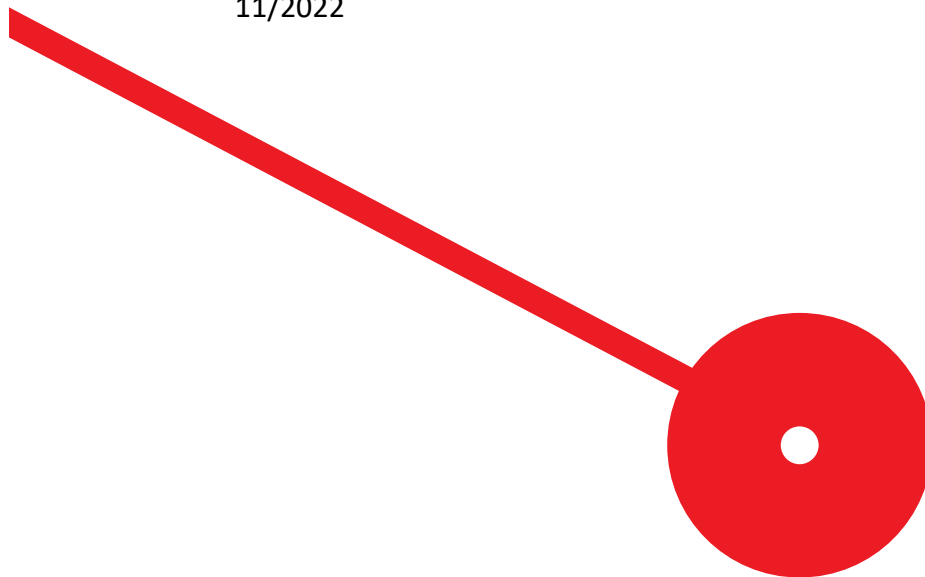




Business Intelligence: Projeto de criação de *dashboards* para a empresa NDrive

Hugo Filipe de Sousa Vieira

11/2022

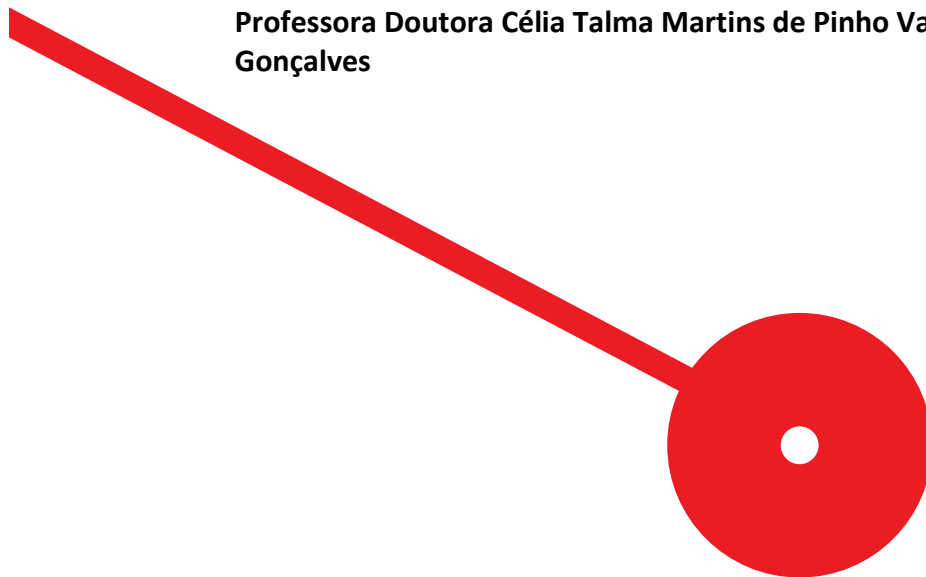




Business Intelligence: Projeto de criação de *dashboards* para a empresa NDrive

Hugo Filipe de Sousa Vieira

Trabalho de Projeto apresentado ao Instituto Superior de Contabilidade e Administração do Porto para a obtenção do grau de Mestre em Assessoria de Comunicação Digital, sob orientação de Professora Doutora Célia Talma Martins de Pinho Valente Oliveira Gonçalves



Agradecimentos

A concretização deste trabalho de projeto só foi possível com a ajuda de todas as pessoas que de uma forma ou de outra contribuíram com a sua ajuda.

Aos meus pais, o vosso apoio e força que tanto fortaleceram a minha vontade para conseguir superar mais uma etapa académica. Obrigado por todas as alturas em que tiveram de lidar com os meus desabafos e devaneios. Não existem palavras que consigam expressar o sentimento que nutro por vocês.

À Sara, a minha companheira para a vida toda, obrigado por caminhares mais outra etapa comigo e obrigado por me teres feito acreditar em mim mesmo e que nada é impossível, basta querer.

Agradeço à minha orientadora Professora Doutora Célia Talma Gonçalves por toda a orientação dada para conseguir alcançar os objetivos propostos.

Ao ISCAP, por me ter recebido de braços abertos e aos seus docentes, que me permitiram obter o conhecimento necessário que tanto foi valioso para o meu percurso académico, profissional e pessoal.

Agradeço à NDrive por ter aceiteado esta proposta de implementação e por toda a ajuda prestada durante a sua conceção.

Aos meus amigos de turma, obrigado por toda a motivação e disponibilidade que prestaram durante estes 2 anos de percurso académico.

A todos vocês, um obrigado de coração!

Resumo:

Cada vez mais os dados são uma mais-valia tanto para as grandes empresas como para as PMEs, e como tal, os processos de tomada de decisão desempenham um papel fundamental. Com a evolução das ferramentas de Business Intelligence a permitirem um acesso cada vez mais eficiente, torna-se imperativo para as empresas apostarem neste tipo de soluções para continuarem a ser competitivas no mercado atual.

Neste trabalho de projeto foi proposto à NDrive Navigation Systems, S.A., a criação de uma solução de *Business Intelligence* através da ferramenta Microsoft Power BI para gerir e monitorar os indicadores mais eficientes para a empresa. A implementação desta solução teve como base a criação e modelação de um Data Warehouse que implementasse os KPI mencionados e posterior exploração através da elaboração de dashboards.

A implementação desta solução de BI foi aceite e validada pela administração, assim como pelos elementos constituintes do departamento de Marketing da empresa, onde foi possível auferir a vantagem de se ter os principais indicadores disponíveis em dashboards, sem necessidade de recorrer a diferentes ferramentas para análise conjunta dos dados, resultando assim num processo de análise muito mais rápido e eficiente.

Palavras chave: *Business Intelligence, Dashboards, Sistemas Apoio à Decisão, Power BI*

Abstract:

Data is increasingly an asset for both large companies and SMEs, and as such, decision-making processes play a key role. With the evolution of Business Intelligence tools allowing an increasingly efficient access, it becomes imperative for companies to invest in this type of solutions to remain competitive in today's market.

In this project work it was proposed to NDrive Navigation Systems, S.A., the creation of a Business Intelligence solution using the Microsoft Power BI tool to manage and monitor the most efficient indicators for the company. The implementation of this solution was based on the creation and modeling of a Data Warehouse that would implement the mentioned KPI's and subsequent exploration through the elaboration of dashboards.

The implementation of this BI solution was accepted and validated by the administration, as well as by the constituent elements of the company's Marketing department, where it was possible to benefit from having the main indicators available in dashboards, without the need to use different tools for joint analysis of the data, resulting in a much faster and efficient analysis process.

Key words: Business Intelligence, Dashboards, Decision Support Systems, Power BI

Índice

Capítulo I – Introdução.....	1
1.1 Introdução.....	2
1.2 Enquadramento e motivação	2
1.3 NDrive, S.A.	3
1.4 Questão de Investigação e Objetivos.....	3
1.5 Estrutura do documento.....	4
Capítulo II – Revisão de literatura	6
2.1 Business Analytics.....	7
2.2 Business Intelligence	12
2.3 O Business Intelligence como suporte ao processo de tomada de Decisão nas Organizações.....	15
2.4 Relação entre o Business Intelligence e Data Warehouses	18
2.4.1 Conceito de Data Warehouse e Data Marts.....	19
2.4.2 Arquitetura de Inmon vs Arquitetura de Kimball.....	22
2.4.3 Extração, Transformação e Carregamento de Dados	26
2.4.4 Modelação Multidimensional – Cubo OLAP – Factos e Dimensões.....	28
2.4.5 Data Science	35
2.4.6 Machine Learning.....	36
2.4.7 Ferramentas de BI.....	37
2.5 A importância da Visualização de Dados e dos Dashboards.....	44
2.6 Análise e Seleção de Indicadores	50
Capítulo III - Metodologia.....	54
2.7 Metodologia.....	55
Capítulo IV – Planeamento.....	58
3.1 Diagnóstico.....	59
3.2 Planeamento das Ações	59

Capítulo V – Implementação	65
4.1 Execução das Ações	66
4.1.1 Processo de Negócio.....	66
4.1.2 Granularidade dos Dados.....	66
4.1.3 Tabelas de Dimensão.....	66
4.1.4 Tabelas de Facto	69
4.2 Fase ETL (Extract-Transform-Load).....	70
4.2.1 Extract.....	70
4.2.2 Transform	72
4.2.3 Load.....	77
4.2.4 Modelação Multidimensional	77
4.3 Dashboards	79
4.3.1 Apresentação dos Dashboards criados	81
Capítulo VI – Avaliação	90
5.1 Validação e discussão de dados.....	91
5.2 Aprendizagem Específica.....	97
Capítulo VII – Conclusão	98
6.1 Conclusão	99
6.2 Contribuições.....	99
6.3 Dificuldades e Limitações	100
6.4 Trabalho Futuro	100
Referências Bibliográficas	103
Apêndices.....	113
Apêndice A: Código em linguagem M usado no editor avançado do Power BI para criação da tabela calendário DimTime na fase ETL.....	113
Apêndice B: Questionário de System Usability Scale (Sus) para avaliação das dashboards	114

Índice de Figuras

Figura 1: As 3 áreas de Business Analytics.....	8
Figura 2: As 4 eras de Business Analytics	10
Figura 3: Benefícios do BI nas organizações	14
Figura 4: Papel do Business Intelligence na tomada de decisão	16
Figura 5: Arquitetura de software com foco na relação de BI com DW	18
Figura 6: Abordagem top-down de Inmon.	23
Figura 7: Abordagem bottom-up de Kimball	23
Figura 8: Arquitetura ETL para um data warehouse	26
Figura 9: Arquitetura OLAP.....	29
Figura 10: Esquema em estrela (Star Schema).....	29
Figura 11: Esquema Floco de Neve (Snowflake Schema)	30
Figura 12: Exemplo de uma tabela de dimensão	32
Figura 13: Exemplo de uma tabela de facto	32
Figura 14: Exemplo de Cubo OLAP tridimensional	33
Figura 15: Operação Drill-down/Drill-up no cubo.....	34
Figura 16: Operação Slice no cubo	34
Figura 17: Operação Slice and Dice no cubo.	35
Figura 18: Operação Pivot no cubo	35
Figura 19: Quadrante Mágico Gartner para BI.....	38
Figura 20: Versões do Power BI	39
Figura 21: Visualizações possíveis de usar no Power BI Desktop e Service	40
Figura 22: Linha de produtos tableau	42
Figura 23: Exemplo de Dashboard	46
Figura 24: Exemplo de visualização de indicadores	52
Figura 25: Representação dos indicadores numa ficha	53
Figura 26: As 5 etapas do modelo Investigação-Ação	56
Figura 27: Funcionalidade “Obter Dados” e conteúdo do Google Analytics.....	71
Figura 28: Funcionalidades do Power Query no Power BI.....	72
Figura 29: Passos feitos no Power Query para a DimTempo em linguagem M	72
Figura 30: Funcionalidade “Referência” para referência de tabelas no Power BI....	73
Figura 31: Passos feitos no Power Query para a FactAppAnalytics em linguagem M..	75
Figura 32: Função Intercalar consultas no Power BI	76
Figura 33: Carregamento dos dados para o modelo multidimensional	77
Figura 34: Modelo de dados multidimensional proposto em formato star schema.....	78
Figura 35: Visão geral das relações entre tabelas no modelo de dados.....	79
Figura 36: Marcadores criados para as dashboards.	80
Figura 37: Cabeçalho final criado para os dashboards.....	81
Figura 38: Menu Principal.....	81
Figura 39: Dashboard: Secção General Overview.	82
Figura 40: Sessions by Country.....	84
Figura 41: Users by Country	85
Figura 42: New Users by Country	86
Figura 43: Sessions by Continent	87
Figura 44: Users by Continent.....	88
Figura 45: New Users by Continent.	89

Índice de Tabelas

Tabela 1: Comparação das abordagens de Inmon e Kimball	25
Tabela 2: Tipos de Dashboards	49
Tabela 3: Indicador Sessions	62
Tabela 4: Indicador Users	62
Tabela 5: Indicador New Users	63
Tabela 6: Cronograma de atividades	64
Tabela 7: DimTempo	67
Tabela 8: DimCountry	67
Tabela 9: DimDeviceCategory	67
Tabela 10: DimDeviceModel	68
Tabela 11: DimOperatingSystem	68
Tabela 12: DimOSVersion	68
Tabela 13: DimTime	69
Tabela 14: FactAppAnalytics	69
Tabela 15: Fórmulas DAX da dashboard General Overview	83
Tabela 16: Valores totais e média global do SUS	91
Tabela 17: Avaliação da frequência de uso das dashboards	92
Tabela 18: Avaliação da complexidade das dashboards	92
Tabela 19: Avaliação da facilidade de uso	93
Tabela 20: Análise à capacidade do participante	93
Tabela 21: Avaliação da integração das funcionalidades do dashboard	94
Tabela 22: Avaliação das inconsistências nos dashboards	94
Tabela 23: Avaliação da rapidez de aprendizagem no uso das dashboards	95
Tabela 24: Avaliação do grau de complexidade das dashboards	95
Tabela 25: Avaliação do grau de confiança durante o uso das dashboards	96
Tabela 26: Análise da capacidade de aprendizagem do participante	96

Lista de abreviaturas

API – *Application Protocol Interface*

BA – *Business Analytics*

BI – *Business Intelligence*

DSS – *Data Science*

DSS – *Decision Support Systems*

DM – *Data Mart*

DW – *Data Warehouse*

EDW – *Enterprise Data Warehouse*

ETL – *Extract, Transform and Load*

GPS – *Global Positioning System*

KPI – *Key Performance Indicators*

MOLAP – *Multidimensional Online Analytical Processing*

OLAP – *Relational Online Analytical Processing*

PND – *Portable Navigation Device*

ROLAP – *Relational Analytical Processing*

SAD – *Sistemas de Apoio à Decisão*

SUS – *System Usability Scale*

SQL – *Structured Query Language*

TI – *Tecnologias da Informação*

CAPÍTULO I – INTRODUÇÃO

1.1 Introdução

Nos dias de hoje, devido ao enorme crescimento de dados, para além das grandes empresas, as pequenas e médias empresas (PMEs) também necessitam do *Business Intelligence* para as auxiliar nas mais variadas tarefas de tomada de decisão, redução de custos, otimização de tarefas, na previsão de crescimento empresarial e para a contribuição de elaboração de estratégias (Primak, 2008). Cada vez mais é necessário que as empresas saibam como tratar e tirar partido da enorme quantidade de dados disponíveis e tomar decisões efetivas de modo a atingir as metas estabelecidas e a estar a um passo à frente em relação à concorrência.

O Business Intelligence nasceu dentro do mundo industrial no início dos anos 90 com o objetivo de satisfazer os pedidos dos gestores para analisar os dados empresariais de forma eficiente e eficaz de modo a poderem entender o estado dos seus negócios e melhorar os processos de tomada de decisão (Golfarelli et al. 2004).

Dado que o mercado de Business Intelligence continua a ser cada vez mais valorizado ano após ano, tendo sido valorizado em 18,3 mil milhões de Euros em 2017 e com uma previsão de crescimento que pode chegar aos 22.8 mil milhões de Euros num futuro próximo (Ain et al., 2019), as empresas começam a aperceber-se das vantagens do BI nos processos de tomada de decisão, tornando-se efetivamente uma solução vantajosa para as organizações, especialmente para as PMEs onde o uso de soluções de BI ainda não são tão predominantes quando comparadas com as grandes organizações.

Neste trabalho de projeto, que foi proposto à empresa NDrive Navigation Systems, S.A., procurou-se otimizar o processo de tomada de decisão, proporcionando uma solução de *Business Intelligence* com *dashboards* para fácil identificação dos indicadores usados pela empresa NDrive, possibilitando assim uma solução de acesso aos dados recebidos diariamente através da plataforma *Google Analytics* de forma mais rápida e sem a necessidade de redigir relatórios adicionais.

1.2 Enquadramento e motivação

O Business Intelligence ainda permanece uma componente forte nas grandes empresas, contudo, nas Pequenas e Médias Empresas (PMEs) o conceito de BI ainda é território praticamente desconhecido, pouco trabalhado, ou então por não considerarem o investimento justificativo. Ferramentas como o Microsoft Power BI, que começam com

patamares gratuitos e que o seu uso pode ser feito por qualquer pessoa, podem ajudar as PMEs a dar os primeiros passos na correta gestão dos seus dados e otimizar os processos de tomada de decisão.

A principal motivação foi integrar todos os indicadores numa única plataforma de análise, uma vez que a NDrive não tinha integrado nos seus serviços uma solução de Business Intelligence, especialmente que devido à natureza da empresa, lida diariamente com um grande volume de dados que são vitais para os processos de tomada de decisão diários.

1.3 NDrive, S.A.

A NDrive é uma empresa portuguesa, fundada em 2007, sendo um spin-off da empresa InfoPortugal, tendo a sua sede no Porto e escritórios em Shenzhen, China. A NDrive concentra os seus esforços no desenvolvimento de sistemas de navegação para as plataformas digitais Android e iOS, tendo assim uma presença digital em mais de 50 países graças às lojas Play Store e App Store.

Os seus primeiros produtos foram os GPS de bolso (também conhecidos por PND ou simplesmente GPS) denominados de NDrive GO, tendo sido comercializados inicialmente em Portugal e no Brasil, e posteriormente a sua oferta foi alargada para todo o mundo. Atualmente todos os recursos da NDrive estão no desenvolvimento de soluções baseadas na localização para clientes nas mais diversas áreas de negócio, deixando de ser assim unicamente um fornecedor de software de navegação por GPS, a NDrive está a contribuir para a criação de um mundo cada vez mais ligado e a ajudar os parceiros de negócio no processo de tomada de decisão baseados em dados de localização.

1.4 Questão de Investigação e Objetivos

A questão de investigação formulada para este trabalho de projeto foi: “Quais os benefícios do Business Intelligence nos processos de tomada de decisão de uma PME?”

Para dar-se resposta à questão de investigação, foram delineados diversos objetivos, sendo o principal objetivo a conceção de uma solução de *Business Intelligence* que consiste na criação e modelação de um *Data Warehouse* e posterior exploração através de *dashboards* com recurso à ferramenta Power BI de modo a apoiar o processo de tomada de decisão por parte do departamento de Marketing da NDrive.

A fonte de dados principal usada pela empresa NDrive não permite obter de uma forma célere a informação desejada, pelo que muitas vezes para se chegar a este objetivo é necessário filtrar, processar e redigir relatórios para serem apresentados no período necessário.

De modo que seja possível alcançar o objetivo principal que foi proposto neste trabalho de projeto, foram definidos os seguintes objetivos específicos:

- Compreender as metodologias de implementação de um sistema de *Business Intelligence* recorrendo à revisão da literatura existente.
- Através da análise documental dos relatórios existentes, no departamento de Marketing da empresa, fazer um mapeamento dos KPIs atuais e os KPIs necessários para apresentação.
- Conceber e implementar uma solução de *Business Intelligence* para apoio ao processo de tomada de decisão desenvolvendo um Data Mart (DataWarehouse) para a área de Marketing.
- Desenvolvimento e exploração de *dashboards* para otimização do acesso e análise dos indicadores acordados como prementes para a organização para suporte à decisão.
- Apurar a situação futura do projeto, garantindo o desenvolvimento contínuo de novas funcionalidades conforme as necessidades da empresa NDrive.

Pretende-se assim com esta implementação que os processos de tomada de decisão sejam mais céleres, através da apresentação dos dados através de diversas formas visuais, incluídos nos *dashboards* desenvolvidos.

1.5 Estrutura do documento

Relativamente à estrutura deste trabalho de projeto, o mesmo foi estruturado em 4 capítulos, os quais são aqui explicados de uma forma resumida:

- **Capítulo I – Introdução:** Este capítulo começa com uma breve introdução ao estado da arte e a ligação com o *Business Intelligence*, seguido de um enquadramento com as PME's e a adoção do BI nas mesmas. Seguidamente é feita uma introdução à empresa NDrive, S.A., dando a conhecer um pouco da história da mesma, assim como foram delineados os objetivos gerais e específicos para

este trabalho de projeto. Para terminar, é feita uma análise à metodologia investigação-ação usada para conceber e implementar a solução de BI na empresa e explicado a estrutura de todos os capítulos deste documento.

- **Capítulo II – Revisão de literatura:** Neste capítulo é feita uma revisão de literatura, a qual aborda os conceitos de *Business Analytics* e *Business Intelligence* associados ao tema, assim como o seu papel nos processos de apoio à decisão. É também feita uma revisão de literatura acerca das arquiteturas existentes de um Data Warehouse e todos os componentes que fazem parte dele, assim como as ferramentas de BI que foram alvo de análise. O capítulo termina com a importância da visualização de dados e na escolha dos indicadores de negócio mais indicados.
- **Capítulo III – Metodologia:** No capítulo III é referida a metodologia que foi escolhida para a solução de BI (investigação-ação), detalhando todos os passos que foram usados para o projeto.

Capítulo IV – Planeamento: Este capítulo aborda as etapas de diagnóstico e planeamento das ações tomadas para dar início ao desenvolvimento, tais como avaliação dos requisitos, arquitetura escolhida, indicadores necessários e cronograma do projeto.

- **Capítulo V – Implementação:** Aqui são explicados todos os passos de implementação da solução de BI para a empresa, que vai desde a execução das ações utilizando a metodologia de modelação de Kimball, passando pelas fases de modelação do *Data Warehouse* e criação das tabelas de dados necessárias, terminando com o desenho e apresentação dos *dashboards* criados na ferramenta Power BI Desktop.
- **Capítulo VI – Implementação:** No capítulo VI os dados obtidos através de questionário são validados e discutidos. Há também uma reflexão acerca do conhecimento adquirido do projeto.
- **Capítulo VII – Conclusão:** O capítulo VII conclui este trabalho de projeto e é onde é feita uma análise conclusiva dos objetivos alcançados com este trabalho, assim como a viabilidade da solução proposta e implementada para a empresa. São também abordadas as principais dificuldades e limitações que existiram durante o desenvolvimento deste trabalho, terminando com propostas para enriquecimento e melhoramento da solução de BI no futuro.

CAPÍTULO II – REVISÃO DE LITERATURA

2.1 Business Analytics

O conceito de *Business Analytics* (BA) ainda é algo vago, dado que não existe uma definição única e amplamente aceite de BA, sendo que se encontra em constante evolução e muito dificilmente será possível ter uma única definição tendo em conta a aplicação de BA em contexto académico, científico e profissional (Power, Heavin, McDermott, & Daly, 2018). Turban et al. (2014) referem que a palavra “*analytics*” substituiu os componentes individuais das tecnologias usadas nos *Decision Support Systems* (Sistemas de Apoio à Decisão) que eram usadas no passado, definindo assim “*analytics*” como a análise do processo de desenvolvimento de decisões acionáveis ou recomendação de ações com base em insights gerados a partir de dados históricos.

Segundo Davenport and Harris (2007), o termo *Analytics* é definido como “*o uso extensivo de dados, análises quantitativas e estatísticas, modelos explanatórios e preditivos e gestão baseada em factos para mover decisões e ações*”. Evans and Lindner (2012) definem *Business Analytics* como o “uso de dados, tecnologias de informação, análises estatísticas, métodos quantitativos e modelos matemáticos ou baseados em computação para ajudar gestores a obter uma melhor compreensão acerca das suas operações de negócios e poderem tomar melhores decisões baseadas em factos. O *Business Analytics* consiste assim no uso de ferramentas, técnicas e metodologias de análise com o objetivo principal na obtenção de valor a partir de dados (Gonçalves,2021).

De acordo com Sun et al. (2017), as ferramentas e recursos de *Business Analytics* permitem às empresas obter apoio na tomada de decisão nos seus negócios, disponibilizando dados, informações e conhecimentos valiosos. Entre 60% e 73% de todos os dados de uma organização não são utilizados para análise (Microsoft, 2019).

O *Business Analytics* engloba 3 grandes áreas (Evans & Lindner, 2012):

- Análise Descritiva
- Análise Preditiva
- Análise Prescritiva



Figura 1: As 3 áreas de Business Analytics

Fonte: “Advanced analytics advance your business” (Peregud, 2019)

A Microsoft (2019) descreve em maior detalhe estas 3 grandes áreas do *Business Analytics*:

Análise Descritiva: Permite aprofundar o conhecimento sobre os dados e usa indicadores de negócio (KPIs) para mostrar o estado atual de um negócio. A nível de exemplo indicam a informação em tempo real acerca de dados demográficos, interesses e comportamento de compra, os quais podem ser de natureza financeira ou mesmo de métricas de social media (likes, tweets ou seguidores).

Nesta análise, exemplos como relatórios, OLAP, *dashboards/scorecards* e visualizações de dados são cenários onde o BI é aplicado de forma mais comum (Watson, 2015)

Análise Preditiva: Permite prever ações futuras baseadas em tendências de dados históricos. Alguns dos exemplos de análise preditiva são:

- Uso de informação passada para entender que tipo de produtos os clientes de um dado negócio possam ter interesse com base nos dados mais recentes, e se estão propensos a adquirir novamente.
- Num cenário de um negócio que tenha um orçamento baixo para as campanhas de marketing, é possível usar as análises preditivas, e em conjunto com as análises descritivas saber quais os clientes que poderão vir a comprar um dado produto.

Os algoritmos e métodos para análise preditiva incluem análise de regressão, *machine learning* e redes neurais. Estas técnicas já existem há algum tempo e são tradicionalmente conhecidas como *Data Mining/Machine Learning* (Watson, 2015).

Análise Prescritiva: Na análise prescritiva é possível saber qual a melhor ação a tomar para uma dada situação. Enquanto na análise descritiva é exibido o que já aconteceu e a análise preditiva tenta prever o que irá acontecer depois, a análise prescritiva usa essa informação para demonstrar potenciais soluções com base em situações semelhantes (dados de ano a ano, dados sazonais, ou de lançamento de produtos). Por exemplo, vendas de bilhetes para um evento festivo estão abaixo das vendas do ano passado. A análise prescritiva poderá sugerir que os preços sejam diminuídos ou sejam tomadas outras decisões para responder a essa quebra de vendas.

Para Davenport & Harris (2007), os sistemas de Business Analytics permitem obter benefícios e vantagem competitiva nas organizações de diferentes áreas, incluindo apoio ao cliente, Marketing, manufatura, planeamento de produção e operações de cadeia de fornecimento (*supply chain management*).

Davenport (2016), definiu 4 Eras de Business Analytics denominadas de: *Business Analytics* 1.0, 2.0, e 4.0. Chen et al. (2012) também fizeram uso da mesma nomenclatura para referenciar a evolução de *Business Analytics* ao longo do tempo (figura 2)

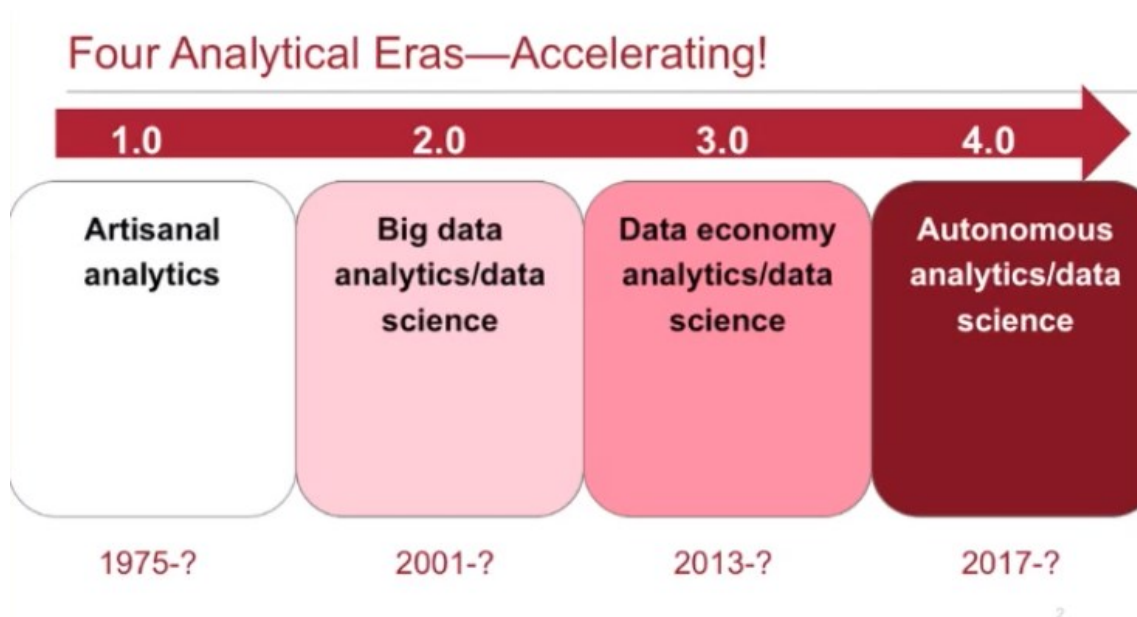


Figura 2: As 4 eras de Business Analytics

Fonte: Adaptado de Davenport, (2016)

Business Analytics 1.0:

Segundo Aleksic, Krasavac and Karamata (2019), a era do BA 1.0 refere-se ao período em que foram usadas todas as ferramentas analíticas disponíveis, sendo que este período abrangeu meados do século passado e terminou em 2005. Esta era provém da área de gestão de base de dados, dependendo completamente de várias tecnologias de recolha, extração e análise de dados (Chaudhuri et al. 2011; Turban et al. 2008). O valor desta fase foi na sua grande maioria impulsionado por objetivos centrados nas técnicas de apoio na tomada de decisão, em vez do uso de recursos preditivos avançados ou insights estatísticos granulares (Davenport, 2018).

Business Analytics 2.0:

O fenómeno do *Big Data* aconteceu no período de 2005 até 2012, onde este tipo de dados passou a ser uma única entidade denominada de *Big Data Analytics*. Este período foi considerado assim como *Analytics 2.0* onde empresas como a Google, Microsoft, Amazon, ou Facebook fizeram (e continuam a fazer) uso do *Big Data* para recolha de dados, manipulação e análise (Aleksic et al., 2019).

