

Universidade do Minho
Escola de Engenharia

Sara Maria Martins Faria

Melhoria do processo de migração e digitalização na adoção de um *software* inteligente de gestão da manutenção

Dissertação de Mestrado

Mestrado em Engenharia e Gestão Industrial

Trabalho efetuado sob a orientação do(s)

Paulo Alexandre Costa Araújo Sampaio

Carlos Filipe da Silva Portela

Outubro de 2022

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição

CC BY

<https://creativecommons.org/licenses/by/4.0/>

AGRADECIMENTOS

Agradeço, inicialmente ao professor e coorientador Filipe Portela pelo apoio, pela confiança e pelos conselhos ao longo de todo este processo. Agradeço ao meu orientador, o professor Paulo Sampaio pela sua disponibilidade e por acreditar sempre em mim.

Agradeço à Infraspark por me acolherem tão bem desde o início, por tudo o que me ensinaram e, acima de tudo, por me fazerem entender como deve realmente ser uma empresa.

Um obrigada aos meus amigos porque tornarem sempre o processo mais fácil: à Gui, à Dani, à Telma, à Rita, ao Carlos, à Estefany, à Nádia, ao Tone, à Aninhas, à Ru, à Rita Gomes, à Joana, à Margarida, ao Diogo Silva, ao Diogo Figueiredo, à Catarina, ao Luís Paulo e ao Luís Fernandes e à Catarina Gonçalves. Sem vocês, não teria tido tanta piada e não tinha sido tão bom.

Por último, um obrigada muito especial à minha família pela paciência, pelo amor e pelo apoio incondicional.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

Melhoria do processo de migração e digitalização na adoção de um software inteligente de gestão da manutenção

RESUMO

A tecnologia e o poder da informação orientam, cada vez mais, uma sociedade que por um lado, se encontra em mudanças constantes e que, por outro lado, cresce de forma exponencial devido aos avanços tecnológicos. Esta era é marcada também pelo crescimento da quantidade de dados que uma organização armazena e convertê-los em informações claras e concretas é uma chave fundamental para o sucesso de uma empresa. Assim, estes fatores levam as organizações a adotar estratégias que lhes permitam extrair dos dados, informação relevante que suporte a tomada de decisão para o sucesso do negócio.

A *startup* tecnológica Infraspæk detém e estima deter um aumento do número de clientes e, por isso, um grande volume de dados. Cada vez que a Infraspæk adquire um novo cliente, este passa por um processo de *onboarding*, que pretende melhorar a sua experiência na adoção do produto. Como tal, o principal objetivo da presente dissertação reside no desenvolvimento de um sistema de apoio à tomada de decisão, mais concretamente na conceção de um sistema de *Business Intelligence*, que inclui um *Data Warehouse* para o repositório dos dados. Esta solução aplicada à área de *Customer Success* da Infraspæk pretende agregar os dados relativos aos clientes e proporcionar ferramentas orientadas à tomada de decisão.

Através da metodologia estudo de caso, foi necessário conhecer em profundidade o contexto da empresa, para selecionar e descrever os dados a serem utilizados. Após a recolha desta informação, os dados foram tratados, foi realizada uma análise estatística univariada, implementou-se o *Data Warehouse*, seguida de uma *dashboard* e realizou-se uma análise dos dados obtidos. Através dos dados obtidos, conclui-se que a Infraspæk tem uma média de duração do *onboarding* de 60,41 dias, mas esta média foi variando e variáveis como o segmento da empresa e o tipo de atividades de manutenção que executam, o mercado e/ou país a que pertencem podem influenciar o processo de *onboarding*. A solução implementada permitiu um acesso simples e eficiente dos dados para a decisão.

PALAVRAS-CHAVE

Análise de Dados, *Business Intelligence*, *Customer Success*, *Data Warehouse*, *Onboarding*

Improving the migration and digitisation process in the adoption of an intelligent maintenance management software

ABSTRACT

Technology and the power of information increasingly guide a society that on the one hand is in constant change and on the other hand is growing exponentially due to technological advances. This era is also marked by the growth in the amount of data that an organisation stores and converting it into clear and concrete information is a fundamental key to a company's success. Thus, these factors lead organisations to adopt strategies that allow them to extract from data, relevant information that supports decision making for business success.

The technological startup Infraspak holds and estimates to hold an increase in the number of customers and, therefore, a large volume of data. Each time Infraspak acquires a new customer, this one goes through an onboarding process, which aims to improve their experience in adopting the product. As such, the main objective of this dissertation lies in the development of a decision support system, more specifically in the design of a Business Intelligence system, which includes a Data Warehouse for data repository. This solution applied to Infraspak's Customer Success area intends to aggregate customer data and provide decision-oriented tools.

Through the case study methodology, it was necessary to know in depth the context of the company in order to select and describe the data to be used. After collecting this information, the data was processed, a univariate statistical analysis was carried out, the Data Warehouse was implemented, followed by a dashboard and an analysis of the data obtained was carried out. Through the data obtained, it was concluded that Infraspak has an average onboarding duration of 60.41 days, but this average varied and variables such as the company's segment and the type of maintenance activities they perform, the market and/or country they belong to can influence the onboarding process. The solution implemented allowed simple and efficient access to data for the decision.

KEYWORDS

Business Intelligence, Customer Success, Data Analysis, Data Warehouse, Onboarding

ÍNDICE

1.	Introdução	13
1.1	Enquadramento e Motivação	13
1.2	Apresentação da empresa	14
1.3	Objetivos da Investigação	15
1.4	Estrutura da dissertação	16
2.	Revisão da Literatura	17
2.1	Estratégia de pesquisa	17
2.2	<i>Software-as-a-Service</i>	17
2.3	Sucesso do Cliente	19
2.3.1	Experiência do Cliente	22
2.3.2	<i>Onboarding</i> de Clientes	23
2.4	<i>Business Intelligence</i>	25
2.5	<i>Data Science</i>	27
2.6	<i>Data Warehouse</i>	28
2.6.1	Dimensões e Factos	31
2.6.2	Modelos de <i>Data Warehouse</i>	34
2.6.3	Tipos de Servidores OLAP	37
3.	Métodos e Materiais	38
3.1	Delineamento da pesquisa	39
3.2	Desenho da pesquisa	39
3.3	Preparação e recolha dos dados	40
3.4	Análise dos casos e relação entre eles e elaboração dos relatórios	41
3.5	Matriz de relacionamento das metodologias	42
3.6	Ferramentas	42
3.7	Arquitetura tecnológica	44
4.	Estudo de Caso	45
4.1	Fase 1 Contexto do estudo	45
4.1.1	Caracterização do processo de <i>onboarding</i> da empresa	45

4.1.2	Observação direta do processo de <i>onboarding</i>	49
4.2	Fase 2 Definição e descrição dos dados recolhidos	50
4.3	Fase 3 Preparação e tratamento de dados	61
4.3.1	Limpeza de dados	62
4.3.2	Análise estatística descritiva univariada	63
4.3.3	Implementação do <i>Data Warehouse</i>	68
4.4	Fase 4 Apresentação dos resultados do obtidos	77
4.4.1	Tabelas dinâmicas.....	78
4.4.2	Visualização da Informação em <i>dashboards</i>	83
4.5	Fase 5 Discussão de resultados	86
5.	Conclusões.....	89
5.1	Considerações finais	89
5.2	Limitações.....	91
5.3	Trabalho futuro	92
	Referências Bibliográficas	93
	Apêndice I – <i>Boxplot</i> das variáveis	98
	Apêndice II – Criação das tabelas do modelo físico	100
	Apêndice III – <i>View</i> tempo e <i>View</i> área de negócio.....	108

ÍNDICE DE FIGURAS

Figura 1 - Ciclo de <i>Customer Success</i> . Adaptado de (Nirpaz & Pizarro, 2016).....	20
Figura 2 - Atividades de <i>Customer Success</i> . Adaptado de (Tran, 2020).....	21
Figura 3 - Evolução das diferentes fases da experiência do cliente	22
Figura 4 - Aumento do interesse em customer <i>Onboarding</i> . Fonte: Google Trends	23
Figura 5 - Pirâmide tradicional dos dados de BI para tomar melhores decisões. Adaptado de (Kalathas et al., 2021)	26
Figura 6 - Cubo com 3 dimensões. Retirado de (Alves, 2015).....	31
Figura 7 - Esquema em Estrela	35
Figura 8 - Esquema em Floco de Neve. Retirado de (Han et al., 2012)	36
Figura 9 - Constelação de Factos. Retirado de (Han et al., 2012).....	36
Figura 10 - Fases do estudo de caso. Retirado e adaptado de (Branski et al., 2010).	39
Figura 11 - Atividades para definição e desenho da pesquisa. Retirado de (Branski et al., 2010)	40
Figura 12 - Arquitetura tecnológica do projeto	44
Figura 13 - Etapas do processo de onboarding da Infraspak	46
Figura 14 - Distribuição Normal da Variável <i>Setup Fee</i>	56
Figura 15 - Distribuição Normal da Variável Duração <i>Onboarding</i> (em dias)	58
Figura 16 - Modelo conceptual.....	73
Figura 17 - Modelo lógico.....	74
Figura 18 Passos realizados para a obtenção da <i>dashboard</i>	83
Figura 19 - Exemplo de dashboard desenvolvida em Microsoft Power BI.....	84
Figura 20 - Exemplo da adaptabilidade da <i>dashboard</i>	85
Figura 21 - Aplicação de filtros na dashboard desenvolvida.....	85
Figura 22 - <i>Boxplot</i> da variável <i>Setup Fee</i>	98
Figura 23 - <i>Boxplot</i> da variável Total <i>Technicians</i>	98
Figura 24 - <i>Boxplot</i> da variável Total <i>Admins+ Managers</i>	99
Figura 25 - <i>Boxplot</i> da variável <i>Onboarding Duration</i>	99

ÍNDICE DE TABELAS

Tabela 1 - Matriz de relacionamento das fases da metodologia	42
Tabela 2 - Descrição das ferramentas utilizadas na dissertação	43
Tabela 3 - Enumeração das variáveis utilizadas	51
Tabela 4 - Objetos da Variável Classificação Fase	52
Tabela 5 - Objetos da Variável Classificação CSM <i>Score</i>	53
Tabela 6 - Objetos da Variável Classificação <i>Health Score</i>	54
Tabela 7 - Objetos da Variável Classificação <i>Setup Fee</i>	56
Tabela 8 - Objetos das variáveis Classificação de <i>Technicians</i> e Classificação <i>Admins + Managers</i>	57
Tabela 9 - Objetos da Variável Classificação Duração <i>Onboarding</i>	59
Tabela 10 - Objetos da Variável <i>Customer Classification</i>	59
Tabela 11 - Atribuição de países pela variável <i>Deal Market</i>	60
Tabela 12 - Objetos das variáveis <i>Segment, Industry, Macro-Industry</i> e <i>Deal Type</i>	61
Tabela 13 - Estatística descritiva univariada inicial.....	64
Tabela 14 - Estatística descritiva univariada final	67
Tabela 15 - Descrição dos atributos da tabela de factos	69
Tabela 16 - Descrição dos atributos da dimensão local.....	70
Tabela 17 - Descrição dos atributos da dimensão tempo	71
Tabela 18 - Descrição dos atributos da dimensão área de negócio	72
Tabela 19 - Resultados relativos à média da duração de onboarding consoante o <i>Segment</i>	78
Tabela 20 - Resultados relativos ao número de onboardings que cumpriram a duração esperada divididos por <i>Segment</i>	79
Tabela 21 Resultados relativos à duração média do onboarding consoante a classificação da fase do cliente	79
Tabela 22 - Resultados relativos à média da duração de onboarding consoante a <i>Macro-industry</i> e o <i>Deal Market</i>	80
Tabela 23 - Resultados relativos à média da duração de onboarding consoante o <i>Deal Type</i> e o <i>Deal Market</i>	80
Tabela 24 - Resultados relativos à média da duração de onboarding consoante o <i>Segment</i> , as rondas de investimento e a pandemia Covid-19.....	81

Tabela 25 - Resultados relativos ao número de onboardings consoante a Classificação da Duração do <i>onboarding</i> e a Classificação do <i>Setup Fee</i>	82
Tabela 26 - Resultados relativos à média da duração de <i>onboarding</i> consoante a <i>Health Score</i> e a CSM <i>Score</i>	82
Tabela 27 - Matriz de relação entre Objetivos do Projeto e Resultados Obtidos.....	89
Tabela 28 - Riscos associados à dissertação	92

LISTA DE ABREVIATURAS, SIGLAS E ACRÓNIMOS

BI – *Business intelligence*

B2B - *Business-to-business*

CS - *Customer Success*

CSM - *Customer Success Manager*

DW – *Data Warehouse*

DS - *Data Science*

HOLAP - *Hybrid On-line Analytical Processing*

IaaS - *Infrastructure-as-a-Service*

IT - *Information technology*

KPI - *Key Performance Indicator*

MOLAP - *Multidimensional On-line Analytical Processing*

MRR – *Monthly Recurring Revenue*

NPS – *Net Promoter Score*

OLAP - *On-line Analytical Processing*

PaaS - *Platform-as-a-Service*

PK – *Primary Key*

ROLAP - *Relational On-line Analytical Processing*

SaaS – *Software-as-a-Service*

SQL - *Structured Query Language*

1. INTRODUÇÃO

A presente dissertação enquadra-se no âmbito do Mestrado em Engenharia e Gestão Industrial na Universidade do Minho, tendo sido desenvolvida ao longo de um estágio curricular na Infrasppeak, uma *startup* tecnológica com uma plataforma inteligente de gestão da manutenção.

Este capítulo pretende fornecer um enquadramento do tema, uma visão geral da empresa, refere também os objetivos do trabalho desenvolvido e a motivação, a metodologia de investigação utilizada para a conceção do problema e, por fim, é explicada a estrutura da restante dissertação.

1.1 Enquadramento e Motivação

A tecnologia e a informação fazem parte do dia-a-dia das organizações, o que leva a um maior conhecimento nestas áreas (Alves, 2015).

Atualmente, as empresas deparam-se com desafios na análise dos dados que possuem para gerar conhecimento para suportar os seus negócios. Com a grande quantidade de dados disponíveis, as empresas de quase todas as indústrias focam-se na exploração dos mesmos para obter vantagens competitivas (Fawcett & Provost, 2013).

Por isso, as organizações consideram cada vez mais a utilização de sistemas de *Business Intelligence* (BI) como um ativo, cujo principal objetivo é fornecer acesso à informação de forma atempada, a fim de apoiar o processo de tomada de decisão (R. Teixeira et al., 2014).

Para além disso, recentes avanços tecnológicos permitiram assim uma mudança a nível mundial na venda de *software* em discos físicos para a venda em registo *online* como um serviço. Esta mudança deu origem a um modelo de negócio baseado em assinaturas onde as empresas de *software* oferecem os seus serviços, baseados em planos de pagamento recorrentes. Isto significa que, o valor económico dos clientes é realizado, simultaneamente, durante a relação entre empresas e clientes (Kshatri, 2022).

Consequentemente, uma nova função comercial, denominada Sucesso do Cliente, começou a expandir-se entre as empresas SaaS (*Software-as-a-Service*) que depende da duração da vida útil do cliente e da sua capacidade de maximizá-la (Forrester, 2014).

A Infrasppeak é uma *startup* que se define como uma empresa SaaS e uma das equipas presentes nesta empresa é a equipa de *Customer Success* (CS) que tem a responsabilidade de gerir as relações com os clientes, maximizando a duração de vida útil destes e o valor monetário derivado destes. Por

outras palavras, CS facilita vários processos de sucesso do cliente para alcançar resultados específicos. Esta equipa é também responsável, por numa fase inicial, capacitar o cliente das funcionalidades do *software* que este acabou de adquirir, este processo designa-se de *onboarding*. Com o aumento do número de clientes por parte da Infraspak, tornou-se necessário melhorar este processo, porque muitas vezes os clientes migram do papel ou de uma outra ferramenta para a plataforma Infraspak. Tendo por base o aumento do número de clientes da empresa ao longo do tempo, este valor continuará a aumentar, o que resultará numa exigência constante para a equipa de *Customer Success* que, conseqüentemente, leva a um número elevado de *onboardings* que terá de conduzir. Dada a importância que este processo apresenta, deve manter-se eficaz e com a menor duração possível, no sentido de reduzir custos, garantindo um bom serviço que deixe o cliente satisfeito. Posto isto, é necessária uma solução que ajude na tomada de decisão, o que levou à motivação para desenvolver esta dissertação, melhorar o processo de *onboarding* com o crescimento da Infraspak.

Assim, o tema deste projeto consiste em melhorar este processo através de uma solução de *Business Intelligence*. A presente dissertação visa lidar com informação relativa aos clientes, em que os dados serão recolhidos, tratados e analisados num sistema BI para que seja possível tomar melhores decisões. Para tal, existem sistemas de suporte à decisão como o *Data Warehouse*, sistema esse que será desenvolvido.

A tecnologia tornou-se uma mais-valia na vida das pessoas, por isso usá-la e manipulá-la para conseguir tomar melhores decisões através dos dados que se produzem é outra grande motivação para a realização desta dissertação.

1.2 Apresentação da empresa

A Infraspak é uma *startup* portuguesa, fundada em 2015, com o objetivo de revolucionar e ser a plataforma de referência e líder no mercado da gestão da manutenção de instalações.

A Infraspak tem vindo a crescer através de financiamento. Em 2018, recebeu um investimento no total de 1,6 milhões de euros, proveniente de vários fundadores de unicórnios (conceito criado por Aileen Lee em 2013 que representa uma empresa que atingem uma avaliação de 1.000 milhões de dólares sem ter presença na bolsa de valores (Iberdrola, 2022)) internacionais entre os quais: a britânica Firstminute Capital, a polaca Innovation Nest, a brasileira Construtech Ventures, a norte-americana 500 *Start-ups* e a portuguesa Caixa Capital. Já em 2019, a capital de risco portuguesa Indico Capital Partners investiu 3 milhões de euros. Por último, em 2021, angariou 10 milhões de euros numa ronda de financiamento de Série A que foi coliderada pela Indico Capital e pela Knight

Capital, da Holanda, e que contou também com a participação dos investidores atuais Innovation Nest e Caixa Capital.

Atualmente, a empresa conta com escritórios em Portugal, Reino Unido, Brasil e Espanha e a tecnologia Infraspak já é utilizada em mais de 15 países.

A Infraspak desenvolve e comercializa uma plataforma inteligente de gestão da manutenção, fornecendo um *software* como um serviço a outras empresas. Por conseguinte, opera no setor *Business to Business* (B2B).

O *software* Infraspak recorre à tecnologia *Near Field Communication* (NFC), que associa uma *tag* de um sensor a determinado equipamento/ativo fazendo com que um técnico de manutenção, através de uma interface móvel, consiga visualizar e armazenar toda a informação relativa do mesmo e as manutenções que é ou será sujeito.

A plataforma contempla um conjunto de aplicações principais como a *Dashboards*, Ativos, Trabalhos, Pedidos e Calendário. Simultaneamente, podem ser ativadas outras aplicações, consoante as necessidades do cliente, sendo elas: Gestão de *Stock*, Análises, Compras, Documentos, *Planner*, Requisições, Consumos de edifícios, Vendas, *Housekeeper*, o *Maps*, Análise Económica, Gestão de Custos, Assinaturas, Auditorias, Gestão de Fornecedores, Acordos de Nível de Serviço, Satisfação do Clientes, entre outras.

Deste modo, a Infraspak constrói um sistema operativo flexível e inteligente que facilita a comunicação entre as várias equipas de uma empresa, transformando dados que são recolhidos através das tarefas de manutenção, que ajudarão na tomada de decisão.

Com base na visão e missão da empresa, a Infraspak é movida pela tecnologia e procura acrescentar valor aos seus clientes com recurso à utilização da mesma, de forma simples e promovendo um elevado nível de automatização de tarefas de manutenção, afirmando-se como uma fonte de boa vida para os clientes.

1.3 Objetivos da Investigação

O principal intuito desta dissertação passa por melhorar a experiência global da implementação da plataforma Infraspak pelos clientes, através do desenvolvimento de uma solução de BI.

Por conseguinte e de forma a cumprir o objetivo principal exposto, estabeleceram-se os seguintes objetivos parciais:

1. Realizar revisão bibliográfica sobre o tema abordado;
2. Analisar processo de *onboarding* de novos clientes;

3. Extrair dados quantitativos a partir da análise dos mesmos;
4. Modelar um banco de dados dimensional para armazenamento de dados;
5. Apresentar uma *dashboard*;
6. Avaliar os resultados obtidos;

A questão de investigação formulada, tendo em conta o contexto em que esta dissertação se desenvolve e para atingir os objetivos propostos para a mesma, é definida como:

Quais os fatores que influenciam o processo de onboarding de um novo cliente?

Como resultado, pretende-se obter uma solução que permita a visualização e o acesso a dados de forma simples e acessível que suporte a tomada de decisão.

A metodologia de investigação utilizada na presente dissertação é a abordagem “Estudo de Caso”.

1.4 Estrutura da dissertação

A presente dissertação está organizada em cinco capítulos: Introdução, Revisão da Literatura, Métodos e Materiais, Estudo de Caso e Conclusões.

O primeiro capítulo revela a finalidade da tese, são mencionados e definidos os objetivos e a metodologia abordada e é feita uma breve explicação da empresa onde o projeto foi realizado, sendo descrita a motivação para o mesmo.

No capítulo seguinte, Revisão da Literatura apresenta-se uma base teórica dos tópicos relevantes para a execução deste projeto. Aborda-se, primeiramente, o conceito de *Software-as-a-Service*, seguido de uma revisão sobre Sucesso do Cliente. Termina-se com a revisão de literatura acerca de *Business Intelligence*, *Data Science* e *Data Warehouse*.

De seguida, no capítulo 3 descreve-se a metodologia e as ferramentas utilizadas, bem como é apresentada a arquitetura tecnológica.

No quarto capítulo é aplicada a metodologia anteriormente descrita. A fase inicial deste capítulo tem como finalidade a obtenção de conhecimentos sobre o modo de operação da empresa. Nesta fase, será também apresentado um conjunto de passos para a análise de dados obtidos, a criação de um *Data Warehouse* e de uma interface visual que suporta a visualização de dados e que auxilia a tomada de decisões, seguida de uma discussão de resultados.

Por último, na Conclusão é realizada uma análise geral do trabalho realizado e são apresentadas as principais dificuldades encontradas, apontando o trabalho futuro a prosseguir à investigação.

2. REVISÃO DA LITERATURA

O presente capítulo apresenta os conceitos teóricos abordados ao longo da dissertação. O mesmo encontra-se dividido em cinco temáticas: *Data Science*, *Data Warehouse*, *Business Intelligence*, *Software as a Service* (SaaS) e Sucesso do cliente.

2.1 Estratégia de pesquisa

Na revisão da literatura foram utilizadas várias fontes bibliográficas, com ênfase em artigos científicos, livros, dissertações e páginas *web*. Os motores de pesquisa utilizados foram o ScienceDirect, Scopus, IEEE Xplore, ResearchGate e Google Scholar.

Para obter os resultados pretendidos, os termos mais utilizados na pesquisa foram os seguintes: *Customer Success*, *Customer Onboarding*, *Business Intelligence*, *Data Science*, Análise de dados, *Data Warehouse*, Tabela Dimensões, Tabela de Factos, Esquema em Estrela, Estudo de Caso, Estatística Univariada, Modelo Conceptual, Modelo Lógico, Tabelas dinâmicas e *Power BI*.

A estratégia seguida para a seleção do conteúdo, passou pela leitura dos resumos dos artigos científicos, livros, dissertações de modo a entender se seria ou não relevante. Recorrer, sempre que possível, ao conteúdo mais recente da data de publicação e considerar a relevância científica, tendo sido esta última analisada considerando o número de citações.

2.2 *Software-as-a-Service*

Nos últimos anos, *cloud computing* transformou a indústria *Information Technology* (IT) através da entrega de diferentes serviços através da Internet (Armbrust et al., 2010). A partir desta tecnologia, surgiram três modelos de serviços distintos: *Infrastructure-as-a-Service* (IaaS), *Platform-as-a-Service* (PaaS) e *Software-as-a-Service* (SaaS) (Kshatri, 2022). IaaS diz respeito a serviços relacionados com o *hardware*, tais como servidores, armazenamento e gestão de bases de dados, já PaaS refere-se a *software* e ferramentas de desenvolvimento de produtos. Por último, o modelo de serviço SaaS retrata aplicações completas pré-construídas entregues aos utilizadores finais através da Internet (Kshatri, 2022).

Uma das principais diferenças entre o modelo de negócios SaaS e o modelo tradicional, conhecido como *software on-premises*, é a forma como o *software* é entregue. Enquanto as aplicações SaaS são alojadas pelo fornecedor SaaS e acessíveis a partir de vários dispositivos do cliente através de um

navegador da *Web*, o *software on-premises* é instalado e executado em computadores nas instalações do cliente que utilizam o *software* (Mell & Grance, 2011).

Este novo conceito introduzido no ano de 1999, perturbou a indústria de *softwares* e levou ao licenciamento baseado em assinaturas (Mehta et al., 2016). Em contraste com o modelo clássico de licença de *software* em que o cliente pagava o *software* antecipadamente, o *software* é agora licenciado com base numa assinatura (Bhardwaj et al., 2010). Os clientes pagam uma taxa anual ou mensal, muitas vezes dependente do número de utilizadores que utilizam o *software* (Bhardwaj et al., 2010). Este novo modelo de pagamento é atrativo para os clientes uma vez que elimina os grandes custos iniciais associados ao antigo modelo de licença (Bhardwaj et al., 2010). Também para os fornecedores de SaaS, o modelo de subscrição é vantajoso, uma vez que proporciona um fluxo recorrente de receitas. Isto, por sua vez, dá certeza e previsibilidade ao negócio SaaS, atraindo potenciais investidores (Kshatri, 2022).

