta temática. A história apresentada pretendeu abranger todos os tópicos principais dos dados, ressaltando alguns pontos de destaque.

Como referido anteriormente, o questionário aplicado apresenta questões de carácter positivo (1, 3 e 5) e questões de carácter negativo (2, 4 e 6). As questões de carácter positivo inquerem o utilizador acerca da sua opinião quanto à capacidade de identificar tópicos em análise, facilidade em interpretar os dados apresentados e facilidade de navegação pelos *dashboards*, através do acesso à sua história de dados. Abordando especificamente as questões de carácter positivo, nestas, praticamente todas as respostas indicam concordância com as afirmações apresentadas, nomeadamente com valores entre 4 (concordo) e 5 (concordo totalmente). Assim, pode inferir-se que a história de dados apresentada foi capaz de transmitir os principais tópicos em análise, sendo responsável por apresentar eventuais padrões relevantes relativamente à evolução de vendas, faturação e perfil de clientes e produtos; foi responsável por facilitar a interpretação dos dados apresentados, através das informações transmitidas ao utilizador, encaminhando-o de forma a otimizar o seu tempo de exploração dos *dashboards*; e, foi capaz de facilitar a navegação pelos diferentes *dashboards*, orientando o utilizador

pelos diferentes gráficos e elementos visuais, guiando a navegação de forma a encontrar os dados desejados no mais curto espaço de tempo.

As questões de caráter negativo inquerem o utilizador acerca da sua opinião quanto ao fator de confusão criado pela história de dados, relevância dos tópicos apresentados e influência do *data storytelling* na interpretação dos dados. Ao contrário das questões de caráter positivo, nas questões de caráter negativo as conclusões a retirar não se apresentam tão “visíveis”, uma vez que as respostas conferidas pelos utilizadores se apresentam menos homogêneas. Ainda assim, na questão 2, os utilizadores apresentam respostas orientadas no mesmo sentido, onde um utilizador discorda da afirmação e cinco dos utilizadores discordam totalmente. Neste caso, é possível entender que nenhum dos utilizadores considera que a história de dados seja um fator que contribua para o aumento da confusão na interpretação dos dados. Não é possível afirmar que o *data storytelling* contribua para a redução da confusão, porém é possível afirmar que não constitui um fator que aumenta este fenómeno.

Relativamente à questão 4, as conclusões a retirar não se apresentam muito assertivas. Neste caso, a questão aborda a temática da apresentação de tópicos pouco relevantes para a análise dos dados. É possível verificar que três dos utilizadores discordam totalmente da afirmação, enquanto dois destes concordam e um nem concorda nem discorda. Desta forma, pode entender-se que a história apresentada deu ênfase a um conjunto de tópicos cuja relevância não é unânime entre os utilizadores. Para alguns, todos os tópicos apresentados são revestidos de relevância para a análise subsequente dos dados, enquanto para outros, alguns dos tópicos poderiam ser alterados por outros de maior relevância ou, simplesmente, omissos. Neste caso, a relevância da história relaciona-se com o perfil do utilizador e com a informação que procura explorar.

Na questão 6, as respostas conferidas pelos utilizadores concentram-se entre os valores 1 (discordo totalmente), 2 (discordo) e 3 (nem concordo nem discordo), com uma distribuição igual entre os três valores. Neste caso, à exceção de dois utilizadores, que nem concordam nem discordam da afirmação, todos os restantes quatro utilizadores discordam da ideia de que a história de dados não influenciou a sua interpretação dos dados. Isto indica que a maioria dos utilizadores considera que se sentiu influenciado pela história de dados, ao analisar e interagir com os *dashboards*. Ainda assim, alguns utilizadores não apresentam uma opinião convicta relativamente a este tema, o que pode indicar que o *data storytelling* não impacta os utilizadores da mesma forma. Pode entender-se que apesar dos utilizadores

se sentirem influenciados, essa mesma influência pode encaminhá-los num sentido mais ou menos acertado, no ponto de vista dos seus interesses.

Estabelecendo uma relação entre as questões 4 e 6, é possível verificar que os dois utilizadores que, na questão 4, concordam com o facto de a história de dados dar ênfase a tópicos pouco relevantes, são os mesmos que na questão 6 não concordam nem discordam com o facto de não se sentirem influenciados pela história de dados. A partir destas respostas é possível propor que os utilizadores que entendam que a história de dados apresenta tópicos de menor relevância, também se sentem menos influenciados por esta aquando da interpretação dos dados.

## **6.2. Conclusão, Limitações do estudo e Futuras linhas de investigação**

Em primeiro lugar importa ressaltar que a elaboração do estudo prévio, e o enquadramento teórico associado, permite sintetizar as principais contribuições relacionadas com a interseção entre visualização de dados, *data storytelling* e a adaptação dos *dashboards* ao perfil de cada utilizador. A partir do enquadramento teórico, foi possível reconhecer que a visualização de dados, quando integrada com o *data storytelling*, permite não apenas a apresentação de informações de forma acessível e compreensível, mas também o fortalecimento do impacto das mensagens transmitidas. Ao longo deste estudo, verificou-se que o *data storytelling* desempenha um papel fundamental na construção de narrativas que facilitam a interpretação de dados, especialmente quando essas narrativas são ajustadas de acordo com as necessidades, preferências e características de cada perfil de utilizador.

As abordagens, que conseguem combinar as técnicas visuais com o apoio das narrativas, são especialmente relevantes em cenários onde a tomada de decisão baseada em dados é essencial. A capacidade de transformar dados complexos em histórias significativas demonstra-se como um diferencial estratégico, potencializando a capacidade de absorção e retenção de informação.

Relativamente ao projeto aplicado, um dos principais pontos positivos a ressaltar prende-se com o facto do projeto UniBench não apresentar, até à presente data, nenhum sistema de representação visual que permita a apresentação da informação ao público-alvo interessado, com a vantagem de fornecer informação concreta e de rápida absorção, sobre o desempenho da loja online (Lu et al., 2023).

Quanto à história de dados pode inferir-se que esta constitui um fator de influência sobre os utilizadores, e promove a interpretação dos dados e a navegação pelos *dashboards*, desde que os tópicos abordados sejam considerados relevantes para o cômputo geral. Verifica-se que ao serem apresentados temas menos interessantes, para os utilizadores, de igual forma estes sentem-se menos influenciados pela história de dados apresentada, o que permite inferir que a análise prévia do perfil dos utilizadores constitui um aspeto preponderante no sucesso dos *dashboards* desenvolvidos e da história apresentada.

Pelos testes realizados, pode inferir-se que a história de dados apresentada foi capaz de transmitir os principais temas em análise, sendo capaz de expor padrões relevantes relativamente à evolução de vendas, faturação e perfil de clientes e produtos; foi também um motivo facilitador na interpretação dos dados apresentados, através da informação transmitida ao utilizador, encaminhando-o de forma a otimizar o seu tempo de exploração dos *dashboards*; e, foi ainda capaz de facilitar a navegação pelos diferentes *dashboards*, promovendo a orientação dos utilizadores, minimizando o tempo despendido em busca de determinado conteúdo.