Neste período também começaram a ser utilizados métodos de *machine-learning*, os quais eram usados para gerar modelos a partir da maior quantidade de dados disponibilizada, o

que levou a uma mudança nas competências necessárias para a era 2.0. Os analistas quantitativos passaram a ser chamados de cientistas de dados e necessitavam de ter competências tanto ao nível analítico como ao nível computacional, o que levou a que os mesmos quisessem por iniciativa própria trabalhar em novas ofertas de produtos e ajudarem a melhorar os negócios. (Davenport, 2013)

Business Analytics 3.0:

Esta era que começou em 2013 e que ainda se estende até aos dias de hoje, envolve o conceito do Big Data ter sido aceite pelas grandes empresas e o *analytics* ter-se tornado uma peça fundamental para o negócio das organizações, tendo sido usado para o apoio na tomada de decisões e na criação de produtos focados em *analytics* (Davenport, 2018). De acordo com Aleksic et al. (2019), que identificam a era 3.0 ter iniciado em 2012 até aos dias de hoje, referem que os esforços analíticos focaram-se no uso combinado das duas capacidades analíticas anteriores. Para estes autores, o objetivo das empresas nesta era não é só o uso de BA para apoio na tomada de decisão, mas sim também para a possibilidade de criação de novos produtos e serviços.

Business Analytics 4.0:

Segundo Davenport (2016), nesta era que ainda mal começou, o *analytics* será usado para automatizar tarefas e decisões. O uso do Facebook com o *analytics* para detetar e resolver situações básicas de hardware, tais como automatizar uma solução de instalação de servidores. Com esta automatização, o Facebook só necessita de um engenheiro de redes para cada 25.000 servidores.

Aleksic et al. (2019), referem que a era 4.0, quando comparada com as eras anteriores, será mais autónoma e sem a dependência do fator humano, sendo que novos tipos de empregos baseados em *analytics*, inteligência artificial e machine learning e *machine learning* irão estar disponíveis. Davenport (2018) também estima que grande parte das grandes empresas ainda se encontre a usar *Analytics 1.0*, mas denota o uso potencial de *Analytics 3.0* e *4.0* nas empresas de TI.

Para Aleksic et al. (2019), todos os tipos de *Business Analytics* estão, em prática, a funcionar lado a lado sem qualquer hipótese de conflito entre os diferentes tipos.

2.2 Business Intelligence

O termo *Business Intelligence* (BI) foi criado por Richard Millar Devens em 1865, tendo sido posteriormente reformulado por Hans Peter Luhn, investigador na IBM em 1958. (Oey et al.,2021). Em 1989 Howard Dresner, investigador e analista de BI no Gartner Group, descreveu BI como “um termo abrangente que engloba aplicações, infraestruturas, ferramentas, e as melhores práticas que permitem o acesso e a análise de informações para melhorar e otimizar decisões e desempenho”.

Analisando esta descrição de BI, Watson (2009), considera que a mesma é uma definição ampla, na medida em que o BI não engloba apenas aplicações, mas também tecnologias e processos. Acrescenta também que o processo de BI inclui “retirar dados” (através de ferramentas e aplicações), mas também a “entrada de dados” (para um *data mart* ou *data warehouse*).

Turban et al. (2014) consideram como objetivo principal de *Business Intelligence* o acesso interativo aos dados, permitindo assim a exploração dos mesmos pelos gestores e analistas de dados das organizações. Definem também o BI como um termo genérico que inclui arquiteturas, ferramentas, bases de dados, e metodologias. Na perspetiva de Chen et al. (2012), o BI é considerado como as técnicas, tecnologias, sistemas, práticas, metodologias e aplicações que analisam dados de negócio críticos para ajudar as organizações a poderem compreender melhor os seus negócios e mercados, assim como para tomarem decisões relativas aos negócios de forma mais rápida.

Segundo Power (2018) O BI, tal como é entendido nos dias de hoje, evoluiu a partir dos sistemas de suporte à decisão (DSS) que se iniciaram na década de 1960 e desenvolveram-se por volta da década de 1980. O DSS originou-se nos modelos auxiliados por computador criados para auxiliar na tomada de decisões e planeamento. A partir do final dos anos 80, tecnologias como DSS, DW, *Executive Information Systems* (EIS), OLAP e BI já eram predominantes.

De acordo com Visinescu et al. (2015), O BI disponibiliza aos elementos responsáveis pela tomada de decisão com dados, informação, ou conhecimento, para que possam resolver decisões acerca de problemas específicos, e que podem também ser ‘agregados’ para apoiar a tomada de decisão em níveis mais amplos de uma organização.

Nos dias de hoje, o *Business Intelligence* não deve ser considerado uma nova tecnologia, mas sim uma solução integrada para organizações e empresas, nas quais as necessidades

de negócio são o fator principal que motiva a inovação tecnológica (Ranjan, 2005). Para Hočevár and Jaklič, (2008), um sistema de BI não consiste numa única aplicação, mas sim em diversos componentes que estão fortemente relacionados entre si, permitindo assim aos utilizadores escolher os dados a serem analisados, agregados, e publicar os dados numa forma fácil de ser compreendida. Jaklič ressalva ainda que de um ponto de vista arquitetural, um sistema de BI consiste nos seguintes elementos:

- Bases de dados operacionais e externas como fontes de dados
- Processo ETL (*Extract, Transform, Load*)
- *Data Warehouse*
- Ferramentas analíticas, as quais consistem em ferramentas de consulta e de relatórios, ferramentas OLAP, *data mining*, painéis de controlo, soluções avançadas de análise (Cenários *What-if*, otimização, análises estatísticas, entre outros.)

A *Commonwealth Scientific and Industrial Research Organisation* (CSIRO), (citada por Niu & Zhang, 2009), referiu em 2003 as 5 etapas associadas aos processos de BI:

- **Aquisição de Dados:**

Na fase de aquisição de dados, o sistema de BI terá que conseguir extrair dados provenientes de diferentes fontes de dados tais como marketing, recursos humanos ou financeiros. Após a sua extração, os dados terão que ser limpos, transformados e integrados (fase ETL - *Extract, Transform, Load*) para que possam ser devidamente analisados.

- **Análise de Dados:**

Utilizando as diversas técnicas de análises de dados existentes, tais como *reporting*, modelação, visualização e *data mining*, os dados poderão ser convertidos em informação nesta fase. Os resultados provenientes da análise de dados ajudam as organizações a poderem obter uma melhor compreensão analítica e a possibilidade de melhorarem o seu processo de tomada de decisão.

- **Situation Awareness:**

Com base nos resultados provenientes da etapa anterior (análise de dados) será possível obter uma compreensão mais eficiente e profunda das decisões atuais. Esta fase

denomina-se como *situation awareness*, e é um requisito fundamental para a tomada de decisão.

- **Avaliação de Risco:**

Uma avaliação de risco feita em condições permitirá às organizações identificar possíveis ameaças e oportunidades no futuro, assim como poderem atuar sobre as mesmas de forma eficiente. O processo de tomada de decisões nas organizações está constantemente em risco de ser alterado devido a fatores externos, assim como fatores dentro das próprias organizações.

Devido a estas situações, a avaliação de risco é uma fase importante num sistema de BI.

- **Apoio à Decisão:**

Com base nos dados atuais, o objetivo final do BI é proporcionar às organizações uma estrutura sólida com a qual possam tomar decisões mais eficientes.

Watson, (2009), descreve uma série de benefícios do BI para as organizações (figura 3):

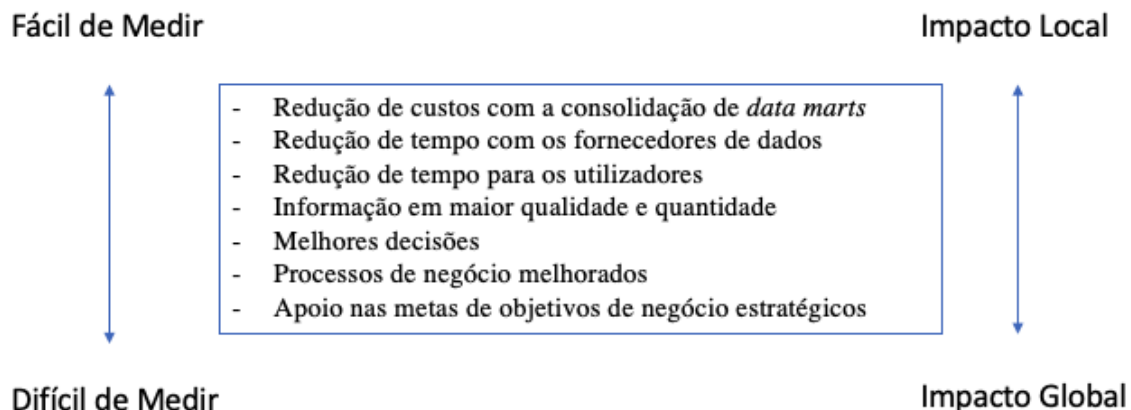


Figura 3: Benefícios do BI nas organizações

Fonte: Adaptado de Watson, (2009)

Algumas das vantagens são a redução de custos ao consolidar múltiplos *data marts* num *data warehouse* centralizado. Outros benefícios, tais como o apoio nas metas de objetivos de negócio estratégicos, já são mais difíceis de se alcançar, sendo que por norma, é mais fácil avaliar os benefícios após a implementação do projeto de BI do que antes (Watson, 2009).

Já Irani and Love, (2001), concluem que o BI é uma das áreas das TI nas quais as técnicas tradicionais de avaliação podem ter resultados insatisfatórios, pois muitos dos benefícios são estratégicos, e conseqüentemente, não são facilmente quantificáveis.

2.3 O Business Intelligence como suporte ao processo de tomada de Decisão nas Organizações

Em finais dos anos 60 e inícios da década de 70, os primeiros Sistemas de Apoio à Decisão (SAD) ou Decision Support Systems (DSS) começaram a ser discutidos. Uma das definições de SAD mais antigas provém de Keen and Scott Morton (1978), na qual definem um SAD como uma fusão de recursos intelectuais de indivíduos com as capacidades dos computadores para melhorarem a qualidade de decisões. Para estes autores, os computadores desempenham um papel importante na gestão do processo de tomada de decisão, sendo principalmente destinados para:

- Auxiliar os gestores no processo de tomada de decisão em tarefas semiestruturadas.
- Apoiar, ao invés de substituir, as decisões da organização.
- Melhorar a eficácia do processo de tomada de decisão.

Segundo Sharda et al. (2014) ao analisar dados históricos e atuais, situações e desempenhos, os tomadores de decisão obtêm informações valiosas que lhes permitem tomar decisões mais informadas e melhores. O processo de BI é baseado na transformação de dados em informações, depois em decisões e, finalmente, em ações.

Shim et al. (2002) definem um SAD como sistemas e subsistemas interativos para ajudarem os gestores na interação com tecnologias de comunicação, dados, documentos, conhecimento e/ou modelos para complementarem tarefas de tomada de decisão. Na opinião de Rippen (2005), um SAD é qualquer ferramenta usada para melhorar o processo de tomada de decisão em sistemas complexos, principalmente onde a informação seja apresentada de forma incoerente ou incompleta.

Para Vercellis (2009), o SAD é o processo pelo qual um indivíduo tenta preencher uma falha ou aproveitar uma oportunidade nas condições operacionais atuais de um determinado sistema (como ele é), e as condições mais vantajosas a serem alcançadas no futuro (o que vai ser). A natureza do processo de tomada de decisão depende das

características da organização à qual o sistema é introduzido, das atitudes dos gerentes responsáveis pelos fatores decisórios, da disponibilidade de metodologias apropriadas para resolução do problema e da disponibilidade de tecnologias eficazes que visem apoiar a tomada de decisão (Vercellis, 2009).

De acordo com Power (2003), os SAD ajudam os gestores a tomarem decisões fora do comum. Os SAD, para além de usarem os dados da organização, podem ainda usar dados que não sejam provenientes da própria organização. Desta forma podem ser usados os modelos qualitativos e quantitativos do processo de tomada de decisão e análise. (Power, 2003). Atualmente, o uso dos sistemas baseados em BI tem-se tornado mais comum e como tal, o processo de tomada de decisão tem vindo a melhorar nas organizações. Para que tal possa continuar a ser feito, é necessário analisar o BI e a sua importância na melhoria das decisões estratégicas da organização (Power, 2003).

Para Olszak and Ziembra, (2007) os sistemas de BI são soluções que são responsáveis pelo processo de transformar dados em informação e conhecimento, recorrendo a diversas tecnologias (ETL, DW, OLAP, *Data mining*, entre outros) para além de poderem criar um ambiente para tomadas de decisão mais eficazes, pensamento estratégico e ações nas organizações, conforme exemplificado na figura 4.

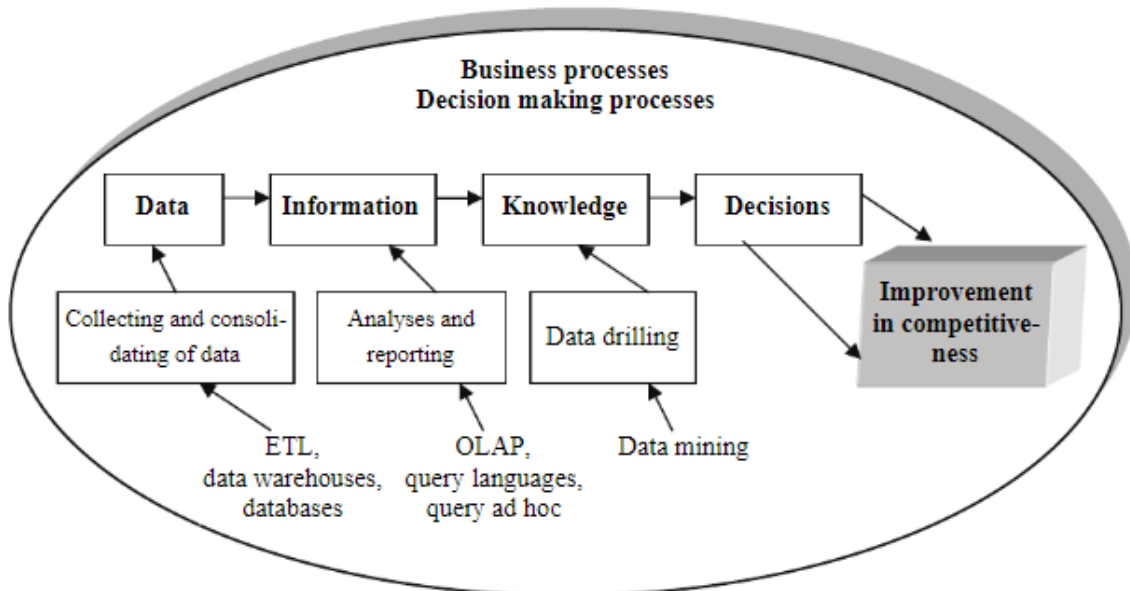


Figura 4: Papel do Business Intelligence na tomada de decisão

Fonte: Olszak & Ziembra (2007).

Olszak and Ziemba (2003) referem que o BI pode apoiar o processo de tomada de decisão em todos os níveis de gestão, independentemente do nível de estruturação em que se encontram e fazem esta distinção em 3 níveis (estratégico, tático e operacional).

- **Estratégico:** Neste nível é possível delinear objetivos de forma bastante precisa e acompanhar a realização dos mesmos. O BI permite a criação de diferentes relatórios comparativos, e.g.: podemos referir resultados históricos, rentabilidade de uma oferta em particular, a eficácia dos canais de distribuição, assim como a realização de simulações de desenvolvimento ou previsão de resultados futuros baseados em presunções (Olszak & Ziemba, 2003).

Para O'Brien et al. (2009), no nível estratégico não é possível especificar de antemão a maioria das decisões a seguir.

- **Tático:** A nível tático, os sistemas de BI poderão possibilitar uma base inicial para a tomada de decisão nos departamentos de marketing, vendas, financeiro, gestão de capital, entre outros. Os sistemas permitem ainda otimizar ações futuras assim como modificar aspetos de desempenho das organizações que sejam de teor organizacional, financeiro ou tecnológico de forma que seja possível ajudar as organizações a aperceberem-se dos seus objetivos estratégicos de forma mais eficaz (Olszak & Ziemba, 2003).

Os procedimentos de tomada de decisão podem ser especificados previamente, mas não o suficiente de modo a levar à decisão correta (O'Brien et al., 2009).

- **Operacional:** No nível operacional, os sistemas de BI são usados para efetuarem análises *ad-hoc* e responderem a perguntas relacionadas com as operações dos departamentos em análise, posições financeiras atualizadas, vendas e cooperação com fornecedores, clientes, etc... (Olszak & Ziemba, 2003).

Podem ser especificados de antemão os procedimentos necessários a seguir quando é necessária uma decisão (O'Brien et al., 2009).

Antonelli, (2010) conclui no seu artigo que o BI ganhará um status indispensável nas organizações mais relevantes de mercado, pois se houver um processo de tomada de decisões com base em informações úteis e confiáveis, será uma mais-valia para os empreendedores, o que resultará num maior número de decisões acertadas. Por sua vez,

Silva (2010), conclui no seu trabalho que o uso das ferramentas de BI torna-se uma opção viável e de diferencial competitivo no atual mercado, que se encontra bastante competitivo para as empresas.

2.4 Relação entre o Business Intelligence e Data Warehouses

O *Business Intelligence* e *Data Warehouses* são considerados dois processos que dependem um do outro, uma vez que a arquitetura de BI necessita de meios eficientes para armazenar e obter dados das organizações. Os dados necessitam de ser extraídos das fontes de dados e armazenados em bases de dados internas ou externas. Para além disso, os dados têm de ser acedidos facilmente pelas ferramentas de BI que são usadas para auxiliar os processos de tomada de decisão (Bhardwaj, 2021). A figura 5 detalha todos os componentes e etapas de um sistema de *Business Intelligence* e dos processos em paralelo, com destaque no *Data Warehouse* e nos sistemas ETL. Estes componentes serão mencionados em maior detalhe nas secções seguintes.

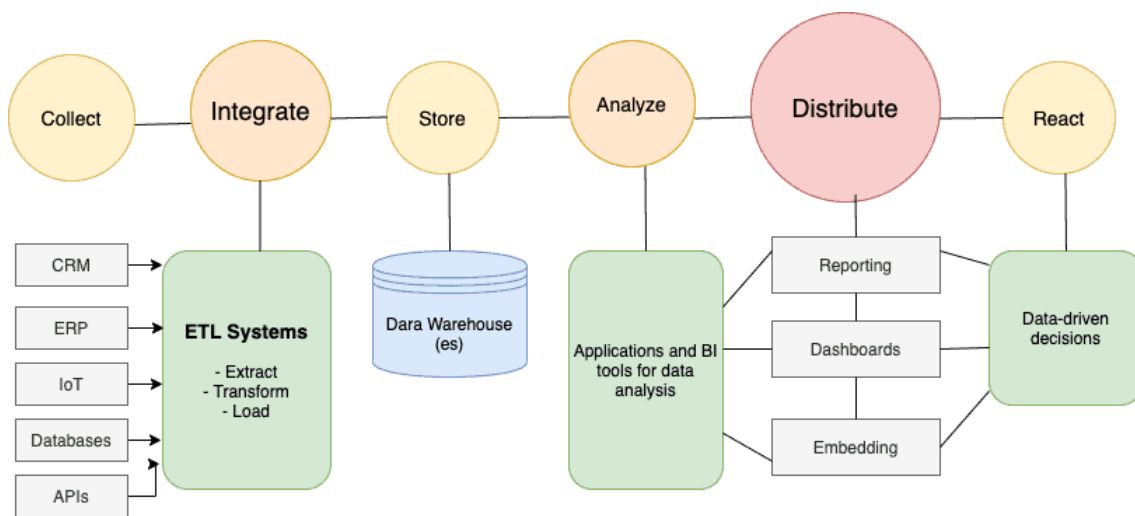


Figura 5: Arquitetura de software com foco na relação de BI com DW

Fonte: “Importance of data warehousing in BI software development services”

(Bhardwaj, 2022)

Para Monterey (2022), existe uma grande necessidade de armazenar dados tanto em BI como nos *data warehouses*. Para além do processo de obtenção, catalogação e análise de dados, o BI envolve uma metodologia de análise e obtenção de dados, uma vez que os processos de BI se baseiam nos *data warehouses* que por sua vez armazenam e organizam os dados recolhidos.

2.4.1 Conceito de Data Warehouse e Data Marts

Desde o início dos anos 90, os *Data Warehouse* têm estado na vanguarda do uso das tecnologias de informação para que as organizações usem de forma eficaz informações de negócio e planeamento de negócios e no apoio à tomada de decisão. (Jawadekar, 2010). O DW é uma das tecnologias mais poderosas que emergiram nas últimas décadas (Ramamurthy et al., 2008). O DW recolhe dados a partir de vários departamentos de uma organização para que possam ser tratados e organizados recorrendo ao sistema OLAP, o qual será explicado mais à frente.

Segundo Inmon (1993), o qual é considerado por muitos como o pai do *Data Warehouse*, um Data Warehouse é definido como “uma coleção de dados orientada por assunto, integrada, variável no tempo e não-volátil, que proporciona apoio ao processo de tomada de decisões”. Lane, and Schupmann (2002) descrevem em maior detalhe as características de um DW:

- **Orientada por Assunto:** Os DW são desenhados com o objetivo de ajudar na análise de dados, como por exemplo, para saber mais acerca da informação de vendas de uma dada empresa, cria-se uma DW focada nas vendas.
- **Integrada:** Dados provenientes de diferentes fontes necessitam de ser integrados na DW e torná-los num formato heterogéneo numa única fonte de dados. Deverão também resolver problemas de conflitos e inconsistências entre as unidades de medida. Ao atingirem estes objetivos serão consideradas integradas.
- **Variável no Tempo:** A fim que seja possível descobrir tendências históricas ou atuais nos negócios, é necessário obter grandes quantidades de dados. O foco de uma DW na mudança ao longo do tempo é o que é explicado pelo termo “variável no tempo”.
- **Não-volátil:** Assim que os dados sejam introduzidos na DW, estes não poderão voltar a ser modificados. Este é um procedimento lógico dado que o propósito de uma DW é de analisar o que ocorreu no passado.

Já Kimball (1996) define um DW como "uma cópia dos dados de transação estruturados especificamente para consulta e análise" e apresenta uma definição mais concreta sob a forma de requisitos necessários:

1. O Data Warehouse possibilita acesso empresarial ou organizacional
2. Os dados do DW devem ser consistentes.
3. Todos os dados presentes num DW podem ser separados e combinados através de qualquer medição num negócio (o requisito clássico de *slice-and-dice*).
4. O DW não consiste exclusivamente em fontes de dados, mas também usa um conjunto de ferramentas para consultar e analisar dados presentes e passados.
5. O DW é o local onde se é publicado dados que já tenham sido usados
6. A qualidade dos dados no DW é um impulsionador da reengenharia de negócios.

Ressalvam Alhyasat and Al-Dalahmeh (2013), que os meios para recuperar e analisar dados, extrair, transformar e carregar dados, e gerir o dicionário de dados, também são considerados componentes essenciais de um data warehouse. De acordo com Kimball (2011), um DW deve cumprir os seguintes objetivos:

- Tornar a informação facilmente acessível.
- Apresentar a informação da organização de forma consistente.
- Capaz de se adaptar a qualquer mudança.
- Proteger a informação.
- Servir como a fundação de um processo de tomada de decisão melhorado.

Lane and Schupmann (2002) referem os seguintes elementos constituintes de um DW tradicional, conforme resumido na figura mais abaixo:

- Uma base de dados relacional (é um tipo de base de dados que armazena e fornece acesso a pontos de dados relacionados entre si) para armazenar e gerir dados.
- Uma solução ETL (*Extract, Transform, Load*) para preparar os dados para análise.
- Análises estatísticas, relatórios e capacidades de *data mining*.

- Ferramentas de análise de cliente para visualizar e apresentar dados para utilizadores empresariais.
- Outras aplicações mais sofisticadas que possam gerir ações acionáveis por meio da aplicação de algoritmos de ciência de dados e inteligência artificial (IA), ou recursos gráficos e espaciais que permitam outros tipos de análise de dados em escala.

Chowdhury (2014) divide o Data Warehouse em 5 camadas (Aquisição de dados, Integração de dados, Repositório de dados, Analítica e Apresentação), sendo que cada camada desempenha uma função específica:

- **Aquisição de dados:** Esta camada consiste em componentes para obter dados a partir de todas as fontes de informação, tais como recursos humanos, finanças, e de faturação.
- **Integração de dados:** Consiste em componentes de integração e transformação de dados provenientes de fontes de dados para o repositório de dados da arquitetura do DW.
- **Repositório de dados:** Armazena dados num modelo relacional para otimizar o desempenho das consultas e a extensibilidade.
- **Analítica:** Armazena dados no formato de cubo para facilitar análises hipotéticas pelos utilizadores.
- **Apresentação:** Aplicações ou portais que providenciem acesso a conjuntos diferentes de utilizadores. As aplicações e portais consomem os dados através de páginas web ou *portlets* que são definidos na ferramenta de relatório ou através de serviços web.

De acordo com a IBM (2020), um *Data Mart* é um subconjunto de um DW que contém dados específicos de uma determinada linha de negócios ou departamento. Ao conterem um subconjunto menor de dados, os DM permitem que um departamento ou linha de negócios possa descobrir *insights* focados de forma mais eficiente ao trabalharem com o conjunto de dados do DW de uma forma mais ampla.

Na opinião de Chenoweth, Schuff, and Louis, (2003) dado que os DM têm como objetivo visualizar os dados de maneira única, o processo começa sempre com a análise das necessidades do cliente. Ressalvam também que o software de virtualização de dados é

aplicável ao criar os *Data Mart*, combinando-os com outros dados necessários para atender aos requisitos de empresas específicas, oferecendo assim às equipas de gestão das bases de dados um nível de controlo sobre os registos da organização ao longo do seu ciclo de vida.

Segundo a Oracle (1999), os Data Marts podem ser divididos em 3 tipos de Data Marts: dependentes, independentes e híbridos, sendo que a categorização é baseada maioritariamente na fonte de dados que alimenta o Data Mart.

1. **Data Mart Dependente:** Um DM dependente permite às organizações unificar os seus dados num DW, levando assim a que os dados fiquem centralizados. Para além disso, ao centralizar-se os dados, o processo de extração de informação de uma enorme quantidade de dados torna-se mais simplificado, uma vez que os dados formatados e resumidos já foram carregados para o DW centralizado (Beri and Rydhn, 2016).
2. **Data Mart Independente:** Um DM independente pode ser criado sem quaisquer dados de referência de um DW centralizado. Para além disso, a extração de informação através de um DM independente torna-se mais fácil quando comparada com um DM dependente. As fontes de dados para DM independentes poderão ser menores e a quantidade de dados associados com o DM será menor (Beri and Rydhn, 2016).
3. **Data Mart Híbrido:** Ao usar-se um DM híbrido a organização terá mais facilidade para combinar a entrada de dados provenientes de diferentes fontes, para além do DW. Este tipo de DM é habitualmente usado para alcançar o melhor desempenho, e disponibilizar os dados com o menor custo relevante para um departamento específico da organização. (Beri and Rydhn, 2016).

2.4.2 Arquitetura de Inmon vs Arquitetura de Kimball

De acordo com Breslin (2004), a abordagem de Bill Inmon num DW defende uma abordagem *top-down*, a qual adapta as ferramentas tradicionais de base de dados relacionais para as necessidades de desenvolvimentos de uma EDW (*Enterprise Data Warehouse*). A partir das *data store* criadas, são então criadas bases de dados individuais para poderem ser usadas na maioria das necessidades de apoio à decisão. No modelo de Inmon, os dados no DW são integrados, o que significa que o DW é a fonte

dos dados que termina nos DM existentes. Desta forma preserva-se a integridade dos dados e a sua consistência na organização (Figura 6).

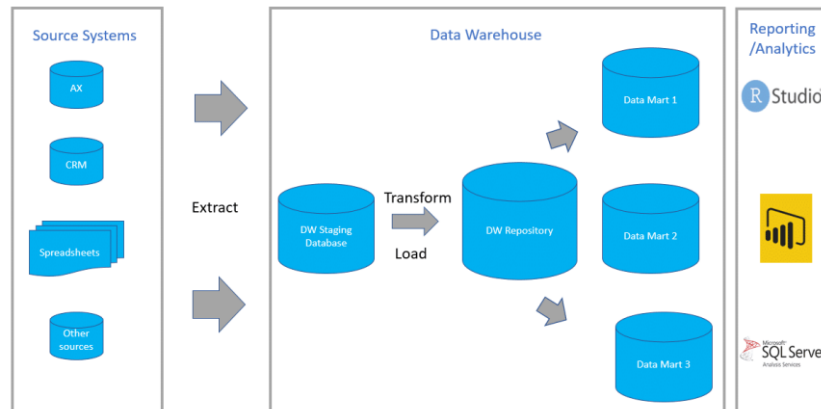


Figura 6: Abordagem top-down de Inmon.

Fonte: “What is a data warehouse?” (Burki, 2018)

Por outro lado, Kimball (1996) sugere uma abordagem *bottom-up* (de baixo para cima) que faz uso de modelação dimensional, que é uma abordagem de modelação de dados exclusiva para *Data Warehousing*, conforme ilustrado na figura 7. Em vez de ser criada uma única base de dados empresarial, Kimball sugere que seja criada uma base de dados (ou *Data Mart*) por cada processo de negócio.

Na abordagem de Kimball, o DW é o conglomerado dos vários data marts existentes. Isso contrasta com a abordagem de Inmon, que cria data marts com base nas informações do warehouse. “O data warehouse nada mais é do que a união de todos os data marts.” (Kimball, 2013).

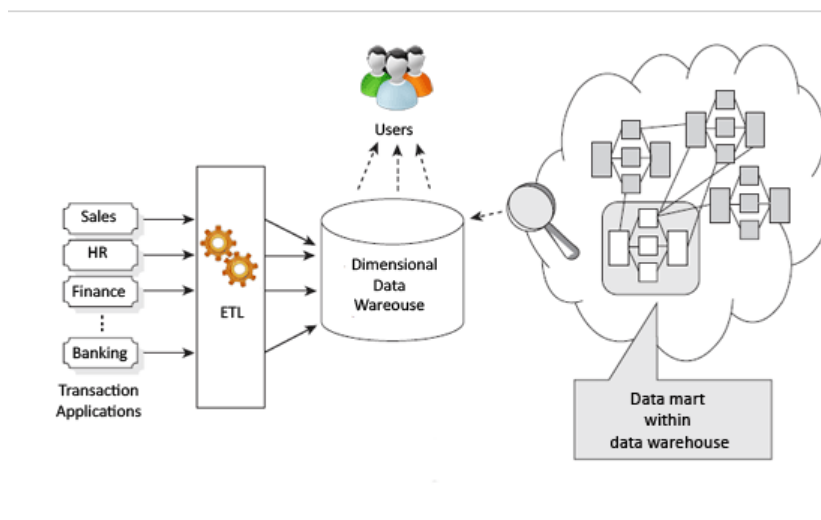


Figura 7: Abordagem bottom-up de Kimball

Fonte: “Ralph Kimball data warehouse architecture” (Zentut, 2021)

As semelhanças que mais se destacam entre os modelos de Inmon e Kimball são o uso de dados com *timestamp* e o processo de extração, transformação e carregamento (*Extract-Transform-Load* ETL). Embora a execução desses dois elementos seja diferente entre ambos os modelos, os atributos de dados e os resultados de consultas são muito semelhantes. (Breslin, 2004)

Os comentários mais comuns acerca da abordagem da Inmon é que sua fase de modelagem de dados requer altos níveis de especialização e um maior tempo de desenvolvimento inicial. Por outro lado, a crítica mais comum à abordagem Kimball é que ela não se foca em toda a empresa e concentra-se principalmente nas unidades de negócios individuais ou grupos de utilizadores (Jukic, 2006).