Este novo modelo de entrega SaaS apresenta inúmeras vantagens aos clientes, explicando o seu rápido crescimento: uma vez que a infraestrutura de nuvem subjacente é alojada e gerida remotamente por fornecedores SaaS, os clientes não precisam de se preocupar com questões de IT, tais como armazenamento, sistemas operativos, servidores ou redes. Como resultado, os clientes enfrentam custos de infraestrutura de IT mais baixos e maior flexibilidade operacional. Além disso, uma vez que as aplicações SaaS já não são instaladas localmente num computador de secretária do utilizador, podem ser facilmente atualizadas para a versão mais recente, proporcionando assim ao cliente acesso imediato a novas funcionalidades e inovações (Bhardwaj et al., 2010).

A empresa SaaS representa uma entidade que cumpre a definição de empresa digital, com grau elevado de digitalização ao longo da cadeia de valor. Atualmente, a tecnologia da plataforma está disponível para todos e não gera por si só valor, mas a questão é se a empresa tem os recursos e capacidades para alavancar a tecnologia, trazê-la ao mercado e gerar valor para os clientes (Cavusgil & Knight, 2015).

Uma das mudanças mais importantes trazidas pela nova era do SaaS, passa pelo seguinte: "Nos negócios tradicionais, a relação com o cliente termina com a compra. Mas num negócio de assinaturas, a relação com o cliente começa com a compra" (Mehta et al., 2016). Essa mudança prepara o cenário para o surgimento do sucesso do cliente.

A aquisição de novos clientes e o *onboarding* destes estão associados a custos que só podem ser justificados se os clientes continuarem a sua subscrição durante um período significativo. É por isso que Kooij e Pizarro sugerem que os fornecedores SaaS devem estar conscientes de quanto tempo leva para a empresa começar a gerar lucros a partir de novos clientes (Pizarro & Kooij, 2017). A duração do tempo de vida do cliente afeta diretamente as receitas recorrentes que dela derivam (Farris et al., 2010).

2.3 Sucesso do Cliente

Segundo Nick Mehta e Dan Steinman e Lincoln Murphy, o sucesso do cliente é a “organização que se concentra na experiência do cliente com o objetivo de maximizar a retenção e o valor do tempo de vida” (Mehta et al., 2016). A empresa de consultoria baseada na investigação, Forrester, no ano de 2014 defendeu que o papel de *Customer Success* (CS), Sucesso do Cliente, depende da fase de crescimento da empresa e que o seu principal objetivo é levar os seus clientes a utilizarem os produtos que adquiriram com sucesso, assegurando que estão a promover valor económico dos seus investimentos a fim de acumular receitas (Forrester, 2014).

Segundo *The Customer Success Association*, CS define-se como “uma estratégia de negócios de longo prazo, cientificamente projetada e direcionada profissionalmente para maximizar de forma sustentável a lucratividade do cliente e da empresa.” (The Customer Success Association, 2021).

O termo CS tornou-se uma expressão cada vez, mais importante e recorrente no campo SaaS (Tran, 2020). No modelo de negócio SaaS, durante toda a jornada do cliente, deve ser fornecido uma prestação de serviços e, simultaneamente, um envolvimento constante com o cliente. O fim do ciclo de Sucesso do Cliente significa que o cliente decidiu deixar de utilizar o produto adquirido, ou seja, o cliente não renovou ou descontinuou a sua subscrição. A esta ação chama-se *churn*, em português, cancelamento (Nirpaz & Pizarro, 2016). Na Figura 1 está apresentado um esquema do Ciclo de CS de um modelo de negócio baseado em assinaturas, uma adaptação da proposta de Nirpaz e Pizarro (Nirpaz & Pizarro, 2016).

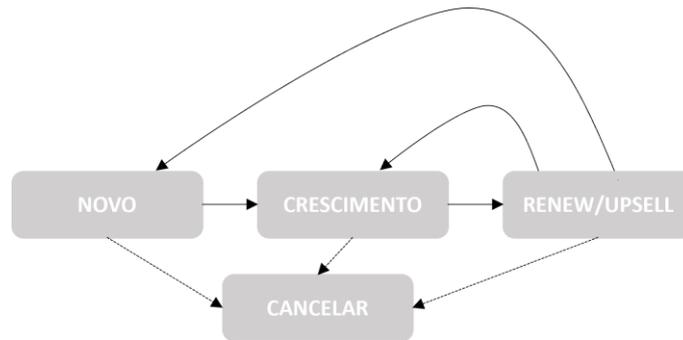


Figura 1 - Ciclo de *Customer Success*. Adaptado de (Nirpaz & Pizarro, 2016)

A etapa “Novo” refere-se à fase em que o cliente subscreveu um novo serviço. Neste ponto, a empresa assegura e conduz um *onboarding*, explicado em *Onboarding de Clientes*, para que o cliente possa adquirir valor o mais rapidamente possível do produto, bem como uma adoção bem-sucedida do mesmo (Tran, 2020).

Na etapa que se segue, “Crescimento”, o cliente deve conseguir perceber o valor da utilização do produto para o seu negócio. Durante esta fase, a empresa implementa várias atividades de envolvimento, com o objetivo de construir uma melhor relação com o seu cliente, aumentando assim a proposta de valor (Tran, 2020).

Durante a fase "*Renew and upsell*", o cliente deve decidir se pretende renovar ou expandir a utilização do produto. Este período costuma ser relativamente curto, mas muito importante, visto que determina as receitas futuras. Esta decisão depende da experiência do cliente na fase de "Crescimento" (Tran, 2020).

Por último, a fase de "Cancelar" indica o fim do ciclo de vida de Sucesso do Cliente. Todas as fases anteriormente designadas do ciclo de vida do CS tem como objetivo retardar a chegada do cliente à fase de cancelar a subscrição (Tran, 2020). O cancelamento não representa, em todos os casos, uma má adoção do produto, pode também ser desencadeada por outros fatores que não estão sob o controlo direto do prestador de serviços (Nirpaz & Pizarro, 2016).

Em cada uma das fases do ciclo de vida do cliente, são realizadas um conjunto de diferentes atividades para maximizar o valor vitalício do cliente. A Figura 2 ilustra, de forma simplificada, as atividades de CS mais comuns (Tran, 2020).

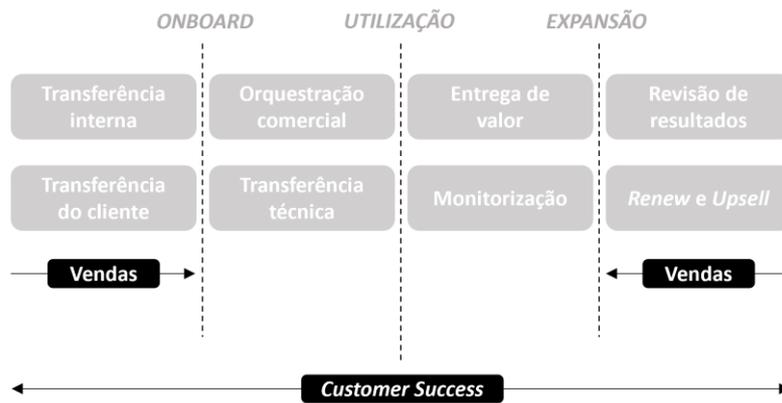


Figura 2 - Atividades de *Customer Success*. Adaptado de (Tran, 2020)

Quando ocorre a subscrição de um novo cliente, representada como *Onboard*, na empresa prestadora de serviços existe uma transferência interna de informação acerca do cliente organizada entre a equipa de vendas e a equipa de CS (Tran, 2020).

O suporte da equipa de sucesso pode, por vezes, ser utilizada como uma ferramenta para influenciar o resultado das negociações de vendas antes de o cliente ter feito uma decisão final, pois é um acréscimo de valor para o cliente. Durante a transferência do cliente, a equipa de vendas passa a responsabilidade e os detalhes de negócio para a equipa de CS, é fornecida uma introdução ao cliente e é realizado um pré-planeamento (Tran, 2020).

Na fase de Utilização, ou seja, *onboarding*, as atividades mais comuns são a orquestração comercial e a transferência técnica, sendo o principal objetivo mapear as ações necessárias para atingir os objetivos esperados e assegurar que a o cliente começa a utilizar o serviço (Tran, 2020).

Na fase seguinte, a empresa corrobora a potencialização de valor para o cliente. Para gerir a jornada do cliente de forma eficaz, a equipa de Sucesso do Cliente controla as taxas de adoção e responde aos alertas dos clientes, através da avaliação das suas métricas (Tran, 2020).

Por fim, os resultados e métricas da performance do cliente são revistos para encorajar a expansão da conta (Pizarro & Kooij, 2017). As operações de *Renew* e *Upsell* podem ser feitas quer pela equipa de Sucesso do Cliente, quer pela equipa de vendas (Tran, 2020).

O modelo descrito, apresenta-se como o mais adequado para empresas que operam no sector B2B (Tran, 2020).

2.3.1 Experiência do Cliente

Lemon e Verhoef definiram a experiência do cliente como um complexo fenómeno que consiste nas reações cognitivas, emocionais, comportamentais, sensoriais, e sociais de um cliente durante a sua jornada de compra (Lemon & Verhoef, 2016). A experiência do cliente tornou-se um elemento crucial nas estratégias empresariais, dado que proporcionar experiências de maior qualidade aos clientes torna-se um fator de diferenciação (Kshatri, 2022).

A interação entre o cliente e o prestador de serviços é um fator decisivo na experiência do cliente. Esta experiência é criada através da comunicação de diferentes canais humanos e não humanos, tornando a jornada sempre interativa (Kshatri, 2022). Este conceito de experiência do cliente não deve ser vago, mas sim algo que possa ser vendido ou integrado na oferta original, tal como qualquer bem ou serviço, visto que é uma característica que acrescenta valor (Kshatri, 2022).

Historicamente, as raízes da investigação da experiência do cliente começaram por volta dos anos 60 e até aos dias de hoje, percorreram diferentes fases centradas em temas diferentes (Lemon & Verhoef, 2016). A Figura 3 apresenta a evolução das diferentes fases da experiência do cliente.

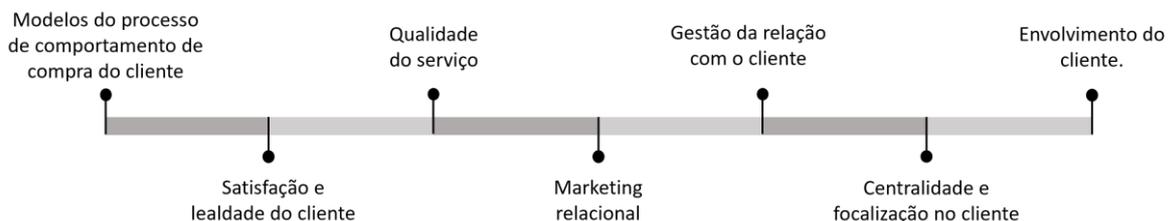


Figura 3 - Evolução das diferentes fases da experiência do cliente

Entre 1960 e 1970, a experiência do cliente é tratada como Modelos do processo de comportamento de compra do cliente, onde compreender a experiência do cliente e a sua tomada de decisão eram vistos como um processo (Kshatri, 2022).

Na década de 1970, o foco passava pela Satisfação e lealdade do cliente. Esta fase passava por perceber como a satisfação do cliente influenciava a sua lealdade e através de avaliações e evoluções das perceções e atitudes do cliente avaliava-se a experiência. Já na década de 1980, a experiência do cliente focava-se na Qualidade do serviço (Kshatri, 2022).

Nos anos de 1990, o *Marketing* relacional ganhou destaque, alargando o âmbito das respostas e *feedback* dos clientes pois estes eram considerados na experiência (Kshatri, 2022).

A Gestão da relação com o cliente surgiu nos anos 2000, onde empresas administravam as suas interações com os clientes utilizando a análise de dados (Kshatri, 2022).

Entre 2000 e 2010, houve uma maior Centralidade e focalização no cliente. Por último, nos anos de 2010, na fase do Envolvimento do cliente, há um reconhecimento do papel do cliente na experiência (Lemon & Verhoef, 2016).

2.3.2 *Onboarding* de Clientes

Atrair novos clientes é um aspeto importante para o sucesso de uma empresa, por isso melhorar a experiência e a jornada do cliente é um aspeto fulcral (Kshatri, 2022). Na Figura 4 é possível observar como o processo de integração de um cliente, começa a ser uma preocupação das empresas (Dilmegani, 2021).

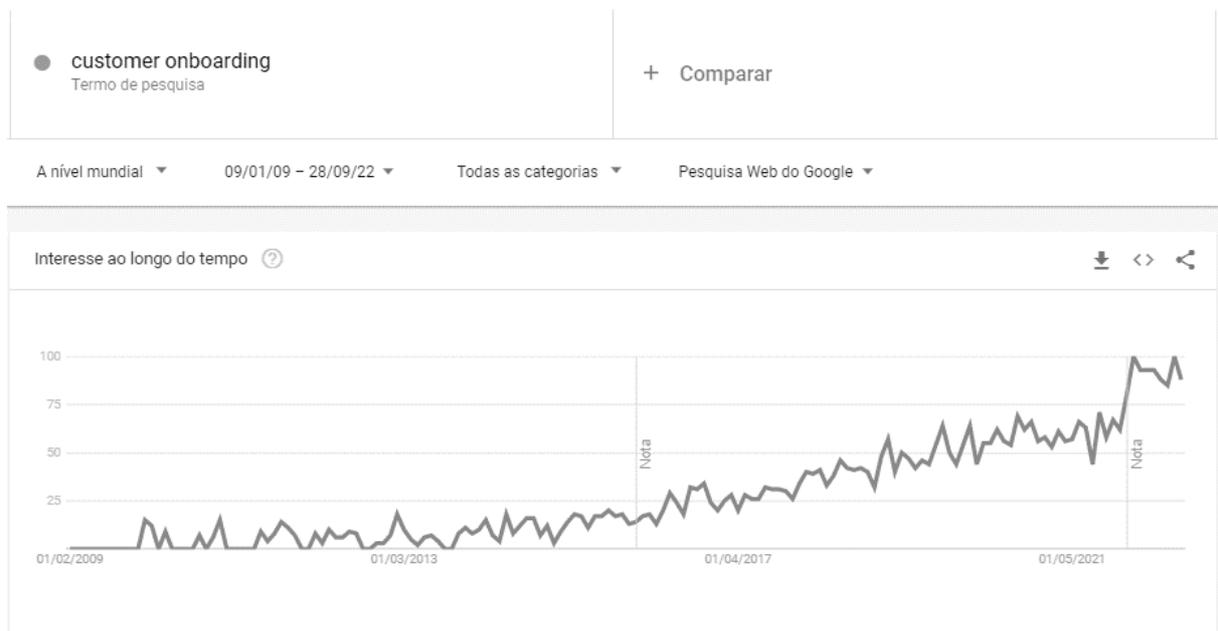


Figura 4 - Aumento do interesse em *customer Onboarding*. Fonte: Google Trends

O *onboarding* de novos clientes diz respeito às ações que uma empresa desencadeia após a compra de um produto por parte do cliente (Kshatri, 2022) . Esta experiência de pós-compra dita a futura relação entre cliente e empresa, uma vez que, uma experiência positiva pode motivar os clientes a tornarem-se clientes fiéis durante um longo período, enquanto uma experiência negativa pode incentivar a optar pela concorrência (Kshatri, 2022).

O termo *onboarding* tem origem na teoria organizacional e refere-se às etapas utilizadas por uma organização para facilitar e integrar da melhor forma um recém-chegado (Sjöberg & Winbäck, 2017). O

processo de *onboarding* pode ser interpretado de várias formas, isto é, como um processo inicial que ajuda os utilizadores a ultrapassar problemas quando inicia, pode ser entendido como uma rápida compreensão do valor do produto por parte dos clientes, ou como uma prática que permite tornar o produto ou serviço o mais fácil possível para os novos clientes, aumentando as probabilidades de os tornar bem-sucedidos (Sjöberg & Winbäck, 2017).

O processo de *onboarding* contempla um conjunto de etapas responsáveis por melhorar a experiência do cliente, que promovam a lealdade e a retenção de clientes. As especificidades do processo de *onboarding* variam em função da natureza de cada produto da empresa, mas em geral, o processo divide-se em 3 fases: acolhimento de novos clientes, definição de expectativas e a formação (Kshatri, 2022).

O acolhimento de novos clientes implica a apresentação de todo o processo de *onboarding* ao cliente, isto é, apresentação do contacto responsável que conduzirá as sessões de *onboarding*, pré-visualização das próximas etapas e uma orientação geral. Durante esta fase, uma transferência bem-sucedida da equipa de vendas para a equipa de sucesso do cliente é de extrema importância, para que CS possa ter uma visão macro do cliente garantindo que os objetivos de implementação estão alinhados com as necessidades comerciais deste (Kshatri, 2022).

A próxima etapa é a definição das expectativas. O *onboarding* deve focar-se, primeiramente, em fornecer valor ao cliente e a construir um plano de sucesso para este alcançar os objetivos delineados à medida que o tempo vai avançando, isto porque é importante perceber as necessidades do cliente e quais os resultados que pretende alcançar. A entrega de valor durante o período de *onboarding*, atenua a relação do cliente com o produto e este também identifica mais rapidamente um maior benefício na sua aquisição. Por outro lado, CS consegue planear os indicadores-chave de desempenho que serão monitorizados ao longo da implementação, consoante o cliente e os seus requisitos (Kshatri, 2022).

Na fase de formação, os clientes são orientados através de um processo de aprendizagem para começarem a experimentar o valor do produto. Para a máxima satisfação, o processo de formação deve ser tão fácil e curto quanto possível. Deve destacar também como o produto ajuda o cliente nas suas necessidades e como este tem impacto nos *Key Performance Indicator* (KPI) que são definidos. Durante esta fase, deve solicitar-se o *feedback* dos clientes sobre o valor que estão a retirar do produto e quaisquer desafios que estejam a enfrentar com a implementação do produto. Esta fase pode também incluir o apoio a clientes na implementação de funcionalidades avançadas ou integrações no produto (Kshatri, 2022).

As empresas podem otimizar sempre o processo de integração dos seus clientes, seguindo cada vez melhores práticas, avaliar procedimentos de *onboarding* de empresas concorrentes e utilizar e monitorizar KPIs para acompanhar o desempenho e performance do *onboarding*.

O *onboarding* de um cliente fornece informações importantes para o *onboarding* de futuros novos clientes, isto porque, este processo de integração pode ser sempre melhorado em torno das expectativas dos clientes (Dilmegani, 2021).

Os benefícios do processo *onboarding* passam por: mostrar os benefícios do produto ou serviço para os clientes, recolher e analisar o *feedback* do cliente sistematicamente, aumentar o uso da plataforma e melhorar a relação cliente-empresa (Dilmegani, 2021).

2.4 Business Intelligence

O conceito de *Business Intelligence* (BI) pode ser definido como um conjunto de metodologias, processos, arquiteturas e tecnologias que convertem dados não tratados em informação útil e com significado, através da identificação, extração, análise e apresentação eficiente, de modo a possibilitar uma tomada de decisão otimizada (Wu et al., 2007). Para gestores executivos tomarem cada vez melhores decisões é utilizado o BI nas organizações (R. Teixeira et al., 2014). Os sistemas BI são dinamizadores da mudança contribuindo para uma boa gestão e tornando a empresa mais ágil num mercado cada vez mais competitivo, sendo, por isso, a análise de dados um ponto chave (Moura, 2017).

BI combina os dados operacionais de uma organização, armazenados em repositórios criados para esse fim, e através da análise, processamento e ferramentas de exploração de dados são manuseados para fornecer informações úteis aos gestores para a tomada de decisão (R. Teixeira et al., 2014). Este conjunto de aplicações de apoio à decisão, permitem um acesso rápido e partilhado dos dados e a sua análise (Moura, 2017).

A importância do BI recai na combinação do processo de entrega de informação com a capacidade que as pessoas apresentam de identificar as necessidades reais de uma organização, uma vez que sem ela, as ferramentas por não têm qualquer valor (R. Teixeira et al., 2014).

Para além do BI possibilitar aos colaboradores de uma organização analisar informações de um negócio com vista a melhorar a qualidade de informação, facilita também a melhoria de processos operacionais (Moura, 2017).

Os sistemas BI apresentam um conjunto de objetivos que refletem as características subjacentes a estes sistemas (Filipe et al., 2013):

- Acesso a dados fiáveis;
- Aumento da transparência e compreensão do negócio;
- Suporte para a tomada de decisão.

Os dados do sistema BI são as pessoas, os processos e os equipamentos, que interagem e colaboram para processar dados e fornecer informações à empresa. Este sistema é constituído por níveis de desenvolvimento constante que compõem uma pirâmide. Na fase inicial, localizam-se os dados históricos brutos e na fase final são tomadas as decisões finais conforme mostrado na Figura 5 (Kalathas et al., 2021).



Figura 5 - Pirâmide tradicional dos dados de BI para tomar melhores decisões. Adaptado de (Kalathas et al., 2021)

Para atingir a fase final, existem processos como preparação dos dados, investigação dos dados, visualização dos dados e *data mining* (Kalathas et al., 2021).

A aplicação de BI permite o controlo do progresso de objetivos delineados por uma empresa, a análise de dados para obter informações sobre o negócio, a criação de relatórios de forma a possibilitar uma gestão estratégica do negócio e, por último, permite a gestão de conhecimento (Bogza & Zaharie, 2008).

O recurso a BI torna-se mais eficiente para tomada de decisões quando se recorre a dados relativos ao mercado onde a empresa opera juntamente com dados internos e para realização deste de análise, em certas ocasiões, recorre-se a um DW para o armazenamento de dados (Aquila et al., 2007).

Uma das áreas de aplicação de BI nas empresas são as vendas e o serviço ao cliente, onde o BI é principalmente utilizado para análises relacionadas com os clientes (Azeroual & Theel, 2018). BI

desempenha um papel importante neste campo pois permite uma melhor compreensão do comportamento do cliente (Wittel, 2017).

A implementação de BI leva ao aumento da eficácia das iniciativas dinamizadas pela empresa com base em ferramentas analíticas, à otimização da produtividade, à otimização de tempo, através da identificação de tendências e padrões leva a análises preditivas e à inovação. Torna também possível conhecer com mais detalhe os clientes da empresa (Wittel, 2017).

2.5 Data Science

Todos os dias são gerados e manipuladas enormes quantidades de dados para se obter informações e extrair valor (Pereira et al., 2020).

Data Science (DS) define-se como um campo interdisciplinar que combina várias áreas, tais como *Computer Science*, *Machine Learning*, Matemática, Estatística, conhecimento de domínio, desenvolvimento de *software* e investigação tradicional (Portela, 2021).

DS é também uma fusão de disciplinas clássicas como estatísticas, *data mining*, bases de dados, e sistemas distribuídos, envolvendo a aplicação de métodos quantitativos e qualitativos para resolver problemas relevantes com o fim de melhorar a tomada de decisão (Pereira et al., 2020).

Através da aplicação de métodos científicos, processos, algoritmos e sistemas para extrair conhecimento e dados estruturados e não estruturados (sendo esta última a base da DS), a DS começa a tornar-se bastante importante e crucial para o processo de tomada de decisão (Portela, 2021). Pela existência de dados cada vez mais heterogêneos e desestruturados, como texto, imagens, vídeos e pelo desenvolvimento rápido da tecnologia, Dhar garante que será cada vez mais comum a tomada de decisão pelos computadores ao invés dos humanos (Dhar, 2012). A tomada de decisão baseada em dados refere-se à prática de tomar decisões baseadas na análise de dados e não na intuição (Fawcett & Provost, 2013).

Um estudo conduzido pelo economista Bryk Brynjolfsson e os seus colegas do MIT e da *Penn's Wharton School* comprovou que, estatisticamente, quanto mais orientada para os dados é uma empresa, mais produtiva ela é (Brynjolfsson et al., 2011).

Todas as tomadas de decisão baseadas em dados referentes aos problemas das empresas são únicas pois dependem dos seus objetivos, desejos e opiniões e através da DS é possível decompor os problemas empresariais para realizar uma melhor análise (Brynjolfsson et al., 2011). Assim, compreender o problema a ser resolvido é bastante importante e esta análise do negócio implica a aplicação de várias tecnologias relacionadas com a análise de dados (Brynjolfsson et al., 2011).

O livro *Data Science for Business*, destaca, assim, a importância de estar familiarizado com técnicas analíticas para melhor compreender o problema, sendo algumas delas: estatística, *Database Querying* e *Data Warehouse* (DW) (Brynjolfsson et al., 2011).

Para a análise de negócios, o termo “estatística” pode ter dois usos distintos. Por um lado, é utilizado para o cálculo de determinados valores numéricos de interesse, como, por exemplo: médias, taxas, mediana, entre outros. Por outro lado, a estatística pode fornecer uma enorme quantidade de dados que ajudam a compreender as diferentes distribuições dos dados, a testar hipóteses e a estimar a incerteza das conclusões (Brynjolfsson et al., 2011).

A consulta de bases de dados (*Database Querying*) é uma técnica analítica que consiste num pedido específico a um conjunto de dados através de uma linguagem técnica, contudo não permite a descoberta de padrões. Estas são indicadas quando já existe uma ideia de qual amostra de dados será interessante para o estudo ou para confirmar uma teoria. O *On-line Analytical Processing* (OLAP), em português Processamento Analítico On-line, ajuda os utilizadores a interagir com um *software* para facilitar a exploração manual ou visual dos dados, clicando, arrastando e manipulando gráficos dinâmicos (Brynjolfsson et al., 2011).

Por fim, o *Data Warehouse* é definido como uma tecnologia facilitadora da prospeção de dados que permite extrai-los de forma ampla e profunda numa organização. Estes armazéns de dados recolhem dados de toda a empresa (Brynjolfsson et al., 2011). Este tipo de técnica permite descrever características individuais de clientes e que características diferenciam os clientes (Brynjolfsson et al., 2011). O próximo capítulo abordará o DW em mais detalhe.

Segundo Foster Provost e Tom Fawcett, técnicas de DS, são utilizadas na relação geral que uma empresa tem com o cliente para analisar o seu comportamento e maximizar o valor esperado por este (Fawcett & Provost, 2013).