Relativamente à usabilidade dos *dashboards*, foi possível constatar que os utilizadores dão primazia à simplicidade de utilização e navegação, sendo este um fator decisivo na sua escolha por determinada plataforma ou *dashboard*. É essencial que os filtros mais relevantes (filtros de data, género e idade, por exemplo) sejam de fácil acesso e que se apresentem bem visíveis, sendo relevante que sejam apresentados em vários painéis, de forma redundante.

Importa clarificar que deve ser desenvolvido um esquema de navegação conciso, cujo acesso aos dados seja coerente e que ocorra mediante processos semelhantes, de forma que o utilizador consiga identificar com facilidade a forma de resgatar determinados dados. Importa ainda que os elementos visuais apresentados sejam consistentes, mantendo um padrão visual e adaptados a cada tipo de análise. Por fim, torna-se relevante que cada *dashboard* (caso existam vários) aborde uma temática específica, tornando possível compartimentar os dados apresentados por segmentos de interesse. Para finalizar, importa ressaltar que a história de dados deve sustentar os elementos visuais conferindo uma fonte adicional de suporte à navegação pelos diferentes painéis, complementando a informação apresentada com a interpretação do próprio utilizador.

Em suma, torna-se relevante referir que os objetivos propostos para este trabalho foram cumpridos, nomeadamente a construção do sistema de *dashboards*, bem como o estudo da

sua usabilidade e da aplicação do *data storytelling*. Importa ressaltar que a classificação da usabilidade atribuída pelos utilizadores concentrou-se entre os valores de 75 e 97,5, ou seja, entre as classificações de Bom e Excelente, o que permite verificar que existe alguma margem para melhorar a qualidade dos *dashboards*, mas que, regra geral, foram apresentadas as características que os utilizadores consideraram necessárias e adequadas para este tipo de sistemas. Sobre a aplicação do *data storytelling* ficou claro que constituiu uma vantagem adicional na retenção de informação pelo utilizador, facilitando a navegação pelos *dashboards* e identificação dos principais tópicos em análise.

Relativamente às limitações apresentadas neste estudo, em primeiro lugar, importa referir que a base de dados que serviu por base a este estudo apresenta algumas lacunas a nível de variedade de dados, isto é, poderia apresentar um conjunto de dados mais relevantes para a construção dos *dashboards* para as equipas de gestão. Desta forma, durante os testes realizados, surgiram algumas sugestões de melhoria dos *dashboards* que se prendem com a adição de informação não existente, e que acrescentariam valor ao sistema desenvolvido. Em segundo lugar, seria interessante que todos os utilizadores que testaram o sistema desempenhassem funções de gestão em setores semelhantes ao apresentado no estudo, ou seja, venda de produtos desportivos, exclusivamente online. Apesar de todos os utilizadores apresentarem formação académica na área de gestão e atividade profissional em funções de gestão, nem todos estão ligados ao setor em análise, e, por este motivo, a análise e a perceção após a utilização do sistema desenvolvido poderá não traduzir todos os benefícios/lacunas apresentados por este.

No que diz respeito a futuras linhas de investigação, seria interessante explorar o comportamento destes sistemas de visualização de dados quando atualizados com dados em tempo real. Poderia ser relevante analisar a forma como os utilizadores interagem não só com os elementos visuais, mas de que forma a história de dados se adapta ao contexto desenvolvido. Além disto, seria também interessante poder dar continuidade ao trabalho desenvolvido, adicionando informações complementares, como sugerido por alguns utilizadores. Entende-se que a informação disponibilizada pode apresentar-se limitada em alguns contextos, não permitindo uma análise global de toda a estrutura e evolução do negócio, e, desta forma, não permite um verdadeiro envolvimento dos utilizadores. Assim, e de forma ainda mais interessante, seria relevante melhorar a qualidade da base de dados apresentada, tornando o sistema apresentado mais completo e dinâmico, complementando-o com atualização dos dados em tempo real.

Além dos pontos referidos, seria interessante aplicar o tema estudado a diferentes casos de estudo, com vista a comprovar as conclusões retiradas com a aplicação deste caso. Só desta forma será possível tentar generalizar as conclusões retiradas, tornando possível a sua aplicação a diferentes áreas, perfis de utilizador e diferentes tópicos.

## Referências Bibliográficas

- Abdalla, H. B. (2022). A brief survey on big data: technologies, terminologies and data-intensive applications. *Journal of Big Data*, 9(1), 107. <https://doi.org/10.1186/s40537-022-00659-3>
- Ackoff, R. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16(1), 3–9.
- Anuncia, S., Gohel, A., & Vairamuthu, S. (2020). *Data Visualization Trends and Challenges Toward Multidisciplinary Perception*. Springer.
- Anuradha, J. (2015). A brief introduction on big data 5Vs characteristics and hadoop technology. *Procedia Computer Science*, 48, 319–324. <https://doi.org/10.1016/j.procs.2015.04.188>
- Behera, R. K., & Swain, A. K. (2019). Big data real-time storytelling with self-service visualization. *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018*, 813, 405–415. [https://doi.org/10.1007/978-981-13-1498-8\\_36](https://doi.org/10.1007/978-981-13-1498-8_36)
- Bertin, J. (2011). *Semiology of Graphics: Diagrams, Networks, Maps* (ESRI Press, Ed.).
- Bharadiya, J. P. (2023). Machine Learning and AI in Business Intelligence: Trends and Opportunities. *International Journal of Computer (IJC)*, 48(1), 123–134. <https://www.researchgate.net/publication/371902170>
- Bimonte, S., Hifdi, Y., Maliari, M., Marcel, P., & Rizzi, S. (2020). To Each His Own: Accommodating Data Variety by a Multimodel Star Schema. In *Proceedings of the 22nd International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data Co-Located with EDBT/ICDT 2020 Joint Conference*. <https://hal.science/hal-03009808>
- Brent Dykes. (2016, March 31). *Data Storytelling: The Essential Data Science Skill Everyone Needs*. Forbes. <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>
- Brooke, J. (2013). SUS: A Retrospective. *Journal of Usability Studies*, 8(2).
- Card, S. T., Mackinlay, J. D., & Scheiderman, B. (1999). *Readings in Information Visualization: using vision to think*. Morgan Kaufmann.
- Carvalho, E. S., & Marcos, A. F. (2009). Visualização da Informação. *Centro de Computação Gráfica (CCG)*. <https://www.researchgate.net/publication/236863622>
- Chang, D., Dooley, L., & Tuovinen, J. (2002). *Gestalt Theory in Visual Screen Design-A New Look at an Old Subject*. <http://crpit.com/abstracts/CRPITV8Chang.html>