A comparação das abordagens de Inmon e Kimball encontram-se resumidas na tabela 1.

Tabela 1: Comparação das abordagens de Inmon e Kimball

Fonte: Adaptado de Breslin, (2004)

	Inmon	Kimball
Metodologia e Arquitetura		
Abordagem	<i>Top-down</i> (de cima para baixo)	<i>Bottom-up</i> (de baixo para cima)
Estrutura arquitetural	O DW de toda a empresa "alimenta" bases de dados de departamentos	Os Data Marts modelam um processo de negócio individual, sendo que a consistência da empresa é alcançada através de um barramento de dados e dimensões em concordância.
Complexidade do método	Bastante complexo	Relativamente fácil
Comparação com metodologias de desenvolvimento existentes	Derivada da metodologia espiral	Processo de 4 passos; uma despedida dos métodos RDBMS (bases de dados relacionais)
Discussão do projeto físico	Bastante detalhada	Bastante leve
Modelação de dados		
Orientação dos dados	Orientados por assunto ou por dados	Orientados por processos
Ferramentas	Tradicional (ERDs, DISs)	Modelação dimensional; abandono da modelação relacional
Acessibilidade do utilizador final	Baixa	Alta
Audiência principal	Profissionais das TI	Utilizadores finais
Lugar na organização	Parte integral da corporate information factory (CIF) (fábrica de informações corporativas)	Transformador e retentor de dados operacionais
Objetivo	Disponibilizar uma solução robusta e técnica baseada em métodos e tecnologias comprovados	Disponibilizar uma solução que possibilite aos utilizadores finais fazerem consultas aos dados e ao mesmo tempo conseguirem tempos de resposta razoáveis

Na opinião de Breslin (2004), o desenvolvimento de um *data warehouse* deverá ter em grande consideração uma equipa com um conhecimento sólido dos modelos de Inmon e Kimball, onde estará mais que apta para propor uma visão de um DW que seja compatível com as características e metas estabelecidas pelas organizações.

2.4.3 Extração, Transformação e Carregamento de Dados

ETL é um acrónimo para *Extract, Transform e Load* (Extrair, Transformar, Carregar), e consiste num processo de integração de dados que combina dados de múltiplas fontes para um modelo de dados que é assim carregado para um *data warehouse*, ou outro tipo de sistema. (Microsoft, n.d.). A figura abaixo resume a arquitetura do processo ETL.

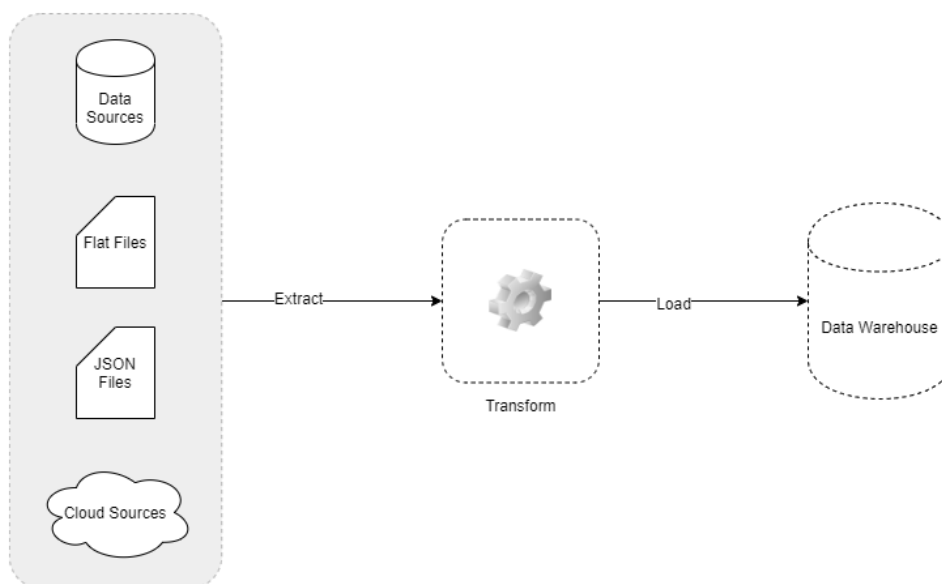


Figura 8: Arquitetura ETL para um data warehouse

Fonte: “An overview of ETL and ELT architecture” (Das, 2020)

Devido ao aumento de popularidade das bases de dados nos anos 70, foi introduzido o processo ETL para integrar e carregar dados para computação e análise, tornando-se eventualmente o método principal para o processamento de dados para projetos de *data warehousing*. (IBM, 2020)

Segundo Jhawar (2022), as organizações enfrentam dificuldades na obtenção de dados provenientes de múltiplas fontes, em diversos formatos. Os dados da fonte de dados poderão não corresponder ao que já está no *data warehouse* ou necessitar de haver

limpeza e revisão aos dados a serem carregados para o DW. Nesse sentido foram criadas diversas ferramentas para colmatar este problema, sendo o processo ETL um deles.

Kimball (2013), define o processo ETL em 3 passos:

- **Extract** (Extração): O processo de extração de dados é o primeiro passo na obtenção de dados para o *data warehouse*, onde os dados das fontes de origem são lidos e copiados para que possam ser manipulados na fase seguinte.
- **Transform** (Transformação): Na fase de transformar os dados as possibilidades de manipulação e alteração dos dados são imensas, tais como:
 - **Limpeza de dados:**
 - Remoção de duplicados
 - Remoção de valores nulos
 - Conversão de tipos de dados (e.g.: texto para números)
 - **Enriquecimento de dados:**
 - Complementar os dados com informação nova (e.g.: junção de fontes de dados)
 - Renomear campos com valores incorretos ou em falta
 - Calcular valores de diversos campos
 - Agregação de dados

A fase de transformação permite assim aos dados extraídos serem alterados e melhorados.

- **Load** (Carregamento): Após os dados terem sido transformados, são então carregados para os modelos dimensionais do *data warehouse*, ficando assim completos e disponíveis para serem usados por ferramentas de *Business Intelligence*.

Kimball and Ross (2013) reconhecem que o sistema ETL consome demasiados recursos tanto a nível de tempo como de esforço necessário nos ambientes de *data warehouses* e de *business intelligence*, os quais poderão influenciar negativamente a qualidade do desenho do processo ETL. Fatores como as exigências de negócio, fontes de dados,

orçamento disponível e o conhecimento humano são fatores externos que não podem ser controlados.

Kimball and Ross (2013), sugerem então um processo de desenho do processo ETL, o qual está dividido em 4 componentes chave.

- **Extração de dados:** Os dados são obtidos a partir das fontes de dados, onde poderão ficar guardados no disco do ambiente ETL antes de haver qualquer reestruturação aplicada nos dados carregados.
- **Limpeza e conformação de dados:** Os dados são melhorados ao passarem pelas várias etapas do sistema ETL. Esta fase é uma das fases mais críticas devido ao seu valor no melhoramento dos dados e o benefício que trará para a organização. Para além disso, esta fase pode ser usada para diagnosticar problemas derivados das fontes de dados.
- **Entrega de dados:** Os dados são estruturados e carregados para os modelos dimensionais do *data warehouse*.
- **Gestão do sistema ETL:** Gestão dos sistemas e processos do ambiente ETL de forma coerente.

2.4.4 Modelação Multidimensional – Cubo OLAP – Factos e Dimensões

A modelação dimensional é uma técnica de modelação que está presente há quase 5 décadas com o objetivo de simplificar as bases de dados. A simplicidade é crítica, dado que garante que os utilizadores conseguem entender corretamente os dados, assim como permite o software navegar e providenciar resultados de forma eficiente e rápida (Kimball & Ross, 2013). O seu propósito principal é disponibilizar navegação através dos dados a utilizadores sem experiência, para que possam realizar consultas sem a ajuda de profissionais das TI. (Kimball & Ross, 2013).

Segundo Cheung et al. (2001), as técnicas de armazenamento de dados mais populares são as MOLAP (*Multidimensional OLAP*) e ROLAP (*Relational OLAP*). Na arquitetura MOLAP os dados e as agregações pré-calculadas são armazenados em conjuntos no servidor OLAP, ou seja, no modelo de dados multidimensional. Na arquitetura ROLAP os dados são armazenados em bases de dados relacionais e com a exigência de ter de usar linguagem SQL (Structured Query Language) (Pirnau & Botezatu, 2010).

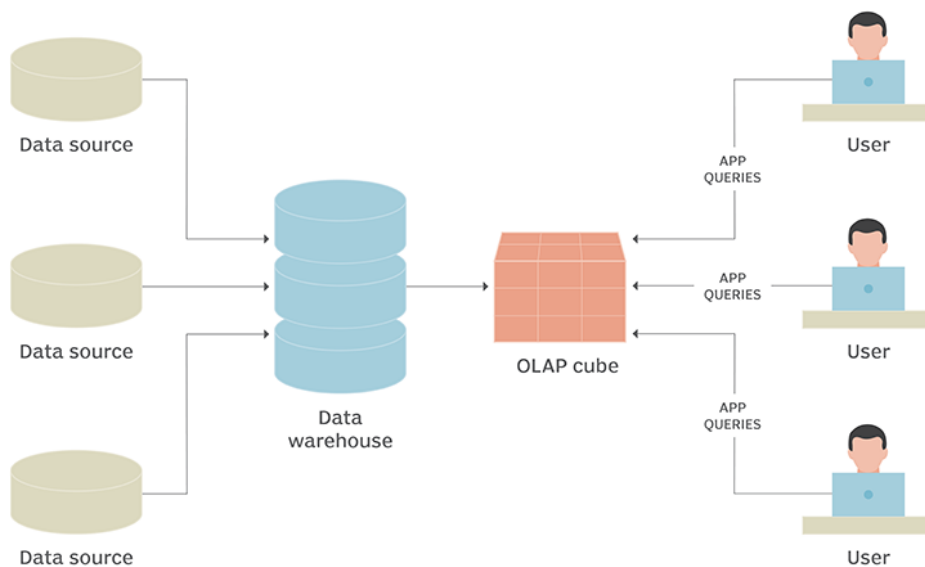


Figura 9: Arquitetura OLAP

Fonte: “What is OLAP (online analytical processing)?” (Biscobing, 2020)

Segundo Pedersen (2009), as bases de dados multidimensionais são normalmente baseadas no esquema em estrela (*star schema*) ou então no esquema floco de neve (*snowflake schema*). Estes esquemas encontram-se explicados em maior detalhe abaixo. Esquema em Estrela (*Star Schema*): Num esquema em estrela a tabela de facto (UnitsSold) está situada no centro, rodeada de diversas tabelas de dimensão. Entre a tabela de facto e as tabelas de dimensão existe uma relação de um-para-muitos. A tabela de facto representa na sua maioria transações de negócio ou eventos. (Martyn, 2004) Esta estrutura pode ser visualizada na figura 10 abaixo.

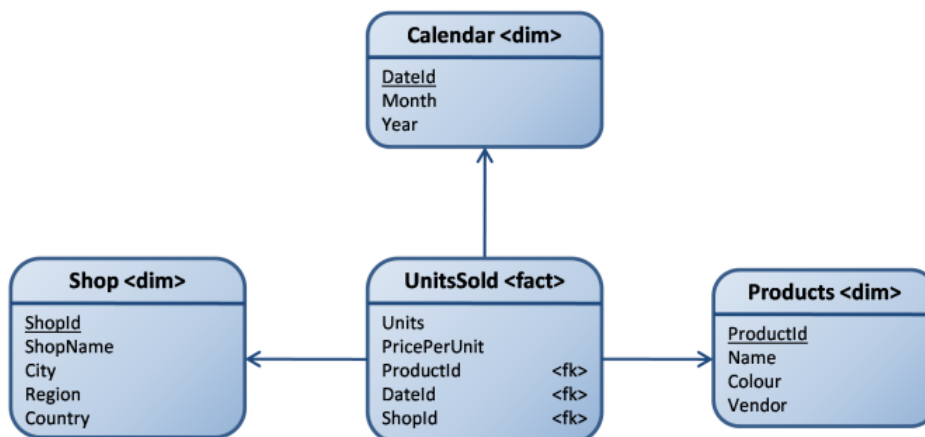


Figura 10: Esquema em estrela (Star Schema)

Fonte: Westerlund, (2008)

Esquema Floco de Neve (*Snowflake Schema*): De acordo com Martyn (2014), o esquema floco de neve é uma extensão do esquema em estrela, sendo que as tabelas de dimensão são normalizadas, permitindo assim a divisão dos dados em tabelas adicionais. Esta estrutura pode ser visualizada na figura 11.

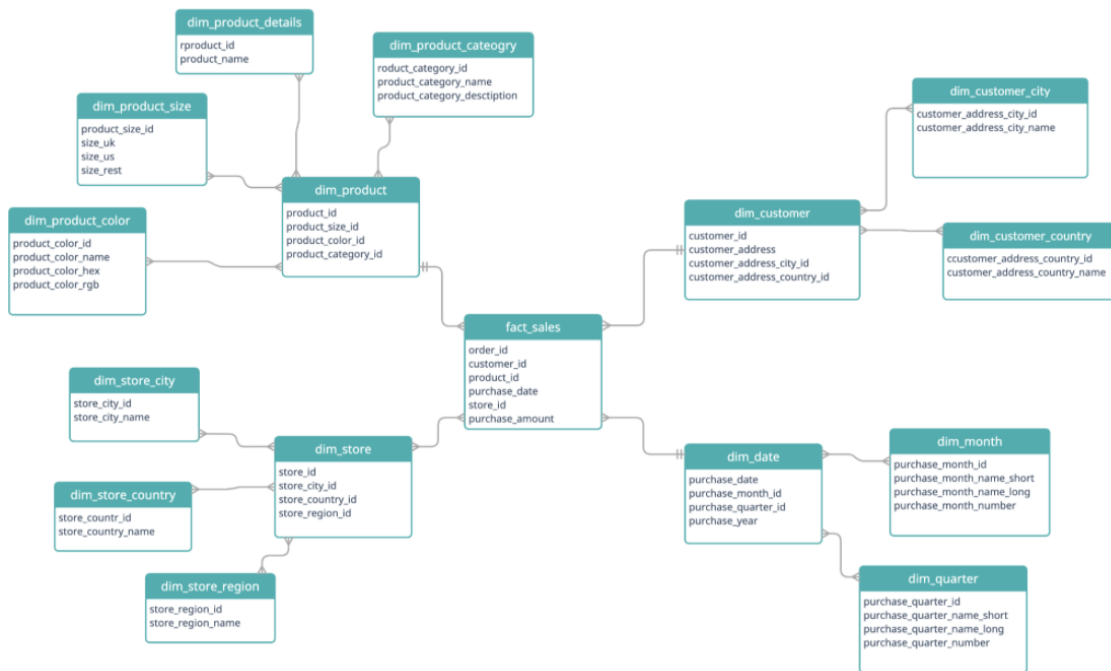


Figura 11: Esquema Floco de Neve (*Snowflake Schema*)

Fonte: “*Star schema vs snowflake schema and the 7 critical differences*”

(Keboola, 2022)

De acordo com Mohammed (2019), no seu trabalho de implementação de um *Data Warehouse* baseado em *snowflake schema* (esquema floco de neve) e *star schema* (esquema em estrela), concluiu que o *star schema* seria o mais apropriado com base nas seguintes conclusões:

- **Complexidade das consultas:** O esquema em estrela é mais fácil e simples de entender, ao passo que o esquema floco de neve tem uma maior complexidade envolvida devido às várias chaves estrangeiras que unem as tabelas de dimensão.
- **Desempenho das consultas:** Há um maior desempenho no esquema em estrela devido à *framework* que este esquema usa baseado em dados preditivos, enquanto o *snowflake schema* usa uma maior quantidade de junções de chaves estrangeiras, fazendo com que o tempo de execução seja maior em comparação com o esquema em estrela.

Para Martyn (2004), o *star schema* é também o esquema mais eficiente pelos seguintes motivos:

- Um esquema que faça uso de tabelas desnormalizadas terá menos operações de junção.
- As ferramentas de otimização possuem lógica suficiente para reconhecer um *star schema* e possibilitarem operações de junção de tabelas mais eficientes.

Kimball propõe um processo de 4 etapas para desenho da modelação multidimensional (Kimball & Ross, 2013):

1. Escolha do processo de negócio:

Os processos de negócio são as atividades operacionais feitas pelas organizações, tais como registo de encomendas, clientes. Os eventos de negócio registam as métricas necessárias que são posteriormente adicionadas a uma tabela de factos. A escolha do processo é importante pois permite definir uma modelação exata que por sua vez permite obter a granularidade, dimensões e factos corretamente. (Kimball & Ross, 2013)

2. Definir a granularidade:

A granularidade estabelece de forma exata como serão representados os campos na tabela de facto. Esta etapa terá de ser definida antes das dimensões e factos pois terão que ser consistentes com a granularidade definida nesta etapa. Um nível de grão de dados alto significa que teremos menos informação dos dados disponíveis, ao passo que se tivermos um nível de granularidade baixo, então significa que teremos mais informação disponível dos dados. (Kimball & Ross, 2013)

3. Identificar as dimensões (Tabelas de Dimensão):

As dimensões (figura 12) esclarecem o contexto de “quem, o quê, quando, onde, porquê e como” de um dado negócio. As tabelas de dimensão contêm assim os atributos descritivos que são usados pelas aplicações de BI para filtrar e agrupar os factos. (Kimball & Ross, 2013).

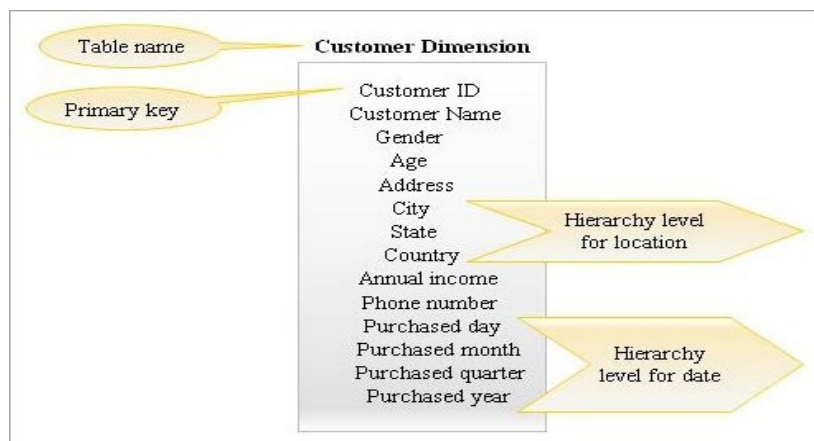


Figura 12: Exemplo de uma tabela de dimensão

Fonte: “Dimensional Data Model in Data Warehouse” (Software Testing Help, 2022)

4. Identificar os factos (Tabelas de Factos):

Os factos (figura 13) são os objetos que representam o sujeito das análises desejadas, e.g.: o evento ou processo considerado interessante para ser analisado com o objetivo de compreender melhor o seu comportamento. (Pedersen, 2009) A tabela de factos é a identidade do *data warehouse* a qual será usada para as necessidades de negócio das organizações. A tabela de facto irá mostrar os resultados dos processos de negócio necessários às necessidades da organização. (Kimball & Ross, 2013)

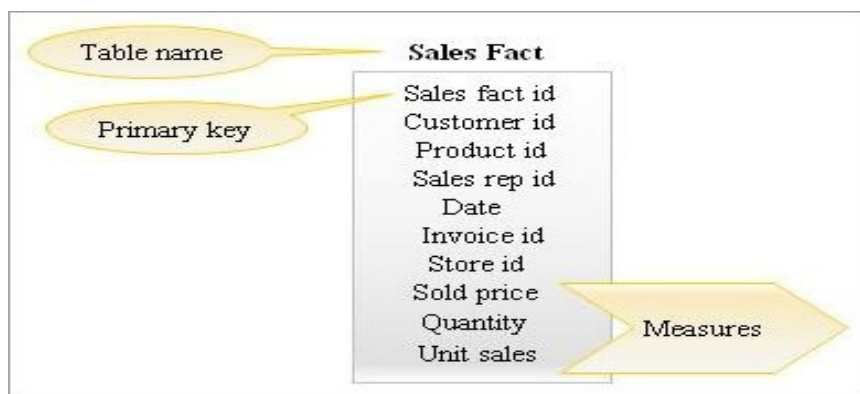


Figura 13: Exemplo de uma tabela de facto

Fonte: “Dimensional Data Model in Data Warehouse” (Software Testing Help, 2022)

De acordo com a definição de Watson (2009), o *Online Analytical Processing*, ou OLAP como é normalmente conhecido, é uma tecnologia que suporta a análise de dados multidimensionais. Codd (1993), cunhou o termo OLAP em 1993, mas até aos dias de hoje ainda não há uma definição formal (Abelló & Romero, 2009). Segundo Codd (1993),

o OLAP representa uma classe de software que possibilita o apoio na tomada de decisão e de relatórios baseados num *data warehouse*.

Para Vassiliadis (1999), o foco do OLAP destina-se a providenciar análises multidimensionais à informação disponível, sendo os dados organizados em cubos OLAP (ou hipercubos) em que consistem em diversas dimensões. Cada dimensão presente compreende um conjunto de níveis de agregação (ou hierarquias). Um hipercubo, ou seja, um cubo com mais de três dimensões (figura 14) é fundamental para que seja possível entender o software que faça uso dos mesmos da mesma forma que as bases de dados usam tabelas (Thomsen, 2002). Kimball and Ross (2013) assemelham o cubo OLAP a um *star schema*, dado que faz uso de atributos dimensionais e factos.



Figura 14: Exemplo de Cubo OLAP tridimensional

Fonte: *Understanding OLAP cubes - A guide for the perplexed* (Keboola, 2022)

Conforme Han et al. (2012), existem várias maneiras de se analisar os dados num cubo OLAP, sendo as operações de análise de dados mais comuns denominadas de *roll-up*, *drill-down*, *slice and dice* e *pivot*:

- **Roll-up:** Esta operação também pode ser denominada de *drill-up*, sendo possível com este procedimento aceder a um aumento do nível da granularidade dos dados

ou diminuição do nível de detalhe através de uma ou mais hierarquias de dimensão (figura 15) (Han et al; Santos & Ramos, 2017).

- **Drill-down:** Conforme ilustrado na figura 15, esta operação é o oposto do *roll-up* (ou *drill-up*), ou seja, permite efetuar uma diminuição do nível da granularidade dos dados ou aumento do nível de detalhe através de uma ou mais hierarquias de dimensão (Han et al, 2012; Santos & Ramos, 2017).

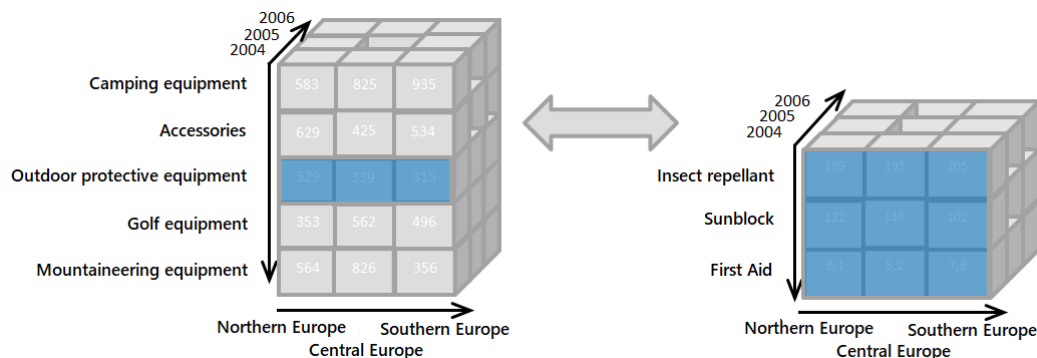


Figura 15: Operação Drill-down/Drill-up no cubo

Fonte: “Understanding OLAP cubes - A guide for the perplexed” (Keboola, 2022)

- **Slice and dice:** A operação *slice* seleciona uma dimensão do cubo, transformando-o num cubo mais pequeno. Já a operação *dice* define um subcubo ao escolher duas ou mais dimensões, conforme ilustrado na figura 16 e na figura 17 (Han et al, 2012; Santos & Ramos, 2017).

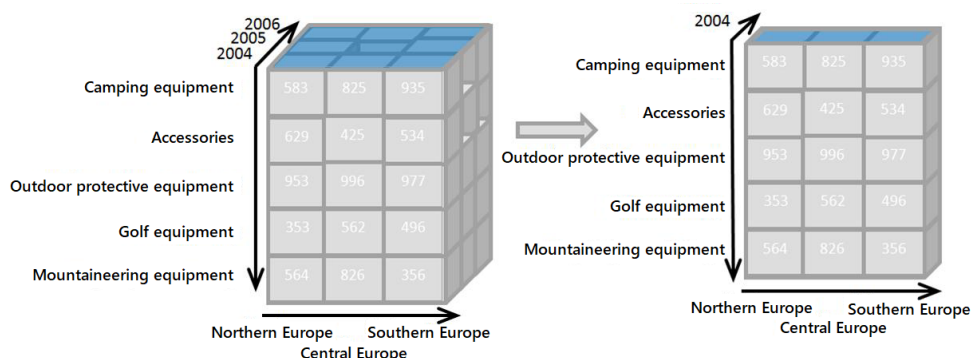


Figura 16: Operação Slice no cubo

Fonte: “Understanding OLAP cubes - A guide for the perplexed” (Keboola, 2022)

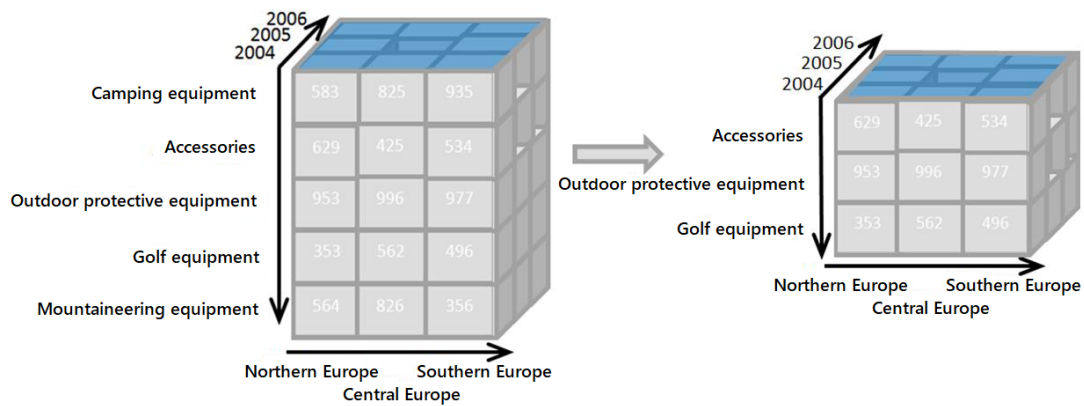


Figura 17: Operação Slice and Dice no cubo.

Fonte: “Understanding OLAP cubes - A guide for the perplexed” (Keboola, 2022)

- **Pivot:** Também conhecido como *rotate*, esta operação permite a orientação dos eixos da visualização multidimensional dos dados, restringindo assim o conjunto de dados em análise, conforme figura 18 (Han et al, 2012; Santos & Ramos, 2017).

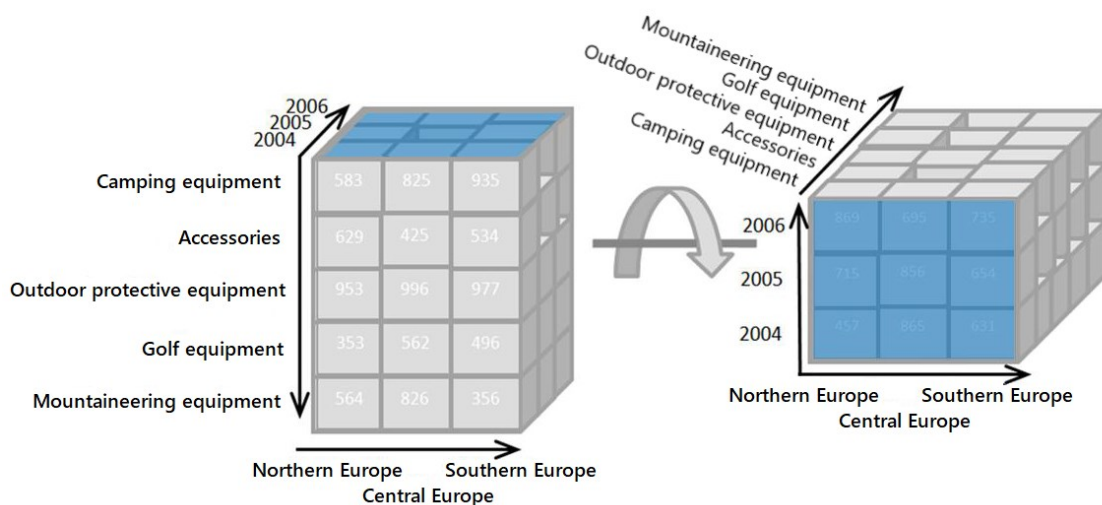


Figura 18: Operação Pivot no cubo

Fonte: “Understanding OLAP cubes - A guide for the perplexed” (Keboola, 2022)

2.4.5 Data Science

O Data Science é o estudo da extração generalizável de conhecimentos a partir de dados (Dvar, 2013). Combina matemática e estatísticas, programação especializada, análise de dados avançada, inteligência artificial e machine learning com conhecimentos específicos sobre o assunto para descobrir percepções acionáveis escondidas nos dados

de uma organização. Estes conhecimentos podem ser utilizados para orientar a tomada de decisões e o planeamento estratégico (IBM, 2022).

Para Provost and Fawcett (2013), o DS é um conjunto de princípios fundamentais que apoiam e guiam as técnicas de extração de informação e dados a partir de dados. De acordo com a IBM (2022), comparativamente ao Business Intelligence, que usa dados descritivos para compreender o que aconteceu no passado para informar uma linha de ação, o Data Science usa os mesmos para determinar variáveis preditivas, que serão posteriormente usadas para a categorização de dados ou para análises preditivas.

Tal como com o Business Intelligence, não existe atualmente uma definição que seja universalmente aceite para definir o que é o Data Science, contudo, as diferentes perspectivas indicam que o DS proporciona oportunidades para tomadas de decisão baseadas em dados, previsões, descobertas, recomendações e uma abordagem diferente para proporcionar soluções tanto na investigação como na prática (Ayankoya et al., 2014).

2.4.6 Machine Learning

O *Machine learning* é um ramo em evolução dos algoritmos computacionais que são concebidos para imitar a inteligência humana aprendendo com o ambiente circundante (El Naqa, 2015). O termo “*Machine Learning*” é de autoria do cientista informático americano Arthur Samuel, o qual em 1959 definiu o ML como uma "capacidade do computador para aprender sem ser explicitamente programado" (Samuel, 1967).