2.6 Data Warehouse

A crescente evolução do ambiente de negócios e das grandes empresas acarretam consigo um grande volume de informações e o correto uso desta informação pode ser um fator crucial para destaque num mercado que é cada vez mais competitivo. Uma das ferramentas mais preciosas de uma organização é a sua informação (Kimball & Roos, 2006).

A origem do conceito de *Data Warehouse* (DW) remota ao início da década 1980 quando os sistemas de gestão de base de dados relacionais emergiram como um produto comercial e começou a ser

introduzido em várias áreas, como indústria, apoio ao cliente, telecomunicações, saúde e serviços financeiros (Ferreira et al., 2017)

DW é uma base de dados utilizada para armazenar um grande número de diferentes fontes de dados, atuais e históricos, relacionados com um tema específico, dimensional ou padronizado, permitindo uma análise seletiva dos dados a fim de ajudar gestores na tomada de decisões. Estes sistemas de armazenamento são projetados para dar suporte às atividades de BI e OLAP (R. Teixeira et al., 2014).

Segundo o conceito de Kimball, um DW é uma fonte de dados à qual as pessoas podem aceder para recolher informações que lhe são úteis (Kimball & Roos, 2006). Inmon afirma que um DW é um banco de dados que reúne dados sobre as operações de uma empresa, desde vendas, compras, informações do cliente, extraídos e recolhidos de uma fonte única ou de várias fontes e, transforma-os em informações úteis para a tomada de decisão (Inmon, 2005).

(Alves, 2015) define um DW como “um armazém de dados, um repositório integrado que permite o armazenamento de informação relevante para a tomada de decisão. Estes repositórios podem ser analisados utilizando ferramentas *On-Line Analytical Processing* e/ou ferramentas de *Data Mining*”.

Segundo Inmon, há algumas vantagens no uso de um DW que passam pela qualidade de dados, acesso rápido, facilidade de uso, separação das operações de decisão das operações de produção, valores quantitativos e a segurança (Inmon, 2005).

Um DW deve ser construído para permitir um acesso aos dados históricos de uma organização, uma visão integrada e completa de toda a organização, ter uma fonte de dados plausíveis e facilitar os processos de tomada de decisão, sem sobrecarregar os sistemas operacionais (Moura, 2017).

Um dos precursores do conceito de DW, William Inmon, defende que estes sistemas de gestão de dados são caracterizados segundo quatro condições fundamentais (Inmon, 2005):

- Orientação para um único assunto ou uma determinada área funcional – informação é organizada por assuntos chave;
- Integrado – as diferentes informações são selecionadas e armazenadas no DW provenientes de diversas fontes onde se juntam para fornecer uma visão coerente;

- Não volátil – um DW permite adicionar nova informação, mas não é possível removê-la do mesmo;
- Varia temporalmente – a informação armazenada deve estar catalogada temporalmente através de uma dimensão tempo para que seja possível obter um registo informacional histórico.

Um DW proporciona uma visão integrada e total da empresa, torna a informação atual e histórica facilmente disponível para a tomada de decisões estratégicas, torna a informação da organização consistente, torna possíveis as transações de apoio à decisão sem prejudicar os sistemas operacionais e apresenta uma fonte flexível e interativa de informação estratégica (Ferreira et al., 2017). Um DW representa um ambiente informático onde os utilizadores podem encontrar informação estratégica, e onde são postos em contacto direto com os dados de que necessitam para melhorar decisões (Ferreira et al., 2017).

Os passos para se criar um DW podem ser compostos em três etapas: aquisição e recolha de dados, armazenamento de dados e acesso e análise de dados (Ferreira et al., 2017). Na etapa de aquisição e recolha de dados, os dados são introduzidos ou tratados pela primeira vez e são armazenados em bases de dados operacionais. Antes de os dados serem carregados da base de dados operacional e de fontes externas para o DW, precisam de ser processados através de três funções principais: extração, transformação e carregamento. Na fase de extração, como o nome indica, os dados são extraídos a partir de fontes internas e externas disponíveis, consoante o problema em contexto. Na fase de transformação, pretende-se melhorar a qualidade dos dados extraídos através da correção de inconsistências, inexatidões e valores em falta. Por último, fase de carregamento, os dados são carregados nas tabelas do DW (Ferreira et al., 2017).

Na etapa seguinte, armazenamento de dados, o desenho do DW é baseado no modelo dimensional, existindo dois tipos de tabelas de dados nesta representação: tabelas de dimensão e tabela de factos (Ferreira et al., 2017). No subcapítulo Dimensões e Factos são descritas com maior detalhe.

O acesso a dados da última etapa consiste num conjunto de ferramentas e técnicas que permitem a um utilizador aceder a muitos dados de forma interativa e direta. As melhores técnicas para análise e descoberta de dados são o OLAP e *Data Mining*. O OLAP, como referido em *Data Science*, fornece os meios para explorar e analisar grandes quantidades de dados, envolvendo complexos cálculos, as suas

relações, e apresente visualmente diferentes perspectivas. *Data mining* é o processo de descoberta de contributos significativos, novos padrões, bem como modelos descritivos, compreensíveis, e preditivos de uma grande escala de dados (Ferreira et al., 2017).

2.6.1 Dimensões e Factos

Um DW é composto por tabelas de dimensão e de factos. Aborda-se em seguida estes conceitos, bem como conceitos de cubo OLAP, para clarificar os conceitos anteriores.

a) Cubos OLAP

Um cubo OLAP é uma estrutura de dados multidimensional que permite o acesso e análise de dados de diversas fontes. Os DW têm esta capacidade através dos cubos OLAP (Alves, 2015). A Figura 6 representa um cubo composto por 3 dimensões que lhe confere diferentes perspectivas e formas de visualização da informação (Jensen et al., 2010).

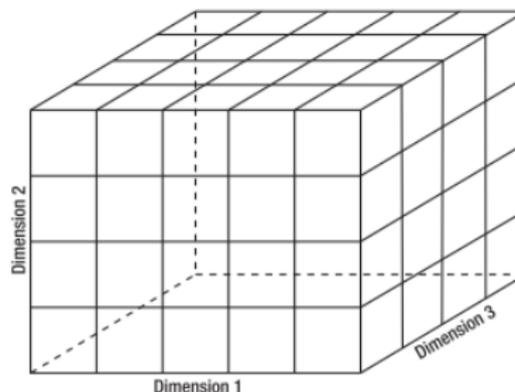


Figura 6 - Cubo com 3 dimensões. Retirado de (Alves, 2015)

O cubo OLAP pode ter mais do que 3 dimensões e através da interseção entre dimensões formam-se diversas células que constituem o cubo, a estas células dá-se o nome de factos (Jensen et al., 2010). Nas dimensões é também possível definir hierarquias para que tudo se conjugue e se obtenha a informação desejada. Esta hierarquia é responsável por conferir uma análise de dados com o nível de detalhes que se deseja (Alves, 2015).

As operações mais comuns efetuadas através da análise do cubo são (Vassiliadis, 1998):

- *Selection/restriction*: esta manipulação restringe os dados.

- *Slice*: esta operação que só pode ocorrer em dimensões, cria um subcubo ao limitar o resultado a um conjunto de valores.
- *Dice*: esta operação origina um novo conjunto de valores, pois realiza uma seleção em duas ou mais dimensões.
- *Drill-down*: este procedimento, que tem como objetivo oferecer uma visão mais detalhada dos dados, permite navegar ao longo de uma dimensão, analisando os valores.
- *Roll-up*: esta ação possibilita a agregação dos dados mais detalhados para os dados mais gerais, através de uma hierarquia.
- *Pivot*: esta manipulação roda os eixos de visualização do cubo, colocando as dimensões sob diferentes orientações, dando um ponto de vista diferente dos dados.

Estas operações descritas podem ser executadas sobre o cubo para analisar os dados em diversas perspetivas.

b) Tabela de Dimensão

A tabela de dimensão alimenta a tabela de factos e os seus componentes descrevem as características de algo tangível. As dimensões são coleções de informações que categorizam e descrevem os factos registados num DW para fornecer respostas significativas, categorizadas e descritivas para perguntas de negócio (Mortier, 2021).

A tabela de dimensão tem como objetivo a descrição dos factos medidos e armazenados nas tabelas de factos. As tabelas de dimensão são altamente desnormalizadas e contêm muitos atributos para se conseguir facultar a maior quantidade de informação sobre os factos recolhidos. As tabelas de dimensão também contêm poucos registos por comparação com as tabelas de factos (Filipe et al., 2013).

Estas tabelas possuem uma chave primária, que pode ser natural ou substituta, e um conjunto de campos denominados atributos, armazenando uma quantidade pequena de dados, em relação à tabela factos, contendo os dados descritivos do negócio (Mortier, 2021). A cada chave primária da tabela dimensão corresponde a uma chave estrangeira na tabela de factos, permitindo a ligação entre ambas

(Alves, 2015). Os atributos das tabelas de dimensão são principalmente textuais ou temporais (Mortier, 2021).

As dimensões de um DW são classificadas de diferentes formas consoante o seu comportamento e uso. Estas podem ser: dimensões conformadas, dimensões de *role-playing*, dimensões de mudança lenta e dimensões *junk* (Mortier, 2021).

Uma dimensão conformada pode ser associada com diferentes tabelas de facto, mantendo o significado com todas elas. Uma dimensão conformada típica é a data, pois esta não varia de acordo com a tabela de factos (Mortier, 2021).

Dimensões de *role-playing* são usadas por diferentes tabelas de factos, assim como dimensões conformadas, todavia estas dimensões têm diferentes significados dependendo da tabela de factos ou até mesmo do campo dentro da tabela de factos (Mortier, 2021).

Dimensões que são suscetíveis a mudanças são chamadas de dimensões de mudança lenta. Aqui, lenta é subjetivo, são dimensões variáveis (Mortier, 2021).

Uma dimensão *junk* reúne combinações possíveis identificando cada combinação com uma única chave primária substituta. Para evitar a criação de pequenas dimensões para cada um desses atributos e aumentar o número e os tamanhos das tabelas de factos desnecessariamente, uma dimensão *junk* é elaborada para reunir todos esses atributos numa única tabela. Neste caso, na tabela de factos, é suficiente incluir uma única chave estrangeira na tabela de dimensão *junk* ao invés de armazenar o valor de cada um dos atributos.

c) Tabela de Facto

Factos são considerados eventos mensuráveis relativos à área funcional coberta por um determinado DW (Mortier, 2021).

A tabela de factos é o local onde são registados os acontecimentos, métricas ou factos que podem ser utilizados para se ter uma perspetiva do processo de negócio. Esta encontra-se ligada às tabelas de dimensões utilizando as suas chaves estrangeiras de forma a poder-se obter informação acerca do facto ocorrido (Filipe et al., 2013).

Nas tabelas de factos, existem dois tipos de atributos: qualitativos e quantitativos. Os atributos qualitativos definem as características de um facto, definidos como uma chave estrangeira para uma tabela de dimensão. Um atributo quantitativo define uma medida de um facto: um valor, uma quantidade, um período ou qualquer outro valor mensurável (numérico) no qual seja possível aplicar cálculos estatísticos como a soma, a média e a variação (Mortier, 2021).

As tabelas de facto, que são o núcleo do modelo, representam um conjunto de fatos (atributos somáveis, não atributos descritivos) em um nível específico de granularidade (detalhe) com o qual a área de negócios se concentra (Tupper, 2011).

As medidas das tabelas de facto podem ser aditivas ou não aditivas, dependendo se podem ou não ser somadas em qualquer dimensão. Estas também podem ser semi-aditivos, o que significa que podem ser somados em algumas das dimensões do DW (Alves, 2015).

2.6.2 Modelos de *Data Warehouse*

A modelação multidimensional é uma técnica usada para projetar bases de dados, desenvolvida para construir DW (Alves, 2015). Este tipo de modelação tem como finalidade desenvolver estruturas de dados fáceis de compreender com capacidade de otimizar o processo de consulta de dados (Alves, 2015).

Desta forma, as tabelas de dimensões e de factos relacionam-se através de um modelo de dados e são apresentadas em forma de esquema (Alves, 2015). A modelação multidimensional pode ser construída através de três tipos de esquema: esquema em estrela, esquema em floco de neve ou um esquema em constelação.

a) Esquema em Estrela

Modelos dimensionais implementados nos sistemas de banco de dados relacionais, são designados esquema em estrela devido à sua semelhança com uma estrutura em estrela (Drkušić, 2016). Este conceito surgiu para atender aos mecanismos de consulta da década de 1980 que exigiam que os dados fossem organizados de maneira a facilitar o processamento rápido (Tupper, 2011).

Com a tabela de factos centrada no esquema, contendo a maior parte dos dados, sem redundância, e as tabelas de dimensão ao redor (Hughes, 2012), estas últimas fornecem a informação necessária

para o cálculo de métricas definidas na tabela de factos (Alves, 2015). A Figura 7 representa um exemplo de um esquema em estrela.



Figura 7 - Esquema em Estrela

O esquema da Figura 7 é constituído por quatro dimensões e uma tabela de factos no centro.

b) Esquema em floco de neve

O esquema em floco de neve é uma variante do modelo esquema em estrela, como ilustrado na Figura 8, a forma do gráfico resultado é semelhante a um floco de neve, daí provém a sua designação (Han et al., 2012) .

A principal diferença entre os modelos esquema em floco de neve e em estrela é que as tabelas de dimensão do modelo em floco de neve podem ser mantidas em formato normalizado para reduzir redundâncias, dissolvendo os dados em tabelas adicionais (Han et al., 2012). Esta nova abordagem permite manter e economizar espaço de armazenamento, contudo pode reduzir a eficácia da navegação, pois mais junções serão necessárias para executar uma consulta. Deste modo, o esquema de floco de neve reduz a redundância, mas o esquema em estrela continua a ser mais popular (Han et al., 2012).

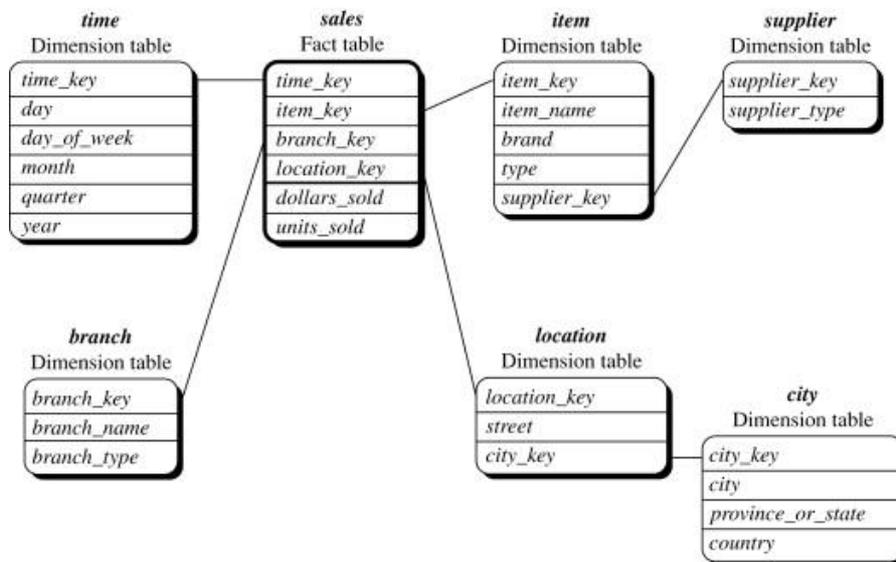


Figura 8 - Esquema em Floco de Neve. Retirado de (Han et al., 2012)

O esquema da Figura 8 apresenta uma maior complexidade em comparação ao esquema apresentado na Figura 7.

c) Constelação de Factos

Em situações mais complexas, pode ser exigido várias tabelas de facto para compartilhar com as tabelas de dimensões (Han et al., 2012). A Figura 9 ilustra um exemplo de um esquema em constelação que pode ser visto como um conjunto integrado de esquemas em estrela (Alves, 2015).

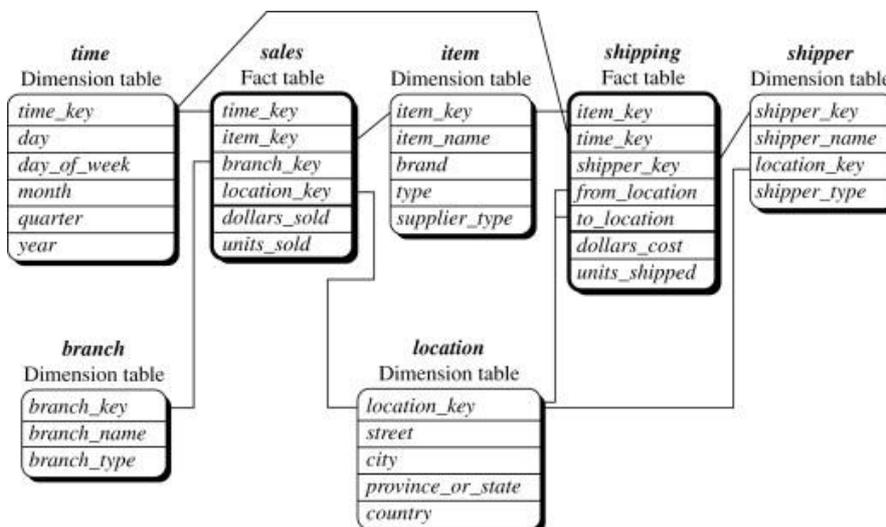


Figura 9 - Constelação de Factos. Retirado de (Han et al., 2012)

Através da Figura 9, observa-se que a tabela *Sales* e *Shipping* estão a interligar as duas estrelas do esquema.

2.6.3 Tipos de Servidores OLAP

Como referido em *Business Intelligence*, OLAP é uma tecnologia que integra BI. As ferramentas OLAP são uma forma interativa de analisar os dados armazenados nas dimensões de um DW para auxiliar os gestores na tomada de decisões estratégicas para o seu negócio (Rainardi, 2008).

Segundo (Rainardi, 2008), os servidores OLAP podem ser divididos em:

- ROLAP (*Relational* OLAP);
- MOLAP (*Multidimensional* OLAP);
- HOLAP (*Hybrid* OLAP).

No caso dos sistemas ROLAP são construídos sobre uma base relacional e realizam a análise de dados. Neste caso, quando os dados armazenados em tabelas podem receber instruções *Structured Query Language* (SQL) e, neste momento, os sistemas OLAP selecionam as tabelas a ser usadas para satisfazer as instruções indicadas (Rainardi, 2008).

Os sistemas MOLAP utilizam bases de dados multidimensionais o que leva a visualizações multidimensionais sobre os dados (Rainardi, 2008).

Por último, os sistemas HOLAP combinam os dois casos anteriores. Contém a alta escalabilidade do ROLAP e a velocidade de processamento do MOLAP (Rainardi, 2008).

3. MÉTODOS E MATERIAIS

A metodologia de investigação utilizada na presente dissertação é a abordagem “Estudo de Caso”, abordagem esta que é bastante benéfica quando existe a necessidade de obter uma compreensão profunda sobre um determinado assunto no seu contexto real (Crowe et al., 2011).

Segundo Robert Yin, a metodologista mais associada à pesquisa de estudo de caso, a conceção de um estudo de caso deve ser considerada quando: (a) o foco do estudo é a obtenção de informação sobre questões mais explicativas do tipo ' como ', 'o que' e ' por que ', como ', (Crowe et al., 2011) b) não manipula o comportamento das pessoas envolvidas no estudo; (c) cobre condições contextuais porque acredita que são relevantes para o fenómeno em estudo; ou (d) os limites não são claros entre o fenómeno e o contexto (Baxter & Jack, 2015).

No caso concreto desta dissertação que pretende explorar quais os fatores que influenciam a duração do processo de *onboarding* de novos clientes na empresa Infraspark, esta investigação não poderia ser considerada sem o contexto específico da empresa em questão. Não seria possível perceber que fatores condicionam esta duração sem considerar o contexto dentro do qual ocorreu.

Esta metodologia é aplicada para maximizar a aprendizagem no período disponível e tende a focar-se numa ou duas questões principais para a perceção do contexto a ser examinado. (Johansson, 2007).

Robert Yin diferenciou também os estudos de caso consoante o seu propósito e as respostas que pretende obter, definindo assim três classes: estudos de caso descritivos, exploratórios e explicativos (Baxter & Jack, 2015). O caso de estudo descritivo tem como principal propósito descrever um fenómeno e o contexto da vida real no qual ocorreu (Baxter & Jack, 2015). O caso de estudo pode ser exploratório se existe pouca informação sobre o problema em questão e, desta forma, procura definir hipóteses, identificar as questões de pesquisa e procedimentos ou adicionar evidências para uma próxima investigação (Drebes Pedron et al., 2008). Por fim, o estudo de caso explicativo procura responder a uma questão construindo ou desenvolvendo uma teoria, apresentando dados que fundamentam a sua análise e a discussão de resultados (Baxter & Jack, 2015).

Assim, uma vez definida a questão de investigação no subcapítulo Objetivos da Investigação e determinado que esta última terá uma melhor resposta utilizando um estudo de caso, passou-se à seleção do seu tipo que irá orientar o objetivo da dissertação. Desta forma, a investigação envolve um estudo descritivo e explicativo.

O método estudo de caso contempla 5 fases, como se verifica na Figura 10: delineamento da pesquisa, desenho da pesquisa, preparação e recolha dos dados, análise dos casos e entre os casos e,

por último a elaboração dos relatórios, sendo que as duas últimas ocorrem em simultâneo (Branski et al., 2010).

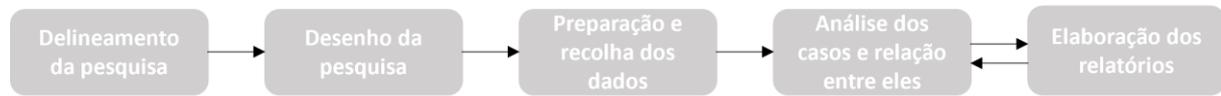


Figura 10 - Fases do estudo de caso. Retirado e adaptado de (Branski et al., 2010).

Apesar da metodologia escolhida ter sido Estudo de Caso, a metodologia Investigação-Ação também se enquadrava no domínio desta dissertação.

3.1 Delineamento da pesquisa

Nesta primeira fase é necessário definir o tema sobre o qual se vai trabalhar, significando isto delimitar, dentro de uma área, o objeto que se pretende estudar (Branski et al., 2010).

Após a definição do tema, são delineados os objetivos e questões da pesquisa que fazem com que o investigador mantenha o foco no objeto de estudo (Branski et al., 2010). Estas questões podem ter como objetivo explorar, se forem do tipo ‘como’ e ‘por que’, como objetivo explicar, ‘como’ e ‘por que’ ou como objetivo descrever, ‘quem’, ‘o que’ ou ‘onde’ (Branski et al., 2010). De seguida, recorre-se a uma revisão da literatura que tem como objetivo apropriar o investigador de conhecimento, ideias e análises importantes que abordará ao longo do estudo de caso, fundamentando o seu trabalho com o conhecimento adquirido de outros autores (Branski et al., 2010).

Neste projeto foi realizada uma Revisão da Literatura que tencionava gerar conhecimento no investigador sobre o atual estado de arte do tópico em foco. Esta abordagem teórica é importante porque permite uma apreciação mais informada sobre os resultados (Crowe et al., 2011). Esta revisão bibliográfica focou-se no sucesso do cliente, no processo de *onboarding* das empresas SaaS e é tida em conta como ponto de partida. Foram utilizadas várias fontes bibliográficas, com ênfase em artigos científicos, a fim de obter resultados valiosos foram utilizadas as seguintes palavras-chave: *Business Intelligence*, *Customer Success*, *Customer Onboarding*, *Data Science* e *Customer Experience*.

3.2 Desenho da pesquisa

Findada a revisão bibliográfica, o investigador deveria possuir as competências necessárias para construir proposições, respostas provisórias às questões da investigação, que guiem o seu estudo de caso e que facilitem a procura de evidências (Branski et al., 2010).

De seguida, a definição e a posterior seleção dos casos implicam que o investigador determine quais os casos que vai abranger pois estes devem apresentar-se vantajosos para o objetivo de investigação e decidir se avança com um único caso ou múltiplos casos (Branski et al., 2010). Nesta parte também é desenvolvido o protocolo que visa orientar o investigador na recolha de dados através de um conjunto de regras que incluem procedimentos e instrumentos de pesquisa (Branski et al., 2010).

Após isto, especialistas e outros investigadores efetuam uma revisão técnica para melhoria da investigação e é então concebido um estudo piloto onde será testado o protocolo de investigação e, por fim, são processados alguns ajustes, resultando assim na metodologia (Branski et al., 2010).

Na Figura 11, encontram-se representadas as atividades para definição e desenho da pesquisa descritas em cima.

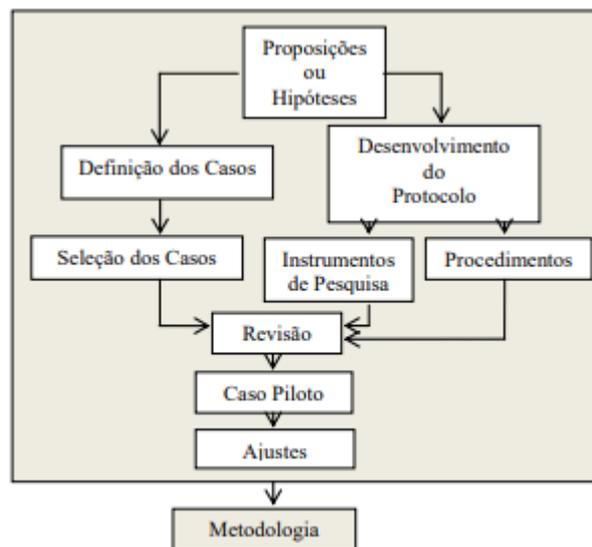


Figura 11 - Atividades para definição e desenho da pesquisa. Retirado de (Branski et al., 2010)

Nesta fase, foram definidas as estratégias de estudo e quais as variáveis e processos a avaliar. Esta é a fase que definiu o resto do trabalho em si e determinou os dados precisos.