- Cox, M., & Ellsworth, D. (1997). Application-controlled demand paging for out-of-core visualization. *Proceedings. Visualization '97 (Cat. No. 97CB36155)*, 235–244. <https://doi.org/10.1109/visual.1997.663888>
- Echeverria, V., Roberto, M.-M., & Buckingham Shum, S. (2017). Towards data storytelling to support teaching and learning. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (pp. 347–351).
- Farias, E. (2020). *Storytelling de dados: Contando histórias com dashboards* [Master's thesis]. Universidade do Sul de Santa Catarina.
- Feigenbaum, A., & Alamalhadaei, A. (2020). *The data storytelling workbook*. Routledge.
- Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press.
- Gina, B., & Budree, A. (2020). A review of literature on critical factors that drive the selection of business intelligence tools. In *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (IcABCD)*, 1–7. <https://doi.org/10.1109/icABCD49160.2020.9183852>
- Hayward, E. (2021, June). *4 Types of Dashboards: Operational, Analytical, Strategic & Tactical*. Klipfolio. <https://www.klipfolio.com/blog/starter-guide-to-dashboards>
- Heer, J., & Shneiderman, B. (2012). Interactive dynamics for visual analysis: a taxonomy of tools that support the fluent and flexible use of visualizations. *Queue*, 10(2), 30–55. <https://doi.org/10.1145/2133806.2133821>
- Hess, K. (2022). *80 types of charts & graphs for data visualization*. <https://www.datylon.com/blog/types-of-charts-graphs-examples-data-visualization>
- ISO 9241-11:2018(en) Ergonomics of human-system interaction*. (2018). ISO.
- Jääskeläinen, P. (2023). *Performance Reporting Tools, Integrability and Information Flow*.
- Jiang, T., Hou, Y., & Yang, J. (2023). Literature Review on the Development of Visualization Studies (2012–2022). *Engineering Proceedings*, 38(1), 89. <https://doi.org/10.3390/engproc2023038089>
- Keim, D. A. (2002). Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 1–8.
- Knaflic, C. N. (2015). *Storytelling with data: a data visualization guide for business professionals*. John Wiley & Sons.
- Lee, B., Riche, N. H., Isenberg, P., & Carpendale, S. (2015). More Than Telling a Story: Transforming Data into Visually Shared Stories. *IEEE Computer Graphics and Applications*, 35(5), 84–90.

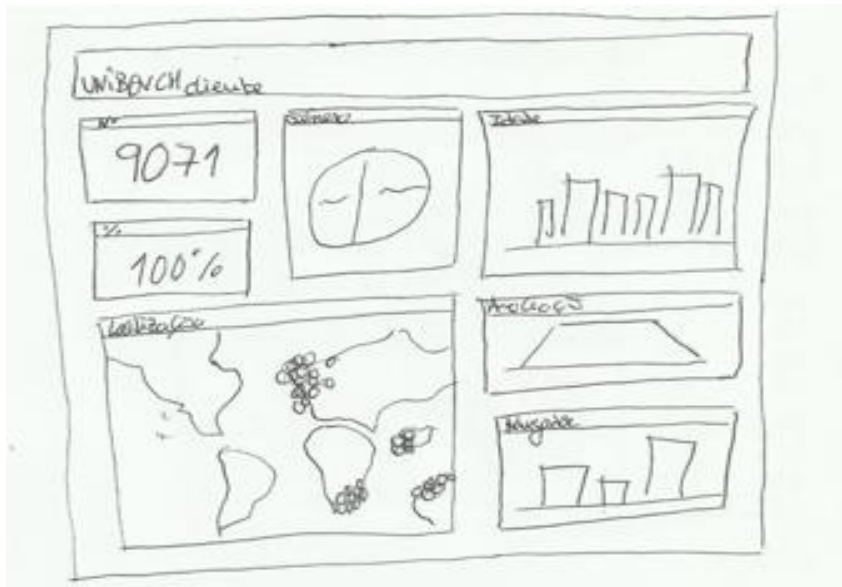
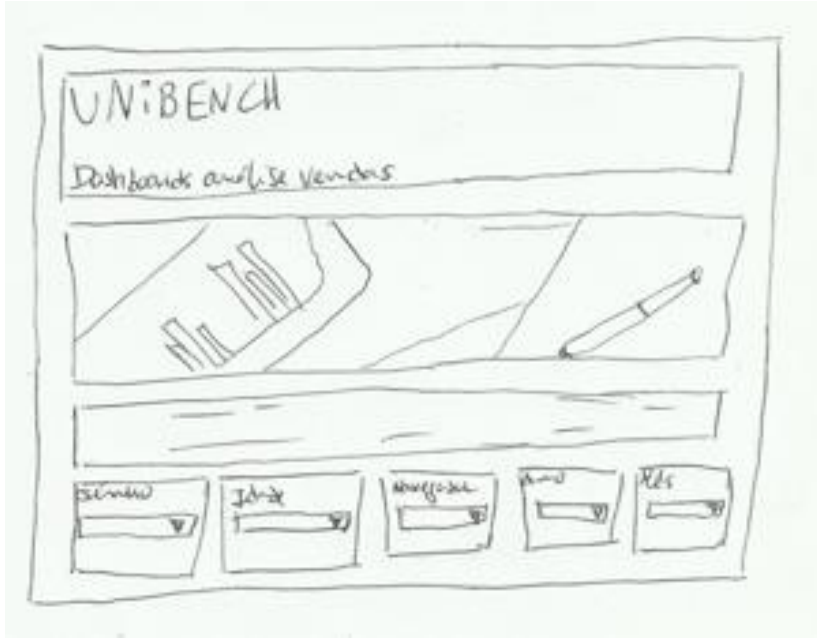
- Lee, S., Kim, S. H., & Kwon, B. C. (2016). VLAT: Development of a Visualization Literacy Assessment Test. *IEEE Transactions on Visualization and Computer Graphics*, 23(1), 551–560. <https://doi.org/10.1109/TVCG.2016.2598920>
- Liem, J., C. Perin, & J. Wood. (2020). Structure and Empathy in Visual Data Storytelling: Evaluating their Influence on Attitude. In *Computer Graphics Forum* (Vol. 39, pp. 277–289).
- Lin, C.-Y., Liang, F.-W., Li, S.-T., & Lu, T.-H. (2018). 5S Dashboard Design Principles for Self-Service Business Intelligence Tool User. *Journal of Big Data Research*, 1(1), 5–19. <https://doi.org/10.14302/issn.2768-0207.jbr-18-2175>
- Lu, J., Guo, Q., Xu, P., & Zhang, C. (2023). *Unified Database Management Systems (UDBMS)*. University of Helsinki.
- MacEachren, A. M. (1992). *Visualizing Uncertain Information*. 13, 10–19.
- MacEachren, A. M. (1995). *How maps work: representation, visualization, and design*. Guilford Press.
- Medeiros, M. M., & Maçada, A. C. G. (2022). Competitive advantage of data-driven analytical capabilities: the role of big data visualization and of organizational agility. *Management Decision*, 60(4), 953–975. <https://doi.org/10.1108/MD-12-2020-1681>
- Morais, L., Jansen, Y., Andrade, N., & Dragicevic, P. (2021). Can Anthropographics Promote Prosociality? a review and large-sample study. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–18).
- Morrison, J. (1974). *A Theoretical Framework for Cartographic Generalization with the Emphasis on the Process of Symbolization*. International Year-book of cartography.
- Munzner, T. (2014). *Visualization Analysis and Design*. CRC Press.
- Nascimento, H., & Ferreira, C. (2011). Uma introdução à visualização de informações. *Visualidades*, 9(2).
- Nielsen, J. (2000). *Why You Only Need to Test with 5 Users*. Nielsen Norman Group. <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>
- Osinska, V. (2018). Visualizing the scientific information nowadays: the problems and challenges. *Informatio et Scientia. Information Science Research*, 1(1), 30–39. <https://www.researchgate.net/publication/323999150>
- Pereira, F. (2015). *Big Data e Data Analysis - Visualização de Informação* [Master's thesis]. Universidade do Minho.
- Ribbecca, S. (2024). *The Data Visualisation Catalogue*. <https://datavizcatalogue.com/>