De acordo com o CGD (2017), o ML, utilizando um vasto conjunto de dados históricos, assim como uma ou mais medidas de desempenho, permite que através da execução dos seus algoritmos, o ML possa criar de forma autónoma modelos de representação de conhecimento. Estando este treino finalizado, o modelo criado possibilitará efetuar previsões de qualidade em situações futuras que estejam associadas a padrões históricos, mediante o acesso à Internet.

Segundo El Naqa (2015), o ML pode ser dividido de acordo com a natureza dos dados em 3 formas:

- *Supervised Learning* (Aprendizagem Supervisionada): Este tipo de aprendizagem usa exemplos desconhecidos de entrada e de saída para estimar um mapeamento

desconhecido de entrada e de saída, no qual a saída é classificada. Os algoritmos de classificação e de regressão fazem uso da aprendizagem não supervisionada.

- *Unsupervised Learning* (Aprendizagem não supervisionada): Na aprendizagem não supervisionada, apenas são disponibilizadas amostras de entrada ao sistema de aprendizagem. Os algoritmos de *clustering* (agregação) e de *density estimation* (estimativa de densidade) são usados com este tipo de aprendizagem.
- *Semi-supervised learning* (aprendizagem semi-supervisionada): A aprendizagem semi-supervisionada é uma combinação dos métodos de aprendizagem supervisionada e não supervisionada, onde parte dos dados são parcialmente identificados e a parte identificada é usada para identificar a parte não identificada. Este tipo de aprendizagem é muito usado por sistemas de recuperação de conteúdo audiovisual.

Sutton (1998) refere ainda uma outra forma de aprendizagem de ML, a qual é denominada de *reinforced learning*, no qual o sistema tenta realizar uma sequência de ações para maximizar uma recompensa cumulativa tal como ganhar um jogo de damas. Este tipo de abordagem é mais indicado para aplicações de aprendizagem online.

Ainda segundo este autor, o *Machine Learning* aliado com o Business Intelligence traz às organizações inúmeros benefícios, sendo alguns deles:

- Eliminar as suposições;
- Alertar as organizações para as crises emergentes;
- Obtenção de informações acerca do comportamento dos clientes;
- Melhorar a satisfação dos clientes;
- Saber quais são os custos reais de fabrico;
- Aumento da receita;
- Gestão mais eficiente do inventário.

Para Tamang et al. (2021), o uso de Machine Learning nas PMEs poderá ser um investimento demasiado caro, contudo, concluem que a implementação de ML com BI será bastante efetiva com o uso dos algoritmos mais eficientes de acordo com as necessidades de cada negócio. Assim sendo, o uso de ML tornar-se-á uma parte integral de todos os sistemas, quer sejam pequenos ou grandes.

2.4.7 Ferramentas de BI

Para que seja possível aumentar o nível de eficácia de funcionamento das empresas, é necessário recorrer ao uso de soluções de BI que estejam disponíveis no mercado. Independentemente do tamanho da organização, as ferramentas de BI são uma grande ajuda na resolução de um dos maiores problemas da gestão moderna - a elevada quantidade de informação (Stecyk, 2018). Segundo Kronz et al. (2021), as plataformas de *Analytics e Business Intelligence* permitem a utilizadores menos experientes ou a empresários modelagem, analisar, explorar, partilhar, gerir os dados, e colaborarem na descoberta de resultados, possibilitados pelas TI e aumentados pela inteligência artificial. Ainda de acordo com estes autores, o foco das plataformas atuais de BI e *Analytics* está a migrar dos analistas de dados para o consumidor ou por quem toma as decisões. Para que tal seja possível, os *insights* automatizados terão de ser relevantes para os objetivos dos utilizadores e não apenas relevantes num contexto estatístico. De uma forma geral, este tipo de plataformas está a começar a obter informações mais relevantes acerca dos comportamentos e interesses dos seus utilizadores de forma a conseguirem fornecer uma experiência mais impactante a todos os consumidores.

2.4.7.1 Power BI

A Microsoft, desenvolvedora dos produtos Power BI, consegue em 2022, pelo 15º ano consecutivo, a posição de líder de mercado no Quadrante Mágico da Gartner para *Analytics e Plataformas de Business Intelligence*, conforme resumido na figura 19:



Figura 19: Quadrante Mágico Gartner para BI.

Fonte: Gartner (2022)

De acordo com a Microsoft (2022), o Power BI é um conjunto de ferramentas que trabalham de forma unificada para converter fontes de dados que não estejam relacionadas em informações coerentes, fazendo uso de aspetos visuais e interativos. Segundo Kronz et al. (2021), O Microsoft Power BI permite realizar operações de preparação de dados e explorar os mesmos recorrendo a diversas opções visuais, painéis interativos e funcionalidades de análise aumentada. Disponível desde 2015, as fontes de dados no Power BI podem ser provenientes de folhas de cálculo do Excel, dados na *cloud* ou provenientes de um data warehouse, facilitando também a partilha destes dados com qualquer pessoa.

Assim sendo, o Power BI é considerado um self-service BI, uma vez que não necessita de conhecimento prévio em áreas de programação e tem um rendimento elevado nas transformações dos dados em visualizações dinâmicas (Santos, 2018), (Ferrari & Russo, 2016).

O desenvolvimento do Power BI foi baseado no Microsoft Excel, e desse modo, a curva de aprendizagem do Excel para o Power BI é pequena, e qualquer pessoa que consiga usar o Excel irá conseguir usar o Power BI, sendo esta última no entanto uma ferramenta muito mais complexa (Wright, 2019).

O Power BI dispõe consiste em três elementos:

- **Power BI Desktop:** Aplicação que é instalada no computador.
- **Power BI Service:** Aplicação na *cloud*, fornecida como um SAAS (*software as a service*).
- **Power BI Mobile:** Aplicações móveis para dispositivos Android, iOS e Windows.

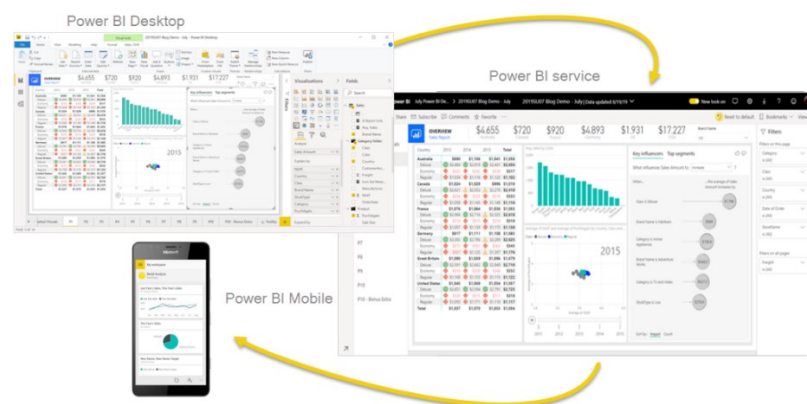


Figura 20: Versões do Power BI

Fonte: Microsoft (2022)

Segundo a Antdata (2022), o Power BI Desktop pode ser usado de forma gratuita, permitindo aos utilizadores ligarem-se a diversas fontes de dados, fazer a análise do mesmo e na criação de *dashboards*. Na versão gratuita não é possível partilhar o conteúdo criado no Power BI no Power BI *Service*, sendo necessário uma subscrição Pro, contudo, é possível obter um trial de 60 dias para testar as funcionalidades da versão Pro. Já a versão *Mobile* destina-se somente à visualização das *dashboards* feitas na versão *Desktop* ou *Service*.

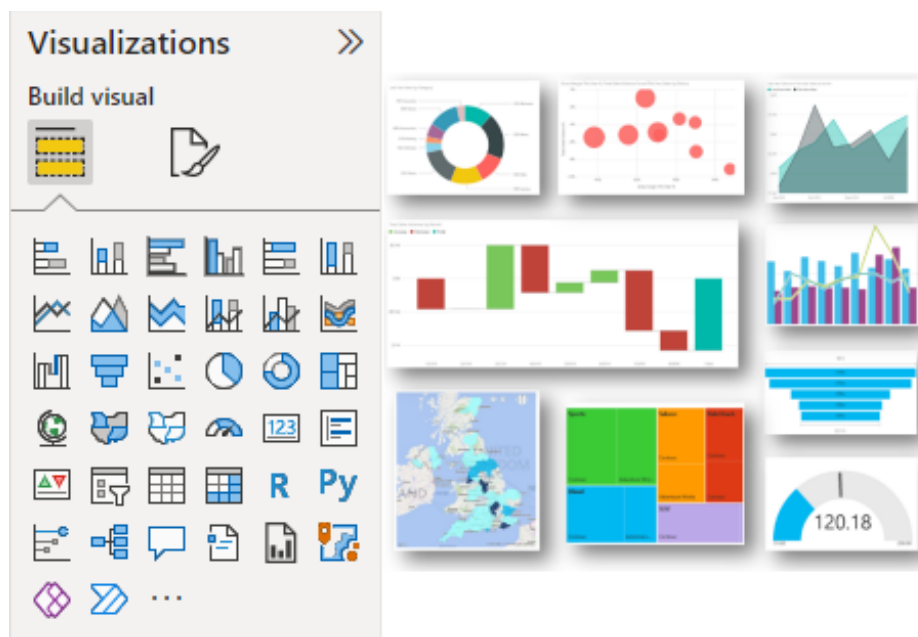


Figura 21: Visualizações possíveis de usar no Power BI Desktop e Service

Fonte: Microsoft, (2022)

Uma funcionalidade que é bastante útil na visualização de dados e que está disponível na versão Desktop e Service do Power BI, é a criação de marcadores de relatório (ou *bookmarks*), que permite guardar todos os filtros criados e ordem de visualização dos elementos visuais, garantindo assim que a *dashboards* apresentem as visualizações conforme o marcador seleccionado. (Microsoft, 2022a).

De acordo com Rad (2019), o Power BI possui diversos componentes:

- Power Query
- Power Pivot
- Power View
- Power Map

- Power Q&A
- Power BI Desktop

O Power Query é responsável por ler, transformar, e carregar os dados para a ferramenta de BI. Além disso, permite filtrar e combinar dados a partir de uma ou mais fontes de dados compatíveis (Rad, 2019). Esta operação é conhecida como mash-up de dados usando o Power Query Formula Language (ou linguagem M) (Klopfenstein & Howell, 2022).

O Power Pivot é um componente de modelação de dados que permite criar modelos de dados, estabelecer relações, e criar cálculos, o qual utiliza a linguagem *Data Analysis Expression* (DAX) para modelar dados simples e complexos (Rad, 2019). Através do *xVelocity in-memory analytics engine*), o Power Pivot proporciona um rápido rendimento em grandes volumes de dados através do uso de uma base de dados em formato de colunas para armazenar os dados (Clark, 2014). Uma outra vantagem do XVelocity engine é a sua capacidade em processar os dados, sendo primeiro carregados para a memória RAM e só depois consultados, o que resulta em tempos de processamento mais rápidos (Clark, 2014).

O Power View permite apresentar os dados de forma visual através das diversas opções de slice and dice existentes. (Rad, 2019).

O Power Map é um componente que permite visualizar informação geo-espacial em 3D através do Bing maps. (Rad, 2019).

O Power Q&A reconhece as palavras introduzidas e vê a melhor maneira para providenciar a resposta mais adequada. Este componente ajuda a formular as perguntas através de reformulação, autopreenchimento, assim como outras ajudas textuais e visuais (Gowthami & Kumar, 2017).

O Power BI Desktop permite que os dados sejam analisados através de relatórios intuitivos, assim como usar funcionalidades de *drag-and-drop* para colocar elementos visuais onde seja necessário, permitindo assim descobrir padrões e explorar os dados em dashboards interativos (Gowthami & Kumar, 2017).

2.4.7.2 Tableau

O Tableau é uma solução proprietária de business intelligence bastante apelativa, sendo uma ferramenta de visualização capaz de fornecer visualizações interativas de forma instantânea devido à sua natureza de *drag and drop* (Nair et al., 2016). Foi fundada em 2003 em Mountain View, Califórnia, tendo esta plataforma sido adquirida em 2019 pela Salesforce (Mackinlay, 2021). Utiliza dados em bruto e apresenta-os em formatos fáceis de entender. A análise de dados feita pelo Tableau é mais rápida e os resultados gerados são guardados como folhas de cálculo usando extensões .twb, ou .twbx (Khedikar, 2021). O Tableau dispõe de uma linha de produtos (figura 23), os quais possibilitam a criação de ficheiros Tableau pelos utilizadores, assim como facilitar a colaboração entre outros utilizadores, na leitura local ou através da Web, e também publicar os resultados na internet (Khedikar, 2021).

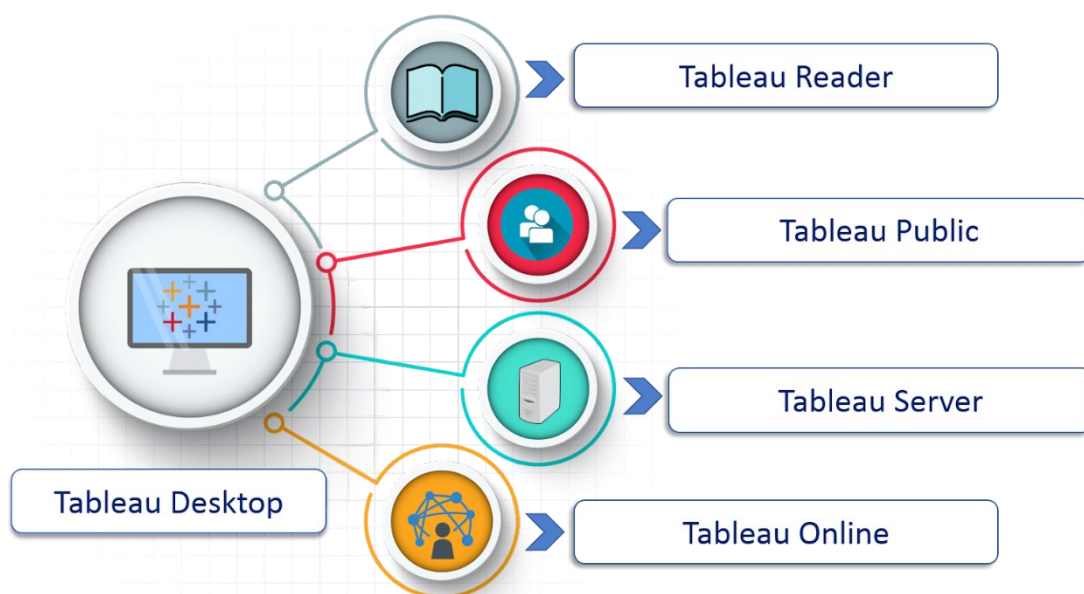


Figura 22: Linha de produtos tableau

Fonte: “Top 50+ tableau interview questions and answers for 2023” (Edureka, 2022)

- Tableau Desktop: Esta versão é disponibilizada em duas versões:
 - Professional Edition: Suporta todas as fontes de dados e liga-se ao Tableau Server para análises baseadas na web.
 - Personal Edition: Só suporta os formatos Excel, Access e CSV de fontes de dados e não é disponibilizado apoio técnico (Gounder et al., 2016). Em 2019 a Tableau anunciou que todas as atualizações disponibilizadas em

2021 deixariam de ser compatíveis com a Personal Edition do Tableau Desktop (Jewett, 2019).

- Tableau Server: O Tableau Server é uma solução de BI online e móvel, que é usada para publicar dashboards recorrendo à ferramenta Tableau Desktop e partilhá-las com a organização. Suporta todas as fontes de dados da versão Desktop e é compatível com todos os navegadores de internet e dispositivos móveis (Jewett, 2019).
- Tableau Online: Esta versão é uma versão alojada do Tableau Server, sendo que é usada para publicar dashboards com o Tableau Desktop e partilhá-las com parceiros comerciais, colaboradores ou clientes. Nesta versão é fácil de encontrar as respostas rapidamente e disponibiliza um armazenamento de 100 Gigabytes para licenças comerciais (Jewett, 2019).
- Tableau Public: O Tableau Public destina-se aos utilizadores que pretendam contar histórias interativas de dados na web. Com esta versão é possível fazer ligações aos dados, criar visualizações interativas dos dados e publicá-los diretamente num website (Jewett, 2019).
- Tableau Reader: Através desta versão é possível abrir, visualizar e explorar os dados presentes em dashboards criadas com a versão Tableau Desktop (Jewett, 2019).

2.4.7.3 Síntese Geral

Para Sekar (2022), o Power BI e o Tableau revelam-se ferramentas que permitem operações de *drag and drop* para a construção de visuais, sendo que todos os passos que fazem parte deste processo são na sua maioria os mesmos para este tipo de ferramentas analíticas. Segundo Mane et al. (2022), tanto o Power BI como o Tableau possuem os seus pontos fortes e fracos para a obtenção, armazenamento, processamento, análise e visualização de dados, onde tudo dependerá das necessidades das organizações e respectivas estratégias, contudo, para análises que façam uso de conjuntos de dados mais pequenos, o Power BI é a melhor escolha, enquanto para conjuntos de dados maiores é recomendado por estes autores o uso do Tableau.

Na análise comparativa de ferramentas BI feita por Lousa et al. (2019), estes autores concluíram que a ferramenta de BI mais apropriada dependerá do tipo de organização, ressaltando que o Tableau e o Power BI são as ferramentas líderes.

2.5 A importância da Visualização de Dados e dos Dashboards

A visualização de dados é uma maneira de apresentar e exibir informações de modo a incentivar a interpretação, escolha e associação. Faz uso das habilidades humanas para reconhecimento de padrões e análise de tendências para exploração da capacidade das pessoas no momento de extrair grandes quantidades de informações num curto período de tempo através de imagens apresentadas num formato padrão. (Yau, 2000).

Segundo Few (2007), a visualização de dados é um conjunto de métodos que fazem uso de representações visuais para explorar, facilitar, e comunicar dados quantitativos. Já Kirk (2016) define este conceito de forma mais simplificada, considerando assim a visualização de dados como a representação e apresentação de dados para facilitar o entendimento. Na opinião de Friedman (2008), o objetivo principal da visualização de dados é de comunicar a informação de forma clara e efetiva através de meios visuais. Ainda de acordo com o autor, a ideia é criar visualizações de dados que sejam tanto estéticas como funcionais de modo a fornecer formas intuitivas de análise e compreensão de dados complexos.

Para Baltzan (2014), a visualização de dados permite às organizações terem noção das avultadas quantidades de informação com que somos bombardeados e criar assim um mecanismo de defesa para auxiliar no processamento de toda a informação recebida. Se os dados forem apresentados num formato gráfico, as organizações serão capazes de encontrar semelhanças e fazer o levantamento de perguntas importantes (Embarak, 2018).

De acordo com Embarak (2018), a visualização de dados ajuda as organizações a alcançar diversas metas:

- Converter os dados de negócio em gráficos interativos para interpretações dinâmicas de modo a auxiliar as metas de negócio.
- Transformar os dados em *dashboards* apelativas de várias fontes de dados para auxiliar os negócios com os indicadores adequados.
- Permitir a criação de *dashboards* adicionais com diversas representações gráficas.
- Tomar decisões corretas ao permitir examinar em profundidade os dados.

- Descobrir padrões, tendências e semelhanças nos dados que estão a ser analisados de forma a poder determinar-se como podem melhorar os seus processos organizacionais, e desta forma, poderem expandir os seus negócios.
- Demonstrar de forma mais completa os dados que estão a ser alvo de análise.
- Permitir a tomada de decisão de forma mais efetiva, rápida e mais informada.

Embora não exista uma definição concreta e formal para um *dashboard*, podemos afirmar que tanto as dashboards como a visualização podem ser caracterizadas como ferramentas cognitivas com o objetivo de melhorar o controlo dos mais variados dados de uma organização (Kitchin et al., 2016). Stephen Few (2007a) define um *dashboard* como “uma representação visual num único ecrã que apresenta informações necessárias para atingir um propósito específico, o qual requer uma resposta oportuna”.

Uma *dashboard* comunica informação acerca de um domínio de conhecimento que já é familiar para a pessoa que a está a visualizar. Assim sendo, o indivíduo já tem um modelo do domínio construído na sua mente, um meio de organizar e pensar acerca do domínio para que faça sentido dele e para acompanhar alterações no que se está a passar (Rasmussen et al., 2009).

De acordo com Presthus and Bergum (2015), o propósito principal para uma *dashboard* de Business Intelligence é de apoiar o processo de tomada de decisão nas organizações. Na visão de Sharda et al. (2018), os *dashboards* permitem que os dados mais importantes sejam possíveis de ser visualizados num único ecrã, facilitando o entendimento por parte dos gestores. A figura 26 abaixo representa um exemplo visual de uma *dashboard* com diversos dados exibidos num único ecrã.

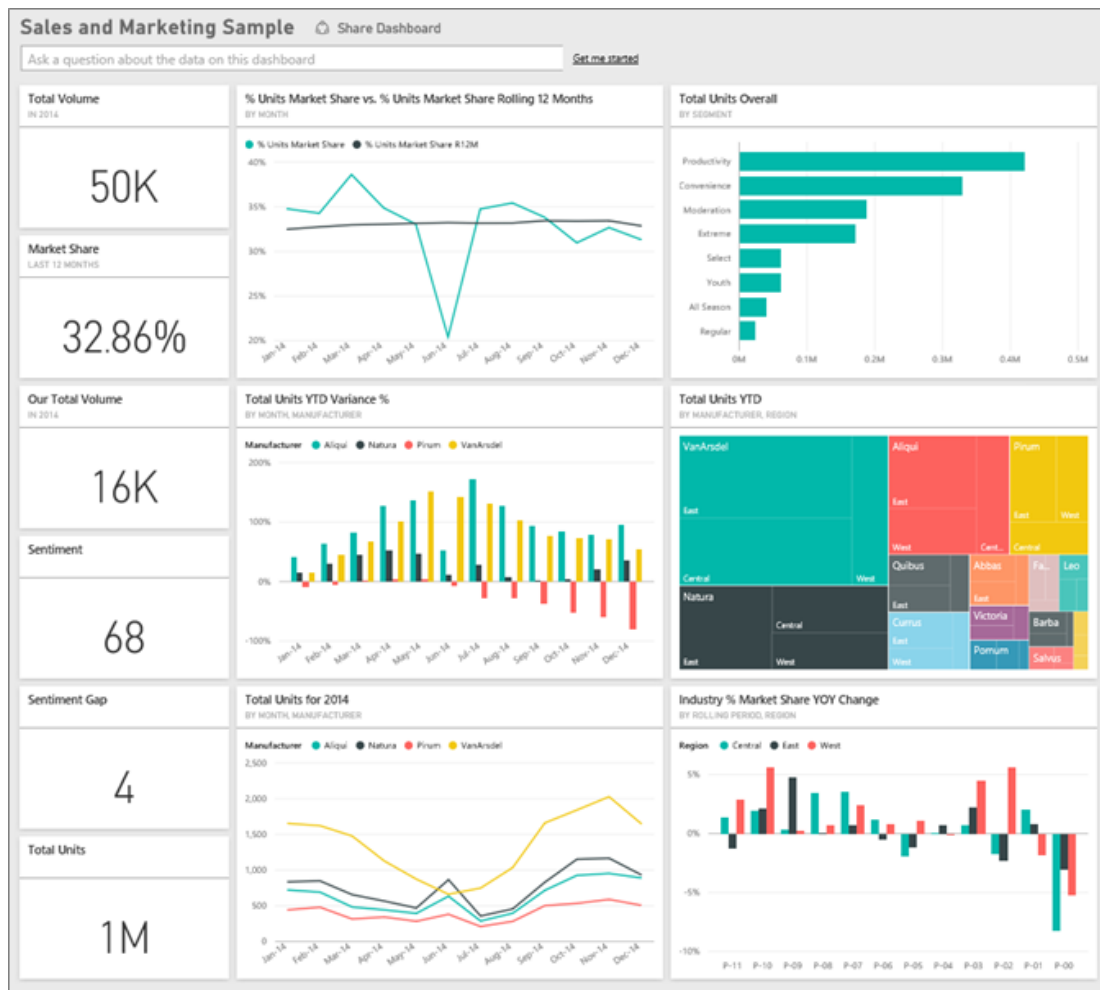


Figura 23: Exemplo de Dashboard

Fonte: Microsoft (2022)

Para a construção de um *dashboard*, Caldeira (2010) propõe 3 fases essenciais:

- Preparação Inicial
- Selecionar, relacionar e posicionar a informação
- Operacionalização da Monitorização

Na fase da preparação inicial do *dashboard* é dado principal destaque à identificação dos objetivos para o projeto, na perceção do negócio e da estratégia da empresa, na identificação dos destinatários, na relevância da informação e na identificação do estado dos dados que serão usados para a criação do *dashboard*.

Na segunda fase, também chamada de desenho da *dashboard*, será feita a seleção, relacionamento e posicionamento da informação pretendida. Por outras palavras, é feita a identificação dos indicadores chave de desempenho que estarão presentes no

dashboard, compreender como é que os dados se relacionam entre si e como é que a informação será apresentada.

Na terceira fase, operacionalização da monitorização, é quando se coloca o *dashboard* em funcionamento, procedendo-se à identificação da informação e como aceder à mesma no dashboard, definindo-se a periodicidade de atualização dos dados e determina-se a necessidade de detalhe da informação.

A escolha de cores tem um papel fundamental no desenho de dashboards. As *dashboards* usadas para mineração visual de dados usam cores como o verde, amarelo e vermelho para destacarem o rendimento bom, satisfatório e mau (Few, 2007). Os mecanismos de visualização também têm um papel fundamental. Ao disponibilizarem informação, a forma como a mesma é apresentada é crucial. (Few, 2013).

Ainda segundo Few, as características que uma *dashboard* corretamente desenhada deverá ter são as seguintes (Few, 2007a):

1. Organizada de forma correta: As *dashboards* deverão auxiliar quem faz uso das mesmas a reconhecer de forma imediata o que necessitam de visualizar.
2. Condensadas sob a forma de sumários e exceções. Os sumários representam um conjunto de números como um único número (somas e médias), enquanto as exceções representam algo incomum que esteja a acontecer; um problema ou uma oportunidade. Desta forma, a audiência da *dashboard* não necessita de ter de percorrer diversos valores para chegar aos resultados que pretende.
3. Específica e customizada para a audiência e objetivos da *dashboard*.
4. Exibida através do uso de ferramentas multimédia que comuniquem os dados e a respectiva mensagem da forma mais direta e clara possível.

Para Eckerson (2010), uma *dashboard* disponibiliza três funcionalidades fundamentais:

- Monitorizar processos críticos de negócio e atividades através do uso de métricas de rendimento de negócio, as quais serão usadas para alertar os utilizadores em caso de problemas.
- Analisar a principal fonte dos problemas através da análise detalhada de informação relevante e oportuna.
- Gerir pessoas e processos de modo a melhorar os processos de tomada de decisão e assim otimizar o rendimento de negócios.

Um dos pontos principais para uma gestão eficaz de rendimento em qualquer organização é o uso de métricas, e uma *dashboard* é uma interface popular para indicar essas métricas aos utilizadores. Contudo, caso uma empresa disponibilize uma série de iniciativas com fim a apoiar a sua estratégia, necessitam de ter ferramentas para poderem medir o sucesso (ou insucesso) dessas iniciativas, caso contrário, a habilidade da empresa em ajudar as suas iniciativas com o passar do tempo irá ficar dependente de opiniões subjetivas e estimativas. (Rasmussen et al., 2009)

Segundo Rasmussen et al., (2009) as dashboards podem ser divididas em 3 categorias:

- Estratégicas: Apoiam a visão de negócio das organizações com metas estratégicas.
- Táticas: Apoiam a medição de progresso em projetos chave ou iniciativas.
- Operacionais: Apoiam a observação de atividades de negócio específicas.

Nas *dashboards* estratégicas, também conhecidas como *scorecards*, permitem aos executivos sénior executarem estratégias, gerirem o rendimento ou a potenciar comportamentos eficientes nas organizações. Este tipo de *dashboards* facilita as estratégias mensais de revisão ou as sessões de planeamento operacional, ajudando assim os executivos a colaborarem na resolução de problemas ou a tirarem partido de oportunidades. As *dashboards* estratégicas usam na sua maioria indicadores provenientes de fontes de dados externas ou manuais via Excel. (Eckerson, 2009)

As *dashboards* táticas permitem às organizações monitorizar o progresso e tendências relacionadas para cada planeamento estratégico. São adequadas para tecnologias que possam fazer análise de dados de forma mais minuciosa através dos métodos *drill down* ou *slice and dice*. (Rasmussen et al., 2009). Os dados são resumidos e detalhados de forma diária ou semanal, proveniente de sistemas operacionais, permitindo aos gestores identificar problemas e elaborar soluções para que consigam alcançar os objetivos estabelecidos a curto e a longo prazo.

Uma vez que as *dashboards* táticas são maioritariamente populadas de forma manual ou automática com KPIs de resultado (*outcomes*) e de ação (*drivers*), este tipo de *dashboards* assemelha-se a um portal de BI onde os gestores podem modificar de acordo com as exigências de negócio e as preferências pessoais. (Eckerson, 2009)

Nas *dashboards* operacionais os funcionários podem monitorizar e controlar os processos centrais em tempo real, onde os dados são provenientes de sistemas operacionais e consistem em KPIs de ação (*drivers*) e métricas operacionais. Estas *dashboards* assemelham-se às dashboards encontradas nos automóveis (uso de manómetros), sendo que podem ser atualizadas de forma contínua sempre que haja algum evento. (Eckerson, 2009). Os funcionários deste tipo de dashboards têm um menor nível de responsabilidade (e.g.: vendas, apoio técnico, etc.), o que faz com que necessitem de acesso a informações com maior detalhe e de uma forte funcionalidade analítica para poderem analisar de forma minuciosa os dados que são apresentados (Rasmussen et al., 2009).

Eckerson (2009), divide os 3 tipos de *dashboards* usados pelas organizações como dashboards Estratégicas, Táticas e Operacionais e consoante diversos fatores enumerados na tabela 2.

Tabela 2: Tipos de Dashboards

Fonte: Adaptado de Eckerson (2009)

Tipos de Dashboards	Estratégicas	Táticas	Operacionais
Foco	Executar estratégias	Otimizar processos	Controlar operações
Uso	Gestão	Análise	Monitorização
Utilizadores	Executivos	Gerentes	Funcionários
Âmbito	Empresarial	Departamental	Operacional
Métricas	KPIs de resultados (<i>outcomes</i>)	KPIs de resultados (<i>outcomes</i>) e de ação (<i>drivers</i>)	KPIs de ação (<i>drivers</i>)
Dados	Resumidos	Detalhados/Resumidos	Detalhados
Fontes	Manuais/Externas	Manuais/Sistemas Centrais	Sistemas Centrais
Ciclo de Atualização	Mensal/Trimestral	Diário/Semanal	Em tempo real

a um..."

Assim sendo, uma implementação de *dashboards* efetiva numa organização poderá reduzir de forma dramática a necessidade de relatórios operacionais e financeiros, assim como irá contribuir para apoiar melhores tomadas de decisão e na ajuda para melhorar o rendimento (Rasmussen et al., 2009).