3.3 Preparação e recolha dos dados

O início desta etapa passa pela seleção dos casos que devem ser analisados e pela rejeição dos casos que não se encontram qualificados, interessados ou capacitados para o estudo em questão (Branski et al., 2010). Esta terceira etapa da metodologia desdobra-se em três sub-etapas: a preparação, a recolha de dados e a organização do material. Na preparação dos dados, o investigador deve proceder ao levantamento do máximo de informação possível sobre o caso, para dominar o conhecimento do objeto de estudo e para o entrevistado (caso hajam entrevistas ou observação direta) deve fornecer uma carta de

apresentação, resumo da pesquisa e uma cópia da entrevista, para que assim estes se encontrem melhor preparados (Branski et al., 2010).

Já na fase da recolha de dados, pode recorrer-se a entrevistas, estas devem ser gravadas e devidamente documentadas, observação direta e/ou participativa e a dados secundários (Branski et al., 2010). O estudo de caso, normalmente, envolve a recolha de múltiplas fontes, usando técnicas quantitativas e qualitativas e este uso da triangulação de dados tem sido defendido como forma de aumentar a validade interna de um estudo e a credibilidade dos dados (Crowe et al., 2011). Isto acontece porque dados recolhidos de fontes diferentes devem levar a conclusões semelhantes, analisando vários ângulos (Crowe et al., 2011). Esta convergência de dados fortifica as conclusões à medida que as várias vertentes de dados são entrelaçadas para promover uma maior compreensão do caso (Baxter & Jack, 2015).

Por último, deve organizar-se toda a informação recolhida para a constituição da base de dados (Branski et al., 2010). A utilização da base de dados tem como vantagem ter os dados em bruto disponíveis para inspeção independente e melhorar a fiabilidade do estudo de caso, uma vez que permite ao investigador localizar e organizar fontes de dados (Baxter & Jack, 2015).

Neste projeto, os dados foram recolhidos de duas formas distintas: através do acompanhamento de sessões de *onboarding*, ou seja, observação direta e dados relacionados com os clientes.

3.4 Análise dos casos e relação entre eles e elaboração dos relatórios

Estas duas etapas ocorrem simultaneamente. Por um lado, a análise dos dados é uma etapa fundamental e é nesta fase que se seleciona as informações recolhidas, evitando distorções e por outro lado, os dados são organizados e trabalhados de modo a realizar um relatório sobre o mesmo (Branski et al., 2010).

Neste relatório, as evidências devem ser apresentadas de forma neutra e os dados devem estar dispostos de forma a facilitar o estabelecimento de relações e a identificação de padrões entre os casos, confrontando sempre os dados obtidos com a revisão realizada (Branski et al., 2010). Este relato das descobertas deve fornecer ao leitor informações suficientes para compreender os processos seguidos e as conclusões alcançadas e, para além disso, deve garantir o anonimato dos locais ou entidades alocando códigos apropriados (Crowe et al., 2011).

Todo o trabalho desenvolvido é reunido no relatório que, neste caso, é a própria dissertação de mestrado. Nesta elaborou-se uma análise crítica dos dados seguida de uma avaliação, discussão e

conclusão dos resultados, validando ou não as questões de investigação referidas no início da investigação.

3.5 Matriz de relacionamento das metodologias

Nesta dissertação, as fases da metodologia apresentadas anteriormente diferem das fases abordadas na restante dissertação. A Tabela 1 apresenta uma matriz de relacionamento entre estas fases e permite obter uma visão geral da sua relação com a respetiva correspondência.

Recorreu-se a este método uma vez que as fases abordadas ao longo do presente projeto englobam mais informação e detalhe que a metodologia apresentada.

Tabela 1 - Matriz de relacionamento das fases da metodologia

	Delineamento das pesquisas	Desenho da pesquisa	Preparação e recolha dos dados	Análise dos casos e relação entre eles	Elaboração do relatório
Revisão da literatura	x				x
Contexto do estudo			x		x
Definição e descrição dos dados recolhidos		x			x
Preparação e tratamento dos dados				x	x
Apresentação dos resultados				x	x
Discussão dos resultados				x	x

As colunas da Tabela 1 dizem respeito às fases de Estudo de Caso segundo (Branski et al., 2010) e as linhas correspondem às fases efetivamente utilizadas.

3.6 Ferramentas

Para o desenvolvimento desta dissertação foram utilizadas diversas ferramentas. Estas encontram-se descritas na Tabela 2, onde é possível ver a sua descrição e para que motivo foram utilizadas.

Tabela 2 - Descrição das ferramentas utilizadas na dissertação

Ferramentas	Descrição
Infraspeak	Plataforma desenvolvida pela Infraspeak. Esta foi utilizada como caso de estudo pois era neste ambiente que os clientes realizavam o processo de <i>onboarding</i> .
<i>Planhat</i>	Plataforma que permite acompanhar o sucesso de um cliente no uso de SaaS e, por isso fornece dados estruturados e organizados de uma forma clara e definida. Os dados utilizados ao longo de toda a dissertação foram extraídos desta plataforma.
PAST	O PAleontological STatistics (PAST) é um <i>software</i> para análise de dados científicos, onde é possível manipular dados, realizar análise estatística univariada e multivariada, correlação de dados e outras análises (Hammer et al., 2001). Este <i>software</i> foi utilizado para a realização de uma análise estatística univariada.
<i>Microsoft Excel</i>	Programa de <i>software</i> de folha de cálculo, sendo este uma ferramenta avançada para a análise e visualização de dados (Divisi et al., 2017). Esta possibilitou a análise dos dados, com elaboração de gráficos e a criação de tabelas dinâmicas que auxiliaram a análise e apresentação dos dados.
TerraER	Ferramenta projetada para auxiliar na criação de modelos de entidade-relacionamento (H. Rocha & Terra, 2022). Esta foi utilizada para o desenvolvimento do modelo conceptual.
MySQL <i>Workbench</i>	O MySQL é um sistema de gestão de banco de dados, que utiliza a linguagem SQL (<i>Structured Query Language</i>) como interface (Suehring, 2002). Este <i>software</i> foi utilizado para a criação do <i>data warehouse</i> .
<i>Microsoft Power BI</i>	Plataforma que permite a ligação a diferentes dados, transformá-los e visualizá-los. Este modelo de dados permite que a criação de elementos visuais em relatórios para a posterior partilha (Microsoft, 2022b). Esta ferramenta foi utilizada para apresentar os dados de diversas formas em <i>dashboards</i> customizáveis.

De seguida, está descrito como estas ferramentas são aplicadas ao longo da dissertação.

3.7 Arquitetura tecnológica

Neste trabalho, utilizou-se a informação proveniente do *Planhat*, em seguida os dados são selecionados e tratados em *Microsoft Excel*. Nesta fase, também se utilizou o *software* PAST para realizar uma análise descritiva univariada dos dados. De seguida, os dados foram moldados e transformados em TerraER e *MySQL Workbench*, onde ficam organizados em tabelas que contém informação que poderão ajudar na avaliação do processo de *onboarding*. O *MySQL Workbench* desempenha um papel importante pois é nesta plataforma que se tem acesso ao DW para analisar e devolver os dados. Por fim, os dados são devolvidos em formato *Microsoft Power BI* de modo a ajudar a obter respostas a dúvidas de forma visual. Paralelamente, para analisar os dados optou-se por utilizar tabelas dinâmicas para auxiliar nas decisões, utilizando as funcionalidades do *Microsoft Excel*. A Figura 12 ilustra a arquitetura tecnológica desta dissertação.

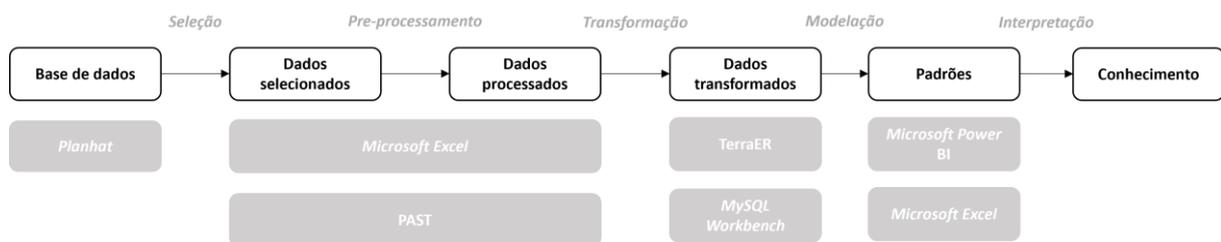


Figura 12 - Arquitetura tecnológica do projeto

Depois de descrita a metodologia, as ferramentas utilizadas na presente dissertação e a arquitetura de todo o projeto, o próximo capítulo demonstra de que forma estas últimas foram aplicadas no contexto do projeto.

4. ESTUDO DE CASO

Neste capítulo descreve-se através da metodologia estudo de caso, todo o trabalho desenvolvido na presente dissertação com o objetivo de compreender quais os fatores que podem influenciar o processo de *onboarding* da Infraspeak, sendo, então, fundamental desenvolver uma ferramenta *Business Intelligence* (BI) capaz de dar resposta às necessidades da organização e que auxilia a tomada de decisões.

Numa primeira fase, é descrito o funcionamento do processo de *onboarding* na empresa em questão, numa fase seguinte e já com o contexto do caso a ser tratado são definidos os dados a serem recolhidos bem como a respetiva explicação e organização dos mesmos.

Na terceira fase, os dados são preparados e tratados para que seja possível criar o *Data Warehouse* (DW). Assim, nesta etapa, estão também descritos os passos para a implementação do DW de forma faseada e segundo o modelo dimensional.

De seguida, são apresentados os dados obtidos através da utilização de tabelas dinâmicas e uma solução BI em formato de *dashboard*. Por fim, realiza-se uma discussão dos dados obtidos.

4.1 Fase 1 | Contexto do estudo

A Infraspeak apresenta um plano de *onboarding* padrão para todos os clientes que é visto como terminado ou completo se aumentar as hipóteses de sucesso de um cliente. De seguida, está descrito de forma detalhada o processo de *onboarding* de um cliente aquando do fecho de uma nova compra da plataforma Infraspeak. Esta descrição diz respeito ao retrato do contexto real do caso de estudo.

4.1.1 Caracterização do processo de *onboarding* da empresa

O processo de *onboarding* da empresa é orientado por um documento denominado *Onboarding Playbook*, este é utilizado para orientar o cliente no projeto de implementação tendo, assim, conhecimento completo da agenda relativamente a todas as sessões a que terá acesso. O *onboarding* de novos clientes contém um conjunto de passos e tarefas, tanto para o CS (*Customer Success*) como para o cliente que passam por definir prazos, audiência e formação. O principal objetivo da empresa no processo de adoção por parte do cliente é criar um plano padrão de *onboarding*, no qual sabem que se for concluído, aumenta as hipóteses de sucesso do cliente, garantindo que este cumpre todos os seus objetivos iniciais e aprende as noções básicas do *software*.

Uma das estratégias utilizadas pela Infraspark é começar de forma simples e aumentar a complexidade ao longo do *onboarding* de modo a ensinar as bases que permitam ao cliente cumprir o resultado desejado o mais cedo possível. A principal finalidade é concentrar-se nos objetivos do cliente e agir de forma a proporcionar valor ao cliente desde o primeiro dia. O CS ao estudar os objetivos do cliente, estabelece um conjunto de primeiros marcos e constroi-se um caminho de ganhos rápidos para os alcançar. Algumas das boas práticas utilizadas pela equipa de CS passam por:

- Fornecer uma agenda de reuniões;
- Ser pontual;
- Enviar um *e-mail* dez minutos antes de uma reunião remota, com o URL da reunião;
- Enviar um *e-mail* de seguimento após cada reunião com uma breve descrição daquilo em que se trabalhou, os próximos passos acordados, e qualquer informação adicional ou material de apoio que se considere útil;
- Registar e guardar todas as anotações da reunião em ferramentas como o Planhat ou Hubspot.

O processo de *onboarding* da Infraspark começa pela passagem de pasta do cliente da equipa de vendas para a equipa de CS com toda a informação relevante recolhida durante o processo de venda via *e-mail*. Após este *e-mail*, o CS fica responsável por contactar o novo cliente, apresentando-se e agendando uma reunião de *kick-off*. Paralelamente, regista o novo cliente nas plataformas utilizadas pela Infraspark para gestão e monitorização da performance do cliente. A Figura 13 ilustra as etapas do processo de *onboarding*.



Figura 13 - Etapas do processo de *onboarding* da Infraspark

O processo *onboarding* é dividido em 6 etapas: reunião de *kick-off* e cinco sessões de trabalho, cada uma das sessões realizadas semanalmente, idealmente. A seguir estão descritas com maior pormenor cada uma das etapas. Perante as necessidades de cada sessão e de cada cliente, é possível adiar sessões e agendar sessões de trabalho intermédias específicas para abordar os problemas encontrados até à data.

a) Reunião de *kick-off*

Esta reunião inicial tem como objetivo fornecer ao cliente uma visão clara e geral do funcionamento do *onboarding* e criar um plano de implementação aprovado por este último. No final desta sessão, o CS deve ser capaz de perceber quais os objetivos do cliente ao adquirir o *software*.

Nesta sessão, designa-se também o gestor interno do projeto, isto é, o principal ponto de contacto do cliente, analisa-se o estado atual de operação da manutenção da empresa cliente, planeiam-se as próximas reuniões e solicita-se um conjunto de informações básicas acerca do funcionamento da empresa do cliente que serão necessárias para submeter na plataforma Infraspak.

A duração desta reunião deve ser entre trinta a quarenta e cinco minutos, consoante a necessidade do cliente, a sua dimensão e a sua experiência.

b) Sessão de trabalho 1: Adicionar utilizadores e localizações

Na segunda sessão de trabalho, o CS é responsável por ensinar aos clientes os termos utilizados na plataforma Infraspak, tais como utilizador (e as suas funções), contacto, especialidade técnica, edifícios e localização e tipo de localização, bem como orientar a criação de cada um dos termos referidos.

Para além disso, o CS também explica e incentiva ao preenchimento da Ficha de Importação de Dados, este é um documento em suporte *Excel* onde o cliente preenche a informação relativa ao seu negócio para que o CS possa submeter na plataforma Infraspak.

c) Sessão de trabalho 2: Avarias, tipos de avarias e reportar e resolver avarias

Após a formação e importação de dados relativos a edifícios e clientes, o cliente recebe instruções sobre o que são avarias, tipos de avarias e como estas funcionam no *software* para que se possa adaptar ao modelo de negócio do cliente. Nesta sessão, o cliente é também convidado a reportar avarias utilizando o ambiente *Web* e a partir dos dispositivos móveis *Android* e/ou *iOS*, os ambientes em que a Infraspak opera.

Por último, também se mostra ao cliente os primeiros relatórios que o *software* é capaz de criar, encorajando-o a realizar análises de manutenção corretivas por conta própria para começar a ficar familiarizado com o *software*.

Assim, no final desta sessão de trabalho, o cliente deve ser capaz de perceber o que são avarias, áreas de avarias e tipos de avaria, como reportar avarias e resolvê-las e é também incentivado a explorar a plataforma para além das sessões de treino.

d) Sessão de trabalho 3: Categorias de manutenção e gestão de equipamentos

Nesta sessão são explicadas as configurações necessárias para iniciar a manutenção preventiva, sendo abordadas as categorias e características de manutenção e como estas se relacionam com os equipamentos sujeitos a manutenção. Também é abordada a forma de adicionar equipamentos.

Desta forma, o cliente deve ser capaz de definir categorias da manutenção e adicionar características, adicionar equipamentos e editar as suas características, saber o que é uma tarefa e uma intervenção e como as relacionar para criar uma manutenção preventiva.

e) Sessão de trabalho 4: Manutenção preventiva e agendamentos

Esta sessão destina-se a importar dados de tarefas e intervenções, abordados anteriormente, e ensinar o cliente a criar ordens de trabalho. São explicadas as diferenças entre ordens de trabalho cíclicas, planeadas e especiais e auditorias e como criá-las. Nesta fase, apresenta-se também o Calendário, o *planner*, os relatórios e como associar estes à manutenção preventiva.

A seguir a esta sessão, o cliente já deve ser capaz de operar todo o seu processo de manutenção na Infraspak.

f) Sessão de trabalho 5: Avaliação global do projeto (recolha de *feedback*)

Esta sessão, apresentada na figura como *Onboard*, é importante para avaliar com o cliente o estado geral da implementação e recolher *feedback* sobre todo o processo de *onboarding*. Com esta sessão pretende-se também perceber se o cliente está satisfeito com a plataforma Infraspak e se está a ajudar no dia-a-dia da empresa. Por último, também é retificado e reforçado algum conceito acerca do *software* que não esteja tão claro para o cliente.

Posto isto, o *onboarding* do novo cliente está completo. Nesta fase, o cliente já deveria conhecer os ensinamentos centrais sobre a Infraspak que lhe permitem cumprir os seus objetivos iniciais.

g) Considerações finais

No final, é também enviado ao cliente um formulário para recolha de *feedback*. A intenção é compreender como os clientes percecionam o processo de *onboarding* e compreender como é possível continuar a melhorá-lo. O principal objetivo é maximizar a experiência do cliente ao longo da jornada de *onboarding* e os dados recolhidos neste inquérito são partilhados trimestralmente com a equipa de CS, e o inquérito será revisto e melhorado no mesmo momento.

No geral, todas as sessões são bastante flexíveis, isto é, perante a curva de aprendizagem do cliente e as suas necessidades, o CS pode adaptar as sessões para irem de encontro ao sucesso definido inicialmente pelo cliente. Uma das vantagens do *onboarding* da Infraspak passa pela postura do CS

que durante as sessões de *onboarding* tenta compreender os processos e os maiores problemas do cliente e atuar sobre eles.

Como referido anteriormente, na sessão de *kick-off* do *onboarding* é definido um plano de sucesso tendo em conta os objetivos descritos pelo cliente. Assim, o plano de sucesso é um documento partilhado entre a Infraspark e o cliente, com o plano de implementação da plataforma Infraspark que deve fornecer uma visão macro de um caminho de sucesso, com datas e objetivos alinhados com os resultados desejados pelo cliente. Este documento deve ser revisitado e alterado sempre que justificado pelo CS e pelo cliente, sendo que este se encontra dividido em três fases distintas: *Onboarding*, Adoção e Expansão.

Assim, após a fase *Onboarding* estar completa, o cliente passa pela fase de adoção e pela fase de expansão.

Durante a fase de adoção, espera-se que a plataforma seja adotada a 100% por toda a organização da empresa cliente, e alguns conceitos e formação dada durante o *onboarding* serão aprofundados.

Já a expansão será o período em que novos objetivos serão acrescentados e coincidirão com a utilização de novas funcionalidades ou a inclusão de novas equipas que utilizem a plataforma. É uma fase que não tem fim uma vez que, os novos objetivos e desafios serão sempre avaliados e sugeridos tanto pelo cliente como pela equipa de CS. Esta fase será marcada por reuniões de revisão de sucesso, onde serão feitos pontos da situação global com o cliente. É a partir da fase de adoção que o cliente é transferido para um CSM (*Customer Success Manager*) que será responsável por trabalhar no sucesso do cliente.

4.1.2 Observação direta do processo de *onboarding*

Para compreender o processo pelo qual um novo cliente passa, as principais dificuldades sentidas, a abordagem utilizada pelo CS e como este desempenha o seu papel na melhoria da experiência do cliente e os processos que segue e como se organiza, foram acompanhadas sessões de *onboarding* de novos clientes. Esta abordagem permitiu ao investigador um maior domínio do conhecimento que influenciou o desenrolar do restante trabalho.

Nesta observação direta, foram acompanhados catorze processos de *onboarding* de empresas distintas e as anotações do processo de *onboarding* foram registadas num diário de bordo. Desta amostra, oito empresas eram do mercado português, quatro do mercado do britânico, uma do mercado brasileiro e uma do mercado espanhol e atuam em diversos ramos, desde o ramo de saúde, hospitalidade, indústria e fornecedores de manutenção de equipamentos, sendo assim, uma amostra diversificada.

Do acompanhamento das sessões, é possível afirmar que os clientes adquirem a Infraspark com um dos seguintes objetivos:

- Melhorar a gestão total da manutenção;
- Melhorar a gestão de informação da empresa;
- Aumentar a perceção geral do trabalho realizado;
- Aumentar o controlo do trabalho realizado;
- Aumentar fluidez dos processos;
- Aumentar rastreabilidade dos pedidos de manutenção;
- Aumentar produtividade;
- Migrar do papel e/ou *Microsoft Excel*;
- Culminar a falta de informação, isto é, plataforma que reúna toda a informação.

Por outro lado, as principais dificuldades encontradas nesta amostra foram as seguintes:

- Falta de tempo por parte da empresa para preenchimento da Ficha de Importação de Dados;
- Falta de informação por parte da empresa para submeter no *software*;
- Adaptação dos termos utilizados no *software* Infraspark para os termos utilizados por cada empresa;
- Compreensão da finalidade das diferentes *apps* da Infraspark;
- Compreensão das sessões de trabalho 1 e 2, nomeadamente distinção entre área e tipo de avaria, associação de um cliente a um edifício;
- Parametrização geral de dados da empresa no *software*.

No geral, o nível de conhecimento tecnológico e de manutenção que o gestor interno do projeto apresenta também influencia a dinâmica destas sessões, pois um maior conhecimento destas áreas corresponde a uma melhor adoção da plataforma. O nível de trabalho da empresa também influencia o processo de *onboarding*, uma vez que muitas reuniões foram adiadas por falta de tempo e, por último, a falta de uma visão clara por parte da empresa cliente do que pretende do *software* também pode ser um entrave a este processo.

4.2 Fase 2 | Definição e descrição dos dados recolhidos

Os dados relativos ao processo de *onboarding* foram recolhidos na plataforma *PlanHat*, tendo estes sido exportados em formato *Microsoft Excel* e são referentes desde janeiro de 2015 até à data de 11 de abril de 2022, inclusive. Estes dados relevam informação sobre todas as empresas clientes da

Infraspeak e funcionam como base de dados para o desenvolvimento do restante trabalho. A base de dados inicialmente exportada continha 559 linhas, o que representa uma base de dados de 559 empresas cliente e 137 colunas, que correspondem a 137 variáveis de dados recolhidos dos clientes.

Após uma maior compreensão do contexto do projeto, as variáveis foram reduzidas a 25 variáveis, que serão analisadas de forma a responder à questão de investigação desta dissertação definida nos Objetivos da Investigação na Introdução, isto é, “*Quais os fatores que influenciam o processo de um onboarding de um novo cliente?*”. Algumas das variáveis sofreram agrupamentos para que a análise dos dados fosse mais simples e acessível. Na Tabela 3 encontram-se enumeradas as variáveis utilizadas ao longo do projeto.

Tabela 3 - Enumeração das variáveis utilizadas

Variáveis	
Entidade	<i>Phase</i>
<i>Country</i>	<i>City</i>
<i>CSM Score</i>	NPS
<i>Health Score</i>	MRR
<i>Setup Fee</i>	<i>Customer since</i>
<i>Total Technicians</i>	<i>Total Admins + Managers</i>
<i>Loving Customer</i>	<i>Kick-off date</i>
<i>Expected Onboarding date</i>	<i>Onboarded date</i>
<i>Onboarding Duration</i>	<i>Alignment with Product vision</i>
<i>Customer Classification</i>	<i>Apps Installed</i>
<i>Deal Market</i>	<i>Segment</i>
<i>Industry</i>	<i>Macro-industry</i>
<i>Deal type</i>	

Em seguida, estão apresentadas e detalhadas as variáveis que serão analisadas e mencionadas ao longo da dissertação.

A variável **“Entidade”** resulta da substituição do nome das empresas clientes da Infraspark por Entidade seguida de uma numeração incremental, isto deve-se ao Regulamento Geral sobre a Proteção de Dados (Diário da República Eletrónico, 2019) para assim proteger as métricas e dados de cada empresa. Para além disso, a abordagem estudo de caso exige tal anonimato, como referido anteriormente. Este processo foi realizado através da função Procv do *Microsoft Excel*.

Dentro da variável **“Phase”**, em português Fase, foi realizado um agrupamento de dados semelhantes através do domínio do conhecimento adquirido na fase de descoberta do processo de *onboarding*. Esta variável caracteriza os clientes como: Em risco, Cancelado, Fechado, Expansão, *Onboard*, *Onboarding*, Piloto e Renovado. Desta forma, criou-se três classes que agrupam os dados em Cliente, Não Cliente e Cliente em risco, numa nova variável designada **“Classificação Fase”**. Na Tabela 4 encontra-se a distribuição dos objetos descritos pela nova variável.

Tabela 4 - Objetos da Variável Classificação Fase

Classificação Fase	Descrição
Cliente	Cliente em expansão, <i>onboard</i> , em <i>onboarding</i> , cliente piloto e renovado.
Não Cliente	Cliente cancelado e fechado.
Cliente em risco	Cliente em risco

A variável que se segue **“Country”**, em português País, corresponde ao país em que cada empresa cliente está sediada. A partir desta informação, gerou-se a variável **“País por Continente”** que associa a cada país o seu respetivo continente: Europa, África, América do Norte, América Central, América do Sul, Ásia. Esta associação foi realizada através da importação de um ficheiro CSV que continha todos os países existentes ligados ao respetivo continente. Após isso, voltou a recorrer-se à função do *Microsoft Excel/Procv* para correlacionar o ficheiro com a nova variável.

Relativamente à variável **“City”**, em português Cidade, esta reflete a cidade onde a empresa cliente se encontra instalada.

Para a variável **“CSM Score”**, recorreu-se, novamente, ao agrupamento de dados similares, para que seja mais eficiente e eficaz descrever cada um dos grupos identificados. A *CSM Score* é uma pontuação atribuída por cada CS ao cliente e esta traduz o sentimento que o CS tem perante o cliente,

isto é, se este é organizado, proativo, nível e curva de aprendizagem, sendo, por conseguinte, bastante subjetiva. Esta pontuação mensal é escalada de 1 a 5, onde o 1 expressa o baixo potencial do cliente no uso da Infraspeak e o 5 expressa o máximo de potencial do cliente. A Tabela 5 ilustra a distribuição da pontuação da CSM *Score* pelo novo agrupamento.