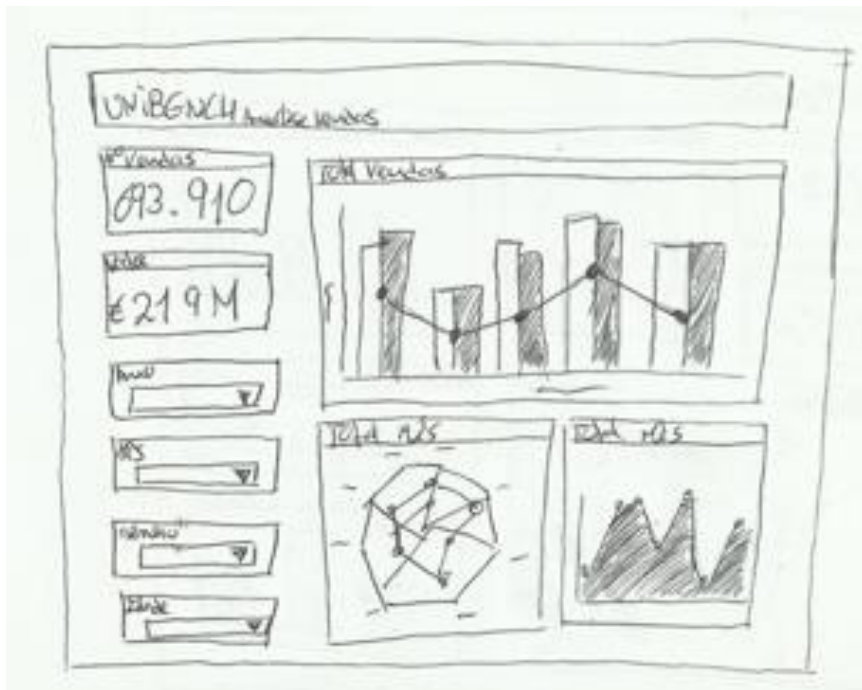
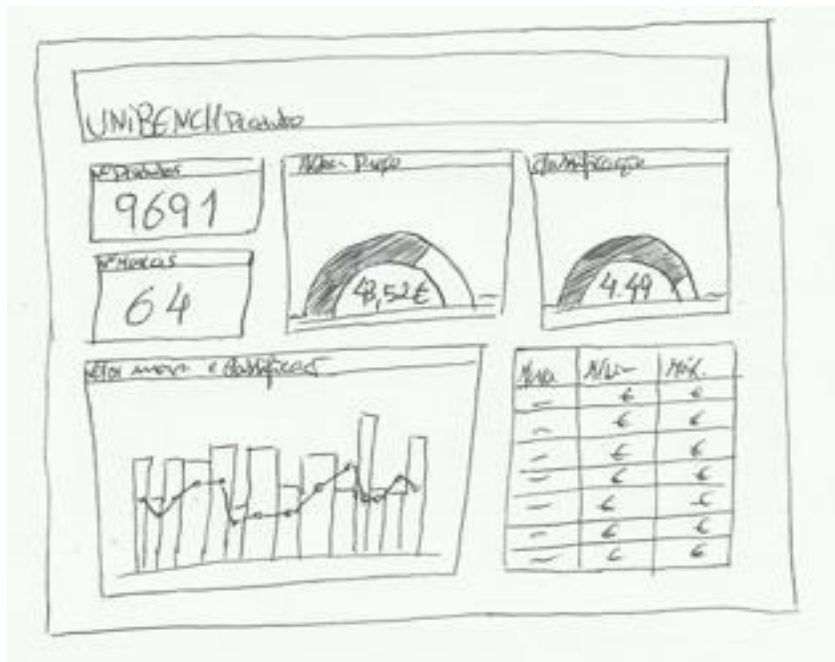
- Rodrigues, A. A., & Dias, G. A. (2017). Estudos sobre visualização de dados científicos no contexto da Data Science e do Big Data. *Pesquisa Brasileira Em Ciência Da Informação e Biblioteconomia*, 12(1), 219–228. <https://doi.org/10.22478/ufpb.1981-0695.2017v12n1.34774>
- Roth, R. E. (2017). Visual Variables. *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, 1–11. <https://doi.org/10.1002/9781118786352.wbieg0761>
- Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), 163–180. <https://doi.org/10.1177/0165551506070706>
- Ryan, L. (2016). *The Visual Imperative : Creating a Visual Culture of Data Discovery*. Morgan kaufmann. <https://doi.org/https://doi.org/10.1016/C2015-0-00786-9>
- Sauro, J. (2010). *How Many Users Do People Actually Test?* Measuring U. <https://measuringu.com/actual-users/>
- Shao, H., Martinez-Maldonado, R., Echeverria, V., Yan, L., & Gasevic, D. (2024). Data Storytelling in Data Visualisation: Does it Enhance the Efficiency and Effectiveness of Information Retrieval and Insights Comprehension? In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–21). <https://doi.org/10.1145/3613904.3643022>
- Spence, R. (2009). Information Visualization Design for Interaction. In *Information Design Journal* (Vol. 17, Issue 3, pp. 279–280). John Benjamins Publishing Company. <https://doi.org/10.1075/idj.17.3.15ehr>
- Steele, J., & Iliinsky, N. (2011). *Beautiful Visualization: Looking at data through the eyes of experts*. O'Reilly Media, Inc.
- Tufte, E. (2007). *The visual display of quantitative information*. Graphics Pr.
- Van Meter, H. J. (2020). Revising the DIKW Pyramid and the Real Relationship between Data, Information, Knowledge, and Wisdom. *Law, Technology and Humans*, 2(2), 69–80. <https://doi.org/10.5204/lthj.1470>
- Ware, C. (2019). *Information Visualization: Perception for Design*. Morgan kaufmann. <https://www.researchgate.net/publication/224285723>
- Zdanovic, D., Lembcke, T. J., & Bogers, T. (2022). The Influence of Data Storytelling on the Ability to Recall Information. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (pp. 67–77).
- Zhang, Y., Reynolds, M., Lugmayr, A., Damjanov, K., & Hassan, G. M. (2022). A Visual Data Storytelling Framework. In *Informatics* (Vol. 9, Issue 4, p. 73). MDPI. <https://doi.org/10.3390/informatics9040073>