2.6 Análise e Seleção de Indicadores

De acordo com Kosala et al. (2017), os KPI (*Key Performance Indicators*), ou indicadores-chave de desempenho, são métricas importantes para as organizações no sentido em que ajudam na escolha da melhor decisão a tomar, no *benchmarking* (ou comparação de desempenho) e no prazo associado com atividades específicas para que as organizações possam alcançar os objetivos a longo prazo. Os indicadores acabam por assumir-se como um elemento gerador de consenso dentro da organização, já que constituem o veículo aceite por todos para a quantificação objetiva dos resultados realizados (Caldeira, 2012). O verdadeiro propósito dos KPIs é a identificação do que necessita de ser feito para melhorar o desempenho e manter a estratégia no rumo certo (Kerzner, 2011).

Para Mendes (2013) os KPIs permitem aos colaboradores compreenderem a importância do seu trabalho na organização, uma vez que este autor considera os KPI como uma medida de comunicação, resultando assim no sucesso ou insucesso de uma dada organização. Parmenter (2010) considera que os KPI representam “um conjunto de medidas com foco nos aspectos do desempenho organizacional que são os mais críticos para o sucesso atual e futuro da empresa”. Segundo Peterson (2006), os Key Performance Indicators nunca são apenas números, podendo representar proporções, taxas, médias ou porcentagens de um dado valor. Dessa forma os KPI resumem de forma efetiva a comparação de valores e transmitem o maior número de informação possível aos utilizadores. As organizações podem usar os KPIs para focar-se no que é verdadeiramente importante para o seu trabalho (Eckerson, 2009).

Na opinião de Doran (1981), idealmente, os objetivos definidos a partir dos KPIs pelas organizações e departamentos devem ser SMART (específicos, mensuráveis, atribuíveis,

realistas e temporais) e define as seguintes características deste acrónimo que é de sua autoria:

- S: *Specific* (Específico) – deve especificar uma área/processo/objetivo a ser melhorado.
- M: *Measurable* (Mensurável) – deve especificar quais os objetivos que poderão ser alcançados com os recursos disponíveis para o efeito.
- A: *Assignable* (Atribuível) – os objetivos deverão ser delegados a um ou vários elementos da equipa.
- R: *Realistic* (Realista) – metas que sejam realisticamente possíveis de atingir
- T: *Time-related* (Temporal) – deve especificar quando será possível obter os resultados.

A regra SMART foi originalmente desenvolvida para estabelecer objetivos concretos para os projetos e mais tarde adaptada para identificar métricas e KPIs (Kerzner, 2011) Ao serem apresentados os objetivos de KPIs e metas claras, os mesmos serão capazes de criar expectativas e de melhorar os processos de tomada de decisão nas organizações (Peterson, 2006). Os KPIs diferem de organização para organização conforme as prioridades de negócios estabelecidas. (Rouse, 2017)

Rouse (2017) salienta também o desafio na seleção de indicadores de modo a serem usados para o sucesso nas organizações. Para este autor haver demasiados KPIs definidos pode levar à atenção dos utilizadores não ficar focada nos indicadores que são mais importantes exibir. Como solução, Rouse propõe que os indicadores sejam separados por categorias, criando assim diversos grupos distintos (Rouse, 2017).

Caldeira (2012), define os seguintes atributos que devem ser representados nos indicadores de negócio:

- Nome do indicador – Descrição do indicador.
- Para que serve? - Aqui é descrito a utilidade do indicador selecionado para a empresa.
- Como se calcula? – Descrição da fórmula de cálculo utilizada para se obter o resultado assim como a unidade de medida (por exemplo: nº, dias, euros, %, etc.).

- Onde se vai buscar a informação? – A partir de que fonte se vai obter os dados a serem calculados para apurar os resultados necessários.
- Quando se deve apurar? – Frequência de uso para o indicador (semanal/mensal/anual)
- Qual a polaridade? – Descreve como é que o indicador será interpretado caso o resultado seja maior ou inferior que o esperado
- Notas adicionais – Possibilita indicar informação complementar que poderá ser necessária à análise de resultados.
- Visualização – Os resultados serão visualizados de acordo com o método de visualização escolhido, por exemplo, através de um gráfico dentro de uma *dashboard*.

As figuras abaixo demonstram um gráfico feito em Excel e a respetiva ficha de indicadores que resumem os atributos definidos por Jorge Caldeira.

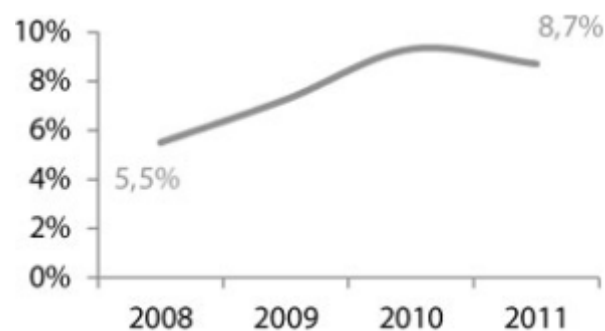


Figura 24: Exemplo de visualização de indicadores

Fonte: Caldeira (2012)

041 INDICADOR	041 % da execução física do projecto																																							
Para que serve?	Este indicador pretende identificar a % de concretização do projecto. Permite perceber se um projecto ainda está no início, no meio ou perto do fim. No caso de ter 100%, significa que está concluído. Para que seja apurado, é necessário que para todas as tarefas estejam identificadas as cargas de duração.																																							
Como se calcula?	Duração das tarefas já concluídas / Duração total das tarefas do projecto Unid.: %																																							
Onde se vai buscar a informação?	Área de projectos																																							
Quando se deve apurar?	Semanalmente/Mensalmente																																							
Qual a polaridade?	Positiva (Quanto maior o valor, melhor)																																							
Notas adicionais	Existir um atraso no início de um projecto não é tão grave como haver um atraso no fim de um projecto. No fim, significa que já se consumiu a maior parte dos recursos e é sempre mais difícil mudar as equipas, ou seja, existe sempre menor margem de manobra para tomar decisões com vista à correcção do atraso. Importa cruzar este indicador com a % de execução física prevista, de forma a identificar o desvio.																																							
Visualização	<p>Gráfico de barras e linhas que mostra a execução física (área sombreada) e a meta (linha tracejada) ao longo dos meses de um ano. A execução física atinge 57% em dezembro, enquanto a meta é de 45%.</p> <table border="1"> <thead> <tr> <th>Mês</th> <th>Execução física (%)</th> <th>Meta (%)</th> </tr> </thead> <tbody> <tr><td>J</td><td>5%</td><td>5%</td></tr> <tr><td>F</td><td>7%</td><td>7%</td></tr> <tr><td>M</td><td>12%</td><td>12%</td></tr> <tr><td>A</td><td>15%</td><td>15%</td></tr> <tr><td>M</td><td>35%</td><td>35%</td></tr> <tr><td>J</td><td>45%</td><td>45%</td></tr> <tr><td>J</td><td>50%</td><td>50%</td></tr> <tr><td>A</td><td>55%</td><td>55%</td></tr> <tr><td>S</td><td>60%</td><td>60%</td></tr> <tr><td>O</td><td>65%</td><td>65%</td></tr> <tr><td>N</td><td>70%</td><td>70%</td></tr> <tr><td>D</td><td>57%</td><td>45%</td></tr> </tbody> </table>	Mês	Execução física (%)	Meta (%)	J	5%	5%	F	7%	7%	M	12%	12%	A	15%	15%	M	35%	35%	J	45%	45%	J	50%	50%	A	55%	55%	S	60%	60%	O	65%	65%	N	70%	70%	D	57%	45%
Mês	Execução física (%)	Meta (%)																																						
J	5%	5%																																						
F	7%	7%																																						
M	12%	12%																																						
A	15%	15%																																						
M	35%	35%																																						
J	45%	45%																																						
J	50%	50%																																						
A	55%	55%																																						
S	60%	60%																																						
O	65%	65%																																						
N	70%	70%																																						
D	57%	45%																																						

Figura 25: Representação dos indicadores numa ficha

Fonte: Caldeira (2012)

No artigo acerca de indicadores de desempenho escrito por Domingues et al. (2020), estes autores concluem que o uso de indicadores chave de desempenho conferem às empresas uma grande utilidade para que possam usar como uma ferramenta para alcançarem os seus objetivos.

2.7 Metodologia

A metodologia escolhida foi a de Investigação-ação, um dos métodos qualitativos que é frequentemente aplicado na área de Sistemas de Informação e que se enquadra neste trabalho de projeto. Como método de recolha de dados usou-se a observação participante e análise documental. Segundo Abib et al. (2013) a observação participante é um tipo de investigação qualitativa, fundamentada em descobertas no campo, envolvendo a participação do investigador no quotidiano na área que está a investigar. Assim sendo, o investigador terá que personalizar a sua investigação consoante a área de investigação, havendo a necessidade do investigador ter que assumir um compromisso a longo prazo.

Segundo Tripp (2005, p.477) a Investigação – ação ou pesquisa - ação “é uma forma de investigação - ação que utiliza técnicas de pesquisa consagradas para informar a ação que se decide tomar para melhorar a prática”. A autoria deste tipo de metodologia científica de carácter qualitativo está normalmente associada ao psicólogo alemão Kurt Lewin (Lewin, 1951), tendo sido usada como base para diversas técnicas de investigação-ação.

A metodologia investigação-ação, na sua definição mais comum, apresenta-se como um processo cíclico de cinco fases Susman and Evered (1978) e Baskerville (1999):

- **Diagnóstico.**
- **Planeamento da ação.**
- **Execução das ação.**
- **Avaliação.**
- **Aprendizagem específica.**

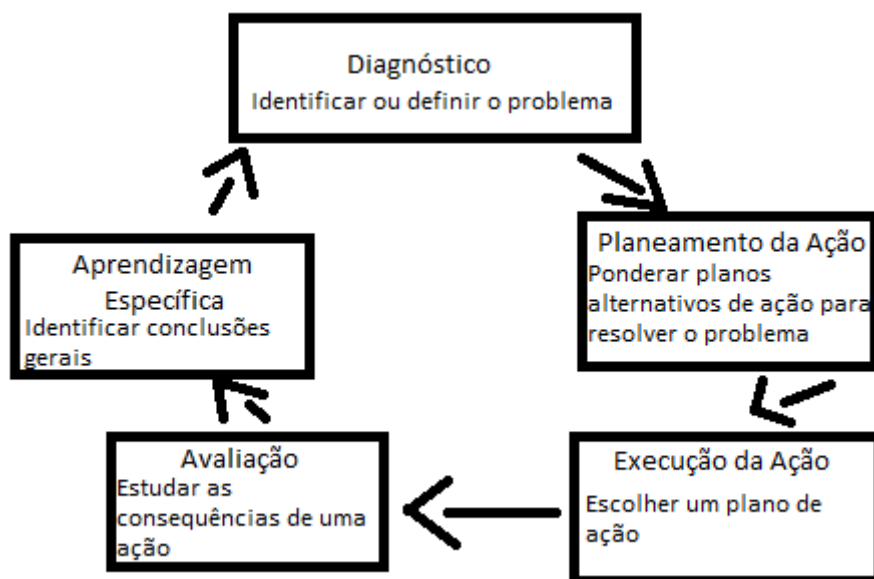


Figura 26: As 5 etapas do modelo Investigação-Ação

Fonte: Adaptado de Susman and Evered (1978)

Cada uma destas etapas é mais detalhada por Baskerville (1999):

Diagnóstico:

Na fase de diagnóstico, o investigador identifica os problemas principais das causas que levaram ao desejo de mudança de uma organização. O processo de diagnóstico envolve a autointerpretação de um complexo problema organizacional através de uma forma holística, onde certas suposições teóricas (hipótese de trabalho) serão desenvolvidas acerca da natureza da organização e o seu domínio do problema. (Baskerville, 1999).

Planeamento da Ação:

Segundo Baskerville (1999), na fase de planeamento de ação são delineadas as ações previamente definidas para resolver ou melhorar os problemas inicialmente encontrados. São assim guiadas pelo referencial teórico as ações planeadas que visam a mudança da organização, assim como as mudanças que permitirão chegar a esse fim.

Execução da Ação:

Baskerville (1999), explica que nesta fase será implementada a ação com base no planeamento definido previamente. De forma que a fase de planeamento de ação seja concluída, o investigador terá que colaborar ativamente com os profissionais, através da

intervenção na organização, de modo que as alterações possam ser feitas. Baskerville também refere que existem várias estratégias de intervenção que poderão ser adotadas, tais como: intervenção diretiva, a qual conduz a mudança, ou a intervenção não diretiva, na qual a mudança é pedida de forma indireta.

Avaliação:

Após as ações terem sido concluídas, os investigadores e profissionais irão efetuar a avaliação dos resultados (Baskerville,1999). A avaliação inclui determinar se os efeitos teóricos da ação foram cumpridos, e se estes efeitos atenuaram os problemas. Quando a ação é concluída com êxito, a avaliação deve questionar de forma crítica se a ação foi a única causa para a resolução do problema. Na eventualidade da mudança não ter surtido qualquer efeito, deverá ser criada uma nova estrutura para o próximo ciclo da investigação-ação.

Aprendizagem Específica:

Embora a atividade de aprendizagem específica seja formalmente o último passo, é normalmente um processo contínuo. O conhecimento adquirido na investigação-ação (independentemente de a ação ter sido bem-sucedida ou não), pode ser dirigida a três audiências (Baskerville, 1999):

1. Reestruturação de normas organizacionais para refletirem o novo conhecimento adquirido pela organização durante a investigação.
2. Quando a ação não é bem-sucedida, o conhecimento adicional poderá servir como base de diagnóstico e para preparação para intervenções futuras de investigação-ação.
3. O sucesso ou fracasso da estrutura teórica proporciona conhecimento importante para a comunidade científica de modo a lidar com futuras ações de investigação.

3.1 Diagnóstico

Na primeira fase da metodologia de investigação-ação, começou-se por avaliar a situação atual da empresa.

Atualmente, a empresa NDrive não possui uma solução de Business Intelligence para o volume de dados que recebe diariamente a partir do Google Analytics. Isto leva a:

- Processos de tomada de decisão morosos.
- Informação insuficiente para análise.
- Necessidade de redação de vários relatórios com os indicadores necessários para análise, causando assim uma dispersividade de informação e perda de tempo com esta etapa, uma vez que as opções de customização de relatórios do Google Analytics acabam por ser um fator limitativo.

3.2 Planeamento das Ações

Na fase de planeamento de ações da metodologia investigação-ação, procedeu-se à análise aos requisitos da solução, a arquitetura que seria usada, a análise e seleção de indicadores atuais, assim como o uso dos mesmos com os requisitos necessários para apresentação nas dashboards, terminando com os prazos delineados e o organograma.

3.2.1.1 Requisitos

De modo a poder-se proceder ao desenvolvimento da solução pretendida, foi feito um levantamento dos requisitos necessários através da análise aos relatórios existentes e analisar o que pode ser apresentado e melhorado, para que possa assim ser possível apurar o maior número de informação possível e condensar tudo numa única proposta. Assim sendo, os requisitos para a solução atual são os seguintes:

1. Os dados a serem analisados devem poder ser obtidos a nível histórico, últimos 30 dias, últimos 28 dias, últimos 7 dias, e últimas 24 horas.
2. Deve haver um dashboard que apresente a informação resumida e outros que permitam a exploração mais aprofundada dos dados em análise.
3. Navegação entre os dashboards deve ser fácil e intuitiva.
4. As dashboards devem usar aproximadamente a mesma paleta de cores usadas pela marca Karta GPS.
5. Uso de bookmarks (marcadores) para que seja possível apresentar diferentes visualizações no mesmo dashboard.

6. Implementar um cabeçalho com a data da última atualização e um *date slider* para escolha da data de início/data de fim dos dados apresentados.
7. Menu principal com escolha das dashboards.

3.2.1.2 Arquitetura

Nesta secção é definida a arquitetura escolhida para este projeto, a qual consiste no seguinte:

- Fonte de dados
- Processo ETL
- Data Warehouse
- Dashboards

Fonte de dados

Conforme indicado no capítulo 1, a fonte de dados usada para este projeto foi proveniente da plataforma Google Analytics, a qual regista diariamente todos os eventos necessários do produto que a NDrive deseja analisar (Karta GPS), e que serão visualizados nos dashboards desenvolvidos.

Processo ETL

Para o processo ETL dos dados recebidos do Google Analytics, foi escolhida a ferramenta de BI Power BI Desktop. Embora o Tableau consiga fazer operações ETL, ligação direta a fontes de dados como o Google Analytics através de conectores oficiais, assim como uma interface *drag and drop* fácil de usar, a escolha de usar o Power BI Desktop deveu-se às seguintes conclusões:

- O Power BI Desktop permite operações ETL graças ao uso nativo do Power Query;
- É possível usar o Power BI Desktop gratuitamente ou com um trial de 2 meses na versão PRO para novos utilizadores Office 365;
- A NDrive possui subscrições Office365 e licenças PRO disponíveis para o uso desta ferramenta para uso futuro com as funcionalidades acrescidas que a versão PRO proporciona.
- A experiência do autor prévia com esta ferramenta.

Data Warehouse

Foi escolhida a metodologia de Ralph Kimball de modelação dimensional, em formato *star schema*, para a modelação do Data Mart para o departamento de Marketing da NDrive.

Dashboards

Para criação das dashboards e exploração dos indicadores necessários, foi escolhida a ferramenta Power BI Desktop, tanto pela sua simplicidade de construção dos dashboards como pelo facto de esta ferramenta já estar a ser usada para as anteriores etapas.

3.2.1.3 Análise e Seleção dos Indicadores de Negócio para a Empresa

Para Peterson (2006), os *key performance indicators* visam transmitir o maior número de informações possível de forma concisa, sendo que os KPIs não podem ser considerados apenas números, mas sim interpretados como percentagens, valores ou médias. A eficácia dos KPIs será alcançada após serem devidamente definidos e apresentados concisamente. Petersen et al. (2009) entendem que existem muitas formas de avaliação dos resultados no Marketing Digital, sendo que a medição em si não é o desafio, mas sim na seleção de um conjunto eficiente de métricas verdadeiramente relevantes para o negócio de uma organização.

Usando como base o modelo de indicadores de Caldeira (2010), os seguintes indicadores foram recolhidos que são atualmente usados na conta Google Analytics da empresa para os produtos Karta GPS:

- Sessions
- Users
- New Users

Tabela 3: Indicador Sessions

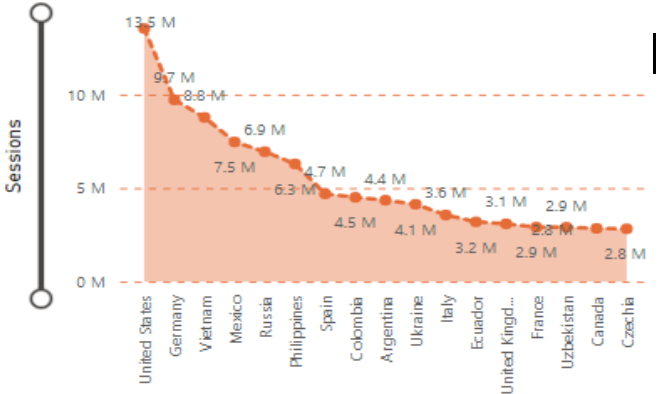
Sessions	
Para que serve?	Contabilizar o número de sessões por utilizador
Como se calcula?	Sempre que um utilizador iniciar uma sessão, o contador de sessões incrementa para esse utilizador específico
Onde é que se vai buscar a informação?	Google Analytics
Quando se deve apurar?	Diariamente
Qual a polaridade?	Positiva (Quanto maior for o número, melhor)
Notas adicionais	Uma <i>session</i> (sessão) representa uma visita do utilizador na app que ocorre num determinado período de tempo.
Visualização	

Tabela 4: Indicador Users

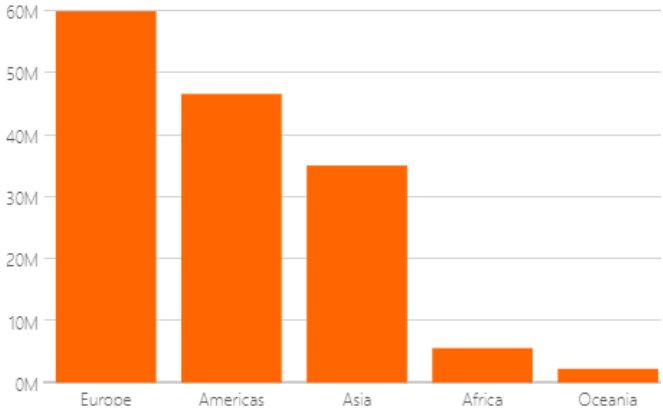
Users	
Para que serve?	Contabilizar o número de utilizadores na app
Como se calcula?	Sempre que um utilizador aceder à app, a contagem incrementa para esse utilizador específico.
Onde é que se vai buscar a informação?	Google Analytics
Quando se deve apurar?	Diariamente
Qual a polaridade?	Positiva (Quanto maior for o número, melhor)
Notas adicionais	Um user (utilizador) é um indivíduo que acedeu à app previamente, enquanto uma sessão representa uma dessas visitas na app.
Visualização	

Tabela 5: Indicador New Users

New Users																			
Para que serve?	Contabilizar o número de utilizadores que visita a app pela primeira vez																		
Como se calcula?	Sempre que um utilizador aceder à app pela primeira vez, o contador incrementa para esse utilizador específico																		
Onde é que se vai buscar a informação?	Google Analytics																		
Quando se deve apurar?	Diariamente																		
Qual a polaridade?	Positiva (Quanto maior for o número, melhor)																		
Notas adicionais	Um 'New User' (Novo Utilizador) é um visitante que nunca iniciou a app e está a iniciar a sua primeira sessão.																		
Visualização	<p>New Users by Continent</p> <table border="1"> <caption>New Users by Continent Data</caption> <thead> <tr> <th>Continent</th> <th>Users (M)</th> <th>Percentage (%)</th> </tr> </thead> <tbody> <tr> <td>Europe</td> <td>9 M</td> <td>46.03%</td> </tr> <tr> <td>Asia</td> <td>5 M</td> <td>22.7%</td> </tr> <tr> <td>America</td> <td>5 M</td> <td>25.22%</td> </tr> <tr> <td>Africa</td> <td>0 M</td> <td>0.11%</td> </tr> <tr> <td>(not set)</td> <td>0 M</td> <td>0.11%</td> </tr> </tbody> </table>	Continent	Users (M)	Percentage (%)	Europe	9 M	46.03%	Asia	5 M	22.7%	America	5 M	25.22%	Africa	0 M	0.11%	(not set)	0 M	0.11%
Continent	Users (M)	Percentage (%)																	
Europe	9 M	46.03%																	
Asia	5 M	22.7%																	
America	5 M	25.22%																	
Africa	0 M	0.11%																	
(not set)	0 M	0.11%																	

Para complementar a informação destes indicadores, os dashboards desenvolvidos irão apresentar os seguintes filtros:

- Quantidade de *Users* registados na aplicação Karta GPS a nível:
 - Histórico
 - Diário
 - 7 dias
 - 15 dias
 - 28 dias

- Quantidade de *New Users* registados na aplicação Karta GPS a nível:
 - Histórico
 - Diário
 - 7 dias
 - 15 dias
 - 28 dias

- Quantidade de *Sessions* registadas na aplicação Karta GPS a nível:
 - Histórico
 - Diário
 - 7 dias
 - 15 dias
 - 28 dias

- Cada indicador será filtrado a nível:
 - Geográfico
 - Por tipo de equipamento
 - Por sistema operativo
 - Por versão de sistema operativo

3.2.1.4 Cronograma

De modo a poder haver uma gestão correta das etapas necessárias ao desenvolvimento do projeto, elaborou-se um cronograma de atividades que indica as etapas que foram feitas durante o período de janeiro a agosto de 2022 (tabela 6).

Tabela 6: Cronograma de atividades

Etapa	Janeiro	Fevereiro	Março	Abril	Maior	Junho	Julho	Agosto
Revisão Literatura								
Análise de Indicadores								
Requisitos								
Definição da Arquitetura								
Modelo star schema								
Processo ETL								
Criação dos Dashboards								
Apresentação e dos dashboards								
Validação Resultados								

4.1 Execução das Ações

Nesta secção da metodologia investigação-ação é feita a descrição das etapas de construção do *Data Warehouse* para a empresa, tendo em consideração os KPIs selecionados no capítulo anterior. Começa pela aplicação das 4 etapas de desenho de um DW/DM segundo Kimball (2011) e terminando no modelo multidimensional escolhido para o *DW*, assim como as relações criadas no modelo.

4.1.1 Processo de Negócio

O processo de negócio escolhido para análise foi o de estatísticas de uso da aplicação Karta GPS a nível mundial, tendo como objetivo a apresentação das métricas relevantes ao nível de sessões, utilizadores, e novos utilizadores. O objetivo principal será providenciar ao departamento de Marketing da empresa uma visão mais clara do uso da aplicação para auxiliar nos processos de tomada de decisão.

4.1.2 Granularidade dos Dados

De modo que se possa providenciar dados com bastante qualidade, é necessário determinar a granularidade necessária. Para esta implementação foi escolhida uma granularidade diária. Embora este tipo de detalhe represente um maior volume de dados na Tabela de Factos do *Data Warehouse*, o mesmo não representa qualquer problema para esta implementação.

4.1.3 Tabelas de Dimensão

A partir das decisões tomadas durante a análise dos indicadores, foram escolhidas as seguintes dimensões:

- **DimTempo:** A DimTime contém o registo de datas que vai iniciar em 2016 e vai incrementando consoante a data (tabela 7).

Tabela 7: DimTempo

Nome Tabela		
DimTempo		
Nome Coluna	Descrição	Tipo Dados
IDData	Coluna de índice - Chave primária	PK
Ano	Ano (numérico)	int64
Dia	Dia (numérico)	int64
Dia Semana	Dia da Semana (extenso)	string
Dia Semana	Dia da Semana (numérico)	Int64
Mês	Mês (numérico)	int64
MêsNome	Nome do Mês (extenso)	string
Trimestre	Número do trimestre	int64
TrimestreNome	Nome do Trimestre	string
Trimestre Ano	Trimestre (T1 2022)	string
Semestre	Semestre (numérico S1)	Int64
Semestre	Semestre (S1 2022)	string

- **DimCountry:** A DimCountry contém o registo de países de cada utilizador das apps (tabela 8).

Tabela 8: DimCountry

Nome Tabela		
DimCountry		
Nome Coluna	Descrição	Tipo Dados
CountryID	Coluna de índice - Chave primária	PK
Country	Nome do país	string
Country ISO Code	Nome do país (abreviado)	string

- **DimDeviceCategory:** A DimDeviceCategory (tabela 9) representa o tipo de dispositivo que os utilizadores usam para aceder à app (Mobile ou Tablet).

Tabela 9: DimDeviceCategory

Nome Tabela		
DimDeviceCategory		
Nome Coluna	Descrição	Tipo Dados
DeviceCategoryID	Coluna de índice - Chave primária	PK
Device Category	Categoria do dispositivo	String

- **DimDeviceModel:** A DimDeviceModel (tabela 10) regista os dispositivos móveis usados pelos utilizadores. Para identificar os dispositivos é usado o nome da marca do equipamento em conjunto com a referência interna do equipamento

(e.g.: Samsung Galaxy 404SC Galaxy 6). Esta combinação permite obter uma granularidade mais específica.

Tabela 10: DimDeviceModel

Nome Tabela		
DimDeviceModel		
Nome Coluna	Descrição	Tipo Dados
DeviceModelID	Coluna de índice - Chave primária	PK
Mobile Device Info	Informação do dispositivo	string

- **DimOperatingSystem:** A DimOperatingSystem (tabela 11) contém os dados relativos ao tipo de sistema operativo do dispositivo usado pelo utilizador.

Tabela 11: DimOperatingSystem

Nome Tabela		
DimOperatingSystem		
Nome Coluna	Descrição	Tipo Dados
OperatingSystemID	Coluna de índice - Chave primária	PK
Operating System	Nome do sistema operativo	String

- **DimOSVersion:** A DimOSVersion (tabela 12) regista a versão de sistema operativo dos dispositivos que acedem às apps Karta GPS.

Tabela 12: DimOSVersion

Nome Tabela		
DimOSVersion		
Nome Coluna	Descrição	Tipo Dados
OSVersionID	Coluna de índice - Chave primária	PK
Operating System Version	Versão do sistema operativo	int

- **DimTime:** A DimTime (tabela 13) permite associar atributos das tabelas de dimensão e de facto ao tempo. Esta dimensão é essencial para podermos obter análises de dados ao longo do tempo (seja a nível diário, semanal, mensal, trimestral ou anual).

Tabela 13: DimTime

Nome Tabela		
DimTime		
Nome Coluna	Descrição	Tipo Dados
IDData	Coluna de índice - Chave primária	PK
Day	Dia	int
DiaSemana	Nome do dia da semana (extenso)	text
MesNome	Nome do mês (extenso)	text
Trimestre	Trimestre	int
TrimestreNome	Nome do Trimestre	text

4.1.4 Tabelas de Facto

De acordo com as decisões anteriormente tomadas, foi escolhida a seguinte tabela de factos (tabela 14), a qual armazena os indicadores relevantes e métricas de acesso à app para o departamento de Marketing da NDrive.

Tabela 14: FactAppAnalytics

Nome Tabela		
FactAppAnalytics		
Nome Coluna	Descrição	Tipo Dados
IDData	Coluna de índice - Chave estrangeira	FK
DimContinent.ContinentID	Coluna de índice - Chave estrangeira	FK
DimCountry.CountryID	Coluna de índice - Chave estrangeira	FK
DimDeviceCategory.DeviceCategoryID	Coluna de índice - Chave estrangeira	FK
DimDeviceModel.DeviceModelID	Coluna de índice - Chave estrangeira	FK
DimOperatingSystem.OperatingSystem ID	Coluna de índice - Chave estrangeira	FK
DimOSVersion.OSVersionID	Coluna de índice - Chave estrangeira	FK
Operating System Version	Coluna de índice - Chave estrangeira	FK
Sessions	Número de sessões	int64
New Users	Número de novos utilizadores	int64

Number of Sessions per User	Número de sessões por utilizador	int64
Users	Número de Utilizadores	int64

4.2 Fase ETL (Extract-Transform-Load)

Após a escolha dos dados referidos no capítulo anterior, procedeu-se à fase ETL onde todos os dados previamente escolhidos foram extraídos, transformados e carregados para o *Data Warehouse*. Nesta fase deu-se uso ao recurso *Power Query* do Power BI, que é uma ferramenta ETL de preparação e transformação de dados e que permite conectar-se a múltiplas fontes de dados fontes de dados, independentemente da sua origem.

De salientar que, devido a questões de privacidade de dados, todos os dados aqui exibidos não correspondem aos dados reais da empresa e são meramente usados para fins exemplificativos da fase ETL.

4.2.1 Extract

Na primeira fase do processo ETL, procedeu-se à ligação ao serviço Google Analytics, usando o conector oficial do Power BI. Para tal, usou-se a opção “*Obter Dados*” do Power BI (figura 26).

Seguidamente fez-se a extração dos dados através de uma ligação direta ao serviço Google Analytics, usando a integração nativa do Power BI com esta plataforma, conforme ilustrado na figura 27.