Tabela 5 - Objetos da Variável Classificação CSM *Score*

Classificação CSM <i>Score</i>	Descrição
CSM Baixo	CS atribui uma pontuação de 1 ou 2 ao cliente.
CSM Médio	CS atribui uma pontuação de 3 ao cliente.
CSM Alto	CS atribui uma pontuação de 4 ou 5 ao cliente.

Outra métrica amplamente utilizada como indicador de desempenho é o *Net Promoter Score* (NPS) (Lewis & Mehmet, 2020) sendo uma variável explorada nesta investigação. Esta pontuação de 0 a 10 manifesta quão provável seria um cliente recomendar a Infraspeak a um amigo, parceiro ou colega, no qual 0 é certo que não recomendaria e 10 muito provável de recomendar. O NPS é calculado da seguinte forma (Lewis & Mehmet, 2020):

$$NPS = \frac{(Promotores - Detratores)}{(Promotores + Detratores + Passivos)}$$

Assim, as respostas dos clientes são classificadas em promotores (9-10), passivos (7-8) e detratores (0-6) (Lewis & Mehmet, 2020). A partir desta classificação, foi criada uma variável “**Classificação NPS**” que tem como objetos: Promotores, Passivos e Detratores.

No que diz respeito a métricas, a equipa de CS também dispõe da “**Health Score**”, uma métrica importante que determina se os clientes estão saudáveis ou em risco (Pendo.io, 2022). Esta métrica trimestral combina vários tipos de informação que é aglomerada num único algoritmo, desde o nível de atividade na plataforma, NPS, CSM *Score* ao último contacto com o cliente. Toda esta informação é convertida num único número de 3 a 10 que permite classificar a saúde do cliente. Portanto, com o objetivo de reduzir o número de objetos desta métrica contínua, organizou-se a *Health Score* dos clientes numa nova classe “**Classificação Health Score**”. A Tabela 6 apresenta a descrição dos objetos da nova variável gerada.

Tabela 6 - Objetos da Variável Classificação *Health Score*

Classificação <i>Health</i>	
Score	Descrição
<i>Health Score</i> baixa	Corresponde a pontuação menor que 4, significando que a empresa está em risco e, muito provavelmente, irá cancelar a subscrição da Infraspeak.
<i>Health Score</i> média	Corresponde a uma pontuação entre 5 e 7, inclusive, e reflete um uso da plataforma adequado, mas não no seu potencial máximo ou com algumas dificuldades.
<i>Health Score</i> alta	Corresponde a uma pontuação maior que 8 e traduz um uso muito confortável da plataforma.

Esta nova classificação foi definida com base no conhecimento do processo.

A variável “**Monthly Recurring Revenue (MRR)**”, em português Receita Recorrente Mensal, corresponde, de forma simplificada, ao cálculo das receitas geradas num mês de uma empresa cliente, em euros. Esta métrica é importante para as empresas SaaS visto que fornece uma métrica da saúde da empresa do ponto de vista das receitas (ProductPlan, 2022). Como está associada a valores contínuos, foi feita uma análise logarítmica aos dados para a construção de classes. De modo a proteger os valores de faturação da empresa, nesta variável não são apresentados valores.

Inicialmente, foi calculada o intervalo dos valores, *Int*, isto é, a diferença entre o maior e o menor valor de MRR, sendo a fórmula:

$$Int = \text{Maior valor} - \text{Menor valor}$$

Posto isto, para o cálculo do número de classes, *K*, nas quais os dados serão agrupados, utiliza-se a seguinte fórmula para determinar:

$$K = 1 + 3,322 \log N$$

Sendo N o número total de observações, que no caso é igual a 559. Assim sendo, $K \approx 10,1269$ o que revela que existirão 10 classes. O próximo passo é determinar o tamanho aproximado do intervalo da classe, h, obtido da seguinte forma:

$$h = \frac{Int}{K}$$

Obtendo assim, um tamanho de intervalo de classe. No caso de resultados fracionários, o próximo número inteiro mais alto é considerado o tamanho do intervalo de classe, nesta situação o intervalo da classe é de 1220 (Learn Kraft, 2022).

Após a aplicação da função logarítmica, foi possível perceber que não se adequa ao problema, visto que 94% dos valores de MRR se encontravam na 1.^a classe identificada, assim esta classe é tendencial em virtude dos extremos dos dados.

Testou-se, desta forma, a Distribuição normal, para a criação de classes. Através das fórmulas do *Microsoft Excel*, calculou-se o valor da média e do desvio padrão. Aplicando a fórmula do *Microsoft Excel* da distribuição normal, isto é, DIST.NORMAL(x; média; desvio padrão; cumulativo) (Microsoft, 2022), onde o x representa o valor cuja distribuição pretende obter e o “cumulativo” é um valor lógico que determina a forma da função, que neste caso, o cumulativo é “VERDADEIRO” e assim devolve a função de distribuição cumulativa. Após isto, elaborou-se um gráfico a partir da função aplicada, obtendo um gráfico da Distribuição Normal da Variável MRR.

A partir da análise deste gráfico, conclui-se que não existe um padrão, por isso, a agrupamento da variável MRR será feito com base no domínio do conhecimento. Tendo a Infraspark metas específicas para o MRR que gostaria de alcançar, a criação das novas classes baseia-se nisso. A nova classificação não é representada neste projeto por questões de anonimato referidas anteriormente.

A variável “**Setup Fee**” diz respeito ao orçamento inicial que um cliente tem de pagar à Infraspark, em euros, para esta fazer a configuração da conta e pelas sessões de *onboarding*. Recorrendo à metodologia abordada na variável “MRR”, começou-se por criar classes a partir da função logarítmica, todavia esta também não funcionou, dado que 89,64% dos dados encontravam-se na primeira classe, no caso entre 0 e 3 210. Testou-se a criação de classes através da Distribuição Normal, contudo não existe um padrão que permitisse avançar com esse método, como mostra a Figura 14.

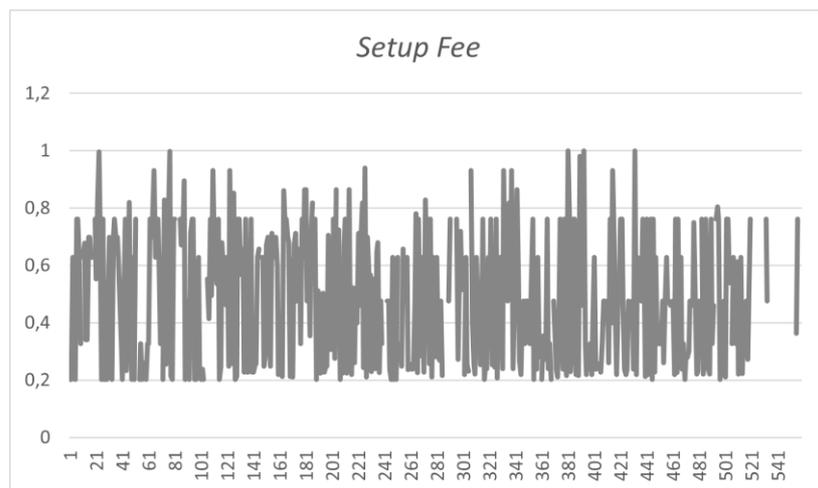


Figura 14 - Distribuição Normal da Variável *Setup Fee*

Como a empresa oferece um conjunto de pacotes que o cliente pode escolher para esta configuração inicial, estes servirão de base para a classificação da nova variável “**Classificação Setup Fee**”. A Tabela 7 representa a distribuição de valores dos objetos desta nova variável.

Tabela 7 - Objetos da Variável Classificação *Setup Fee*

Classificação Setup	
Fee	Descrição
<i>Remote Basic</i>	Contempla valores entre 0€ a 2000€.
<i>Remote Premium</i>	Contempla valores entre 2000€ e 4000€.
<i>Onsite Basic</i>	Contempla valores entre 2000€ e 4000€.
<i>Onsite Premium</i>	Contempla valores superiores a 5000€.

Uma das outras variáveis utilizadas é a “**Customer since**”, em português Cliente Desde, que representa a data, em formato aaaa-mm-dd, em que os clientes fecharam negócio com a Infraspak. A partir desta variável, gerou-se 3 novas variáveis:

- “**Customer Since Ano**” que identifica o respetivo ano em que a empresa fechou negócio;
- “**Customer Since (Covid-19)**”, variável que agrupa os negócios que foram fechados antes e após pandemia, como referência foi utilizada a data 2020-03-10 uma vez que foi o primeiro dia de teletrabalho na empresa Infraspak.
- “**Customer Since (rondas de investimento)**” variável que agrupa o fecho de um cliente de acordo com as rondas de investimento que a empresa teve, isto é, se o cliente fechou

negócio até à 1.^a ronda de investimento, entre a 1.^a e a 2.^a ronda, entre a 2.^a e a 3.^a ronda ou a partir da 3.^a ronda de investimento.

As próximas variáveis **“Total Technicians”**, em português Total Técnicos, e **“Total Admins + Managers”**, em português Total Admins + Gestores, referem-se ao tipo de conta do sistema da plataforma Infraspark que o cliente comprou, isto é, no ato de compra o cliente adquire contas para os técnicos (*technicians*), as mais baratas, e contas para os gestores e administradores da plataforma (*admins+managers*), contas mais caras. Novamente, recorreu-se aos pacotes que a Infraspark comercializa, referidos na variável **“Setup Fee”**, que incluem também os preços das contas de sistema, e formaram-se duas novas variáveis **“Classificação Technicians”** e **“Classificação Admins + Managers”**. Assim, e segundo o pacote de preços, os objetos da primeira variável criada são: menos de 10 técnicos, mais de 10 técnicos, mais de 25 técnicos e mais de 50 técnicos, classificando assim o número de contas que o cliente comprou. A segunda variável criada, comporta-se de forma semelhante, porém diz respeito às contas de administrador e gestor ao invés de contas de técnico. Na Tabela 8 encontram-se os objetos das variáveis anteriormente descritas.

Tabela 8 - Objetos das variáveis Classificação de *Technicians* e Classificação *Admins + Managers*

Classificação <i>Technicians</i>	Classificação Admins+Managers
Menos de 10 técnicos	Menos de 10 administrados + gestores
Mais de 10 técnicos	Mais de 10 administrados + gestores
Mais de 25 técnicos	Mais de 25 administrados + gestores
Mais de 50 técnicos	Mais de 50 administrados + gestores

A variável que se segue denomina-se **“Loving Customer”** e representa uma designação criada pela Infraspark para diferenciar os clientes mais ativos e com maior utilização da plataforma, esta depende de valores como *Health Score*, NPS e número de *apps* e *add-ons* instalados. Se os clientes cumprirem os requisitos para serem *loving customers* é-lhes atribuído o valor “Verdadeiro”, caso não o seja, é atribuído o valor “Falso”.

As próximas variáveis **“Kick-off date”**, **“Expected Onboarding date”** e **“Onboarded date”** apresentadas também no formato *aaaa-mm-dd*, correspondem, respetivamente, à data em que o cliente iniciou o seu *onboarding*, a data expectável de término definida pela equipa de CS e a data em que terminou efetivamente. A partir destas duas últimas variáveis, gerou-se a variável **“Expected date**

(Sim/Não)” que diferencia aqueles clientes que efetuaram o *onboarding* no tempo esperado dos que não atingiram essa meta. Os objetos desta última são: "Cumpriu expectável", "Não cumpriu o expectável" e "Sem dados".

A variável seguinte **“Onboarding Duration”** evidencia a duração, em dias, que cada cliente demorou a realizar as suas sessões de *onboarding* e, baseada nesta, originou-se a variável **“Classificação Duração Onboarding”** que pretende agrupar as durações mais similares em classes para que a análise futura se torne mais fácil. Primeiramente, foi realizada uma análise logarítmica para a construção de classes, todavia esta não é válida porque 64,4% dos valores se situavam na 1.ª classe, que no caso era de 0 a 66. Posto isto, foi usada a Distribuição Normal e o resultado encontra-se na Figura 15. Como se pode observar não há um padrão facilmente identificável, por isso não é uma abordagem válida.

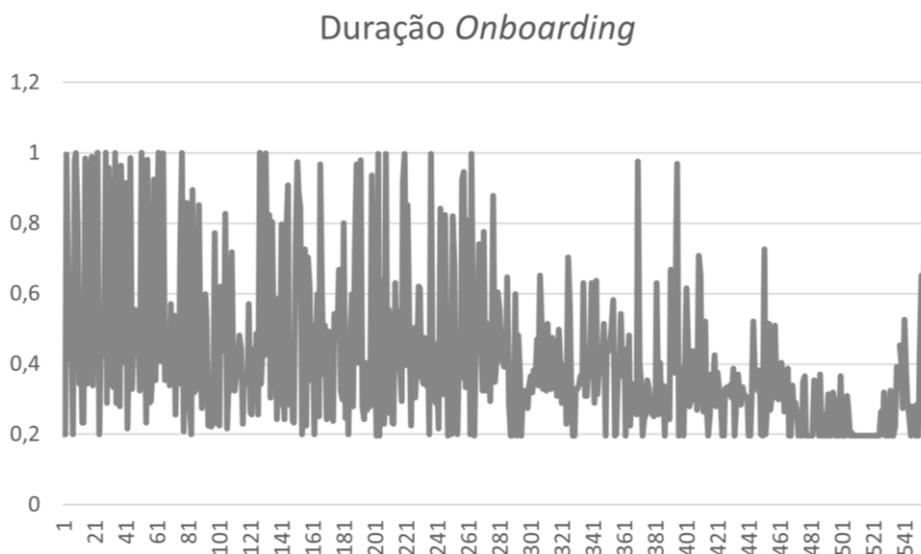


Figura 15 - Distribuição Normal da Variável Duração *Onboarding* (em dias)

Com base na análise da Figura 15, conclui-se que não há concretamente um padrão (não há padrões facilmente identificáveis), por isso não é uma abordagem válida.

Assim e com base nos objetivos da empresa, os objetos da nova variável são cinco: Duração Extraordinária, Duração Excelente, Duração Boa, Duração Elevada e uma Duração *Outlier*. Na Tabela 9 encontra-se a explicação de cada um dos objetos criados. Esta classificação foi realizada com base no conhecimento de domínio.

Tabela 9 - Objetos da Variável Classificação Duração *Onboarding*

Classificação Duração	
<i>Onboarding</i>	Descrição
Duração Extraordinária	Classifica um <i>onboarding</i> com duração de 15 dias ou menos
Duração Excelente	Classifica um <i>onboarding</i> com duração entre 16 e 34 dias
Duração Boa	Classifica um <i>onboarding</i> com duração entre 35 e 68 dias
Duração Elevada	Classifica um <i>onboarding</i> com duração entre 69 e 103 dias
Duração <i>Outlier</i>	Classifica um <i>onboarding</i> com duração maior a 104 dias

Em seguida, a variável “***Alignment with Product vision***”, em português Alinhamento com a Visão do Produto, agrupa os clientes consoante o seu alinhamento com a visão do produto, isto é, se lhes foi atribuído o valor 1 significa que a plataforma Infraspak consegue responder aos problemas da empresa cliente e esta irá beneficiar do desenvolvimento das próximas funcionalidades do produto. Caso lhe seja atribuído o valor 0, ou a plataforma responde ao problema atual ou as próximas funcionalidades irão conseguir responder, no último caso, -1, a plataforma não responde ao problema nem os futuros desenvolvimentos o farão.

A variável “***Customer Classification***”, em português Classificação do Cliente, que a empresa utiliza para classificar os clientes distingue-se em três grupos: *Excellence customer*, *Key customer* e *Leading customer*. A Tabela 10 apresenta o significado deste tipo de clientes.

Tabela 10 - Objetos da Variável *Customer Classification*

<i>Customer Classification</i>	Descrição
<i>Excellence customer</i>	Cliente de pequena dimensão e, por isso, não há perspetiva de crescimento do negócio. Desta forma, não são clientes influentes para trazer mais clientes para a empresa.
<i>Key customer</i>	Cliente considerado estratégico e pode ajudar a abrir mercado.
<i>Leading customer</i>	Cliente que corresponde a um intermédio entre o <i>Excellence customer</i> e o <i>Key customer</i> .

A próxima variável denomina-se “***Apps Installed***” e enumera todas as *apps* e *add-ons* instalados pelo cliente presentes no Infraspak Hub. Com o intuito de facilitar o tratamento desta informação, criou-se a variável “**Número de apps instaladas no Hub**” que contabiliza o número de *apps* e *add-ons*

instalados. A nível de exemplo, se um cliente instalou as seguintes *apps* ou *add-ons*: Avarias, Trabalhos, *Stocks* e Requisições, na nova variável é atribuído o valor 4. Esta variável é importante pois permite ajudar a determinar o nível de interesse e compromisso que um cliente tem com a plataforma e melhor perceber quais as suas necessidades.

A variável que se segue “**Deal Market**”, em português Mercado de Negócios, separa os países em que os clientes estão sediados por diversos mercados, na Tabela 11 estão apresentados os diferentes países distribuídos pelos respetivos mercados. Esta nomenclatura é utilizada pela Infraspark e personaliza o mercado onde cada empresa opera.

Tabela 11 - Atribuição de países pela variável *Deal Market*

Deal Type	Países
BR	Brasil.
ES	Espanha
LATAM	Argentina, Bolívia, Brasil, Chile, Colômbia, Costa Rica, Cuba, Equador, El Salvador, Guatemala, Haiti, Honduras, México, Nicarágua, Panamá, Paraguai, Peru, República Dominicana, Uruguai e Venezuela.
PALOP	Angola, Cabo Verde, Guiné-Bissau, Moçambique e São Tomé e Príncipe.
PT	Portugal
ROW	Países referentes ao resto do mundo que não estão referidos nos outros objetos.
UK	Reino Unido e Irlanda

As últimas variáveis dizem respeito à área de negócio em que cada cliente opera, sendo elas: “**Segment**”, “**Industry**”, “**Macro-industry**” e “**Deal type**”. Na Tabela 12 estão discriminados cada um dos objetos das variáveis referidas.

Tabela 12 - Objetos das variáveis *Segment*, *Industry*, *Macro-Industry* e *Deal Type*

Variável	Objetos
<i>Segment</i>	Assistência Técnica, Centro Comercial, Educação, Hospitalidade, Indústria, <i>Facility Manager</i> (Gestão de instalações), Retalho
<i>Industry</i>	Educação, Energia, <i>Hard Facilities Management</i> , <i>Hard Facilities Services</i> , Cuidados de saúde, Hospitalidade, Industrial, IT & Telecom, Lazer e Cultura, Escritórios, Sector Público, Parques Recreativos e Centros Desportivos, Residencial, Retalho, Centros Comerciais, <i>Soft Facilities Management</i> , <i>Soft Facilities Services</i> , <i>Total Facilities Management</i>
<i>Macro-Industry</i>	<i>Facilities Management Providers</i> (Prestadores de Serviços de Gestão de Instalações), <i>Facilities Services Providers</i> (Prestadores de Serviços em Instalações), <i>Facilities Owners</i> (Proprietários das instalações)
<i>Deal Type</i>	Comprador, Parceiro, Vendedor

As variáveis relativas ao “**Segment**”, em português Segmento, e a “**Industry**”, em português Indústria, são bastante semelhantes, contudo a última referida apresenta um grau de maior detalhe, mas ambas representam o ramo em que a empresa opera. Por exemplo, se uma universidade compra a plataforma Infraspark, o seu objeto de segmento e a indústria será igual, que no caso seria “Educação”. Contudo se a empresa em questão, fornecer serviços de manutenção de ar condicionados, o seu segmento será “Assistência Técnica” e a indústria corresponde a “*Hard Facilities Services*” pois este tipo de serviços traduz uma grande carga de manutenção preventiva. A variável “**Macro-Industry**”, em português Indústria Macro, classifica ainda de forma mais abrangente o cliente, tendo em conta a sua relação com as instalações, isto é, se é proprietário, se fornece manutenção ou se faz a sua gestão. Por último, a variável “**Deal Type**”, em português Tipo de Negócio, distingue se o cliente compra ou vende manutenção ou se é um parceiro.

Depois da descrição das variáveis a serem manuseadas ao longo do projeto, passa-se para a preparação e tratamento das mesmas.

4.3 Fase 3 | Preparação e tratamento de dados

Após a obtenção de um entendimento mais aprofundado do processo a ser analisado e definição e descrição das variáveis a serem estudadas, é fundamental desenvolver uma estrutura que seja capaz

de responder às necessidades da empresa. Assim, após a recolha de dados em bruto descrita no subcapítulo anterior, é agora efetuada a preparação dos dados para, posteriormente, realizar a sua análise. A análise de dados consiste em examinar, categorizar, classificar e/ ou recombinaer evidências (Santos, 2020).

Portanto, nesta etapa, os dados passaram pelas seguintes fases:

- Limpeza dos dados de modo a potencializar a sua utilização;
- Análise estatística univariada com vista a analisar, de modo geral, o que aconteceu no passado no processo de *onboarding*;
- Criação do *Data Warehouse* (DW), uma forma de *Business Intelligence*, que permite às empresas e organizações tomar decisões inteligentes com base no que os dados lhes fornecem.

4.3.1 Limpeza de dados

Uma vez recolhidos os dados, o passo seguinte passa por capacitá-los para análise, limpando-os para garantir que se trabalha com dados de qualidade (Microsoft, 2002). Após o acesso à base de dados, foram realizadas as seguintes ações com a finalidade de prevenir e corrigir potenciais erros (Hellerstein, 2008):

- **Remoção de colunas desnecessárias:** alguns dos dados recolhidos não são relevantes para a investigação em questão. Como referido anteriormente, das 137 variáveis disponíveis inicialmente, apenas se trabalhou sob 25, o que representa uma redução de 82%. Esta redução significativa foi realizada com o conhecimento adquirido ao longo do projeto e necessária para que houvesse um número de variáveis com que fosse possível trabalhar. Por exemplo, variáveis relativas ao CS associado ao negócio ou o número de avarias reportadas num dia por uma empresa, não são relevantes para a presente dissertação;
- **Correção de erros ortográficos de dados textuais:** revisão da capitalização inconsistente e convenções de nomenclatura irregulares. Foram realizadas 56 alterações entre as colunas de “*Country*” e “*City*”, que corresponde a 6% dos dados. A título de exemplo, na coluna “*Country*” que apresenta valores como Brasil, Brazil e brasil, foi necessário realizar a uniformização;
- **Preenchimento de dados:** uma vez identificadas lacunas, pode proceder-se ao seu preenchimento, que, neste caso, se traduziu em 274 alterações em 3 colunas diferentes,

equivalendo a 16% de dos dados alterados. Exemplificando a coluna “*City*” contava com alguns valores em falta que poderiam ser preenchidos, caso se soubesse a morada completa.

- **Preenchimento dos campos sem dados:** aquando da falta de uma informação descritiva e não numérica da empresa, foi preenchido o campo como “Sem dados”, nesta situação foram acrescentados 406 campos como “Sem dados”.

Como já abordado, a análise de dados requer dados efetivamente limpos para produzir perceções precisas e fiáveis. Mas os dados limpos têm também uma série de outros benefícios como a organização, evita potenciais erros e melhora a produtividade pois o acesso a dados é facilitado.

4.3.2 Análise estatística descritiva univariada

A análise estatística descritiva univariada é fundamental, não só no contexto de exploração inicial de dados, mas também no contexto de validação de dados e preparação dos dados para análises mais complexas (A. L. Teixeira, 2016). Esta análise estatística permite verificar todas as variáveis numa primeira estância e verificar a existência de dados mal inseridos (A. L. Teixeira, 2016). A análise estatística descritiva resume a informação numérica de uma forma estruturada com o objetivo de obter uma visão geral das variáveis medidas numa determinada amostra (Fortin, 1999).

Assim, esta investigação inicial do conjunto de dados procura perceber e resumir os mesmos nas suas principais características. A análise é realizada para entender como os dados estão estruturados, identificar possíveis padrões e tendências bem como detetar anomalias. Assim, a análise abrange a dimensão da amostra (N), os valores mínimos, máximos, somatório (soma), média, erro padrão, variância, desvio padrão, mediana, percentil 25 e percentil 75.

Na Tabela 13, está presente as estatísticas descritivas dos dados, usada para sumariar os dados obtidos e para permitir uma leitura clara e objetiva dos mesmos, de forma a analisá-los.

Tabela 13 - Estatística descritiva univariada inicial

	N	Mínimo	Máximo	Soma	Média	Erro padrão	Variância	Desvio padrão	Mediana	Percentil 25	Percentil 75
<i>CSM Score</i>	452	1	5	1869,0000	4,1350	0,0480	1,0394	1,0195	4,0000	4,0000	5,0000
<i>Health Score</i>	465	3	10	3946,0000	8,4860	0,0788	2,8883	1,6995	9,0000	8,0000	10,0000
<i>NPS</i>	299	1	10	2503,4070	8,3726	0,1009	3,0440	1,7447	9,0000	8,0000	10,0000
<i>Setup Fee</i>	483	0	32310	1041736,0000	2156,8040	117,5346	6672346	2583,0880	2000,0000	374,0000	3360
Total Técnicos	559	0	495	5786,0000	10,3506	1,2701	901,6940	30,0282	4,0000	1,0000	9,0000
Total Admins + Gestores	559	-1	207	2623,0000	4,6923	0,6120	209,3604	14,4693	1,0000	0,0000	4,0000
Duração Onboarding (dias)	486	0	647	29358,0000	60,4074	3,2105	5009,4250	70,7773	38,0000	23,0000	69,0000

De uma análise breve desta Tabela 13, há umas conclusões iniciais que já podem ser retiradas, sendo elas:

- Na variável *CSM Score*, sendo o percentil 25 igual ao valor 4, 75% dos valores desta variável são iguais a 4 ou a 5. Estes valores representam um elevado potencial do cliente.
- Nas variáveis *Health Score* e *NPS*, sendo o percentil 25 igual ao valor 8, também 75% dos valores estão compreendidos entre os valores 8 e 10, que se traduz num uso muito confortável da plataforma.
- Nas restantes variáveis, observam-se amplitudes muito elevadas (valor máximo a subtrair pelo valor mínimo) o que significa uma grande variação entre os limites do conjunto numérico e, por isso, pouca estabilidade. Desta forma, foi construído, em PAST, um *boxplot* para cada uma das variáveis para perceber o centro e a dispersão da distribuição de cada variável. No Apêndice I – *Boxplot* das variáveis, estão presentes Figura 22, Figura 23, Figura 24 e Figura 24 que representam, respetivamente, o *boxplot* da variável *Setup Fee*, o *boxplot* da variável *Total Technicians*, o *boxplot* da variável *Total Admins+ Managers* e o *boxplot* da variável *Onboarding Duration*.