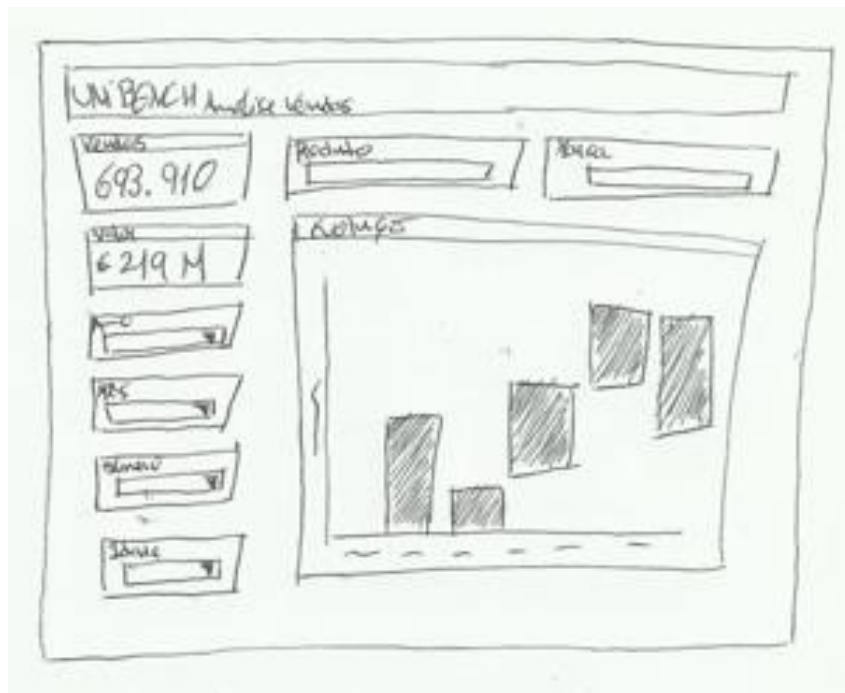
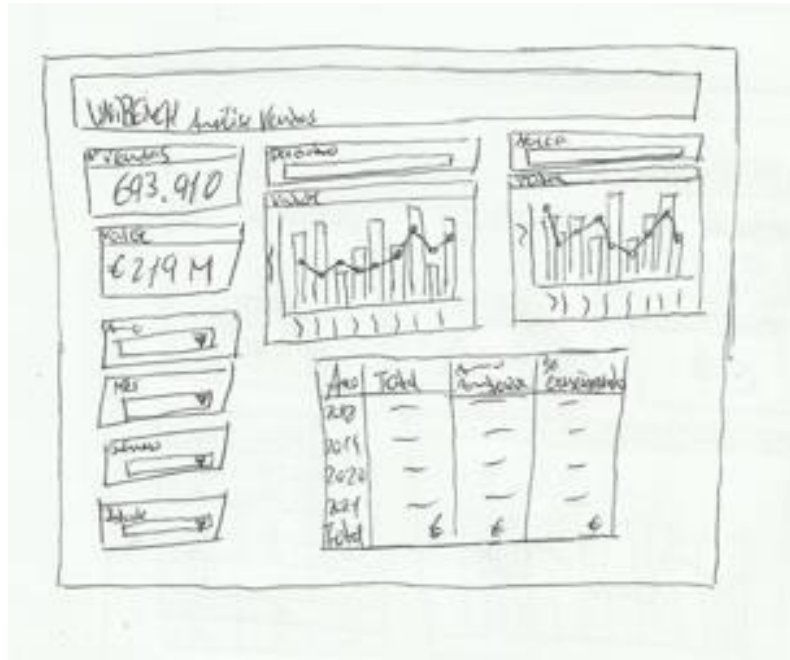
Zhao, Z., Marr, R., Shaffer, J., & Elmqvist, N. (2019). Understanding Partitioning and Sequence in Data-Driven Storytelling. In *Information in Contemporary Society: 14th International Conference, iConference 2019, Washington, DC, USA, March 31–April 3, 2019, Proceedings 14* (pp. 327–338). Springer International Publishing.

# Anexos

## Anexo A – Mockups








## Anexo B – Dashboards

### UNIBENCH

*Dashboards para análise de vendas*

Neste menu inicial pode seleccionar o *dashboard* que pretende visualizar. Pode regressar a qualquer momento e alterar o *dashboard* em análise.



Menu Inicial  
Seleção  
Dashboards

Página Inicial

Cliente

Produto

Evolução temporal

Análise Vendas

Evolução Faturação

### UNIBENCH

*Dashboards para análise de vendas*



Selecione o/os filtros apresentados que considera necessários. Estes serão aplicados a todos os *dashboards* seguintes. Pode alterá-los a qualquer momento.

**Género**

Tudo ▼

**Idade**

Tudo ▼

**Navegador**

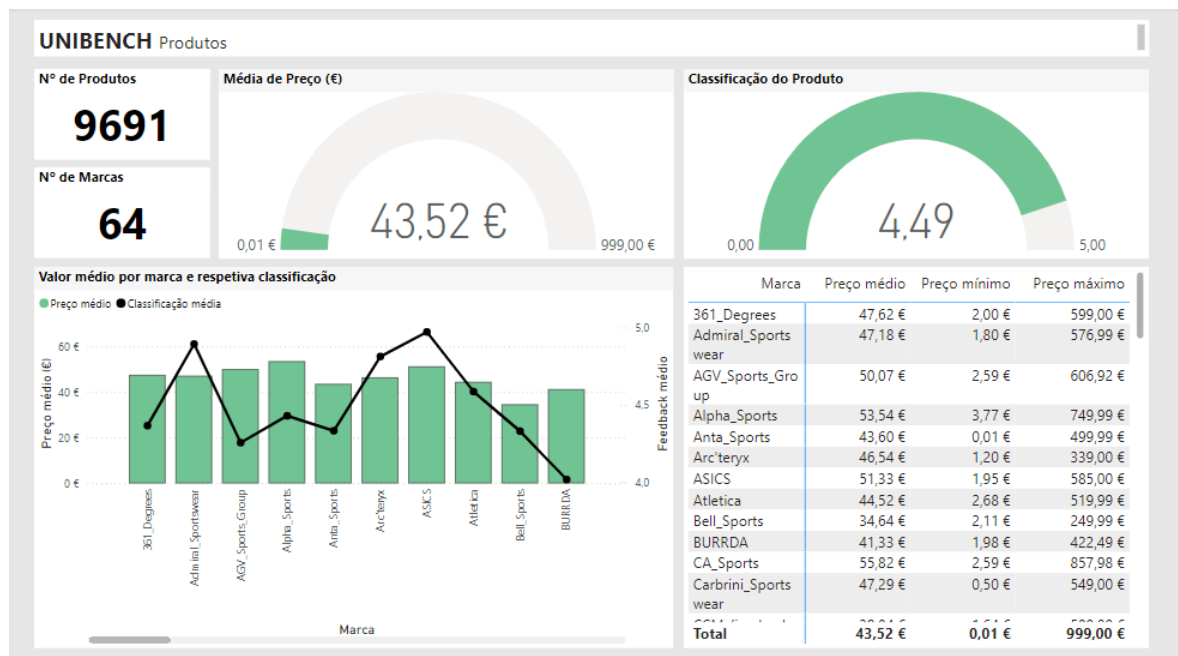
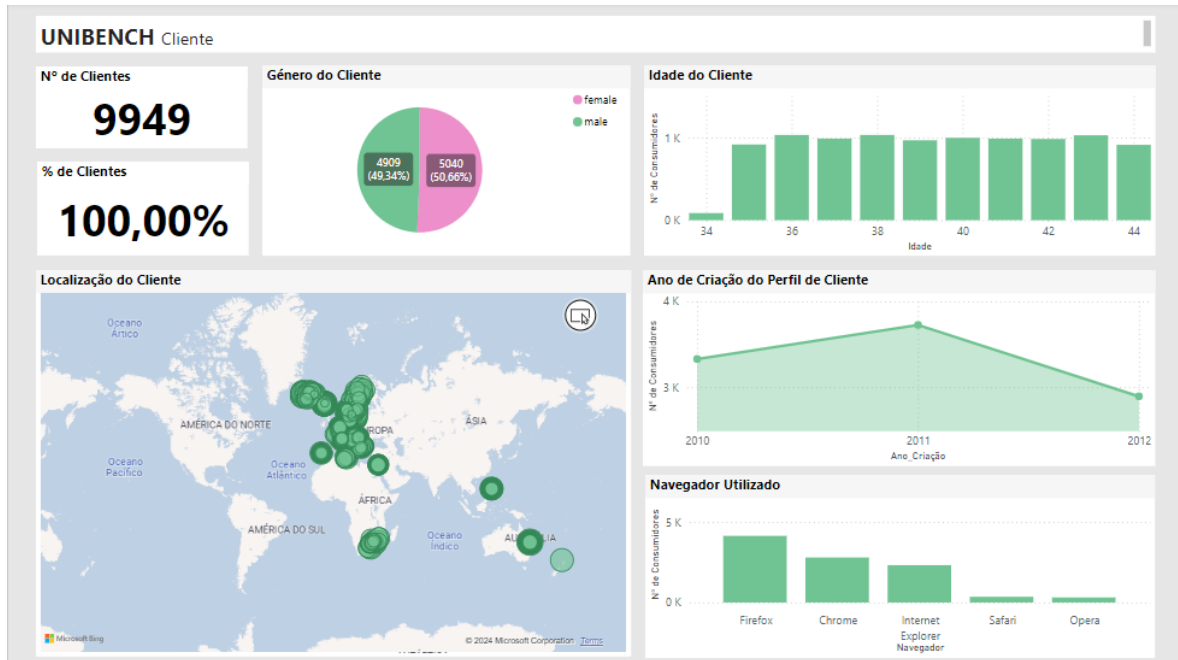
Tudo ▼