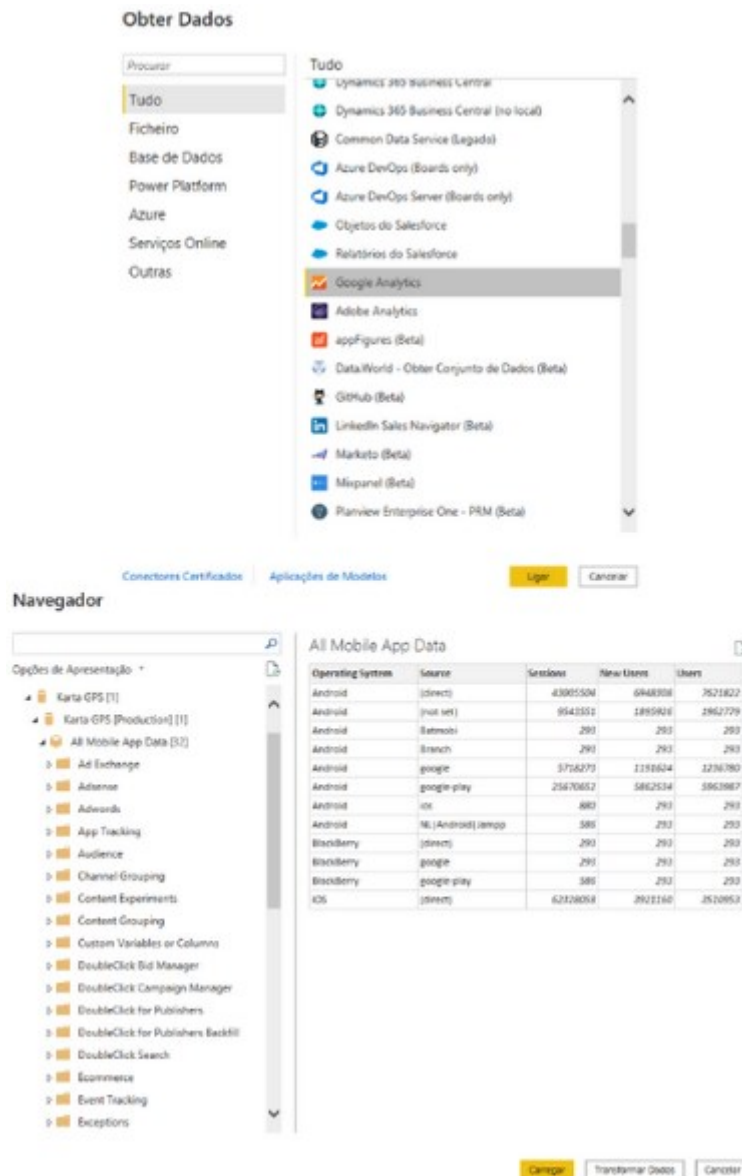


Figura 27: Funcionalidade “Obter Dados” e conteúdo do Google Analytics

Ao ter sido feita a ligação ao *Google Analytics*, procedeu-se à escolha das métricas necessárias para posterior transformação na etapa seguinte, tendo sido possível agrupar os dados necessários numa única tabela “*All Mobile App Data*”, para cumprir os seguintes objetivos:

- Não haver mais que uma ligação feita no Analytics para respeitar os limites da API de 7 dimensões e 10 métricas e evitar duplicação desnecessária de dados.
- Facilitar o processo de transformação dos dados.
- Reduzir o tempo de atualização das tabelas.

4.2.2 Transform

Na segunda fase é necessário efetuar a transformação dos dados (*transform*). A transformação refere-se à limpeza e agregação a ser feita nos dados para que fiquem corretamente preparados para análise. Para tal, deu-se uso à funcionalidade *Power Query*, onde foi possível efetuar todas as transformações necessárias para a análise dos dados. A figura 28 demonstra algumas das funcionalidades que são possíveis de realizar na ferramenta Power BI:

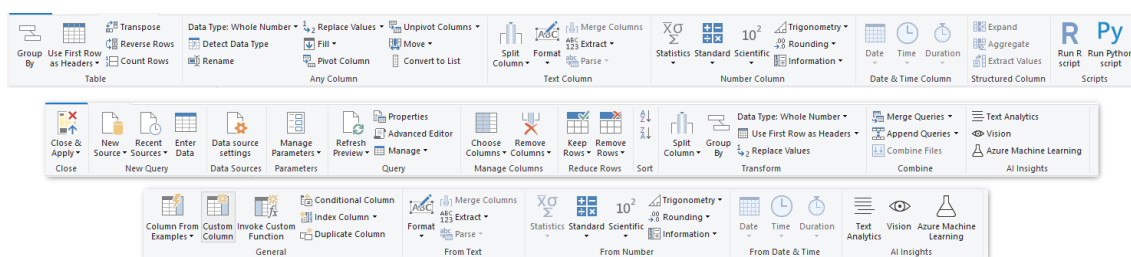


Figura 28: Funcionalidades do Power Query no Power BI

A primeira transformação necessária foi a criação da tabela DimTempo como um calendário utilizando o editor avançado do Power Query (Apêndice A). Após o calendário ter sido criado foi definido o campo IdData como número inteiro, uma vez que originalmente estava a ser considerado como texto pelo Power BI, conforme a figura 29 abaixo que demonstra os passos feitos no Power Query através da linguagem de fórmula M, que o Power BI usa durante a fase ETL.

```
let
    Origem = Excel.Workbook(File.Contents("C:\Users\hugov\OneDrive\Documentos\Dimensao tempo excel 2012 a 2031.xlsx"), null, true),
    Data2012a2031_Table = Origem{[Item="Data2012a2031",Kind="Table"]}[Data],
    #"IDData Tipo de dados = Int" = Table.TransformColumnTypes(Data2012a2031_Table,{{"IDData", Int64.Type}, {"Data", type date}, {"Ano", Int64.Type}, {"Mês", Int64.Type}, {"Day", Int64.Type}, {"MesNome", type text}, {"Trimestre", Int64.Type}, {"TrimestreNome", type text}, {"DiaSemana", type text}})
in
    #"IDData Tipo de dados = Int"
```

Figura 29: Passos feitos no Power Query para a DimTempo em linguagem M

Para os dados provenientes do *Google Analytics*, a importação dos dados foi feita uma única vez, ficando associados a uma tabela designada de “*All Mobile Data*” que continha todos os dados necessários. Esta tabela serviu de referência para a criação das tabelas de dimensão e de facto e só foi usada para as fases de extração e transformação, não tendo sido carregada para o modelo. O Power BI permite usar uma tabela importada como

referência para as restantes através da opção “Referência” do *Power Query*, que foi usado para este caso (Figura 30).

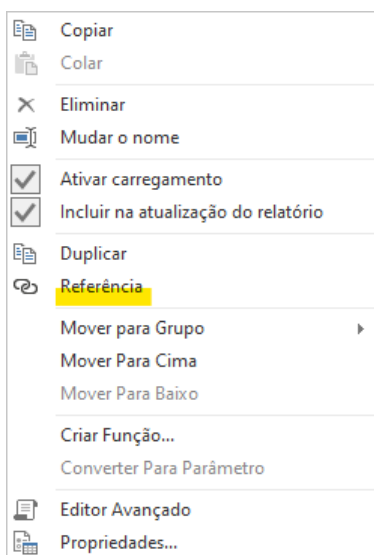


Figura 30: Funcionalidade “Referência” para referência de tabelas no Power BI.

Após referenciar-se uma tabela, a mesma é copiada sem os passos efetuados na mesma, enquanto se fosse usada a opção “Duplicar” do Power BI, seria feita uma cópia integral da tabela, com todos os passos efetuados.

Tendo os dados sido importados com sucesso, procedeu-se à criação das tabelas de dimensão. A primeira tabela criada foi a *DimCountry*, que contém a informação geográfica dos utilizadores, e na qual foram realizadas as seguintes etapas de transformação e limpeza dos dados feitas via *Power Query* (recorrendo ao Editor Avançado, que permite introduzir código M diretamente e desta forma ter um melhor controlo nas transformações realizadas internamente):

Tabela referenciada a partir da tabela original “*All Mobile Data*” e atribuição do nome *DimCountry*.

"Consulta Acrescentada" = Table.Combine({#"Tipo Alterado", AllMobileAppData})

Remoção de colunas irrelevantes para a dimensão *Country*, ficando só as colunas necessárias “*Country*” e “*Country ISO code*” na tabela.

"Colunas Removidas" = Table.RemoveColumns(#"Consulta Acrescentada",{"Coluna1", "Continent", "Device Category", "Mobile Device Info",

"Operating System", "Operating System Version", "Date", "Sessions", "New Users", "Number of Sessions per User", "Users"}})

Os nomes em duplicado dos países foram removidos para que só ficassem valores únicos na tabela.

"Duplicados Removidos" = Table.Distinct(#"Colunas Removidas", {"Country ISO Code"})

Foi criada uma coluna de índice para ser usada como FK da dimensão Country com tipo Int64 e a começar do valor 1.

Índice Adicionado" = Table.AddIndexColumn(#"Duplicados Removidos", "Índice", 1, 1, Int64.Type)

Criado o índice procedeu-se à reordenação das colunas existentes na tabela de dimensão.

Colunas Reordenadas" = Table.ReorderColumns(#"Índice Adicionado", {"Índice", "Country", "Country ISO Code"})

Uma vez que o objetivo seria cada índice de cada tabela de dimensão ter um nome identificável para garantir a correta identificação e associação na tabela de facto, o nome foi mudado para "CountryID". Todos os índices criados nas outras tabelas de dimensão usaram a palavra ID para referenciar o índice.

#"Nome do Índice Alterado" = Table.RenameColumns(#"Colunas Reordenadas", {"Índice", "CountryID"})

Os mesmos passos foram feitos nas restantes tabelas de dimensão *DimContinent*, *DimDeviceCategory*, *DimDeviceModel*, *DimOSVersion*, *DimOperatingSystem* e *DimTime*, de modo a terem só os dados destinados a cada uma destas tabelas de dimensão. Para a tabela de factos *FactAppAnalysis* a transformação teve as seguintes etapas para além do que foi anteriormente feito nas tabelas de dimensão:

A tabela "All Mobile Data" foi usada como a referência para a tabela de factos criada com o nome "FactAppAnalytics", a qual contém os valores numéricos necessários e as chaves de ligação às respectivas Tabelas de Dimensão (figura 31).

```

1 let
2     Origem = GoogleAnalytics.Accounts(),
3     #"██████████" = Origem[{Id="██████████"}][Data],
4     #"UA-██████████" = #"UA-██████████"({Id="UA-██████████"})[Data],
5     #"██████████" = #"UA-██████████"({Id="██████████"})[Data],
6     #"Itens Adicionados" = Cube.Transform(#"██████████",
7     {
8         {Cube.AddAndExpandDimensionColumn, "ga:continent", {"ga:continent"}, {"Continent"}},
9         {Cube.AddAndExpandDimensionColumn, "ga:country", {"ga:country"}, {"Country"}},
10        {Cube.AddAndExpandDimensionColumn, "ga:countryIsoCode", {"ga:countryIsoCode"}, {"Country ISO Code"}},
11        {Cube.AddAndExpandDimensionColumn, "ga:deviceCategory", {"ga:deviceCategory"}, {"Device Category"}},
12        {Cube.AddAndExpandDimensionColumn, "ga:mobileDeviceInfo", {"ga:mobileDeviceInfo"}, {"Mobile Device Info"}},
13        {Cube.AddAndExpandDimensionColumn, "ga:operatingSystem", {"ga:operatingSystem"}, {"Operating System"}},
14        {Cube.AddAndExpandDimensionColumn, "ga:operatingSystemVersion", {"ga:operatingSystemVersion"}, {"Operating System
15        Version"}},
16        {Cube.AddAndExpandDimensionColumn, "ga:date", {"ga:date"}, {"Date"}},
17        {Cube.AddMeasureColumn, "Sessions", "ga:sessions"},
18        {Cube.AddMeasureColumn, "New Users", "ga:newUsers"},
19        {Cube.AddMeasureColumn, "Number of Sessions per User", "ga:sessionsPerUser"},
20        {Cube.AddMeasureColumn, "Users", "ga:users"}
    }
  ),

```

Figura 31: Passos feitos no Power Query para a FactAppAnalytics em linguagem M

Para que fosse possível associar as tabelas de dimensão com as tabelas de facto, intercalou-se cada tabela de dimensão com a tabela de facto *FactAppAnalytics* através das respectivas *Primary Keys* (PK) de cada tabela de dimensão. Este procedimento fez com que as PK ficassem na tabela de facto e se tornassem *Foreign Keys* (FK) ou chaves estrangeiras.

A nível exemplificativo, a tabela de dimensão “*DimContinent*” foi intercalada com a tabela de facto usando o seguinte código no *Power Query*:

"Consultas Intercaladas" = Table.NestedJoin(#"Itens Adicionados", {"Continent"}, DimContinent, {"Continent"}, "DimContinent", JoinKind.LeftOuter)

Alternativamente, também é possível usar a opção “Intercalar Consultas” no meu visual do Power Query do Power BI.

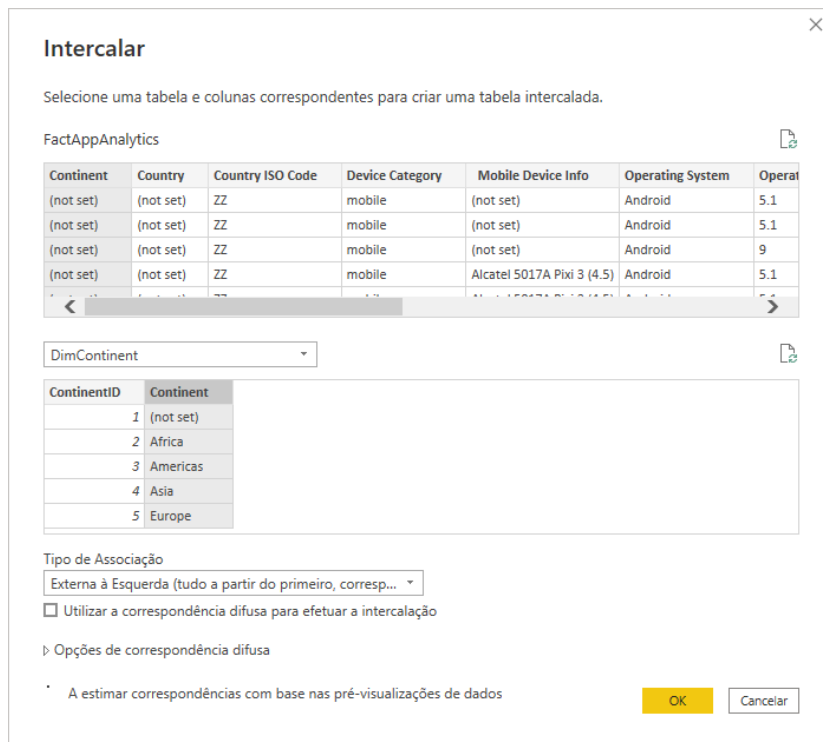


Figura 32: Função Intercalar consultas no Power BI

A intercalação foi feita através da coluna “Continent” da tabela de dimensão “DimContinent” para a tabela de facto “FactAppAnalytics”. Desta forma foi possível obter o índice criado na tabela “DimContinent”. Uma vez que havia uma correspondência exata dos dados, todas as colunas da dimensão foram importadas para a tabela de facto com sucesso.

Seguidamente, removeu-se a coluna “Continent” de modo a só ficar a coluna de índice, tornando-se a chave estrangeira (FK) da “DimContinent”.

"Colunas Removidas" = Table.RemoveColumns("#DimContinent Expandida",{"Continent"})

Estes passos foram repetidos para as restantes tabelas de dimensão *DimCountry*, *DimDeviceCategory*, *DimDeviceModel*, *DimOSVersion*, *DimOperatingSystem* e *DimTime*, tendo no final sido feita a reordenação das colunas na tabela de facto.

"Colunas Reordenadas" = Table.ReorderColumns("#Data2012a2031 Expandida",{"IDData", "DimContinent.ContinentID", "DimCountry.CountryID", "DimDeviceCategory.DeviceCategoryID", "DimDeviceModel.DeviceModelID",

"DimOperatingSystem.OperatingSystemID", "DimOSVersion.OSVersionID", "Operating System Version", "Date", "Sessions", "New Users", "Number of Sessions per User", "Users"}}

Após estas etapas de transformação e limpeza, a tabela de facto obteve as métricas do Google Analytics e as *Foreign Keys* de cada tabela dimensão, sendo agora possível fazer a criação do modelo multidimensional.

4.2.3 Load

Terminada a etapa de transformação e limpeza dos dados importados do processo ETL, é feito o carregamento (*Load*) dos dados para o *Data Mart* no Power BI. Uma vez que a importação dos dados do *Analytics* via API para o modelo multidimensional foi feita ao nível histórico, o tempo de carregamento inicial dos dados para o modelo foi de cerca de 35 minutos (figura 33), sendo que atualizações posteriores ao nível diário dos dados demoram entre 4 e 5 minutos.

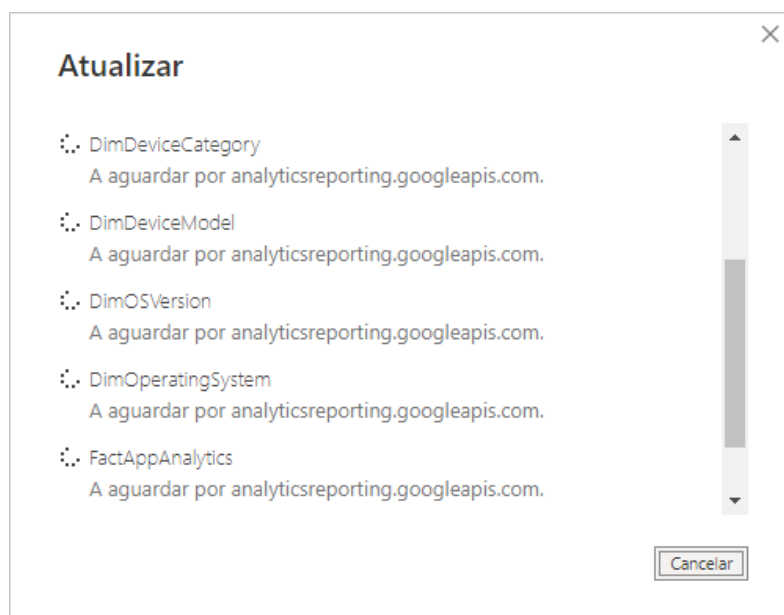


Figura 33: Carregamento dos dados para o modelo multidimensional

4.2.4 Modelação Multidimensional

A modelação multidimensional é uma perspetiva *bottom-up* (de baixo para cima) para a construção do Data Mart de modo que seja possível fazer análises de dados nas mesmas e possibilitar uma interpretação mais intuitiva do cubo OLAP. Terminado o processo

ETL, foi necessário definir as relações entre a tabela de facto e tabela de dimensão, onde foi escolhido o modelo *star schema* (esquema em estrela), para o Data Mart de Analytics.

Para esta implementação foi escolhido o uso do modelo *star schema* (esquema em estrela), onde as tabelas de dimensão estão interligadas com a tabela de facto situada no centro (figura 34).

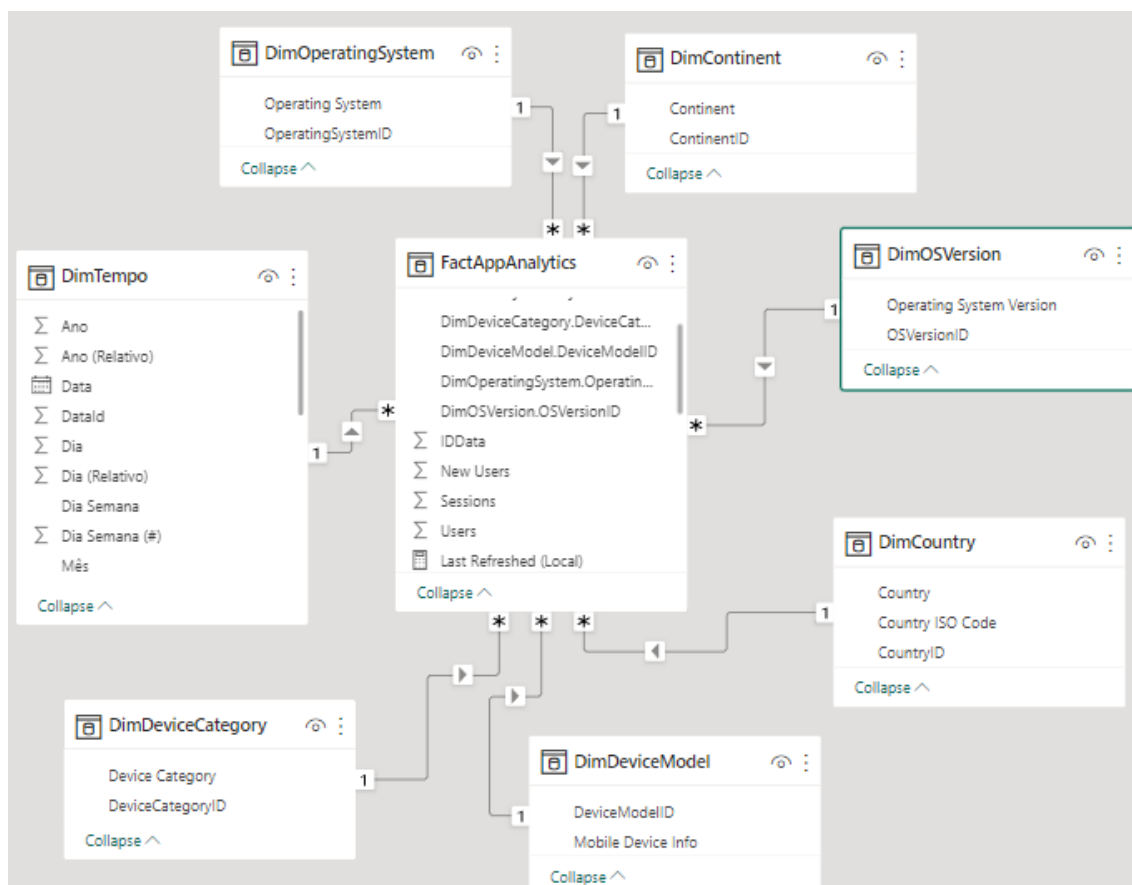


Figura 34: Modelo de dados multidimensional proposto em formato star schema

Conforme o modelo de dados multidimensional ilustrado na figura 34, existe uma tabela de factos principal que está interligada com sete tabelas de dimensão, cada tabela de dimensão possui uma chave primária e os respetivos atributos. A tabela de facto agrega as *Foreign Keys* (FK) que foram criadas durante o processo ETL para as respetivas dimensões de todas as tabelas de dimensão interligadas com a mesma. As tabelas de dimensão estão relacionadas com a tabela de facto numa relação de um para muitos, conforme referenciado nas figuras 34 e 35.

Manage relationships

Active	From: Table (Column)	To: Table (Column)
<input checked="" type="checkbox"/>	FactAppAnalytics (Date)	DimTempo (Data)
<input checked="" type="checkbox"/>	FactAppAnalytics (DimContinent.ContinentID)	DimContinent (ContinentID)
<input checked="" type="checkbox"/>	FactAppAnalytics (DimCountry.CountryID)	DimCountry (CountryID)
<input checked="" type="checkbox"/>	FactAppAnalytics (DimDeviceCategory.DeviceCategoryID)	DimDeviceCategory (DeviceCategoryID)
<input checked="" type="checkbox"/>	FactAppAnalytics (DimDeviceModel.DeviceModelID)	DimDeviceModel (DeviceModelID)
<input checked="" type="checkbox"/>	FactAppAnalytics (DimOperatingSystem.OperatingSystemID)	DimOperatingSystem (OperatingSystemID)
<input checked="" type="checkbox"/>	FactAppAnalytics (DimOSVersion.OSVersionID)	DimOSVersion (OSVersionID)

New... Autodetect... Edit... Delete

Close

Figura 35: Visão geral das relações entre tabelas no modelo de dados

4.3 Dashboards

Finalizado o modelo de dados, o Data Mart pode agora ser explorado. Para que tal seja possível, deu-se início à última fase da execução de ações, que foi a do desenho dos *dashboards*, os quais incluem diversos tipos de visualização para apoio nos processos de tomada de decisão no departamento de Marketing da empresa. Os *dashboards* foram então criados e propostos à empresa, onde serão sempre alvo de feedback contínuo por parte da empresa e ajustados em conformidade.

Todos os *dashboards* aqui apresentados tiveram em conta o uso das cores que identificam a marca Karta GPS, havendo em alguns casos uma ligeira alteração das cores de modo a gerar maior atenção para uma determinada visualização.

Para as cores, usou-se principalmente a cor da marca (laranja) no formato RGB #FC6A52, tendo também usadas as cores preto, cinza, azul e verde para algumas visualizações. Quanto aos marcadores, foram criados 6 marcadores distintos (figura 36), cada um com a sua própria visualização e que serão detalhados na secção seguinte.

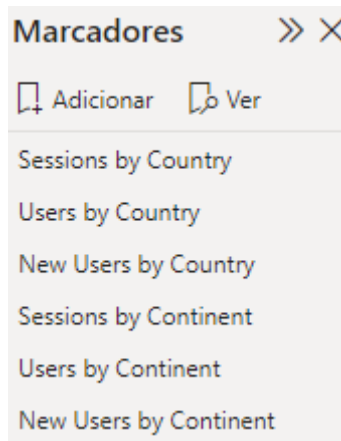


Figura 36: Marcadores criados para as dashboards.

De seguida procedeu-se à criação do cabeçalho para as *dashboards* com o *date slider* para escolha das datas de visualização e data de última atualização dos dados.

Para a informação de quando é que o modelo de dados foi atualizado pela última vez, efetuaram-se os seguintes passos:

1. Criação de uma nova dimensão chamada de *Last_Refresh_Local*, formatada como Data/Hora e com um calendário com o nome *LastRefresh*. Esta dimensão não necessita de ter qualquer relação com as restantes dimensões ou tabela de facto e foi criada com o seguinte código no Power Query:

```
= #table(type table[LastRefresh=datetime], {{DateTime.LocalNow()}})
```

2. De seguida procedeu-se à criação de uma medida em DAX com o seguinte código, ficando alojada na tabela de facto:

```
Last Refreshed (Local) =
FORMAT(LASTDATE>Last_Refresh_Local[LastRefresh]),"mmm dd, yyyy hh:mm:ss
AM/PM")
```

3. Escolheu-se a visualização “*Cartão*” no Power BI, associando o Card à dimensão *Last_Refresh_Local*

Para o *date slider* escolheu-se a visualização “*data segmentation*” no Power BI, ficando associado à tabela *DimTime* para apresentar a data de início / fim no cabeçalho.

Terminada esta etapa, o cabeçalho foi aplicado às *dashboards* (figura 37).

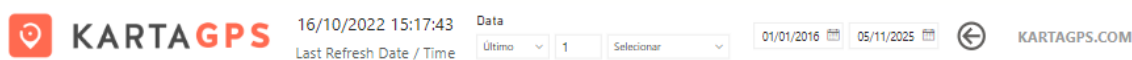


Figura 37: Cabeçalho final criado para os dashboards.

4.3.1 Apresentação dos Dashboards criados

De modo a poder navegar-se entre os *dashboards* criados na ferramenta Power BI, foi criado um menu principal com ligações mediante o uso da funcionalidade “marcadores” do Power BI (figura 38).

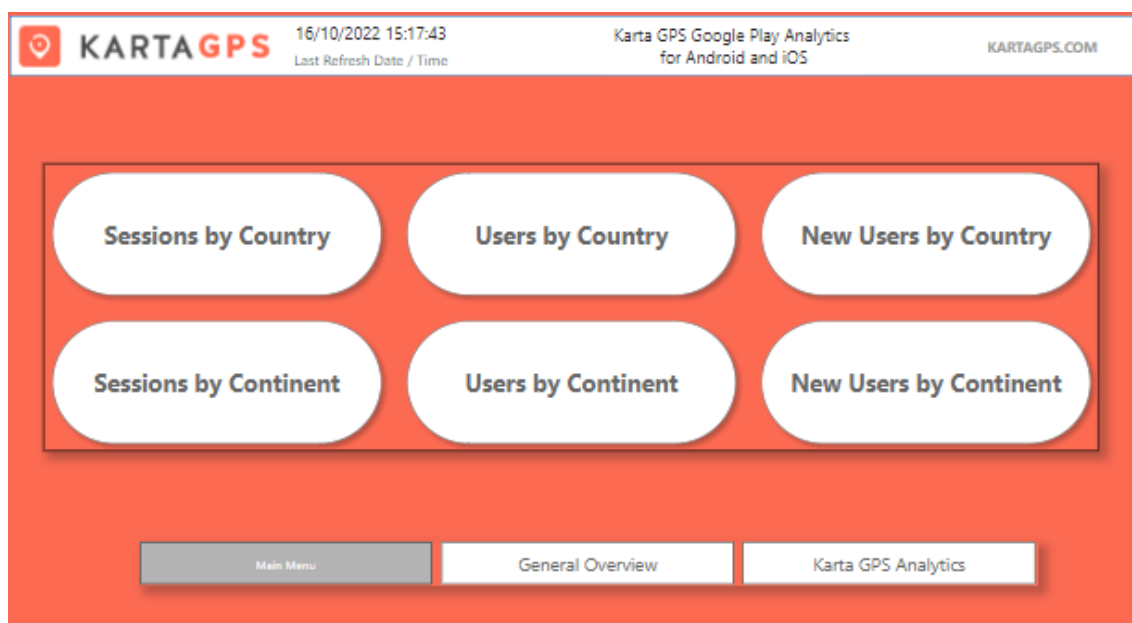


Figura 38: Menu Principal.

A partir deste menu criado é possível aceder às diferentes visualizações de *Sessions*, *Users* e *New Users*, assim como para uma *dashboard* com resumo de alguns indicadores principais e outra que contém todos os restantes indicadores. No menu principal é possível ainda ver a última data em que a *dashboard* foi atualizada.

4.3.1.1 General Overview

Este *dashboard* (figura 39) possui várias medidas que foram calculadas através de fórmulas DAX no Power BI para resumir de uma forma breve todos os indicadores presentes (tabela 15), tendo sido associadas a uma visualização de cartão. Adicionalmente, foram colocadas duas visualizações que referem o número total de

Sessions por Operating System, e outra visualização que contém o número total de Users, Sessions e New Users. Ambas as visualizações usam o formato de gráfico circular.