Através da análise dos gráficos presentes no Apêndice I – *Boxplot* das variáveis, conclui-se que, pelo tamanho da distribuição, um *boxplot* não funciona da melhor maneira, pois todos os gráficos apresentam uma assimetria à direita e, por fim, pode detetar-se a presença de *outliers*. *Outliers* são os dados que estão mais distantes dos restantes valores e, por isso, afetam fortemente os resultados - nos gráficos estão identificados por um asterisco (*).

Como referido, as estatísticas descritivas são úteis para detetar possíveis erros ou ocorrências estranhas no conjunto de dados. Desta forma, identificam-se as seguintes ocorrências:

- A dimensão da amostra (N) para cada variável apresenta valores distintos, o que significa que existem muitos argumentos que não se encontram preenchidos, isto é, estão sem valores.
- Na variável *Total Admins + Managers*, o valor mínimo apresenta-se como -1, esta situação é impossível porque o número de contas no sistema não pode ser negativo.
- Na variável *Onboarding Duration*, o valor mínimo também apresenta um valor igual a 0. Esta situação não é possível porque o número de dias que um *Onboarding* pode durar, no mínimo, é igual a 1 dia.

Combatendo as declarações anteriores, foram efetuadas alterações nos dados das variáveis, isto é, foi realizada uma nova iteração de limpeza de dados para converter os mesmos. Em todas as variáveis,

sempre que não existam dados, estes foram substituídos por “Sem dados”. Os dados das variáveis *Total Admins + Managers* e *Onboarding Duration* foram revistos e alterados para dados que se adequam ao contexto.

Na Tabela 14, está presente a nova estatística descritiva univariada dos dados após as modificações.

Tabela 14 - Estatística descritiva univariada final

	N	Mínimo	Máximo	Soma	Média	Erro padrão	Variância	Desvio padrão	Mediana	Percentil 25	Percentil 75
CSM Score	452	1	5	1869,0000	4,1350	0,0480	1,039396	1,019508	4,0000	4,0000	5,0000
Health Score	465	3	10	3946,0000	8,4860	0,0788	2,888274	1,699492	9,0000	8,0000	10,0000
NP	299	1	10	2503,4070	8,3726	0,1009	3,043957	1,744694	9,0000	8,0000	10,0000
Setup Fee	483	0	32310	1041736,0000	2156,8040	117,5346	6672346	2583,088	2000,0000	374,0000	3360,0000
Total Técnicos	557	0	495	5882,0000	10,5601	1,2719	901,1065	30,0184	4,0000	1,0000	9,0000
Total Admins + Gestores	542	1	207	2908,0000	5,3653	0,6225	210,0548	14,4933	1,0000	1,0000	5,0000
Duração Onboarding (dias)	486	1	647	29358,0000	60,4074	3,2105	5009,4250	70,7773	38,0000	23,0000	69,0000

Pela ausência de alguns dados, o tamanho da distribuição de algumas variáveis diminuiu, como *Total Admins + Gestores* e *Onboarding Duration*.

Uma vez recolhidos os dados e após o seu tratamento estatístico, é possível, agora, proceder à sua organização, sistematização e análise.

4.3.3 Implementação do *Data Warehouse*

Para otimizar os processos descritos, pode recorrer-se ao uso de uma base de dados que, através de uma boa estruturação, consiga criar uma relação lógica entre os dados, controlando a redundância e consistência dos mesmos, facilitando a partilha e exploração dos dados para ajudar na tomada de decisão (Carlos P. Caldeira, 2007).

Assim sendo e tendo em consideração os conceitos teóricos referidos na Revisão da Literatura e todos os aspetos anteriormente abordados, neste subcapítulo pretende-se dar “forma” e estrutura aos dados estudados, criando assim o desenho da solução, isto é, ferramentas que permitam auxiliar na tomada de decisões.

O esquema escolhido para a elaboração da solução *Business Intelligence* (BI) foi o esquema em estrela, construindo-se então um modelo de DW que modele os dados dos clientes da Infraspark consoante o tempo, o local e a área de negócio. A possibilidade de uma consulta simples e uma interpretação acessível dos dados foram os motivos para a seleção deste tipo de esquema. Deste modo, o esquema em estrela elaborado apresenta uma tabela de factos e três tabelas de dimensão que a circundam e de seguida é explicado o processo de criação de cada uma destas tabelas e, por fim, o processo de criação do DW bem como o seu carregamento de dados.

a) Tabela de Factos

No esquema em estrela, a tabela de factos atribui a todas as tabelas de dimensões os dados. As métricas escolhidas para a tabela de factos encontram-se descritas na Tabela 15 que se encontra abaixo.

Tabela 15 - Descrição dos atributos da tabela de factos

Atributo	Descrição
ID_Facto	Chave primária da tabela. Neste campo, o valor é auto incrementado consoante a inserção de um novo registo.
Duracao_Onboarding	Valor da duração do <i>onboarding</i> da empresa. (exemplo, 25)
Onboarding_Data_esperada	Valor lógico se uma empresa atingiu a data esperada de <i>onboarding</i> . (exemplo, "Cumpriu expectável")
MRR	Valor MRR paga pela empresa. (exemplo, 690)
Setup_Fee	<i>Setup Fee</i> adquirido pela empresa. (exemplo, 4900)
Total_Tecnicos	Número de contas Técnico que a empresa adquiriu. (exemplo, 9)
Total_Admins_Gestores	Número de contas Administradores e Gestores que a empresa adquiriu. (exemplo, 4)
Fase	Fase em que a empresa se encontra. (exemplo, "Onboard")
Entidade	Nome atribuído à empresa. (exemplo, "Entidade 336")

b) Tabela de dimensões

Como mencionado, as dimensões costumam ser constituídas por vários atributos e poucas linhas, em comparação com as tabelas de facto, pois isto permite análises mais distintas sobre as tabelas de facto, oferecendo diversas perspetivas (Filipe et al., 2013).

A partir da pesquisa efetuada na empresa e com o domínio do conhecimento, as três dimensões consideradas para a modelação em questão são: Dimensão local, Dimensão Tempo e Dimensão Área de Negócio. A análise das respetivas dimensões é abordada em seguida.

Dimensão Local

A dimensão local (*dimensao_local*) foi idealizada no sentido de englobar todas as cidades existentes em Portugal e as cidades dos restantes países em que estão sediadas as empresas cliente. Na Tabela 16 encontra-se a descrição dos atributos da dimensão local.

Tabela 16 - Descrição dos atributos da dimensão local

Atributo	Descrição
ID_local	Chave primária da tabela. Neste campo, o valor é auto incrementado consoante a inserção de um novo registo.
Cidade	Nome da cidade. (exemplo, Barcelos)
Distrito	Nome do distrito onde se localiza a cidade. (exemplo, para a cidade Barcelos o valor do campo é Braga)
Região	Nome da região onde se localiza a cidade. (exemplo, para a cidade Barcelos o valor do campo é norte)
País	Nome do país onde se localiza a cidade. (exemplo, para a cidade Barcelos o valor do campo é Portugal)
Continente	Nome do mercado onde se localiza a cidade. (exemplo, para a cidade Barcelos o valor do campo é PT)

Desta forma, termina-se a apresentação e descrição da dimensão local.

Dimensão Tempo

A dimensão tempo (*dimensao_tempo*) foi desenhada no sentido de conter todos os dados possíveis das datas entre 2014-01-01 a 2050-12-31, inclusive. De ressaltar que as datas estão de acordo com o calendário gregoriano em vigor. A descrição dos atributos que compõe a dimensão tempo estão apresentados na Tabela 17.

Tabela 17 - Descrição dos atributos da dimensão tempo

Atributo	Descrição
<i>ID_tempo</i>	Chave primária da tabela.
<i>Db_date</i>	Representação da data no formato ano-mês-dia. (exemplo, 2015-01-01)
<i>Year</i>	Ano da data. (exemplo, para a data 2015-01-01 o valor do campo é 2015)
<i>Month</i>	Número do mês da data. (exemplo, para a data 2015-01-01 o valor do campo é 1)
<i>Day</i>	Número do dia do ano. (exemplo, para a data 2015-01-01 o valor do campo é 1)
<i>Quarter</i>	Número do trimestre da data. (exemplo, para a data 2015-01-01 o valor do campo é 1)
<i>Week</i>	Número do dia da semana da data. (exemplo, para a data 2015-01-01 o valor do campo é 1)
<i>Day_name</i>	Nome do dia do ano. (exemplo, para a data 2015-01-01 o valor do campo é <i>Thursday</i> – quinta-feira)
<i>Month_name</i>	Nome do mês da data. (exemplo, para a data 2015-01-01 o valor do campo é <i>January</i> - janeiro)

Desta forma, termina-se a apresentação e descrição da dimensão tempo.

Dimensão Área de Negócio

A dimensão Área de Negócio foi gerada de forma a agregar informação relativa à área de negócio em que um cliente pode atuar, segundo a nomenclatura utilizada pela Infraspak.

Podem ser analisados na Tabela 18 os atributos que integram esta dimensão.

Tabela 18 - Descrição dos atributos da dimensão área de negócio

Atributo	Descrição
ID_Area_Negocio	Chave primária da tabela. Neste campo, o valor é auto incrementado consoante a inserção de um novo registo.
<i>Segment</i>	Nome do segmento da indústria onde o cliente se insere. (exemplo, Assistência Técnica (<i>Technical Assistance</i>))
<i>Industry</i>	Nome do tipo de indústria onde o cliente se insere. (exemplo, <i>Hard Facilities Services</i>)
<i>Macro_Industry</i>	Nome do tipo de indústria mais geral onde o cliente se insere. (exemplo, Prestadores de Serviços em Instalações (<i>Facilities Services (FS) Providers</i>))
<i>Deal_type</i>	Nome do tipo de relação com a manutenção que o cliente apresenta. (exemplo, Vendedor (<i>Seller</i>))

Desta forma, termina-se a apresentação e descrição da dimensão área de negócio.

c) Modelo conceptual

O modelo conceptual consiste na representação de um conjunto de dados e as suas relações que auxiliam a criação de uma base de dados (R. Rocha, 2007). É formado por entidades, relacionamentos e atributos. As entidades são objetos do mundo real sobre os quais se pretende armazenar informação e com uma existência independente. Os relacionamentos representam interações entre entidades, de acordo com as necessidades. Por fim, os atributos são propriedades que caracterizam as entidades e também podem caracterizar entidades relacionamento (Richard, 2017). O modelo conceptual desenvolvido na ferramenta TerraER encontra-se representado na Figura 16. Esta fase considera-se como a representação de alto nível na estrutura de dados.

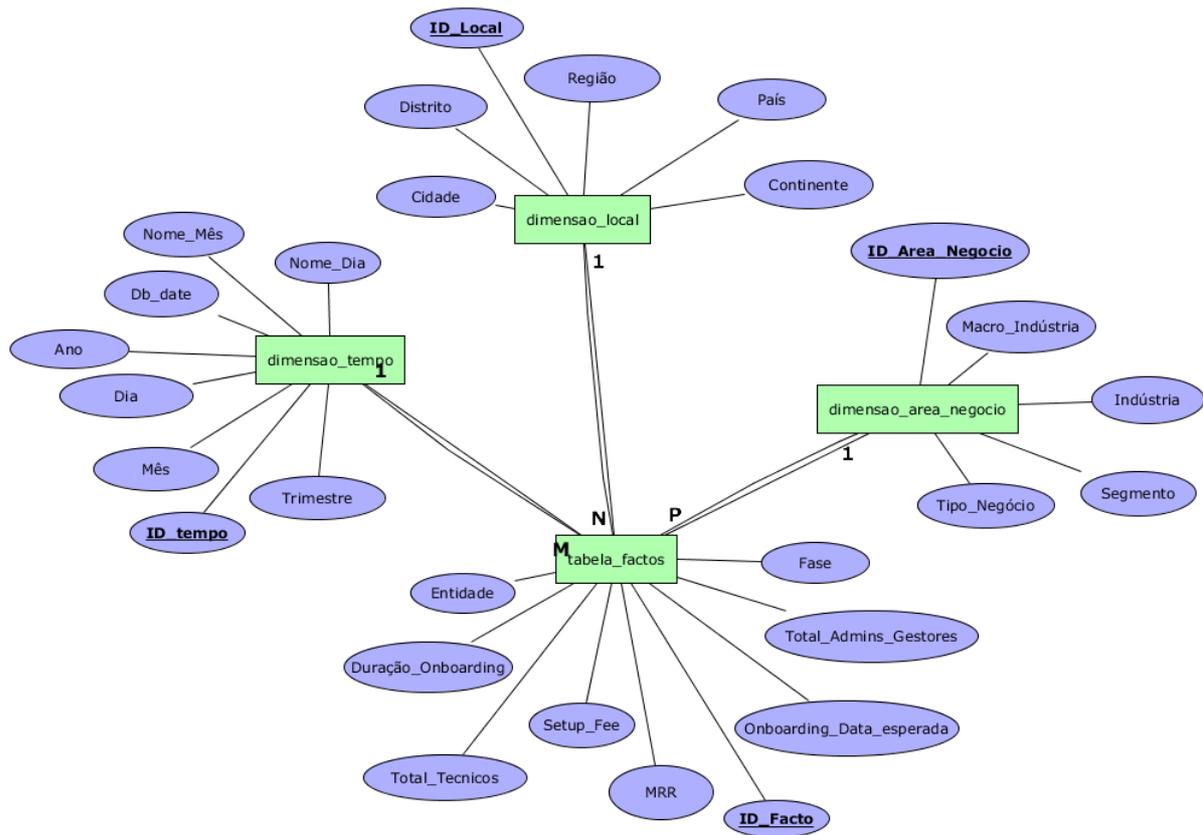


Figura 16 - Modelo conceptual

A partir da Figura 16, observa-se que, a título de exemplo, a entidade `dimensao_local` possui como atributo chave o ID do local e cinco atributos simples, a Cidade, Distrito, Região, País e Continente. Esta relaciona-se com a entidade `tabela_factos` numa relação de obrigatoriedade, onde a tabela de factos pode ter mais que um ID de local igual, mas o local só possui um único, por isso uma relação de “um para muitos” entre `dimensao_local` e `tabela_factos`, por esta razão aparece o valor de 1 junto a `dimensao_local` e a representação N junto a `tabela_factos`. A representação P e M fazem também referência à relação de “um para muitos”, esta nomenclatura é atribuída pelo *software*.

Neste modelo conceptual, as entidades `dimensao_local`, `dimensao_tempo` e `dimensao_area_negocio` dizem respeito às dimensões referidas anteriormente e a entidade `tabela_factos` corresponde à tabela de fatos.

d) Modelo lógico

Para a criação do modelo lógico, que corresponde à transformação do modelo conceptual para a forma de como a base de dados será implementada, recorreu-se à ferramenta de gestão de base de dados MySQL Workbench, obtendo-se o modelo da Figura 17.

Numa primeira fase, representou-se cada uma das dimensões anteriormente definidas por uma tabela, na qual foram colocados os seus correspondentes atributos. Para cada um destes atributos, seleccionou-se o seu formato e procedeu-se à sua caracterização, referindo se são atributos não nulos (NN), caso seja obrigatório o seu preenchimento e *Primary Key* (PK), chave primária, para os atributos chaves.

De seguida, estabeleceu-se relações entre as tabelas através do mecanismo das chaves estrangeiras.

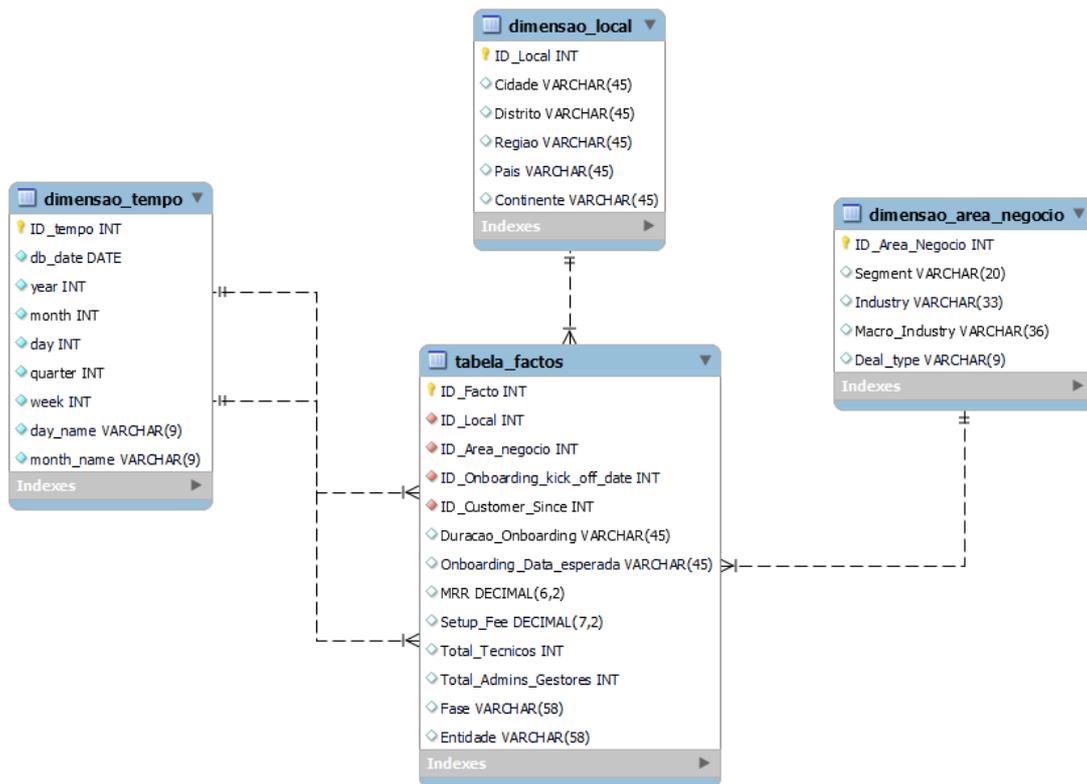


Figura 17 - Modelo lógico

A tabela_factos é construída por uma chave primária ID_Facto, possui quatro chaves estrangeiras ID_Local, ID_Area_negocio, ID_Onboarding_kick_off e ID_Customer_Since, sendo que a primeira se liga à tabela dimensão_local, a segunda à tabela dimensão_area_negocio e as duas últimas à tabela dimensão_tempo. Para além disso, a tabela_factos contém informação acerca da duração do *Onboarding* da entidade (*Duracao_Onboarding*), se atingiu a data de *Onboarding* esperada (*Onboarding_Data_esperada*), do MRR (MRR), do Setup_Fee (Setup_Fee), do número total de técnicos (Total_Tecnicos) e de administradores e gestores (Total_Admins_Gestores), a fase (Fase) em que se encontra cada empresa e o nome da entidade (Entidade). Este modelo permite retirar vários indicadores agrupados pelo tipo de negócio, tempo e local.

e) Modelo físico

Posteriormente, com recurso à opção *Forward Engineer* da ferramenta de base de dados usada, o modelo lógico foi convertido para o modelo físico, tornando-se possível a inserção e manipulação dos dados, assim como o desenvolvimento de *procedures* e funções com o intuito de automatizar a deteção de erros e a atualização de certos dados. O código gerado para a criação das tabelas, passo inicial do modelo físico, encontra-se no Apêndice II – Criação das tabelas do modelo físico do presente trabalho.

f) Criação de *views*

Inicialmente, começou-se por criar uma tabela temporária para facilitar o preenchimento dos dados e criação dos relacionamentos entre tabelas. A *query* desta tabela possuía 583 linhas código, por esse motivo, não se encontra nesta dissertação.

A fim de auxiliar os relacionamentos das tabelas de dimensões com a tabela de factos, recorreu-se ao uso de *views* para facilitar o processo.

Uma *view* funciona como uma tabela virtual (ORACLE, 2022) formada a partir de declarações SELECT, funcionando como uma consulta de uma ou várias tabelas para produzir um conjunto de resultados. Esta tabela virtual não contém registos físicos, apresenta apenas os dados de modo a facilitar a sua visualização.

Relativamente à *view* presente de seguida refere-se ao local. Esta foi criada para identificar os atributos em comum entre a tabela temporária e a dimensão local, sendo o elemento de ligação a cidade.

```
CREATE VIEW esquema_estrela.vw_local AS
SELECT T.Duracao_Onboarding,
T.Onboarding_Data_esperada,
T.MRR,
T.Setup_Fee,
T.Total_Tecnicos,
T.Total_Admins_Gestores,
T.Onboarding_Kick_off_date,
T.Customer_Since,
T.Deal_market,
T.Pais,
T.Entidade,
T.Segment,
T.Industry,
```

```
T.Macro_Industry,
T.Deal_type, L.ID_Local, L.Cidade from temporaria T
LEFT JOIN dimensao_local L ON
T.Cidade = L.Cidade
```

No Apêndice III – *View* tempo e *View* área de negócio, encontram-se as *views* relativas ao tempo e à área de negócio, sendo os elementos de ligação para a *view* tempo *Customer_Since* e *Onboarding_Kick_off_date*. Para a *view* área de negócio os elementos de ligação são: *Industry*, *Macro_Industry*, *Deal_type* e *Segment*.

g) Povoamento das tabelas

Para finalizar e concluir o modelo físico, procedeu-se ao povoamento das tabelas definidas com vista a fornecer uma base de dados correspondente à realidade. Este povoamento foi feito recorrendo ao comando INSERT, através do qual, os valores são inseridos na estrutura tabela pela ordem das suas colunas.

De seguida, apresenta-se um exemplo do código utilizado para o preenchimento das tabelas.

Para introduzir valores na tabela de dimensão da Área de Negócio, utilizou-se o seguinte que faz o povoamento da primeira linha da tabela:

```
INSERT INTO dimensao_area_negocio VALUES
(1, 'Education', 'Education', 'Facilities Owners
(FO)', 'Buyer');
```

Na dimensão Local, o de código usado para preencher a primeira linha:

```
INSERT INTO dimensao_local VALUES
(1, 'Cabinda', 'Congo', 'Norte', 'Angola', 'Africa');
```

Para adicionar os valores na dimensão Tempo, recorreu-se a um procedimento, conjunto de comandos que podem ser executados de uma só vez como uma função (Oracle, 2022), que permitiu preencher os dados da tabela.

```
DROP PROCEDURE IF EXISTS fill_date_dimension;
DELIMITER //
CREATE PROCEDURE fill_date_dimension(IN startdate DATE, IN
stopdate DATE)
BEGIN
    DECLARE currentdate DATE;
    SET currentdate = startdate;
    WHILE currentdate <= stopdate DO
```

```

INSERT INTO dimensao_tempo VALUES (

YEAR(currentdate)*10000+MONTH(currentdate)*100      +
DAY(currentdate),

    currentdate,
    YEAR(currentdate),
    MONTH(currentdate),
    DAY(currentdate),
    QUARTER(currentdate),
    WEEKOFYEAR(currentdate),
    DATE_FORMAT(currentdate,'%W'),
    DATE_FORMAT(currentdate,'%M')
);

SET currentdate = ADDDATE(currentdate,INTERVAL 1
DAY);

END WHILE;

END

//

```

A seguir, apresenta-se o povoamento da primeira linha da tabela temporária, que como referido ajudará a preencher a tabela de Factos.

```

INSERT INTO `esquema_estrela`.`temporaria` VALUES
(1,'Sem dados', 325, 0, 10, 15, 'Onboard', '2015-
01-30 00:00:00', '2015-01-31 00:00:00', 'Porto', 'PT',
'Portugal', 'Entidade 001', 'Education', 'Education',
'Facilities Owners (FO)', 'Buyer');

```

A *query* final realizada permitiu o povoamento da tabela de Factos a partir das *views* e da tabela temporária.

4.4 Fase 4 | Apresentação dos resultados do obtidos

Encontrando-se o Data Warehouse desenvolvido, esta etapa centraliza-se em apresentar os dados obtidos do trabalho realizado anteriormente. Para isso, foi criado um conjunto de tabelas dinâmicas em *Microsoft Excel* utilizando diferentes variáveis e criada um *dashboard* para apoiar visualmente a análise das tabelas dinâmicas.

As tabelas dinâmicas não substituem um Data Warehouse. As duas soluções têm magnitudes de capacidade completamente diferentes e casos de uso apropriados. O uso de tabelas dinâmicas é melhor sucedido quando usado e combinado com um DW (Underwood, 2014).

4.4.1 Tabelas dinâmicas

Uma tabela dinâmica resume grandes quantidades de dados num formato mais acessível e de fácil consulta, na medida em que permite o agrupamento de dados de forma significativa. Este tipo de ferramenta presente no *Microsoft Excel* é utilizada para soluções de BI e, deste modo, com o recurso às tabelas dinâmicas, será possível começar a recolher as primeiras respostas à pergunta da investigação – “*Quais os fatores que influenciam o processo de onboarding de um novo cliente?*”. Posto isto, as seguintes tabelas permitem tornar o processo de análise mais acessível sendo também bastante fácil começar de novo, desmarcando campos ou movendo-os para outro local (Sharib, 2012). Um resultado que foi possível analisar diz respeito à duração média de *onboarding* (em dias) de todos os clientes consoante o seu segmento (“*Segment*”), como apresentado na Tabela 19. Os objetos das variáveis *Segment* estão presentes na Tabela 12.