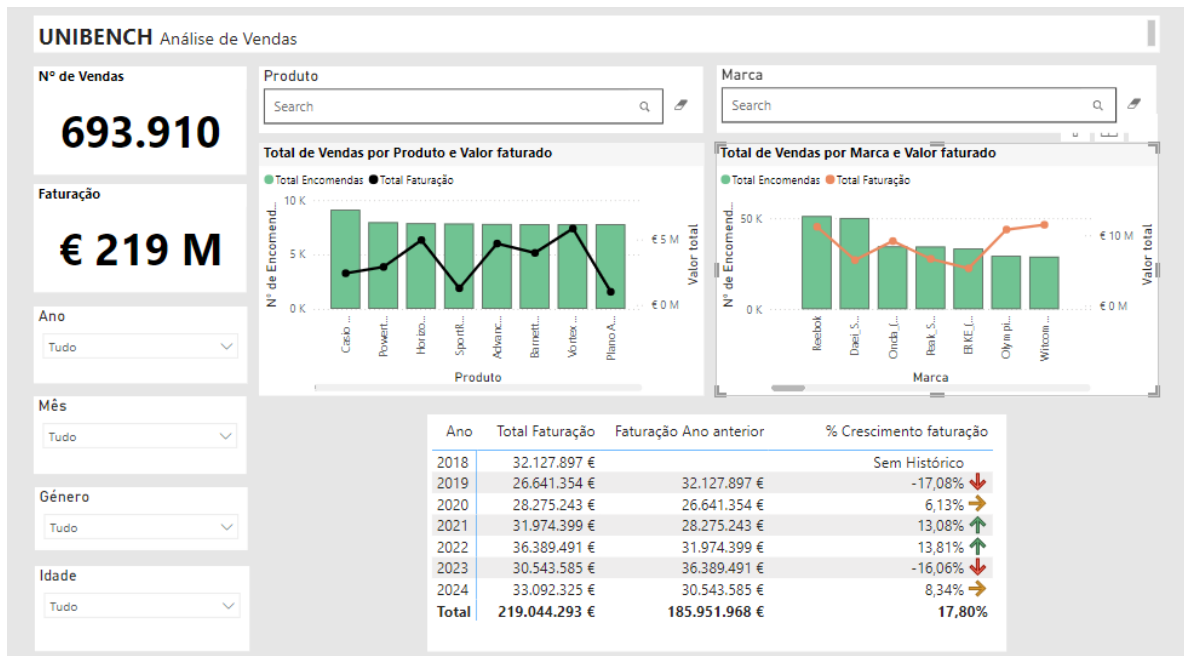
**Ano**

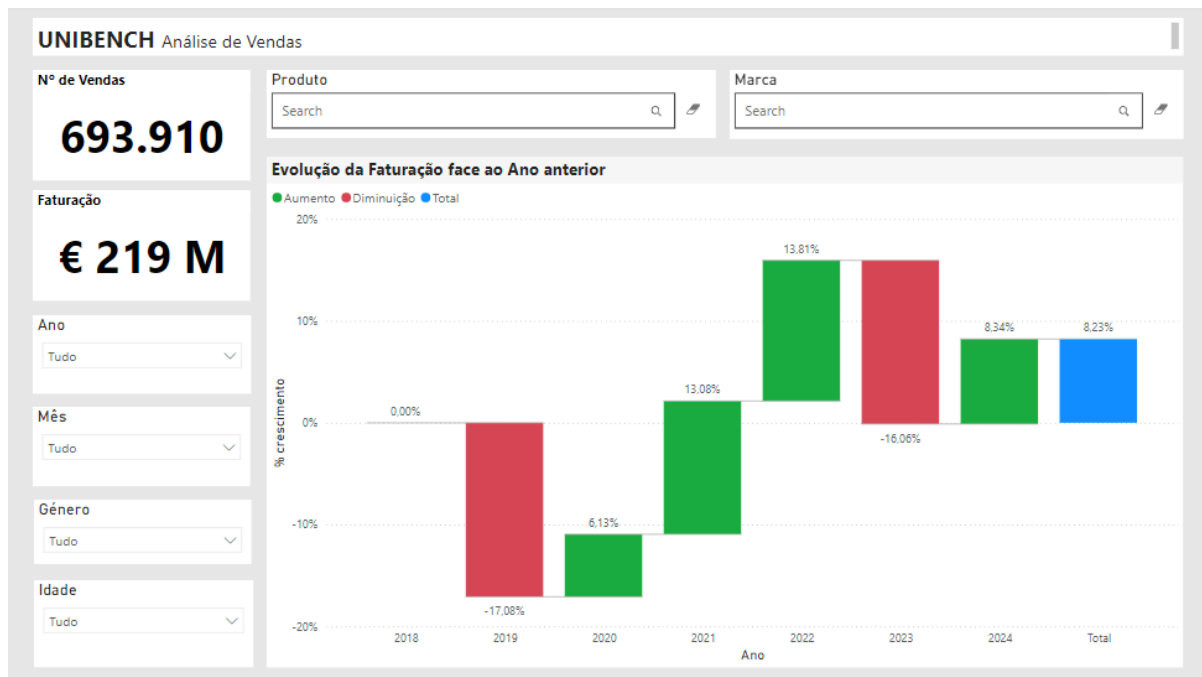
Tudo ▼

**Mês**

Tudo ▼







## Anexo C – Dashboards tipo de Gestor Operacional

**UNIBENCH**  
Dashboards para análise de vendas

Neste menu inicial pode seleccionar o *dashboard* que pretende visualizar. Pode regressar a qualquer momento e alterar o *dashboard* em análise.