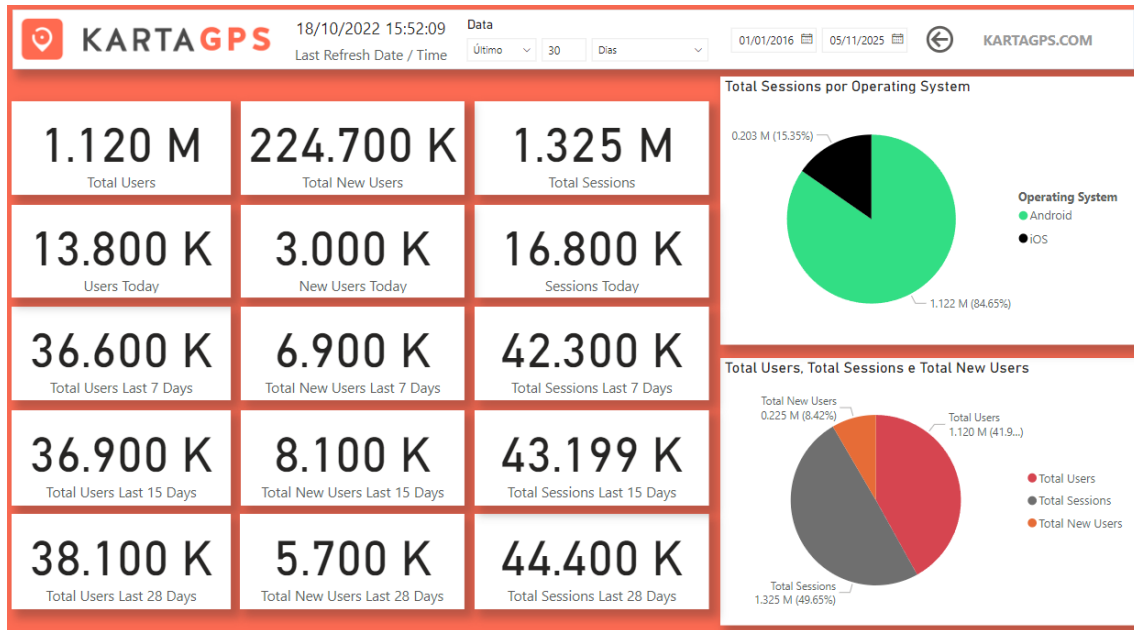


Figura 39: Dashboard: Secção General Overview.

Tabela 15: Fórmulas DAX da dashboard General Overview

Nome da Visualização	Fórmula DAX
Total Users	<code>SUM(FactAppAnalytics[Users])</code>
Users Today	<code>CALCULATE([Total Users], LASTDATE(DimTime[Data]))</code>
Total Users Last 7 Days	<code>CALCULATE([Total Users], DATEADD(LASTDATE(DimTime[Data]), -7, DAY))</code>
Total Users Last 15 Days	<code>CALCULATE([Total Users], DATEADD(LASTDATE(DimTime[Data]), -15, DAY))</code>
Total Users Last 28 Days	<code>CALCULATE([Total Users], DATEADD(LASTDATE(DimTime[Data]), -28, DAY))</code>
Total New Users	<code>SUM(FactAppAnalytics[New Users])</code>
New Users Today	<code>CALCULATE([Total New Users], LASTDATE(DimTime[Data]))</code>
Total New Users Last 7 Days	<code>CALCULATE([Total New Users], DATEADD(LASTDATE(DimTime[Data]), -7, DAY))</code>
Total New Users Last 15 Days	<code>CALCULATE([Total New Users], DATEADD(LASTDATE(DimTime[Data]), -15, DAY))</code>
Total New Users Last 28 Days	<code>CALCULATE([Total New Users], DATEADD(LASTDATE(DimTime[Data]), -28, DAY))</code>
Total Sessions	<code>SUM(FactAppAnalytics[Sessions])</code>
Sessions Today	<code>CALCULATE([Total Sessions], LASTDATE(DimTime[Data]))</code>
Total Sessions Last 7 Days	<code>CALCULATE([Total Sessions], DATEADD(LASTDATE(DimTime[Data]), -7, DAY))</code>
Total Sessions Last 15 Days	<code>CALCULATE([Total Sessions], DATEADD(LASTDATE(DimTime[Data]), -15, DAY))</code>
Total Sessions Last 28 Days	<code>CALCULATE([Total Sessions], DATEADD(LASTDATE(DimTime[Data]), -28, DAY))</code>

Neste *dashboard* também é também o cabeçalho personalizado que foi anteriormente desenvolvido na seção anterior, o qual encontra-se exibido no menu principal. Para além de conter a data da última atualização do modelo de dados, possui também um campo de data relativa que permite obter os dados dos últimos x dias/semanas/meses/anos e outro campo de data que permite colocar uma data de início e de fim. Por último, foi adicionado um botão de retrocesso para o menu principal para facilitar a navegação

4.3.1.2 Sessions by Country

Com este *dashboard* (figura 40) pretendeu-se que houvesse um maior foco no número de *Sessions* registado por cada país. Neste caso, os Estados Unidos têm o maior número de *Sessions* registadas, seguido pela Alemanha, e logo a seguir pelo Vietnam. Para estes resultados foram usadas as visualizações de mapa, gráfico de área e cartão de linhas múltiplas.

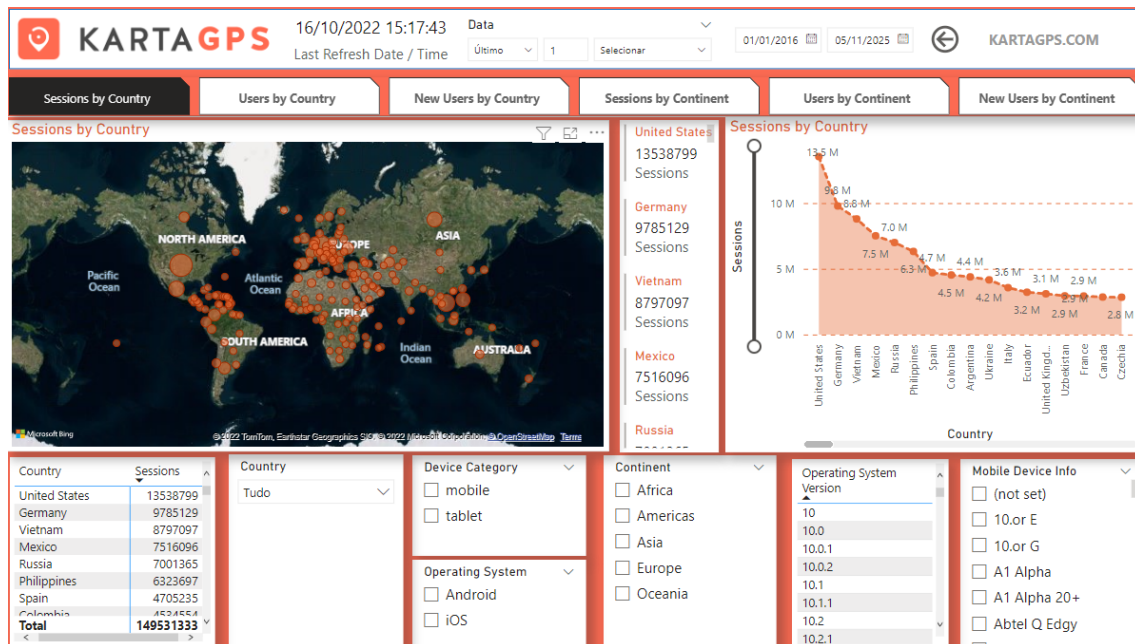


Figura 40: Sessions by Country.

É também possível visualizar diferentes tipos de filtros existentes que permitem filtrar para além do país, permitem filtrar por *Device Category*, *Operating System*, *Continent*, *Operating System Version*, e *Mobile Device Info*. Adicionalmente, existe uma barra de marcadores entre o cabeçalho e as visualizações que permite alternar entre diferentes

dashboards. Estes filtros adicionais, navegador de marcadores e o cabeçalho estão presentes em todas as *dashboards* apresentadas.

4.3.1.3 Users by Country

O objetivo para este *dashboard* (figura 41) foi semelhante ao do anterior, contudo, o objeto de análise foi a análise do número de utilizadores por país (*Users by Country*), que podem registar uma ou mais *Sessions*. O top 3 de países continua a ser entre os Estados Unidos, Alemanha e Vietnam, e as visualizações de mapa, gráfico de área e cartão de linhas múltiplas foram novamente usadas para manter a coerência dos dados filtrados por país. Uma vez mais continua presente o navegador de marcadores, referenciando a *dashboard* selecionada (*Users by Country*).

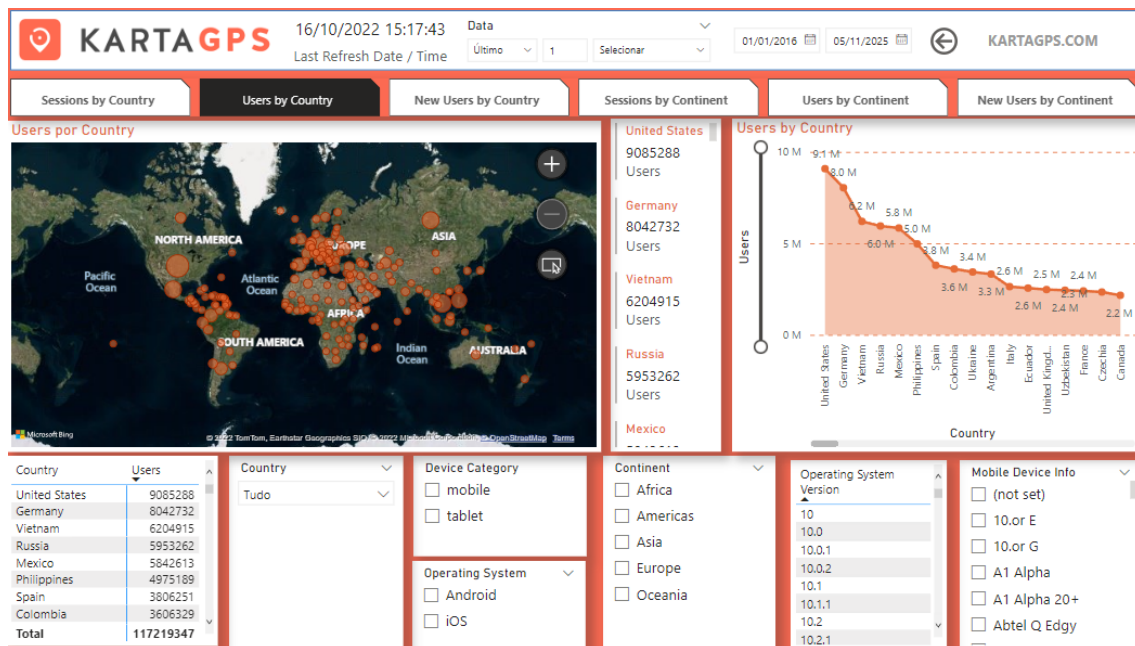


Figura 41: Users by Country

4.3.1.4 New Users by Country

Continuando o mesmo tipo de desenho e filtro das visualizações, este *dashboard* (figura 42) procura dar a entender quais os países que têm um maior número de novos utilizadores. Aqui temos uma grande diferença em que a *dashboard* regista um maior número de novos utilizadores (*New Users*) na Rússia, seguido da Alemanha e a terminar no México. Visto que a Rússia continua a ser o país onde há um maior registo de novos utilizadores, estes dados são coerentes com a realidade atual da NDrive.

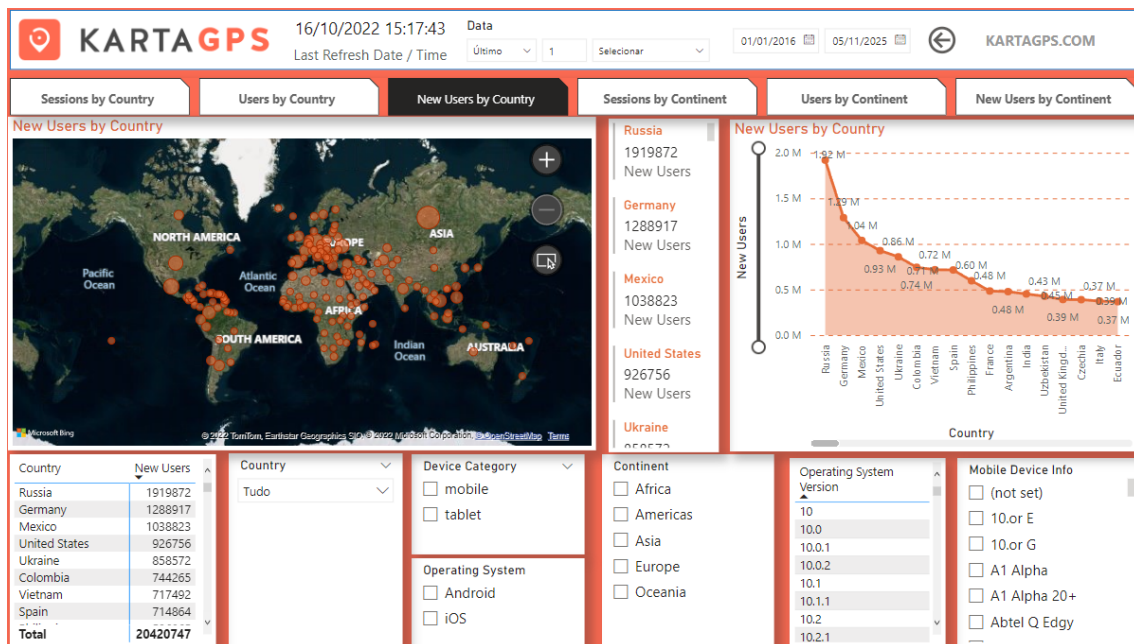


Figura 42: New Users by Country

4.3.1.5 Sessions by Country

Com este *dashboard* (figura 43) o objetivo foi disponibilizar o número de sessões registadas por Continente (*Sessions by Continent*), o qual de acordo com o registado na visualização cartão é de 150 milhões de sessões totais. A Europa detém o maior número de *Sessions* por Continente, registando aproximadamente 60 milhões e 40.05% da totalidade, seguido do continente americano, o qual tem 47 milhões de *Sessions* e 31.13% da totalidade. O Continente Asiático por sua vez regista 35 milhões de *Sessions* o que totaliza 23.41%. Tanto o continente africano como a Oceânia representam menos de 5% e houve também dados que não puderam ser atribuídos a um país (normalmente isto poderá dever-se a utilizadores que nunca tenham usado os serviços de localização na aplicação).

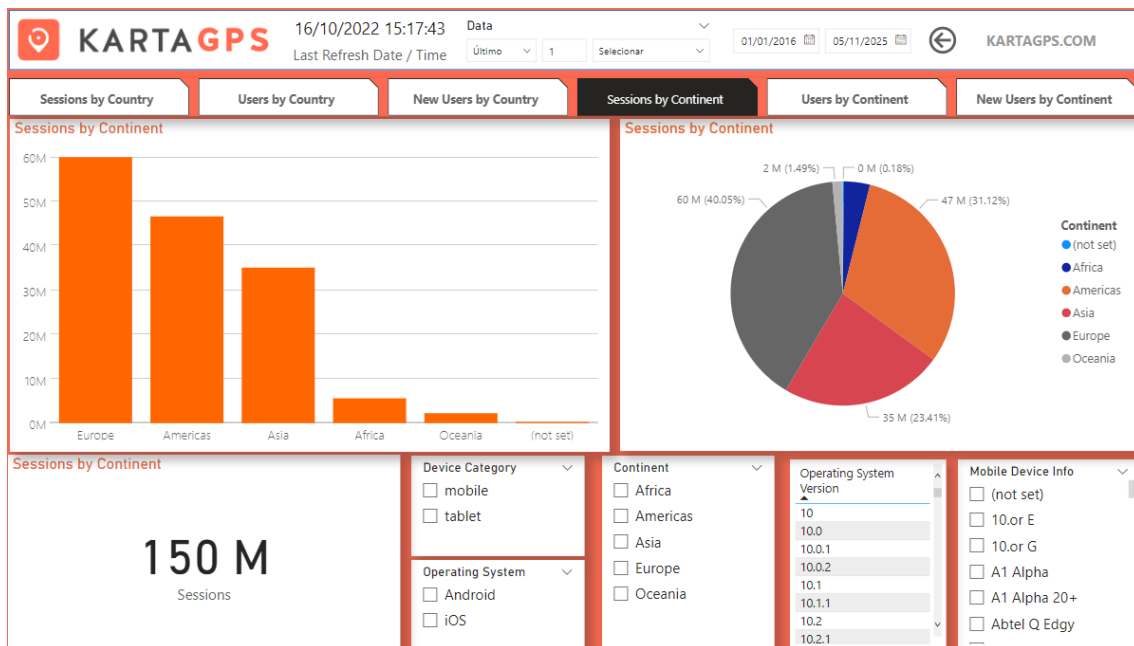


Figura 43: Sessions by Continent

As 3 visualizações foram alteradas para uma visualização de cartão, gráfico de barras e gráfico circular, sendo também utilizadas nos *dashboards* seguintes.

4.3.1.6 Users by Continent

Para este *dashboard* (figura 44) o objetivo proposto foi a visualização do número de utilizadores registado por Continente (*Users by Continent*). Nos dados aqui registados ao nível histórico, foram registados um total de 117 milhões de utilizadores, tendo 41.65% desse registo ter sido gerado por utilizadores residentes no continente europeu, o que equivale a 49 milhões. Seguidamente vem o continente americano que regista um total de 35 milhões de utilizadores (29.59%) e o continente asiático a registar 27 milhões (23.13%).

Os restantes Continentes não permitem aferir uma percentagem aceitável, contudo, não são o foco que a NDrive procura nesta altura.

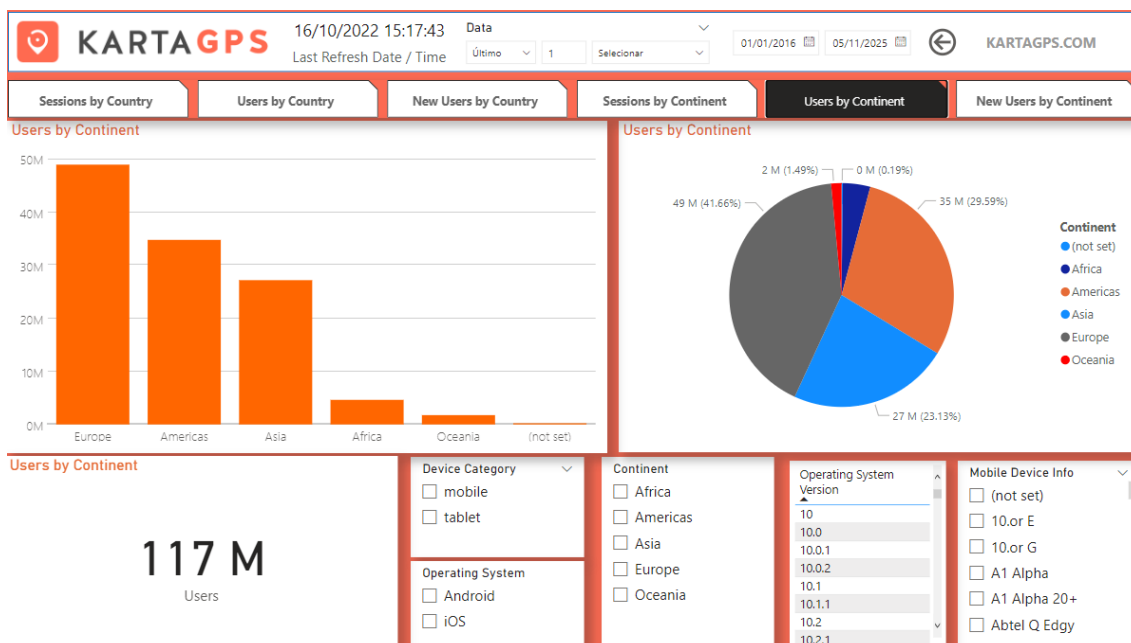


Figura 44: Users by Continent.

4.3.1.7 New Users by Continent

Finalizando com o último *dashboard* (figura 45), procurou-se saber a quantidade de novos utilizadores por continente a usarem a aplicação (*New Users by Continent*), no qual registou-se no período compreendido a partir de 2016, 20 milhões de novos utilizadores (*New Users*) que usaram a aplicação pela primeira vez. Uma vez mais o continente europeu tem o maior número de novos utilizadores, registando assim 9 milhões de novos utilizadores (46.02% do total), seguido do continente americano com 5 milhões, equivalente a 25.25% da totalidade de novos utilizadores e depois o continente asiático com 5 milhões de novos utilizadores, equivalente a 22.68% da totalidade.

Os restantes 6.05% partilhados entre o continente africano, a Oceânia e dados remanescentes, não são considerados como bons resultados, embora o crescimento registado no continente africano possa permitir atuar nesse mercado no futuro.

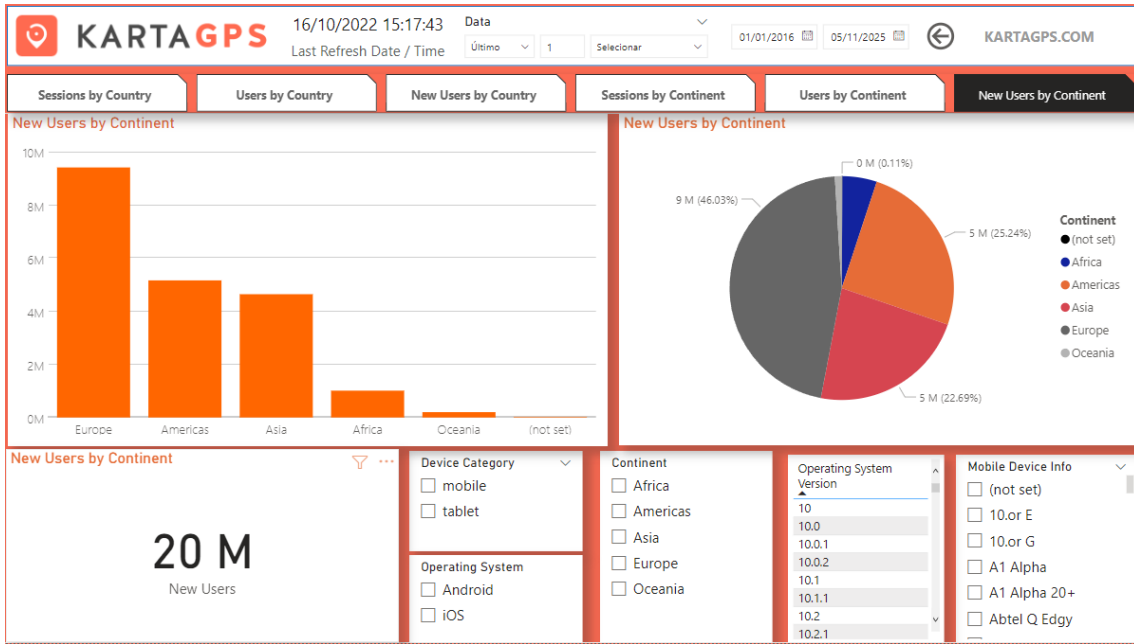


Figura 45: New Users by Continent.

5.1 Validação e discussão de dados

Na etapa de avaliação da metodologia investigação-ação é feita uma avaliação dos resultados obtidos. Para se poder avaliar as dashboards desenvolvidas foi realizado um questionário (APÊNDICE II) que envolveu os elementos da administração e responsáveis pela área de Marketing, de modo a avaliar o grau de usabilidade das dashboards (Brooke,1996).

O tipo de questionário escolhido foi o SUS, ou *System Usability Scale*. Este tipo de inquérito é uma escala de avaliação ad-hoc, permitindo a avaliação de usabilidade de um sistema de forma rápida e fácil. É composto por um questionário com 10 questões que usam a escala de 5 pontos de Likert, que vai de 1 a 5.

Os resultados usam uma escala de 0 a 100, e quanto maior for o resultado, maior será a usabilidade do sistema. É considerado um valor acima da média caso seja superior a 68 e tudo o que seja inferior a 68 será considerado abaixo da média (Sauro, 2011).

Analisando os resultados globais do inquérito apresentado, a média total do questionário SUS foi de 77.75 valores, o que representa um valor registado acima da média (tabela 16).

Tabela 16: Valores totais e média global do SUS

R1	82.5
R2	80
R3	85
R4	45
R5	82.5
R6	65
R7	87.5
R8	82.5
R9	82.5
R10	85
Média total SUS	77.75

Analisando a primeira questão colocada no SUS, 30% dos participantes deram uma resposta neutra (3) quanto à frequência (tabela 17). É uma situação expectável pois nem todos os elementos da administração têm interesse em dar uso frequente às dashboards. Mesmo assim o valor médio foi de 4 valores no grau de concordância.

Tabela 17: Avaliação da frequência de uso das dashboards

Participante	[Acho que gostaria de utilizar este sistema com frequência.]
P1	4
P2	5
P3	4
P4	4
P5	4
P6	3
P7	5
P8	5
P9	3
P10	3
Média	4

Dos 10 participantes que responderam à segunda questão (tabela 18), 30% discordam e só 10% concordam, ficando os restantes com uma opinião neutra.

Tabela 18: Avaliação da complexidade das dashboards

Utilizador	[Considerarei o sistema mais complexo do que necessário.]
P1	2
P2	3
P3	3
P4	4
P5	3
P6	3
P7	3
P8	3
P9	1
*10	2
Média	2.7

90% dos participantes que responderam à questão 3 consideram que o sistema é fácil de usar, tenho apenas um dos participantes deixado uma opinião neutra. (tabela 19)

Tabela 19: Avaliação da facilidade de uso

Participante	[Achei o sistema fácil de utilizar.]
P1	5
P2	5
P3	5
P4	3
P5	4
P6	4
P7	4
P8	5
P9	4
P10	4
Média	4.3

Na tabela 20, que corresponde à quarta questão, temos apenas 10% dos participantes que indicou ter mais dificuldades em usar a dashboard e recorrer a terceiros para ajuda.

Tabela 20: Análise à capacidade do participante

Participante	[Acho que necessitaria de ajuda de um técnico para conseguir utilizar este sistema.]
P1	1
P2	2
P3	2
P4	4
P5	2
P6	2
P7	1
P8	1
P9	1
P10	1
Média	1.7

Na quinta questão (tabela 21), 90% consideram que as funcionalidades do dashboard estavam bem integradas, ficando assim uma média de 4.2 valores no grau de concordância nesta questão.

Tabela 21: Avaliação da integração das funcionalidades do dashboard

Participante	Considere que as várias funcionalidades deste sistema estavam bem integradas.
P1	4
P2	4
P3	5
P4	3
P5	4
P6	4
P7	5
P8	4
P9	5
P10	4
Média	4.2

Na tabela 22, sexta questão, 90% dos participantes discordam que o dashboard tenha muitas inconsistências, resultando assim em 1.6 valores no grau de discordância.

Tabela 22: Avaliação das inconsistências nos dashboards

Participante	[Achei que este sistema tinha muitas inconsistências.]
P1	2
P2	2
P3	1
P4	3
P5	1
P6	1
P7	1
P8	2
P9	2
P10	1
Média	1.6

Na tabela 23, sétima questão, 90% dos participantes concordam que poderiam aprender a utilizar rapidamente o dashboard, equivalendo a 4.1 valores no grau de concordância.

Tabela 23: Avaliação da rapidez de aprendizagem no uso das dashboards

Participante	[Suponho que a maioria das pessoas aprenderia a utilizar rapidamente este sistema.]
P1	4
P2	4
P3	4
P4	4
P5	5
P6	3
P7	4
P8	4
P9	4
P10	5
Média	4.1

Na oitava questão, 80% dos participantes discordam que os dashboards sejam muito complicados de utilizar (tabela 24), enquanto que 10% não concorda nem discorda, e os restantes 10% concordam que sejam complicados.

Tabela 24: Avaliação do grau de complexidade das dashboards

Participante	[Considere o sistema muito complicado de utilizar.]
P1	2
P2	2
P3	1
P4	4
P5	1
P6	3
P7	1
P8	2
P9	2
P10	1
Média	1.9

Na nona questão (tabela 25), 80% dos participantes sente-se muito confiante a usar as dashboards, enquanto 10% sente o oposto e os restantes 10% têm uma opinião neutra.

Tabela 25: Avaliação do grau de confiança durante o uso das dashboards

Participante	[Senti-me muito confiante a utilizar este sistema]
P1	4
P2	4
P3	4
P4	2
P5	4
P6	3
P7	4
P8	4
P9	4
P10	4
Média	3.7

Na décima e última questão (tabela 26), 90% dos utilizadores discordam que tivessem tido que aprender muito para conseguirem lidar com as dashboards, havendo só uma opinião neutra.

Tabela 26: Análise da capacidade de aprendizagem do participante

Participante	[Tive que aprender muito antes de conseguir lidar com este sistema.]
P1	1
P2	1
P3	1
P4	3
P5	1
P6	2
P7	1
P8	1
P9	1
P10	1
Média	1.3

5.2 Aprendizagem Específica

Tendo as dashboards sido apresentadas e avaliadas pela direção e pelo departamento de marketing, conseguiu-se desenvolver uma solução que a empresa decidiu continuar a apoiar o seu desenvolvimento, assim como propor melhorias ao que foi desenvolvido até agora.

Este projeto contribuiu para dar a conhecer à empresa alguns dos benefícios do Business Intelligence e em como pode ajudar a melhorar os atuais processos de tomada de decisão. Compreende-se que esta solução ainda possa suscitar alguma dificuldade ou desinteresse na exploração dos dados, contudo, pode considerar-se como um elemento catalisador para novas oportunidades de tornar esta solução ainda melhor.

6.1 Conclusão

Este trabalho de projeto teve como objetivo principal a conceção de uma solução de Business Intelligence, a qual consistiu na criação e modelação de um *Data Warehouse* e posterior exploração através de *dashboards* para a equipa de Marketing da NDrive. De modo a poder responder-se à questão de investigação e objetivos propostos, foi feita uma revisão de literatura acerca de Business Intelligence e Analytics, cubos OLAP e modelação multidimensional, a qual também permitiu uma maior contextualização e conhecimento acerca do tema.

Os indicadores de negócio necessários foram avaliados com a empresa, tendo sido iniciado o processo ETL utilizando a ferramenta Power BI, onde foram recolhidas as dimensões e métricas necessárias através do Google Analytics para posterior exploração. Posteriormente os dados foram limpos e transformados, tendo sido carregados para o *Data Warehouse*, onde se procedeu à modelação multidimensional dos dados e posterior criação de um conjunto de *dashboards* necessárias para que fosse possível a visualização de toda a informação necessária, complementada por métricas calculadas em linguagem DAX.

A metodologia de investigação-ação permitiu delinear de uma forma bastante efetiva a conceção do trabalho de projeto, permitindo assim entregar uma proposta à empresa com a estruturação adequada e que pode ser alvo de constante mudança e melhoramento. Também houve aqui um enriquecimento pessoal pelo conhecimento obtido na área do *Business Intelligence*, contribuindo para que seja possível trabalhar melhor na plataforma e desenvolver novas soluções baseadas nesta implementação.

Conclui-se assim que a implementação desta proposta de *Business Intelligence* permitiu responder à questão de investigação e aos objetivos propostos, permitindo assim à NDrive obter os dados mais atuais e relevantes diariamente nas *dashboards desenhadas*, proporcionando assim ao departamento de Marketing a capacidade de explorar os dados com *dashboards* intuitivas e fáceis de usar, tornando assim os processos de tomada de decisão mais céleres.

6.2 Contribuições

A solução de Business Intelligence proposta e desenvolvida contribui para as empresas e para o conhecimento científico das seguintes maneiras:

Para as empresas:

- Contribui com importância de ter dashboards que sejam capazes de mostrar os indicadores necessários para os processos de tomada de decisão

Para o conhecimento científico:

- Demonstrar os benefícios da ferramenta Power BI Desktop para a criação de dashboards assim como na modelação dimensional dos dados no setor de marketing de uma PME.
- Demonstrar o uso da metodologia investigação-ação na elaboração de um projeto de Business Intelligence para uma PME para enriquecimento da literatura existente.