Tabela 19 - Resultados relativos à média da duração de *onboarding* consoante o *Segment*

Rótulos de Linha	Média de Onboarding Duration
Educação	57,89
Gestão de instalações	89,96
Hospitalidade	56,69
Indústria	52,31
Outros	54,18
Retalho	89,18
Sem dados	30,45
Centro Comercial	50,33
Assistência Técnica	62,95
Total Geral	60,41

Como é possível observar, o segmento relativo a Gestão de instalações (*Facility Manager*) expõe uma duração de *onboarding* com maior duração com 89,96 dias e o Centro comercial o *onboarding* de menor duração, com 50,33 dias. Contudo a Gestão de instalações e Retalho apresentam uma duração média muito semelhante de 89 dias. O mesmo acontece com Educação e Hospitalidade com 57,89 dias e 56,69 dias, aproximadamente.

Ainda no contexto do segmento da empresa, foram quantificados os *onboardings* que cumpriram o tempo inicialmente estipulado. Tal resultado encontra-se na Tabela 20.

Tabela 20 - Resultados relativos ao número de *onboardings* que cumpriram a duração esperada divididos por *Segment*

Contagem de Onboarding Duration	Rótulos de Coluna		Sem dados	Total Geral
Rótulos de Linha	Cumpriu expectável	Não cumpriu o expectável		
Educação	1	2	6	9
Gestão de Instalações	9	14	30	53
Hospitalidade	12	32	63	107
Indústria	5	10	14	29
Outros	9	10	15	34
Retalho	3	6	13	22
Centro Comercial	1	1	1	3
Assistência Técnica	34	55	83	172
Sem dados	19	32	7	58
Total Geral	93	162	232	487

A partir de uma análise da Tabela 20, constata-se que 33% dos *onboardings* não são concluídos no tempo esperados, principalmente Assistência Técnica (representado a verde) o que também pode ser influenciado pela amostra de 172 casos que apresenta. Conclui-se que as empresas que não cumprem o tempo esperado de *onboarding* (162 casos) é superior àquelas que cumprem (93 casos). Da análise da Tabela 20 também se observa que 47% das empresas não apresentam dados relativamente ao cumprimento do tempo esperado, isto deve-se à falta de preenchimento por parte dos agentes CS. Ambas as análises realizadas anteriormente, englobam casos de clientes que já cancelaram a sua subscrição com a Infraspak, desta forma, a Tabela 21 obtêm informações sobre a duração média de *onboarding* consoante a fase em que o cliente se encontra. A Tabela 4 apresenta os objetos da variável Classificação da fase.

Tabela 21 Resultados relativos à duração média do *onboarding* consoante a classificação da fase do cliente

Rótulos de Linha	Média de Onboarding Duration
Cliente	57,11
Cliente em risco	81,49
Não Cliente	70,46
Sem Resultado	41,00
Total Geral	60,41

De acordo com a leitura que se faz da tabela, clientes em risco e não clientes apresentam uma duração elevado de *onboarding*, de 81,49 dias e 70,46 dias, respetivamente.

Ainda relativamente à *Macro-Industry* a que um cliente pode fazer parte, avaliou-se a duração média de *onboarding* de acordo com o *Deal Market* a que pertence. Esta análise pode encontrar-se na Tabela 22.

Na Tabela 11, encontra-se a atribuição dos países pelo respetivo *Deal Market*.

Tabela 22 - Resultados relativos à média da duração de *onboarding* consoante a *Macro-industry* e *Deal Market*

Média de Duração Onboarding	Rótulos de Coluna					
Rótulos de Linha	Prestadores de Serviços de Gestão de Instalações	Proprietários das instalações	Prestadores de Serviços em Instalações	Sem dados	Total Geral	
BR	47,90	60,38	51,20	21,00	55,14	
ES	77,50	46,85	44,50		49,21	
LATAM		36,88			36,88	
PALOP	108,10	148,50	73,60		96,09	
PT	61,83	56,22	72,15	30,50	61,02	
ROW	70,00	26,75		6,00	30,50	
UK	54,50	56,17	32,50	27,00	46,38	
Total Geral	65,43	56,22	65,81	24,22	59,30	

Da análise deste resultado, concluiu-se que os países PALOP são os que apresentam maior duração média de *onboarding*, de 96,08 dias, e os países ROW e LATAM apresentam os de menor duração, com 30,50 dias e 36,88 dias, respetivamente.

Paralelamente, observa-se que as entidades Prestadores de Serviços de Gestão de Instalações e Prestadores de Serviços em Instalações apresentam médias de *onboarding* bastante semelhantes de 65 dias. Para além disso, para cada *Deal Market* não existe um padrão de qual *Macro-Industry* apresenta maiores e menores valores de *onboarding*.

Tomando como ponto de partida o *Deal Type* voltou-se a comparar a duração média de *onboarding* agora consoante o *Deal Type* a que o cliente pertencia, esta relação encontra-se na Tabela 23. Na Tabela 12 estão presentes objetos das variáveis *Segment*.

Tabela 23 - Resultados relativos à média da duração de *onboarding* consoante o *Deal Type* e o *Deal Market*

Média de Onboarding Duration	Rótulos de Coluna					
Rótulos de Linha	Comprador	Parceiro	Vendedor	Sem dados	Total Geral	
BR	65,07	28,00	55,43	48,68	57,99	
ES	59,22		52,25		58,19	
LATAM	35,83		40,00		36,88	
PALOP	119,50		102,00	63,50	98,64	
PT	58,43		68,79	40,21	60,68	
ROW	40,50	32,00	32,00	6,00	30,50	
UK	55,29		38,38	18,00	44,50	
Total Geral	60,33	30,00	65,36	44,73	60,41	

Desta relação, retira-se que a duração do *onboarding* comparando o Deal Type de Vendedor e Comprador é muito semelhante, há apenas uma diferença de 5 dias, em média. Destaca-se a duração média de *onboarding* elevada em PALOP dos compradores, com 119,50 dias. Os clientes Parceiros apresentam uma duração média de *onboarding* bastante boa de 30 dias.

A Tabela 24 apresenta a duração média de *onboarding* consoante o segmento da empresa, avaliando em que rondas de investimentos o cliente foi adquirido e se foi antes ou depois da pandemia de COVID-19. A Tabela 12 diz respeito aos objetos da variável *Segment*.

Tabela 24 - Resultados relativos à média da duração de *onboarding* consoante o *Segment*, as rondas de investimento e a pandemia Covid-19

Média de Duração de Onboarding (em dias)		Rótulos de Coluna									
Rótulos de Linha		Educação	Gestão de instalações	Hospitalidade	Indústria	Outros	Retalho	Sem dados	Centro comercial	Assistência Técnica	Total Geral
Customer após 3ª ronda de investimento						18,33	40,00	27,67		36,11	29,18
Customer after Covid-19						18,33	40,00	27,67		36,11	29,18
Customer até 1ª ronda investimento		75,00	182,17	75,72	81,33	57,00	129,29	43,75		111,59	104,90
Customer before Covid-19		75,00	182,17	75,72	81,33	57,00	129,29	43,75		111,59	104,90
Customer entre 1ª ronda investimento e 2ª ronda de investimento		53,00	39,43	37,96	56,75	63,50	121,80	127,00	104,00	75,67	60,82
Customer before Covid-19		53,00	39,43	37,96	56,75	63,50	121,80	127,00	104,00	75,67	60,82
Customer entre 2ª ronda investimento e 3ª ronda de investimento		47,50	69,30	53,72	47,55	60,08	36,50	29,17	23,50	52,28	52,26
Customer after Covid-19		46,00	72,56	46,26	50,28	55,39	36,50	29,17	23,50	48,30	49,03
Customer before Covid-19		52,00	54,67	72,60	35,25	74,17				66,67	65,19
Sem dados			19,00	113,00		23,50	63,00	8,50		38,57	46,04
Sem dados			19,00	113,00		23,50	63,00	8,50		38,57	46,04
Total Geral		57,89	89,96	56,69	52,31	54,18	89,18	30,45	50,33	62,95	60,41

Da análise da tabela Tabela 24, é possível concluir que *Facility Manager* (Gestão de Instalações) apresenta a maior duração média de *onboarding* com o valor de 89,96 dias e o Centro Comercial a menor duração com 50,33 dias. Observa-se também que é após a 3.^a ronda de investimento que a duração média de *onboarding* é mais baixa, com 28,19 dias e mais elevada antes da 1.^a ronda de investimento (104,90 dias). Esta tabela mostra bem a diferença que a pandemia COVID-19 fez na duração média de *onboarding*, porque este diminuir, isto deveu-se porque aos *onboardings* começaram a ser realizados em regime *online*. Retalho, por sua vez, apresenta um aumento na duração média de *onboarding* após o último investimento de 36,50 dias para 40 dias.

De modo a entender a relação entre a Classificação da Duração do *onboarding* e a Classificação do *Setup Fee*, tendo como valores a contagem do número de *onboarding* criou-se a Tabela 25. Os objetos da variável Classificação Duração do *onboarding* estão presentes na Tabela 9. Os objetos da variável Classificação *Setup Fee* estão presentes na Tabela 7.

Tabela 25 - Resultados relativos ao número de *onboardings* consoante a Classificação da Duração do *onboarding* e a Classificação do *Setup Fee*

Contagem de Onboarding Duration		Rótulos de Coluna					Total Geral
Rótulos de Linha	Duração Outlier	Sem dados	Duração Extraordinária	Duração Elevada	Duração Boa	Duração Excelente	
Onsite Basic	2		3	4		5	14
Onsite Premium	4		1	1	4	3	13
Remote Basic	29		40	20	74	75	238
Remote Premium	25		16	24	58	40	163
Sem dados	7		11	7	13	21	59
Total Geral	67		71	56	149	144	487

Interpretando a tabela anterior, concluiu-se que 49% dos clientes optam pelo pack *Remote Basic*. Sendo que desses, 48% apresenta um *onboarding* com duração ideal. Por outro lado, 44% dos *onboardings* apresentam uma duração ideal, seguido de 31% que demonstram um *onboarding* de duração média, seguido de 11% de duração alta e, por fim, 14% apresentam uma duração *outlier*.

A tabela dinâmica seguinte - Tabela 26 - permite observar a média da duração do *Onboarding*, consoante a saúde do cliente, *Health Score* e consoante a classificação que o CS lhe atribuiu, *CSM Score*. Os objetos relativos à variável Classificação *CSM Score* e à variável Classificação *Health Score* estão presentes na Tabela 5 e Tabela 6, respetivamente.

Tabela 26 - Resultados relativos à média da duração de *onboarding* consoante a *Health Score* e a *CSM Score*

Média de Onboarding Duration		Rótulos de Coluna				Total Geral
Rótulos de Linha	CSM Score Alta	CSM Score Baixa	CSM Score Média	Sem dados		
Health Score Alta	58,39	57,00	37,83		56,21	
Health Score Baixa		55,40	91,00	132,50	72,80	
Health Score média	52,35	68,00	91,78	23,50	69,73	
Sem dados	44,00	109,25		67,02	69,88	
Total Geral	57,74	69,00	64,39	67,90	60,41	

Da interpretação da tabela conclui-se que a *Health Score* Alta é a classe que apresenta a duração média do *onboarding* mais baixa. Enquanto a *CSM Score* Alta é também a classe que apresenta a

duração média do *onboarding* mais baixa. Também se pode concluir que uma *Health Score* Alta não traduz diretamente numa *CSM Score* melhor.

4.4.2 Visualização da Informação em *dashboards*

No presente trabalho, houve uma abordagem inicial de análise dos dados utilizando ferramentas já conhecidas da entidade *Microsoft Excel*, de seguida foram desenvolvidas *dashboards* em *Microsoft Power BI* tirando partido da interatividade da ferramenta e da sua ligação ao DW por forma a complementar os modelos analíticos desenvolvidos em *Microsoft Excel*.

No seguimento da análise realizada e para que fosse possível visualizar a solução de DW criada, foi desenvolvida uma *dashboard* que disponibiliza ao utilizador uma interface gráfica para projetar os dados presentes no DW. Esta ferramenta de visualização foi gerada no *Microsoft Power BI* que possibilita a criação de relatórios, construídos a partir de gráficos, tabelas ou outro tipo de visualização, havendo a possibilidade de aplicação de filtros. Este tipo de características permite ao utilizador elaborar uma análise dos dados consoante as necessidades e sob diferentes formas.

Os passos executados para a obtenção da *dashboard* estão apresentados na Figura 18.



Figura 18 - Passos realizados para a obtenção da *dashboard*

Como indicado na figura a primeira etapa passa pela obtenção dos dados. O *Microsoft Power BI* permite o carregamento dos dados através da base de dados criada em *MySQL Workbench*, isto é, consome os dados presentes no DW.

Após a ligação ao *Microsoft Power BI*, sucede-se a conceção da *dashboard*, isto é, ao desenho e construção de gráficos e tabelas selecionando a informação que se pretende apresentar e o tipo de visualização que melhor se adequa. O desenvolvimento da *dashboard* baseou-se na análise feita ao longo de todo o projeto.

Por último, e de forma a evidenciar os resultados obtidos a partir desta tecnologia, a Figura 19 apresenta um exemplo de uma *dashboard* concebida.

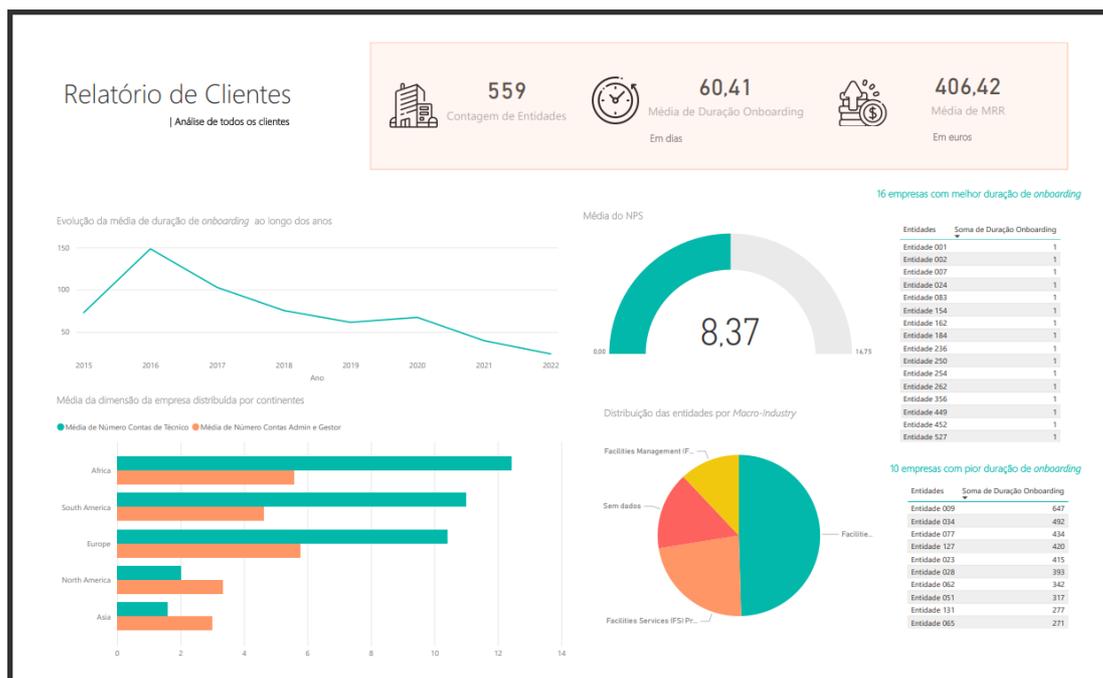


Figura 19 - Exemplo de *dashboard* desenvolvida em *Microsoft Power BI*

Pela análise da *dashboard* anterior, pode concluir-se que a média de duração de todos os *onboardings* até ao dia em que foi extraída a informação é 60,41 dias, a média do MRR é de 406,42 € e existiam até à data, 559 empresas clientes. Paralelamente é possível concluir que a duração média do *onboarding* tem diminuído ao longo do tempo, que 49,55% dos clientes pertencem à categoria de *Facilities Owners* (gráfico circular) e que as empresas de maior dimensão (soma das contas de técnicos, administrados e gestores) se encontram no continente africano.

Um outro exemplo de uma *dashboard* desenvolvida encontra-se na Figura 20. Esta já apresenta a aplicação de filtros.

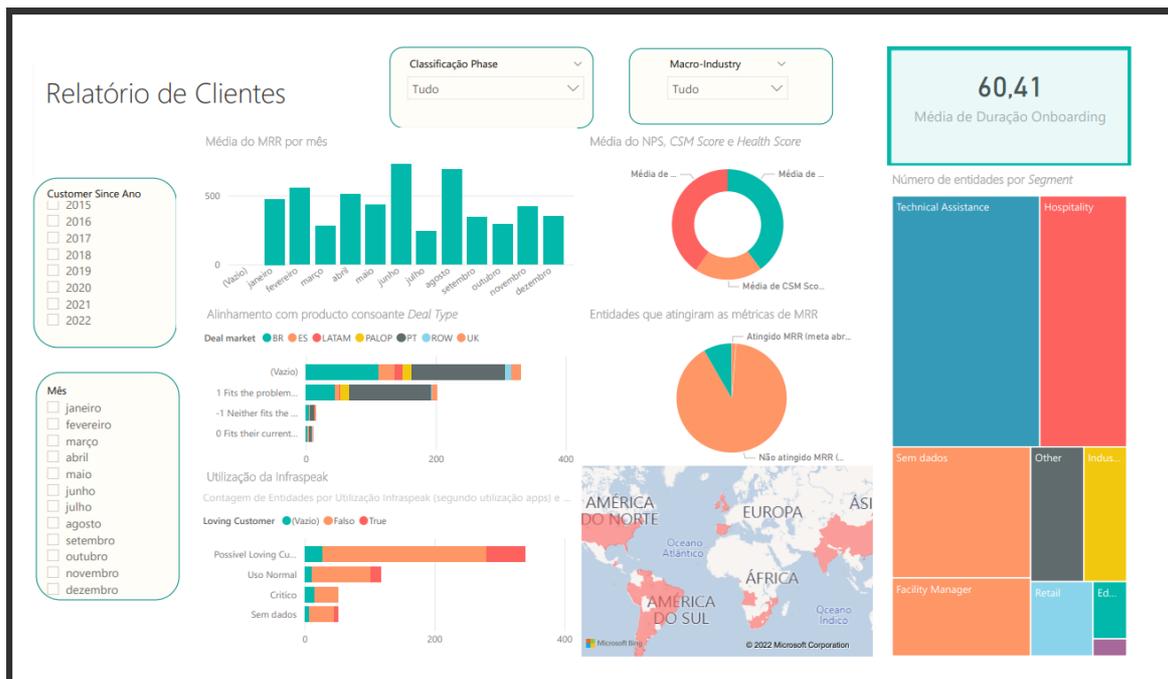


Figura 20 - Exemplo da adaptabilidade da *dashboard*

Este tipo de *dashboard* apresenta uma grande adaptabilidade, uma vez que responde à seleção de vários fatores com atualização dos dados de acordo com os fatores selecionados. Os dados da *dashboard* a seguir apresentados na Figura 21 são referentes aos clientes do tipo *Facilities Management Providers* no ano de 2021.

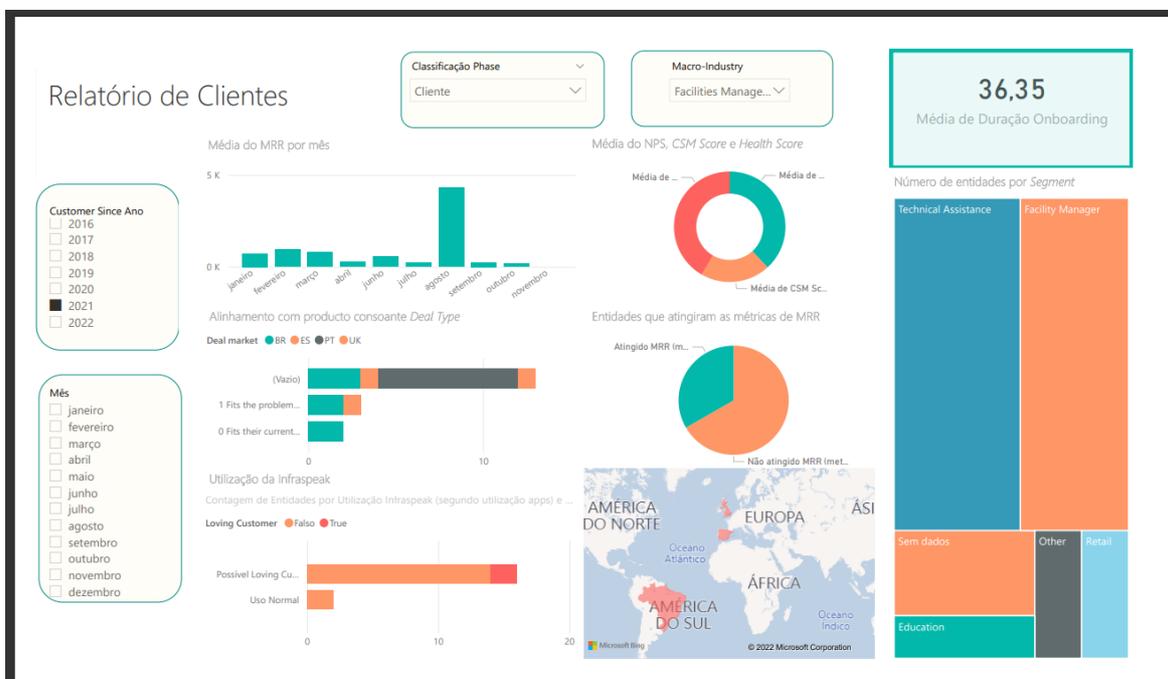


Figura 21 - Aplicação de filtros na *dashboard* desenvolvida

Da análise da Figura 21, juntamente com a *dashboard* da Figura 20 observa-se que a média de duração do *onboarding* foi de 36,35 dias e o MRR médio foi de 1,150 € para as 18 entidades que correspondem aos filtros. Destas 18 entidades, 7 eram de Assistência Técnica, 6 de Gestão de instalações, 1 de retalho, 1 de Educação, 1 na categoria “Outros” e 2 não possuíam dados. Observa-se também que o mês com a média de MRR mais elevado foi agosto com o valor de 4,303,74€.

Esta *dashboard* fornece também a informação da média da *Health Score* que é de 9,28, um NPS médio de 8,5 e uma *CSM Score* média de 4,50. Nota-se também que 14 entidades não são *Loving Customer*, mas têm possível para o ser. Percebe-se também que são clientes de Portugal, Espanha, Reino Unido e Brasil. Como é possível constatar, é possível retirar bastante informação de uma só *dashboard*.

Conclui-se então a última fase deste capítulo: a implementação da solução de BI. Através da utilização desta arquitetura tecnológica é possível tirar conclusões que serão abordadas no capítulo seguinte.

4.5 Fase 5 | Discussão de resultados

De modo a findar a metodologia Estudo de Caso, neste capítulo são discutidos os resultados obtidos de todo o trabalho anteriormente descrito.

Realizando, inicialmente, uma reflexão mais geral da implementação de uma solução BI no contexto da Infraspak, pode identificar-se como principais vantagens a elaboração de uma solução intuitiva com fácil acesso, uma solução que permite aumentar a qualidade de tomada de decisão e, assim, uma redução de custos da análise de dados e uma solução que permite um acesso rápido aos dados. Contudo a solução apresentada, demonstra algumas fraquezas, nomeadamente, a fonte de dados ser apenas uma, já que ficaria mais enriquecida com mais fontes, a falta de dados com qualidade existentes na fonte inicial e de não ser uma solução que seja atualizada diariamente.

Recorrendo à questão de investigação formulada na Introdução, sendo ela “*Quais os fatores que influenciam o processo de onboarding de um novo cliente?*” e toda a investigação realizada, pode apreender-se que variáveis como o segmento da empresa e o tipo de atividades de manutenção que executam, o mercado e/ou país a que pertencem podem influenciar a duração do processo de *onboarding*. Variáveis relativas ao dinheiro que uma empresa investe inicialmente na Infraspak não parece influenciar o seu processo de *onboarding*. Contudo, rondas de investimento ganhas por parte da Infraspak fazem com que consigam melhorar o processo de *onboarding*.

Concretizando, e segundo as dimensões seguidas no subcapítulo Tabela de dimensões, através da presente dissertação é possível uma análise crítica dos principais resultados.

No dia em que foram recolhidos os dados, dia 11 de abril de 2022, até à data, a Infrasppeak tem uma média de duração do *onboarding* de 60,41 dias, mas esta média foi variando. Relativamente ao tempo, foram avaliados os últimos 5 anos e no ano de 2018, a Infrasppeak adquiriu 41 clientes e possuía uma média de duração do *onboarding* de 75,44 dias, já no ano 2019, com 94 clientes a aderirem, a média foi de 61,48 dias. No ano de 2020 a uma média de duração do *onboarding* foi de 67,37 dias (101 clientes) e no ano de 2021 foi de 39,87 dias (173 clientes). No presente ano, a média estava em 24,19 dias correspondendo a 58 clientes.

Foi também avaliado em que mês os *onboardings* apresentavam uma maior duração, apesar de não haver um padrão, nos últimos 5 anos, março e junho foram ambos 3 vezes dos meses com a maior duração. Seguindo-se abril e fevereiro, ambos 2 vezes como dos meses mais lentos. Para chegar a esta conclusão, foram avaliados desde 2018 os 3 meses mais lentos de cada ano.

No que diz respeito ao local, os países que apresentam uma duração média maior de *onboardings* são Moçambique com 100 dias, Angola com 99,81 dias e Irlanda com 71,30 dias. Já em relação aos mercados onde os países estão integrados, PALOP apresenta uma duração média de *onboardings* de 98,64 dias (27 empresas), seguido de PT com 60,68 dias (238 empresas), ES com 58,99 dias (36 empresas) e BR com 57,99 dias de 164 casos. Explorando o caso de Portugal, Lisboa apresenta uma duração média de 67,43 dias (referente a 59 empresas), Vila Nova de Gaia a 91 dias (com 6 empresas) e Matosinhos com 111,80 dias (6 empresas).