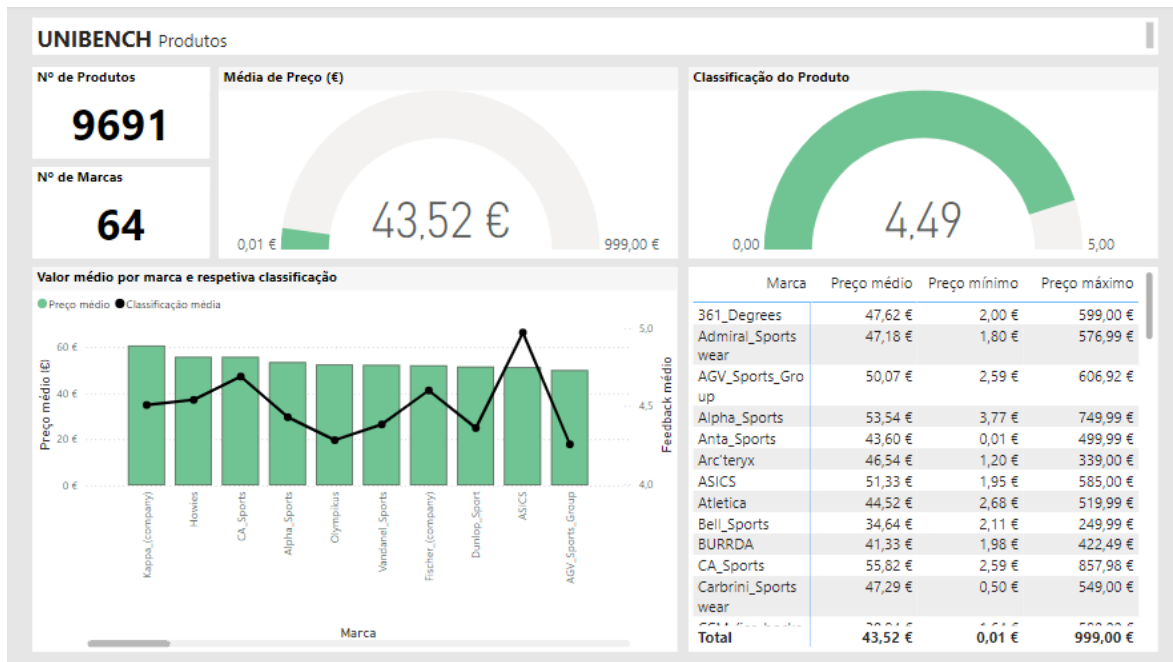
Menu Inicial Seleção Dashboards	Página Inicial	Produto	Evolução temporal	Análise Venda Produto e Marca	Análise Localização
---------------------------------	----------------	---------	-------------------	-------------------------------	---------------------

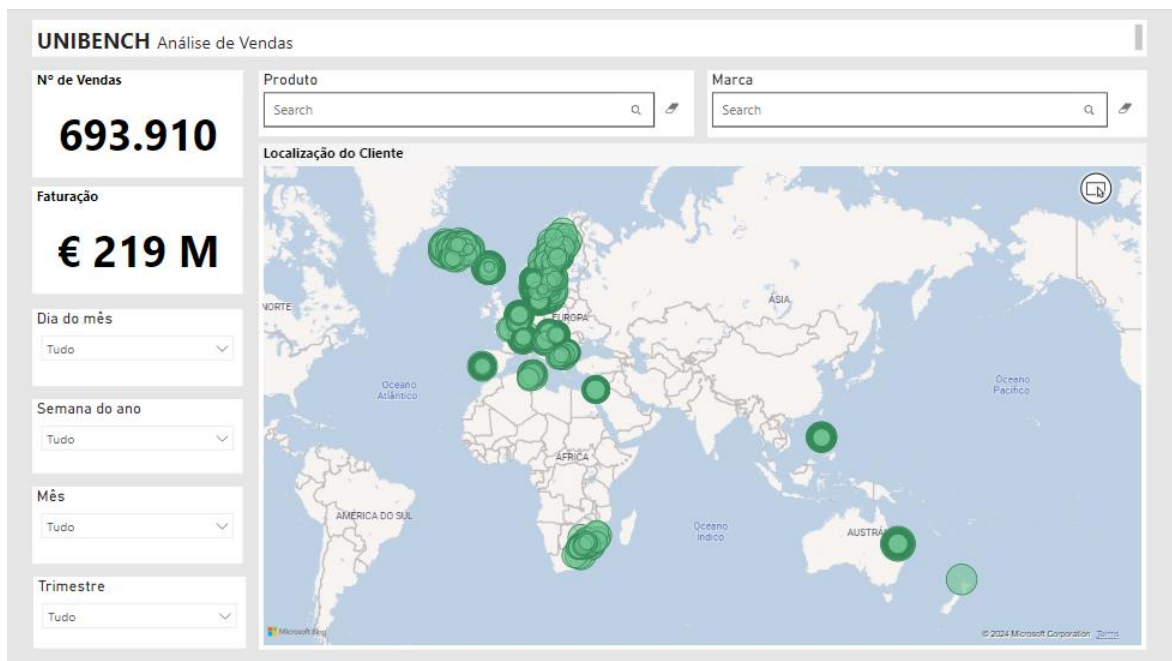
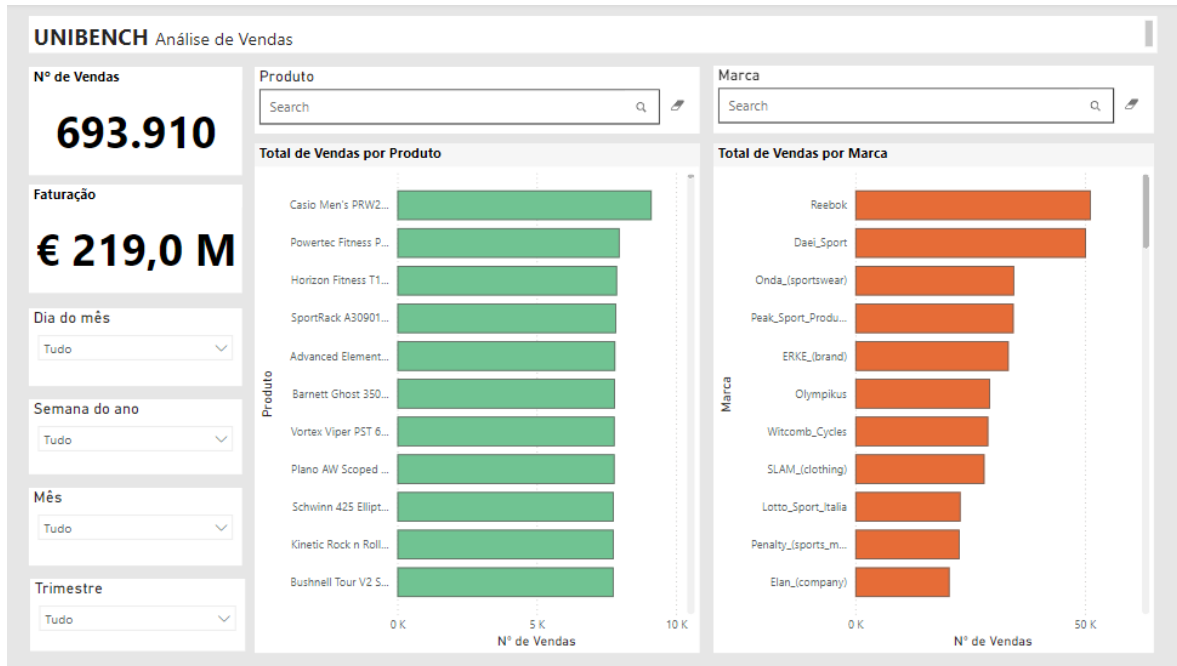
**UNIBENCH**  
Dashboards para análise de vendas