6.3 Dificuldades e Limitações

Durante o desenvolvimento deste trabalho de projeto, foram detetadas algumas limitações que causaram um maior tempo de desenvolvimento do trabalho, contudo, foram ultrapassadas e permitiram também aprender formas como lidar com estas situações no futuro.

O conector usado pelo Microsoft Power BI não permite o uso de dimensões do Google Analytics 4, o qual só usa o *Universal Analytics*, que será descontinuado em 2023. Uma vez que não há qualquer documentação acerca desta limitação na documentação da Microsoft, houve tempo adicional gasto em como tentar resolver a limitação. Optou-se assim por usar as dimensões do Universal Analytics, as quais são perfeitamente usáveis até que sejam descontinuadas pela Google.

Na fase de carregamento do processo ETL dos dados para o modelo, verificou-se que o processo não terminava e os dados carregados excediam os 100GB. Ao analisar-se o modelo de dados através da ferramenta *DAX studio*, verificou-se que havia dimensões que tinham sido duplicadas múltiplas vezes, causando a situação descrita. Como não havia forma de resolver a situação, usou-se uma versão mais antiga do ficheiro pbix criado no Power BI.

6.4 Trabalho Futuro

A proposta apresentada à empresa cumpriu com os objetivos inicialmente propostos para o seu desenvolvimento, contudo, ainda pode ser melhorada de modo a superar algumas

das suas limitações. Deste modo, são enumeradas várias sugestões que podem ser implementadas no futuro caso a empresa assim o decida:

- Integração de outros departamentos da empresa no Data Warehouse, como por exemplo, financeiro e de QA;
- Expandir o atual Data Mart de Marketing com dados provenientes do Google Firebase (nível financeiro);
- Explorar o uso de métodos de Machine Learning para obtenção de métricas de uso das redes sociais, assim como de críticas e avaliações de utilizadores.
- Uso do serviço Google Analytics 4 através do atual conector oficial da Microsoft (não suportado até ao momento, ou procurar uma solução alternativa. Implica a revisão de todo o processo ETL em qualquer um dos casos, dado que o cálculo das métricas e as dimensões diferem do atual Analytics em uso.

REFERÊNCIAS BIBLIOGRÁFICAS

Referências Bibliográficas

- Abelló, A., Romero, O. (2009). On-Line Analytical Processing. In: LIU, L., ÖZSU, M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_252
- Abib, G., Hoppen, N., & Hayashi Junior, P. (2013). Observação participante em estudos de administração da informação no Brasil. *Revista de Administração de Empresas*, 53, 604-616.
- Ain, N., Vaia, G., DeLone, W.H., & Waheed, M. (2019). Two decades of research on business intelligence system adoption, utilization and success - A systematic literature review. *Decis. Support Syst.* 125.
- Aleksic, S., Jasna & Krasavac, C., Biljana & Karamata, Ema. (2019). Business analytics: new concepts and trends. *Management:Journal of Sustainable Business and Management Solutions in Emerging Economies*. 25. 10.7595/management.fon.2019.0013.
- Alhyasat, E., & Al-Dalahmeh, M. (2013). Data Warehouse Success and Strategic Oriented Business Intelligence: A Theoretical Framework. *Journal of Management Research*, 3(3), pp. 169-184. doi:10.5296/jmr.v5i3.3703
- Antdata. (2021). *Power BI License and Subscription Types. What to choose?* .Antdata. https://antdata.eu/power_bi_licenses_and_subscriptions.html
- Antonelli, R. A. (2010). Conhecendo o business intelligence (BI). *CAP Accounting and Management-B4*, 3(3), 79-85
- Ayankoya, K., Calitz, A., & Greyling, J. (2014, Setembro). Intrinsic relations between data science, big data, business analytics and datafication. In *Proceedings of the Southern African Institute for Computer Scientist and Information Technologists Annual Conference 2014 on SAICSIT 2014 Empowered by Technology* (pp. 192-198).
- Baskerville, R. (1999). Investigating Information Systems with Action Research. *Commun. Assoc. Inf. Syst.*, 2, 19.
- Baltzan, P., & Phillips, A. (2014). *Ebook: Business Driven Information Systems*. McGraw Hill.
- Beri, R. (2016). *Data Mining & Data Warehousing: An Exhaustive Elucidation*. 2.
- Breslin, M. (2004). Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models. *Business Intelligence Journal*. 9.

- Bhardwaj, A. (2021, abril, 20). *Importance of Data Warehousing In BI Software Development*. Erp Solutions. <https://erpsolutions.oodles.io/blog/data-warehousing-business-intelligence/>
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- Bhardwaj, A. (2022, Abril 5). *Importance of data warehousing in BI software development services*. Oodles ERP. <https://erpsolutions.oodles.io/blog/data-warehousing-business-intelligence/>
- Burki, I. (2018, Janeiro 11). *What is a data warehouse?* Stoneridge Software. <https://stoneridgesoftware.com/what-is-a-data-warehouse/>
- Caldeira, J. (2010). *Dashboards - Comunicar Eficazmente a Informação de Gestão*. Grupo Almedina.
- Caldeira, J. (2012). *100 Indicadores da Gestão - Key Performance Indicators*. Grupo Almedina.
- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88-98.
- Cheung, D. W., Zhou, B., Kao, B., Kan, H., & Lee, S. D. (2001). Towards the building of a dense-region-based OLAP system. *Data & Knowledge Engineering*, 36(1), 1-27.
- Chowdhury, S. (2014, maio, 26). *Data warehouse augmentation, Part 1*. IBM Developer. <https://developer.ibm.com/tutorials/ba-augment-data-warehouse1/>
- Chenoweth, T. Schuff, D, & Louis, R. S. (2003). Method for developing dimensional data marts. *Communications Of The ACM*, 46(12), 93-98. doi:10.1145/953460.953465
- Codd, F. (1993). Technical Report. San Jose, Calif: Codd EF & Associates. Providing OLAP (Online Analytical Processing) to User-Analysts: An IT Mandate.
- Gonçalves, C.T. (2021). *Business Intelligence and Analytics*. [Powerpoint Slides]. <https://moodle.iscap.pt/ano2021/>.
- Davenport, T. H., and Harris, J. G. 2007. *Competing on Analytics: The New Science of Winning*. Harvard Business School Press.
- Davenport, T.H. (2013). *Analytics 3.0*, Harvard Business Review.
- Davenport, T.H. (2016) *Analytics and IT: New Opportunity for CIOs*, Harvard Business Review. <https://hosteddocs.ittoolbox.com/NewOpportunityforCIOs.pdf>

- Davenport, T.H. (2018) From analytics to artificial intelligence, *Journal of Business Analytics*, 1:2, 73-80, DOI: 10.1080/2573234X.2018.1543535
- Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), 64-73.
- El Naqa, I., & Murphy, M. J. (2015). What is machine learning?. In machine learning in radiation oncology.
- Domingues, R., Pedrosa, I., & Bernardino, J. (2020). Indicadores chave de desempenho em marketing., (E35), 128-140.
- Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review*. 70 (11): 35–36.
- Edureka. (2022, December 13). *Top 50+ tableau interview questions and answers for 2023*. <https://www.edureka.co/blog/interview-questions/top-tableau-interview-questions-and-answers/>
- Eckerson, W.W. (2009). Performance Management Strategies: How to Create and Deploy Effective Metrics. *TDWI Best Practices Report*.
- Eckerson, W. W. (2010). Performance Dashboards: Measuring, Monitoring, and Managing Your Business. Business (2nd ed.). Hoboken, New Jersey: John Wiley & Sons. <https://doi.org/10.2514/6.2008-3494>
- Embarak, O. (2018). The Importance of Data Visualization in Business Intelligence. *Data Analysis and Visualization Using Python*, 85–124. doi:10.1007/978-1-4842-4109-7_2
- Evans, J. R., & Lindner, C. H. (2012). Business Analytics: The Next Frontier for Decision Sciences. College of Business, University of Cincinnati. Decision Science Institute.
- Ferrari, A., & Russo, M. (2016). *Introducing Microsoft Power BI*. Microsoft Press.
- Few, S. (2007). Dashboard design for real-time situation awareness. *Inova Solutions*.
- Few, S., & Edge, P. (2007a). Dashboard confusion revisited. *Perceptual Edge*, 1-6.
- Few, S. (2013). *Information dashboard design: Displaying data for at-a-glance monitoring* (Vol. 5). Burlingame, CA: Analytics Press.
- Friedman, V. (2008). Data visualization and infographics. *Graphics, Monday Inspiration*, 14, 2008.

- Golfarelli, M., Rizzi, S., & Cella, I. (2004). Beyond data warehousing. *Proceedings of the 7th ACM International Workshop on Data Warehousing and OLAP - DOLAP '04*. doi:10.1145/1031763.1031765.
- Gounder, M. S., Iyer, V. V., & Al Mazyad, A. (2016, March). A survey on business intelligence tools for university dashboard development. In *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)* (pp. 1-7). IEEE.
- Gowthami, K., & Kumar, M. P. (2017). Study on business intelligence tools for enterprise dashboard development. *International Research Journal of Engineering and Technology (IRJET)*, 4(04), 2987-2992.
- Han, J., et al. 2012. *Data Mining Concepts and Techniques*. 3rd Edition. USA: Elsevier Inc.
- Hočevár, B., & Jaklič, J. (2010). Assessing benefits of business intelligence systems—a case study. *Management: journal of contemporary management issues*, 15(1), 87-119.
- IBM (2020, maio, 18). What is a Data Mart. IBM Cloud Learn Hub. <https://www.ibm.com/cloud/learn/data-mart>
- IBM. (2022, 2 agosto). *What is data science*. <https://www.ibm.com/cloud/learn/data-science-introduction>
- Inmon W.H., (1993) *Building the Data Warehouse*, A Wiley QED publication, John Wiley and Sons, Inc. New York 123-133
- Irani, Z., & Love, E.D. P. (2000). The Propagation of Technology Management Taxonomies for Evaluating Investments in Information Systems, *Journal of Management Information Systems*, 17:3, 161-177, DOI: 10.1080/07421222.2000.11045650
- Jawadekar, W., (2010) *Knowledge Management*, Tata McGraw-Hill
- Jewett, D. (2019, June 4). *Moving beyond the Personal edition of Tableau Desktop*. Tableau. <https://www.tableau.com/blog/moving-beyond-personal-edition-tableau-desktop>
- Jukic, N. (2006). Modelling strategies and alternatives for data warehousing projects. *Communications of the ACM*, 49(4), 83-88.
- Khedikar, K. A. (2021, April). Data Analytics for Business Using Tableau. In *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*.

- Keboola. (2022, Outubro 28). *Understanding OLAP cubes - A guide for the perplexed*. Keboola - Connect any data source in less than 20 minutes. <https://www.keboola.com/blog/olap-cubes>
- Keboola. (2022a, November 15). *Star schema vs snowflake schema and the 7 critical differences*. Keboola - Connect any data source in less than 20 minutes. <https://www.keboola.com/blog/star-schema-vs-snowflake-schema>
- Keen, P. G. W., & Scott Morton, M. S. (1978). *Decision support systems: An organizational perspective*. Reading, MA: Addison-Wesley.
- Kerzner, H. (2011). *Project management metrics, KPIs, and dashboards: a guide to measuring and monitoring project performance*. John Wiley & Sons.
- Kimball, R., (1996) *The Data Warehouse Toolkit*. John Wiley & Sons. (Last edition: 2nd edition, John Wiley & Sons.
- Kimball, R., & Ross, M. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons.
- Kimball, R. and Ross, M. (2013) *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. 3rd Edition, John Wiley & Sons, Inc., Indianapolis.
- Kirk, A. (2016). *Data visualisation: A handbook for data driven design*. Sage.
- Kitchin, R., Maalsen, S., & McArdle, G. (2016). The praxis and politics of building urban dashboards. *Geoforum*, 77, 93-101.
- Klopfenstein, J., & Howell, D. (2022, 8 setembro). *Power Query M formula language*. Microsoft Learn: Build skills that open doors in your career. <https://learn.microsoft.com/en-us/powerquery-m/m-spec-introduction>
- Kronz, A., Schlegel, K., Sun, J., R., Pidsley, D., and Ganeshan, A. (2021). *Magic quadrant for analytics and business intelligence platforms*. Gartner. <https://www.gartner.com/doc/reprints?id=1-2955ETOT&ct=220215&st=sb?>
- Lane, P. & Schupmann, V. (2002) *Oracle9i Data Warehousing Guide, Release 2 (9.2)*. Oracle. https://docs.oracle.com/cd/B10500_01/server.920/a96520.pdf
- Lewin, K. (1951). *Field Theory in Social Science*. New York: Harper & Bros.

- Lousa, A., Pedrosa, I., & Bernardino, J. (2019, June). Evaluation and Analysis of Business Intelligence Data Visualization Tools. In *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1-6). IEEE.
- Mackinlay, J. (2021, December 2). *Analyzing the history of Tableau innovation*. Tableau. <https://www.tableau.com/blog/analyzing-history-tableau-innovation>
- Martyn, T. (2004). Reconsidering Multi-Dimensional schemas. *SIGMOD Rec.* 33, 1 (March 2004), 83–88. <https://doi.org/10.1145/974121.974136>
- Jhawar, R. (2022, junho, 21). *Extract, transform, and load (ETL)*. Microsoft Learn. <https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl>
- Mendes, J. M. H. (2013). *Balanced scorecard e painel de indicadores: implementação numa entidade pública do sector dos transportes* (Doctoral dissertation, FEUC).
- Microsoft. (2019, outubro 8). *The benefits of business analytics*. Microsoft 365. <https://www.microsoft.com/en-us/microsoft-365/business-insights-ideas/resources/benefits-of-business-analytics>
- Microsoft. (2022). *What is Power BI? | Microsoft Learn*. <https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview>
- Microsoft. (2022a). *Crie marcadores de relatório em Power BI para partilhar insights e construir histórias*. Microsoft Learn. <https://learn.microsoft.com/pt-pt/power-bi/create-reports/desktop-bookmarks?tabs=powerbi-desktop>
- Mohammed, K. I. (2019). Data Warehouse Design and Implementation Based on Star Schema vs. Snowflake Schema. *International Journal of Academic Research in Business and Social Sciences*, 9(14), 25–38.
- Monterey, T. (2022). What Is Data Warehouse In Business Intelligence? <https://www.ictsd.org/what-is-data-warehouse-in-business-intelligence/>
- Nair, L., Shetty, S., & Shetty, S. (2016). Interactive visual analytics on Big Data: Tableau vs D3.js. *Journal of e-Learning and Knowledge Society*, 12(4).
- Niu, L., Lu, J., & Zhang, G. (2009). Cognition-Driven Decision Support for Business Intelligence. *Studies in Computational Intelligence*. doi:10.1007/978-3-642-03208-0
- O'Brien, J. A., & Marakas, G. M. (2009). *Management information systems*. Boston: McGraw-Hill Irwin.

- Oey, E., Harno, S. S. S., & Zain, C. (2021). Developing Integrated Performance Dashboards with Power BI—a Case Study in a Medium-Size Manufacturer. *In 2021 International Conference on Information Management and Technology (ICIMTech)* (Vol. 1, pp. 265-270). IEEE.
- Olszak, C. M., & Ziemba, E. (2003). Business intelligence as a key to management of an enterprise. *Proceedings of Informing Science and IT Education Conference*.
- Olszak, C. M., & Ziemba, E. (2007). Approach to building and implementing Business Intelligence systems. *Interdisciplinary Journal of Information, Knowledge, and Management*, 2, 135–148. DOI:10.28945/105.
- Oracle (1999). *Oracle8i Data Warehousing Guide, Release 2 (8.1.6)*. Oracle. https://docs.oracle.com/cd/A81042_01/DOC/server.816/a76994/marts.htm
- Parmenter, D. (2010). *Key Performance Indicators Developing, Implementing, and Using Winning KPIs*. John Wiley & Sons, Hoboken.
- Pedersen, T. (2009). Multidimensional Modeling. In: LIU, L., ÖZSU, M.T. (eds) *Encyclopedia of Database Systems*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_229.
- Peregud, I. (2019, April 28). *Advanced analytics advance your business*. Indatalabs. <https://indatalabs.com/blog/advanced-analytics-advance-your-business>
- Peterson, E. T., 2006. *The Big Book of Key Performance Indicators*. First Edition.
- Petersen, J. A. McAlister, L. Reibstein, D. J., 2009. Choosing the right metrics to maximize profitability and shareholder value. *Journal of Retailing*, v. 85, n. 1, p. 95-111.
- Pirna, M. & Botezatu, C.P. (2010) General information on business Intelligence and OLAP systems architecture. *Computer and Automation Engineering (ICCAE), 2010. The 2nd International Conference on*, pp. 294.
- Power, D. J. (2003). A brief history of decision support systems. *DSSResources.com*, 3.
- Power, D. J., Heavin, C., McDermott, J., & Daly, M. (2018). Defining business analytics: an empirical approach. *Journal of Business Analytics*, 1(1), 40-53.
- Presthus, W., & Bergum, I. (2015). Business Intelligence to the People. A Case Study of Dashboard Adoption in the Health Care sector. In *Nor. Konf. Organ. bruk av IT* (Vol. 23, No. 1).

- Primak, F. (2008). *Decisões com B.I.-Business Intelligence*. São Paulo: Ciencia Moderna.
- Rad, R. (2019). *From Rookie to Rock Star* (Vol. I). RADACAD Systems Limited.
- Sauro, J. (2011). *A practical guide to the system usability scale: Background, benchmarks & best practices*. Measuring Usability LLC.
- Sekar, M. (2022). Storytelling Tools. In *Machine Learning for Auditors* (pp. 173-179). Apress, Berkeley, CA.
- Rippen, M. (2005). *Tessella Support Services Plc*, Issue V2.R.M0.
- Rouse, M. (2017, setembro). Key performance indicators (KPIs) - [online] Disponível em: <https://www.techtarget.com/searchbusinessanalytics/definition/key-performance-indicators-KPIs>
- Samuel, A. L. (1967). Some studies in machine learning using the game of checkers. II—Recent progress. *IBM Journal of research and development*, 11(6), 601-617.
- Santos, M. Y., & Ramos, I. (2017). *Business Intelligence-Da Informação ao Conhecimento*. FCA—Livros de Informática, Lisboa.
- Santos, R. D. C. D. (2018). Power BI: a experiência de implantação em um escritório de contabilidade.
- Sharda, R., Delen, D., Turban, E., Aronson, J., & Liang, T. (2014). *Business intelligence and analytics: System for Decision Support*. Global Edition.
- Sharda, R., Delen, D., & Turban, E. (2018). *Business intelligence, analytics, and data science: A managerial perspective* (Fourth edition). Pearson.
- Shim, Jung & Warkentin, Merrill & Courtney, James & Power, Daniel & Sharda, Ramesh & Carlsson, Christer. (2002). Past, Present, and Future of Decision Support Technology. *Decision Support Systems*. 33. 111-126. 10.1016/S0167-9236(01)00139-7.
- Silva, A. (2010). *Business Intelligence: auxílio na tomada de decisão*.
- Software Testing Help. (February). *Dimensional Data Model In Data Warehouse – Tutorial With Examples*. <https://www.softwaretestinghelp.com/dimensional-data-model-in-data-warehouse/>
- Stecyk, A. (2018). Business Intelligence systems in SMEs. *European Journal of Service Management*, 3 (27/2), 409–413. DOI: 10.18276/ejasm.2018.27/2-50. #1#

- Sun, Z., Strang, K., & Firmin, S. (2017). Business analytics-based enterprise information systems. *Journal of Computer Information Systems*, 57(2), 169-178.
- Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: an introduction MIT Press. Cambridge, MA, 22447.
- Susman, G. I., & Evered, R. D. (1978), An assessment of the scientific merits of action research. *Administrative science quarterly*, 582-603.
- Tamang, M. D., Shukla, V. K., Anwar, S., & Punhani, R. (2021, April). Improving Business Intelligence through Machine Learning Algorithms. In *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)* (pp. 63-68). IEEE.
- Turban, E., Sharda, R., Aronson, J. E., & King, D. (2008). *Business intelligence: A managerial approach* (pp. 58-59). Upper Saddle River, NJ: Pearson Prentice Hall.
- Turban, E., Sharda, R., Delen, D. (2014). *Business Intelligence and Analytics*, 10th Edition, Prentice Hall, ISBN 10: 0-13-305090-4, ISBN13: 978-0-13-305090-5.
- Vassiliadis, P., & Sellis, T. (1999). A survey of logical models for OLAP databases. *ACM Sigmod Record*, 28(4), 64-69.
- Vercellis, C. (2009). *Business intelligence: data mining and optimization for decision making* (1st ed.). John Wiley and Sons, Ltd.
- Watson, H. J. (2009). *Tutorial: Business Intelligence – Past, Present, and Future*. Communications of the Association for Information Systems, 25, pp-pp. <https://doi.org/10.17705/1CAIS.02539>.
- Watson, H. J. (2015). All About Analytics. *International Journal of Business Intelligence Research*. 4. 13-28. 10.4018/jbir.2013010102.
- Westerlund, P. (2008). Business intelligence: Multidimensional data analysis.
- Wijaya, S. F., Prabowo, H., & Kosala, R. (2017). Identification of key success factors and challenges for ERP systems—A systematic literature review. In *2017 International Conference on Applied Computer and Communication Technologies (ComCom)* (pp. 1-6). IEEE.
- Wright, N. (2019, Janeiro 15). *Everything you ever wanted to know about Microsoft Power BI*. Nigel Frank. <https://www.nigelfrank.com/insights/everything-you-ever-wanted-to-know-about-microsoft-power-bi>

Yau, N. (2011). *Visualize this: the FlowingData guide to design, visualization, and statistics*. John Wiley & Sons.

Zentut. (2021, Junho 5). *Ralph Kimball data warehouse architecture*. Zentut. <https://www.zentut.com/data-warehouse/ralph-kimball-data-warehouse-architecture/>

APÊNDICES

Apêndices

Apêndice A: Código em linguagem M usado no editor avançado do Power BI para criação da tabela calendário DimTime na fase ETL

```

let P_Today = DateTime.LocalNow(), P_StartDate = #date(2018, 1, 1), P_EndDate =
#date(Date.Year(P_Today),12,31), P_Culture = "en-EN", P_FirstDayOfWeek = 1, P_IsCarnivalHoliday =
true, Holiday = if P_Culture = "pt-PT" then "Feriado" else "Holiday", Quarter = if P_Culture = "pt-PT"
then "T" else "Q", Week = if P_Culture = "pt-PT" then "S" else "W", Weekend = if P_Culture = "pt-PT"
then "Fim de Semana" else "Weekend", WorkDay = if P_Culture = "pt-PT" then "Dia Útil" else "Work Day",
DayCount = Duration.Days(Duration.From(P_EndDate - P_StartDate)) + 1, Source = List.Dates(P_StartDate,
DayCount, #duration(1, 0, 0, 0)), TableFromList = Table.FromList(Source, Splitter.SplitByNothing()),
ChangedType = Table.TransformColumnTypes(TableFromList, {{"Column1", type date}}), RenamedColumns =
Table.RenameColumns(ChangedType, {{"Column1", "Date"}}), InsertId = Table.AddColumn(RenamedColumns,
"DateId", each Date.Year([Date])*10000 + Date.Month([Date])*100 +Date.Day([Date])), InsertYear =
Table.AddColumn(InsertId, "Year", each Date.Year([Date])), InsertQuarter = Table.AddColumn(InsertYear,
"Quarter", each Date.QuarterOfYear([Date])), InsertSemester = Table.AddColumn(InsertQuarter,
"Semester", each if [Quarter] < 3 then 1 else 2), InsertMonth = Table.AddColumn(InsertSemester, "Month
(#)", each Date.Month([Date])), InsertWeek = Table.AddColumn(InsertMonth, "Week", each
Date.WeekOfYear([Date], P_FirstDayOfWeek)), InsertDay = Table.AddColumn(InsertWeek, "Day", each
Date.Day([Date])), InsertMonthName = Table.AddColumn(InsertDay, "Month (Long)", each
Date.ToText([Date], "MMMM", P_Culture), type text), InsertShortMonthName =
Table.AddColumn(InsertMonthName, "Month", each try(Text.Range({#"Month (Long)"},0,3)) otherwise
{#"Month (Long)"}) , InsertCalendarWeek = Table.AddColumn(InsertShortMonthName, "Week (Year)", each
Week & Number.ToText([Week]) & " " & Number.ToText([Year])), InsertCalendarMonth =
Table.AddColumn(InsertCalendarWeek, "Month (Year)", each {#"Month"} & " " & Number.ToText([Year])),
InsertCalendarQtr = Table.AddColumn(InsertCalendarMonth, "Quarter (Year)", each Quarter &
Number.ToText([Quarter]) & " " & Number.ToText([Year])), InsertCalendarSem =
Table.AddColumn(InsertCalendarQtr, "Semester (Year)", each "S" & Number.ToText([Semester]) & " " &
Number.ToText([Year])), InsertDayWeek = Table.AddColumn(InsertCalendarSem, "Week Day (#)", each
Date.DayOfWeek([Date], P_FirstDayOfWeek) + 1), InsertDayName = Table.AddColumn(InsertDayWeek, "Week
Day", each Date.ToText([Date], "dddd", P_Culture), type text), InsertWeekYear =
Table.AddColumn(InsertDayName, "WeekYearId", each [Year] * 100 + [Week]), InsertMonthYear =
Table.AddColumn(InsertWeekYear, "MonthYearId", each [Year] * 100 + {#"Month (#)"}) , // InsertStartWeek
= Table.AddColumn(InsertWeekYear, "Start of Week", each Date.StartOfWeek([Date], P_FirstDayOfWeek),
type date), // InsertEndWeek = Table.AddColumn(InsertStartWeek, "End of Week", each
Date.EndOfWeek([Date], P_FirstDayOfWeek), type date), InsertQuarterYear =
Table.AddColumn(InsertMonthYear, "QuarterYearId", each [Year] * 100 + [Quarter]), InsertSemesterYear =
Table.AddColumn(InsertQuarterYear, "SemesterYearId", each [Year] * 100 + [Semester]), #Capitalized
Each Word" = Table.TransformColumns(InsertSemesterYear,{{"Month (Long)", Text.Proper}, {"Month",
Text.Proper}, {"Month (Year)", Text.Proper}, {"Week Day", Text.Proper}}, {"Relative (Year)" =
Table.AddColumn("#Capitalized Each Word", "Year (Relative)", each [Year] - Date.Year(P_Today)),
#Relative (Month)" = Table.AddColumn("#Relative (Year)", "Month (Relative)", each {#"Year
(Relative)" * 12 + ([#"Month (#)" - Date.Month(P_Today))}, {"Relative (Week)" =
Table.AddColumn("#Relative (Month)", "Week (Relative)", each
Duration.TotalDays(DateTime.Date(Date.StartOfWeek([Date])) - DateTime.Date(Date.StartOfWeek(P_Today)))
/ 7), {"Relative (Day)" = Table.AddColumn("#Relative (Week)", "Day (Relative)", each
Duration.TotalDays([Date] - DateTime.Date(P_Today))), // MergedHolidays = Table.NestedJoin("#Relative
(Day)",{"Date"},GetHoliday(P_StartDate,P_EndDate,P_Culture,
P_IsCarnivalHoliday),{"Date"},"Holidays",JoinKind.LeftOuter), // ExpandedHolidays =
Table.ExpandTableColumn(MergedHolidays, "Holidays", {"Holiday"}, {"Holiday"}), // AddedWorkDay =
Table.AddColumn(ExpandedHolidays, "Work Day", each if [Holiday] = null then (if {#"Week Day (#)" > 5
then Weekend else WorkDay) else Holiday), #Reordered Columns" = Table.ReorderColumns("#Relative
(Day)", {"Date", "Day", "Week Day (#)", "Week Day", "Week", "Month (Long)", "Month", "Month (#)",
"Quarter", "Semester", "Year", "Week (Year)", "Month (Year)", "Quarter (Year)", "Semester (Year)",
"WeekYearId", "MonthYearId", "QuarterYearId", "SemesterYearId", "Day (Relative)", "Week (Relative)",
"Month (Relative)", "Year (Relative)"}, {"Changed Type" = Table.TransformColumnTypes("#Reordered
Columns", {{"Day", Int64.Type}, {"Week Day (#)", Int64.Type}, {"Week", Int64.Type}, {"Month (#)",
Int64.Type}, {"Quarter", Int64.Type}, {"Semester", Int64.Type}, {"Year", Int64.Type}, {"Week (Year)",
type text}, {"Quarter (Year)", type text}, {"Semester (Year)", type text}, {"WeekYearId", Int64.Type},
{"SemesterYearId", Int64.Type}, {"MonthYearId", Int64.Type}, {"QuarterYearId", Int64.Type}, {"Day
(Relative)", Int64.Type}, {"Month (Relative)", Int64.Type}, {"Year (Relative)", Int64.Type},
{"DateId", Int64.Type}, {"Week (Relative)", Int64.Type}}, ColumnPT = Table.RenameColumns("#Changed
Type", {{"Date", "Data"}, {"DateId", "DataId"}, {"Day", "Dia"}, {"Week Day (#)", "Dia Semana (#)"},
{"Week Day", "Dia Semana"}, {"Week", "Semana"}, {"Month (Long)", "Mês (Extenso)"}, {"Month", "Mês"},
{"Month (#)", "Mês (#)"}, {"Quarter", "Trimestre"}, {"Semester", "Semestre"}, {"Year", "Ano"}, {"Week
(Year)", "Semana (Ano)"}, {"Month (Year)", "Mês (Ano)"}, {"Quarter (Year)", "Trimestre (Ano)"},
{"Semester (Year)", "Semestre (Ano)"}, {"WeekYearId", "SemanaAnoId"}, {"MonthYearId", "MesAnoId"},
{"QuarterYearId", "TrimestreAnoId"}, {"SemesterYearId", "SemestreAnoId"}, {"Day (Relative)", "Dia
(Relativo)"}, {"Month (Relative)", "Mês (Relativo)"}, {"Year (Relative)", "Ano (Relativo)"}, {"Week
(Relative)", "Semana (Relativa)"}}, result = if P_Culture = "pt-PT" then ColumnPT else # "Changed
Type" in result

```

Apêndice B: Questionário de System Usability Scale (Sus) para avaliação das dashboards

Questionário de avaliação System Usability Scale (SUS) para as dashboards desenvolvidas referente ao projeto Karta GPS

Para cada uma das afirmações que se seguem avalie, de forma espontânea, o seu grau de concordância na escala indicada à direita. *

1- Discordo Plenamente, 2- Discordo, 3- Neutro, 4- Concordo, 5- Concordo Plenamente

	1	2	3	4	5
Acho que gostaria de utilizar este sistema com frequência.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Considere o sistema mais complexo do que necessário.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Achei o sistema 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 sistema.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Considere que as várias funcionalidades deste sistema estavam bem integradas.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Achei que este sistema 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 sistema.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Considere o sistema 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 sistema	<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 sistema.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>