Selecione o/os filtros apresentados que considera necessários. Estes serão aplicados a todos os *dashboards* seguintes. Pode alterá-los a qualquer momento.

<b>Género</b> Tudo	<b>Idade</b> Tudo	<b>Ano</b> Tudo	<b>Trimestre</b> Tudo	<b>Mês</b> Tudo
-----------------------	----------------------	--------------------	--------------------------	--------------------





## Anexo D – *System Usability Scale*

### System Usability Scale (SUS)

	Discordo fortemente				Concordo plenamente
Acho que gostaria de utilizar este produto com frequência.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>				
	1	2	3	4	5
Considerei o produto mais complexo do que necessário.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>				
	1	2	3	4	5
Achei o produto fácil de utilizar.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>				
	1	2	3	4	5
Acho que necessitaria de ajuda de um técnico para conseguir utilizar este produto.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>				
	1	2	3	4	5
Considerei que as várias funcionalidades deste produto estavam bem integradas.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>				
	1	2	3	4	5
Achei que este produto tinha muitas inconsistências.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>				
	1	2	3	4	5
Suponho que a maioria das pessoas aprenderia a utilizar rapidamente este produto.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>				
	1	2	3	4	5
Considerei o produto muito complicado de utilizar.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>				
	1	2	3	4	5
Senti-me muito confiante a utilizar este produto.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>				
	1	2	3	4	5
Tive que aprender muito antes de conseguir lidar com este produto.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>				
	1	2	3	4	5

Created with the SUS PDF Generator (<https://blatigerste.github.io/sus-pdf-generator/>)

## Anexo E – Questionário completo para teste com utilizadores

### Opinião acerca da usabilidade de *dashboards* de gestão


Este questionário, para o qual solicito o preenchimento, enquadra-se no âmbito da realização da dissertação do mestrado em Ciência de Dados, da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria.

O objetivo do presente questionário é auscultar a opinião dos indivíduos acerca do *storytelling* e da usabilidade de *dashboards*, aplicados à gestão de empresas e tomada de decisão por parte dos decisores.

A sua participação neste estudo é voluntária e pode ser interrompida a qualquer momento, sem necessidade de justificação. Além disso, não lhe trará qualquer despesa ou risco.

Todos os dados relativos à identificação dos participantes neste estudo são confidenciais. Os seus dados serão usados, de forma completamente anónima, destinando-se, meramente, a ser utilizados para fins académicos, em estrita obediência ao Regulamento Geral de Proteção de Dados e da sua Lei de Execução Nacional.

Agradeço desde já a sua participação e colaboração, e solicito que responda a cada questão com a opção que melhor representar a sua opinião pessoal.

joaopaiva809@gmail.com [Mudar de conta](#) 

\* Indica uma pergunta obrigatória


Email \*

O seu email

\*

Autorizo a minha participação, de forma voluntária, neste projeto de investigação.

## Opinião acerca da usabilidade de *dashboards* de gestão

joaopaiva809@gmail.com [Mudar de conta](#) 

\* Indica uma pergunta obrigatória

### Dados sociodemográficos

Em cada questão responda com a opção que melhor se adequa à sua realidade.

**Sexo: \***

Masculino

Feminino

Prefiro não responder

**Idade: \***

18-24

25-29

30-39

40-49

50-64

65 ou mais

**Habilitações académicas: \***

Ensino básico

Ensino médio

Ensino secundário

Licenciatura

Mestrado

Doutoramento

**Área de estudo: \***

A sua resposta \_\_\_\_\_

**Atividade profissional: \***

A sua resposta \_\_\_\_\_

**System Usability Scale (SUS)**

Este questionário serve para avaliar a sua experiência com as visualizações anteriores.

De seguida, apresentam-se algumas afirmações que deve responder para avaliar a sua experiência.

Por favor, exprima o seu grau de concordância com cada afirmação, assinalando o círculo correspondente.

Decida de forma espontânea e sem grandes reflexões para transmitir a sua primeira impressão.

Avalie cada afirmação mesmo que considere que não corresponde totalmente à sua experiência.

Não há respostas certas ou erradas - o que conta é a sua opinião pessoal!

Selecione a opção que melhor se adequa à sua experiência. \*

	Discordo fortemente	Discordo	Nem concordo nem discordo	Concordo	Concordo plenamente
Acho que gostaria de utilizar este produto com frequência.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Considereei o produto mais complexo do que necessário.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Achei o produto fácil de utilizar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Acho que necessitaria de ajuda de um técnico para conseguir utilizar este produto.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Considereei que as várias funcionalidades deste produto estavam bem integradas.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Achei que este produto tinha muitas inconsistências.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Suponho que a maioria das pessoas aprenderia a utilizar rapidamente este produto.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Considereei o produto muito complicado de utilizar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Senti-me muito confiante a utilizar este produto.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tive que aprender muito antes de conseguir lidar com este produto.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Storytelling**  
 As próximas questões servem para avaliar a sua experiência relativamente à história de dados apresentada.

Por favor, exprima o seu grau de concordância com cada afirmação, assinalando o círculo correspondente.

Decida de forma espontânea e sem grandes reflexões para transmitir a sua primeira impressão.

Avalie cada afirmação mesmo que considere que não corresponde totalmente à sua experiência.

Não há respostas certas ou erradas - o que conta é a sua opinião pessoal!

---

Selecione a opção que melhor se adequa à sua experiência. \*

	Discordo fortemente	Discordo	Nem concordo nem discordo	Concordo	Concordo plenamente
Através da história de dados apresentada fui capaz de identificar as principais questões em análise.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A história de dados apresentada criou mais confusão na minha interpretação dos dados.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A história de dados apresentada facilitou a minha interpretação dos dados apresentados.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A história de dados apresentada deu ênfase a tópicos que considere pouco relevantes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Se não fosse apresentada a história de dados a minha navegação pelos diferentes dashboards seria mais difícil.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A apresentação da história de dados apresentada não influenciou a minha interpretação dos dados.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>