Por último quanto à área de negócio, o segmento Gestão de instalações (*Facility Manager*), que diz respeito a 10% dos clientes da Infrasppeak, expõe uma duração de *onboarding* com maior duração com 89,96 dias. Relativamente à indústria em que se inserem, Hospitalidade que corresponde a 23,61% dos clientes da Infrasppeak, apresenta um *onboarding* médio de 56,69 dias e *Hard Facilities Services*, correspondente a 20,57% dos clientes, possui um *onboarding* de 67,47 dias. No caso da *Macro-Industry*, 49,55% das empresas são *Facilities Owners* (Proprietários das Instalações) e estas apresentam um *onboarding* de 56,22 dias, em média e 22,90% são *Facilities Services Provides* (Prestadores de Serviços em Instalações) e possuem um *onboarding* médio de 65,81 dias.

Uma outra métrica refere que 47% dos *onboarding* não são concluídos no tempo esperado, o tempo esperado é aproximadamente 30 dias. Estas foram consideradas as métricas mais importantes a serem expostas.

Fatores que parecem influenciar o processo de *onboarding*, numa avaliação mais subjetiva, são o nível de conhecimento tecnológico que o gestor interno da plataforma Infrasppeak tem, a sua proatividade, tempo disponível e vontade de aprender e, por último, a diferença de termos de manutenção que a

plataforma Infraspark utiliza e a que a empresa cliente utiliza. Por último, outro fator que influencia o processo de *onboarding* diz respeito à qualidade e clareza de informação que é submetida na plataforma, isto é, quando um cliente não apresenta a informação relativa aos seus processos de forma documentada, o seu processo de *onboarding* será mais demorado e menos fluído.

5. CONCLUSÕES

O presente capítulo analisa de forma crítica o trabalho desenvolvido, as limitações sentidas e, por fim, são mencionadas sugestões para possíveis desenvolvimentos futuros no campo e na investigação desenvolvida.

A solução de *Business Intelligence* (BI) apresentada tinha como função auxiliar o processo de tomada de decisão pela empresa, apresentando o panorama geral dos dados do cliente.

Todo o processo de estudo do contexto, tratamento de dados, criação do *Data Warehouse* (DW) e a criação de *dashboards* dinâmicas, procuraram obter informações e conhecimento ao utilizador de forma mais simples e intuitiva.

5.1 Considerações finais

A presente dissertação tinha como objetivo a recolha, processamento e análise de dados de modo a compreender que fatores influenciavam o processo de *onboarding* da Infraspark. Posto isto, é necessário compreender quais os resultados obtidos e como estes estão relacionados com os objetivos definidos na Introdução. A Tabela 27 apresenta então a relação entre os Objetivos do Projeto e os Resultados Obtidos.

Tabela 27 - Matriz de relação entre Objetivos do Projeto e Resultados Obtidos

		Resultados Obtidos			
		Descrição detalhado do processo de <i>onboarding</i> com os respetivos problemas	Conceção de um <i>Data Warehouse</i>	Desenvolvimento modelos analíticos	Indicadores para demonstração dos resultados
Objetivos do Projeto	Analisar processo de <i>onboarding</i> de novos clientes	x			
	Extrair dados quantitativos a partir da análise dos mesmos			x	x
	Modelar um banco de dados dimensional para armazenamento de dados		x		x
	Apresentar uma <i>dashboard</i>			x	x
	Avaliar os resultados obtidos			x	x

Como referido também na Introdução, a motivação desta dissertação passava por melhorar o processo de *onboarding* conforme o crescimento da Infraspark, uma vez que um aumento do número de clientes da empresa ao longo do tempo não pode significar diretamente o crescimento da equipa de *Customer Success*. A resposta a este problema pode passar por agrupar *onboarding* de diferentes clientes, mas para isso é necessária uma solução BI que auxilie na tomada de decisão.

Numa fase inicial, realizou-se a revisão da literatura, abordando todos os conceitos teóricos relevantes, nomeadamente *Software as a Service*, *Data Science*, *Data Warehouse*, *Business Intelligence* e Sucesso do Cliente. Assim, através do estudo dos conceitos mencionados foi possível concluir que para trabalhar com processos de *onboarding* é essencial conhecer o meio envolvente. Para além disso para avaliar os fatores que influenciam o processo de *onboarding* é necessário ter acesso aos dados, interpretá-los e extrair informação relevante. Esta premissa exigiu a existência de um DW, onde fosse possível armazenar dados para posterior processamento e a criação de uma *dashboard*.

Desta forma, formularam-se os requisitos necessários para a construção desta solução e, selecionou-se o modelo em estrela implementado com recurso à ferramenta *MySQL Workbench*. Por fim, visualizaram-se os dados presentes na solução através do *Microsoft Power BI*. A partir dos resultados obtidos e da sua análise, pode concluir-se que a solução apresentada contém a informação necessária para auxiliar o apoio na tomada de decisão.

Com este trabalho, os contributos da solução fornecidos à empresa passam por permitir analisar os dados históricos relativos ao *onboarding* de novos clientes, calcular indicadores de desempenho, suportar na tomada de decisão de forma visual e obter uma perceção global dos dados dos clientes.

Posto isto, com o desenvolvimento desta dissertação foram acompanhados 14 processos de *onboarding* de empresas distintas para compreender melhor este processo e foram tratadas 137 variáveis de dados recolhidos dos clientes, das quais 25 tratadas com maior detalhe. Realizou-se 2 vezes uma análise estatística univariada e implementou-se um DW com 3 tabelas de dimensões e uma tabela de factos. Pela análise das tabelas dinâmicas e das *dashboards* foram recolhidas as métricas apresentadas em Fase 5 | Discussão de resultados.

Contudo, não foi possível aplicar esta solução no contexto real por questões internas e, por isso, não foi possível obter um impacto mensurável. Esta solução teria de ser implementada pela equipa *Customer Success* da Infraspark, primeiro com uma explicação por parte do investigador e, por fim, pela equipa para definir os próprios KPIs e projeções futuros e como auxílio na tomada de decisão aquando de um novo *onboarding*.

Com o desenvolvimento desta dissertação adquiriu-se experiência de pesquisa, de construção de um DW e de soluções e conceitos inerentes a BI. Para além disso, o investigador adquiriu maior competência na utilização das ferramentas utilizadas: Infraspark, Planhat, PAST, *Microsoft Excel*, TerraER, *MySQL Workbench* e *Microsoft Power BI*. Por fim, o desenvolvimento de *soft-skills* necessárias no contexto empresarial, como comunicação, resiliência, organização e empatia

Concluindo e como forma a responder ao tema desta dissertação, apresentam-se as seguintes sugestões para o processo de *onboarding* após a análise realizada:

- Recorrer a plataformas de adaptação digital que possam ser integradas no *software* Infraspark, para que seja possível fornecer informações aos clientes sobre as funcionalidades do produto, através de guias, mensagens, tours de *onboarding*, lista de verificação.
- Criar um sistema de revisão para analisar métricas propostas tanto pelo cliente como por *Customer Success*, avaliando o sucesso do processo e melhorá-lo.
- Realizar uma auditoria na revisão dos termos utilizados na plataforma Infraspark para que estão vão em encontro dos termos utilizados pelos clientes.
- Tornar as sessões de *onboarding* mais interativas para que o cliente se sinta mais envolvido.
- Adiar o processo de *onboarding* até que o cliente possua toda a informação necessária para submeter na plataforma Infraspark.

5.2 Limitações

Numa fase inicial do projeto, foram identificados um conjunto de fatores que poderiam influenciar o sucesso da presente dissertação. A Tabela 28 apresenta, de forma organizada pelo nível de gravidade, os riscos associados a esta dissertação.

Para cada linha, que representa um risco diferente, está associada uma probabilidade (P), isto é, a probabilidade estimada do risco acontecer, um impacto (I) que representa o impacto que terá no projeto e uma ação para reduzir a gravidade do risco. A coluna ocorreu permite informar se o risco levantado no início ocorreu ou não ao longo da dissertação.

A coluna da gravidade (G) corresponde ao resultado da multiplicação entre a probabilidade e o impacto. A probabilidade e o impacto apresentam uma escala de 1 a 5, sendo 1 o valor mínimo e 5 o máximo.

Tabela 28 - Riscos associados à dissertação

Número	Risco	P	I	G	Ocorreu	Ações Realizadas
1	Falta de experiência e conhecimento em projetos da área.	4	4	16	Sim	Debater temas com o orientador, bem como esclarecer dúvidas. Aliar também a leitura de artigos científicos, livros, vídeos.
2	Falta de informação	3	5	15	-	Reunir com orientadores e reestruturar plano .
3	Problemas na aplicação	2	5	10	Sim	Identificar problemas e abordar pessoas com mais experiência na área.
4	Os resultados obtidos não irem ao encontro ao expectável.	2	4	8	-	Procurar novas abordagens e adiar a data prevista dos entregáveis.
5	Escolha da metodologia inadequada	1	4	4	-	Ler sobre metodologias mais relevantes para o contexto e discutir o tópico com o orientador.
6	Dificuldade de comunicação com o orientador.	1	4	4	-	Agendar reuniões quinzenais com o orientador e utilizar plataformas de comunicação adequadas.
7	Planeamento inadequado	1	4	4	Sim	Definir uma periodicidade para rever o plano e assim manter as expectativas.
8	Perda de documentos	1	4	4	-	Armazenar documentos na <i>cloud</i> .

Para além dos riscos ocorridos, ao longo deste projeto foram sentidas algumas limitações que influenciaram de forma direta ou indireta o decorrer do projeto e os resultados obtidos. Uma das principais limitações está relacionada a falta de conhecimento e *background* do investigador no *software MySQL Workbench*. Outra limitação passou pelo enquadramento da metodologia em todo o processo. Por último, uma outra limitação foi identificar a melhor estratégia que na prática poderia servir de apoio na tomada de decisões da organização.

5.3 Trabalho futuro

A solução desenvolvida proporciona ao utilizador a capacidade de analisar dados relativos aos clientes. Contudo, esta solução é um protótipo das possíveis funcionalidades que o modelo do sistema BI desenvolvido pode vir a desempenhar. Desta forma e como continuidade desta dissertação, as propostas futuras passariam por incluir mais fontes de dados no sistema e aperfeiçoar a solução implementada e incluir modelos preditivos.

Para além disso, seria importante procurar junto dos utilizadores a sua satisfação da plataforma, recolher o seu *feedback* e realizar as respetivas correções e melhorar a solução.

Por último, alguns dos temas abordados nesta dissertação de mestrado exigem uma investigação mais profunda e, investir na previsão de dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- Alves, J. (2015). *Sistema de Business Intelligence no Projeto Educativo de Guimarães*.
- Aquila, C., Tria, F., Lefons, E., & Tangorra, F. (2007). *Business intelligence systems: a comparative analyses*.
- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., & Zaharia, M. (2010). A view of cloud computing. In *Communications of the ACM* (Vol. 53, Issue 4, pp. 50–58). <https://doi.org/10.1145/1721654.1721672>
- Azeroual, O., & Theel, H. (2018). The Effects of Using Business Intelligence Systems on an Excellence Management and Decision-Making Process by Start-Up Companies: A Case Study. *INTERNATIONAL JOURNAL OF MANAGEMENT SCIENCE AND BUSINESS ADMINISTRATION*, 4(3), 30–40. <https://doi.org/10.18775/ijmsba.1849-5664-5419.2014.43.1004>
- Baxter, P., & Jack, S. (2015). Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. *The Qualitative Report*. <https://doi.org/10.46743/2160-3715/2008.1573>
- Bhardwaj, S., Jain, L., & Sandeep, J. (2010). An Approach for Investigating Perspective of Cloud Software-as-a-Service (SaaS). *International Journal of Computer Applications*. <https://doi.org/10.5120/1450-1962>
- Bogza, R. M., & Zaharie, D. (2008). *BUSINESS INTELLIGENCE AS A COMPETITIVE DIFFERENTIATOR*. <https://doi.org/10.1109/AQTR.2008.4588724>
- Branski, R. M., Lima, O. F., & Franco, R. A. C. (2010). *Metodologia de estudo de casos aplicada à logística*. <https://www.researchgate.net/publication/277598822>
- Brynjolfsson, E., Hitt, L., & Kim, H. (2011). *Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?* https://a51.nl/sites/default/files/pdf/SSRN_id1819486.pdf
- Carlos P. Caldeira. (2007). *A usabilidade das bases de dados relacionais*. <https://dspace.uevora.pt/rdpc/bitstream/10174/867/1/usabilidade.pdf>
- Cavusgil, S. T., & Knight, G. (2015). The born global firm: An entrepreneurial and capabilities perspective on early and rapid internationalization. In *Journal of International Business Studies* (Vol. 46, Issue 1, pp. 3–16). Palgrave Macmillan Ltd. <https://doi.org/10.1057/jibs.2014.62>
- Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011). The case study approach. *BMC Medical Research Methodology*, 11. <https://doi.org/10.1186/1471-2288-11-100>

- Dhar, V. (2012). *Data Science and Prediction*. <http://hdl.handle.net/2451/31553>
- Diário da República Eletrónico. (2019, March 4). *Lei n.º 58/2019, de 8 de agosto*. <https://dre.pt/dre/detalhe/lei/58-2019-123815982>
- Dilmegani, C. (2021, May 15). *Customer Onboarding: Tools, Techniques, & Best Practices in 2022*. <https://research.aimultiple.com/customer-onboarding/>
- Divisi, D., di Leonardo, G., Zaccagna, G., & Crisci, R. (2017). Basic statistics with Microsoft Excel: A review. *Journal of Thoracic Disease*, 9(6), 1734–1740. <https://doi.org/10.21037/jtd.2017.05.81>
- Drebes Pedron, C., Lupi, R. M., & Gab, -. (2008). *O método de investigação ESTUDO DE CASO*.
- Drkušić, E. (2016, April 14). *The Star Schema*. Vertabela. <https://vertabelo.com/blog/data-warehouse-modeling-the-star-schema/>
- Farris, P. W., Bendle, N. T., Pfeifer, P. E., & JReibstein, D. (2010). *MARKETING METRICS - The definitive guide to measuring marketing performance* (2nd ed.). Tim Moore.
- Fawcett, T., & Provost, F. (2013). *Data Science for Business* (M. Loukides & M. Blanchette, Eds.).
- Ferreira, J., Almeida, F., & Monteiro, J. (2017). Building an Effective Data Warehousing for Financial Sector. *Automatic Control and Information Sciences*, 3(1), 16–25. <https://doi.org/10.12691/acis-3-1-4>
- Filipe, J., Albuquerque, L., Fernandes, J., & Bernardino, R. (2013). *Desenho e implementação de um data warehouse para a empresa AdClick*.
- Forrester. (2014). *How To Get Started With Customer Success Management*. http://www.gainsight.com/wpcms/wp-content/uploads/2014/11/WP-Gainsight_Forrester_Getting_Started_Whitepaper.pdf
- Fortin, M.-F. (1999). *O Processo de Investigação - da concepção à realização*.
- Hammer, O., Harper, D. A. T., & Ryan, P. D. (2001). *PAST: Paleontological Statistics Software Package for Education and Data Analysis*. http://palaeo-electronica.org/2001_1/past/issue1_01.htm.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). <https://doi.org/https://doi.org/10.1016/C2009-0-61819-5>
- Hellerstein, J. M. (2008). *Quantitative Data Cleaning for Large Databases*. <http://db.cs.berkeley.edu/jmh>
- Hughes, R. (2012). *Agile Data Warehousing Project Management: Business Intelligence Systems Using Scrum*. Morgan Kaufmann.

- Iberdrola, S. A. (2022). *As “startups” sonham com unicórnios?*
<https://www.iberdrola.com/inovacao/empresas-unicornio-startups-tecnicas>
- Inmon, W. H. (2005). *Building the Data Warehouse* (4th ed.). Wiley.
<https://ia800202.us.archive.org/9/items/2005BuildingTheDataWarehouse4thEditionWilliamH.Inmon/2005%20-%20Building%20The%20Data%20Warehouse%20%284th%20Edition%29%20%28William%20H.%20Inmon%29.pdf>
- Jensen, C. S., Pedersen, T. B., & Thomsen, C. (2010). Multidimensional Databases and Data Warehousing. *Synthesis Lectures on Data Management*, 2(1), 1–111.
<https://doi.org/10.2200/s00299ed1v01y201009dtm009>
- Johansson, R. (2007). On_Case_Study_Methodology. *Open House International* 32(3):48-54.
<https://doi.org/10.1108/OHI-03-2007-B0006>
- Kalathas, I., Papoutsidakis, M., & Drosos, C. (2021). Business Intelligence and Machine Learning Methods for Predictive Maintenance in Greek railways. *Open Journal of Applied Sciences*, 11(01), 20–35. <https://doi.org/10.4236/ojapps.2021.111a003>
- Kimball, R., & Roos, M. (2006). *The Data Warehouse Toolkit: the definitive guide to dimensional modeling* (3rd ed.). John Wiley & Sons, Inc. https://aatinegar.com/wp-content/uploads/2016/05/Kimball_The-Data-Warehouse-Toolkit-3rd-Edition.pdf
- Kshatri, V. (2022). *Customer Success in SaaS Industry-what is it and how can it be achieved?*
<https://webthesis.biblio.polito.it/22881/1/tesi.pdf>
- Learn Kraft. (2022). *How to Construct a Frequency Distribution*. Finance Train.
<https://financetrain.com/how-to-construct-a-frequency-distribution>
- Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, 80(6), 69–96. <https://doi.org/10.1509/jm.15.0420>
- Lewis, C., & Mehmet, M. (2020). Does the NPS® reflect consumer sentiment? A qualitative examination of the NPS using a sentiment analysis approach. *International Journal of Market Research*, 62(1), 9–17. <https://doi.org/10.1177/1470785319863623>
- Mehta, N., Steinman, D., Murphy Lincoln, & Martinez, M. (2016). *Customer Success: How Innovative Companies Are Reducing Churn and Growing Recurring Revenue* (1st ed.). Wiley.
- Mell, P., & Grance, T. (2011). *The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology*.
<http://faculty.winthrop.edu/domanm/csci411/Handouts/NIST.pdf>

- Microsoft. (2002, July 1). *Data Cleaning*. <https://www.microsoft.com/en-us/research/project/data-cleaning/?from=http%3A%2F%2Fresearch.microsoft.com%2Fen-us%2Fprojects%2Fdatacleaning%2F>
- Microsoft. (2022a). *DIST.NORMAL (função DIST.NORMAL)*. <https://support.microsoft.com/pt-pt/office/dist-normal-fun%C3%A7%C3%A3o-dist-normal-edb1cc14-a21c-4e53-839d-8082074c9f8d>
- Microsoft. (2022b). *PowerBI*. <https://learn.microsoft.com/pt-pt/power-bi/fundamentals/power-bi-overview>
- Mortier, G. (2021, December 7). *What Are Facts and Dimensions in a Data Warehouse?* Vertabela. <https://vertabelo.com/blog/facts-dimensions-data-warehouse/>
- Moura, A. H. (2017). *Criação de um Data Warehouse empresarial*. <https://run.unl.pt/bitstream/10362/32048/1/TGI0127.pdf>
- Nirpaz, G., & Pizarro, F. (2016). *Farm Don't Hunt: The Definitive Guide to Customer Success*.
- Oracle. (2022). *CREATE PROCEDURE and CREATE FUNCTION Statements*. <https://dev.mysql.com/doc/refman/8.0/en/create-procedure.html>
- ORACLE. (2022). *MySQL — MySQL 5.7 Reference Manual — 23.5 Using Views*. <https://dev.mysql.com/doc/refman/5.7/en/views.html>
- Pendo.io. (2022). *Customer health score*. <https://www.pendo.io/glossary/customer-health-score/>
- Pereira, P., Cunha, J., & Fernandes, J. (2020). *On Understanding Data Scientists*.
- Pizarro, F., & Kooij, J. van der. (2017). *Blueprints for a SaaS Sales Organization* (2nd ed.). CreateSpace Independent Publishing Platform.
- Portela, F. (2021). Data science and knowledge discovery. In *Future Internet* (Vol. 13, Issue 7). MDPI AG. <https://doi.org/10.3390/fi13070178>
- ProductPlan. (2022). *Monthly Recurring Revenue (MRR)*. <https://www.productplan.com/glossary/monthly-recurring-revenue/>
- Rainardi, V. (2008). *Building a Data Warehouse: with examples in SQL Server* (J. Pepper, Ed.). Apress. www.apress.com.
- Richard. (2017, October 5). *How to Make an Awesome Inventory Management Application in PHP and MySQL*.
- Rocha, H., & Terra, R. (2022). *TerraER - Ferramenta de Modelagem Entidade-Relacionamento*. <https://www.methodsandtools.com/tools/terraer.php>
- Rocha, R. (2007). *Bases de Dados Parte I Conceitos Básicos*.

- Santos, M. (2020). *Can case study methodology be applied to occupational health?* (Vol. 10).
<https://www.rpsa.pt/a-metodologia-de-estudo-de-caso-pode-ser-aplicada-a-nivel-da-saude-ocupacional/>
- Sharib. (2012). *Using Pivot Tables for Data Organization-Data Warehouse-Lab Manual*.
www.doccity.com
- Sjöberg, J., & Winbäck, H. (2017). *Designing the user onboarding process in analytics software Taking an omnichannel perspective*.
<https://publications.lib.chalmers.se/records/fulltext/250057/250057.pdf>
- Suehring, Steve. (2002). *MySQL bible*. Wiley Pub.
- Teixeira, A. L. (2016). *Análise dos dados de inquéritos sociológicos: estatísticas univariada, bivariada e multivariada*.
https://research.unl.pt/ws/portalfiles/portal/3315516/Ana_Lucia_Teixeira_analise_dados_met
- Teixeira, R., Afonso, F., Oliveira, B., Portela, F., & Santos, M. F. (2014, October). *Business Intelligence to improve the quality of Local Government Services: Case-Study in a Local Government Town Hall*. <https://repositorium.sdum.uminho.pt/bitstream/1822/31320/4/2014-KMIS-40.pdf>
- The Customer Success Association. (2021). *The Definition of Customer Success*.
<https://www.customersuccessassociation.com/library/the-definition-of-customer-success/>
- Tran, H. (2020). *Developing a Successful Onboarding for a B2B Software-as-a-Service Company. Case: Company X* [University of Applied Sciences].
https://www.theseus.fi/bitstream/handle/10024/340098/Thesis_Hien%20Tran_Final%20version.pdf?sequence=2&isAllowed=y
- Tupper, C. (2011). *Data Architecture*. Morgan Kaufmann.
- Underwood, J. (2014, March 10). *Does Excel Power Pivot Replace the Data Warehouse?*
<https://www.itprotoday.com/analytics-reporting/does-excel-power-pivot-replace-data-warehouse>
- Vassiliadis, P. (1998). Modeling multidimensional databases, cubes and cube operations. *Scientific and Statistical Database Management - Proceedings of the International Working Conference*, 53–62.
<https://doi.org/10.1109/ssdm.1998.688111>
- Wittel. (2017). *Business Intelligence: como revolucionar o atendimento ao cliente*.
<https://blog.wittel.com/business-intelligence/>
- Wu, L., Barash, G., & Bartolini, C. (2007). *A Service-oriented Architecture for Business Intelligence*. <https://ieeexplore.ieee.org/document/4273437>

APÊNDICE I – *BOXPLOT* DAS VARIÁVEIS

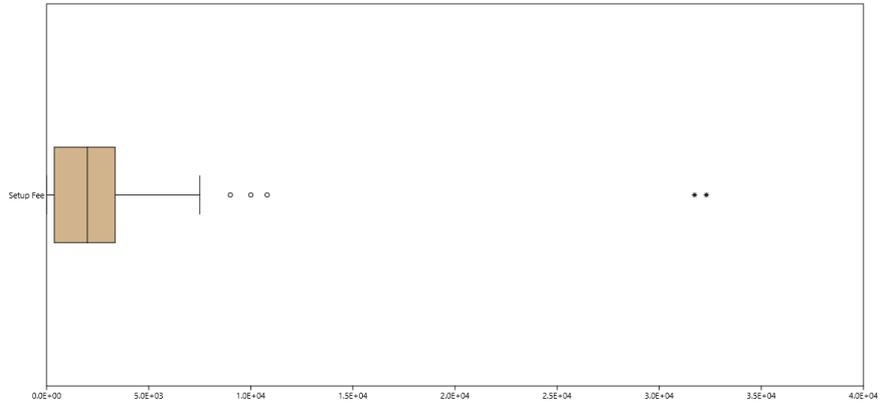


Figura 22 - *Boxplot* da variável *Setup Fee*

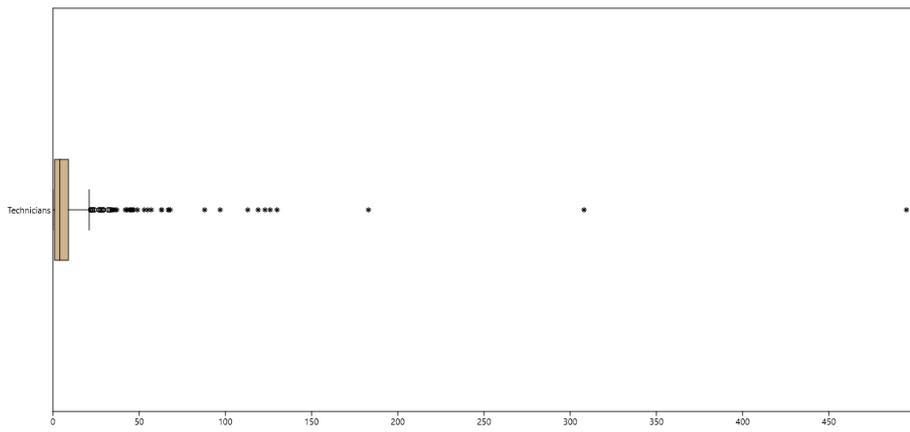


Figura 23 - *Boxplot* da variável *Total Technicians*

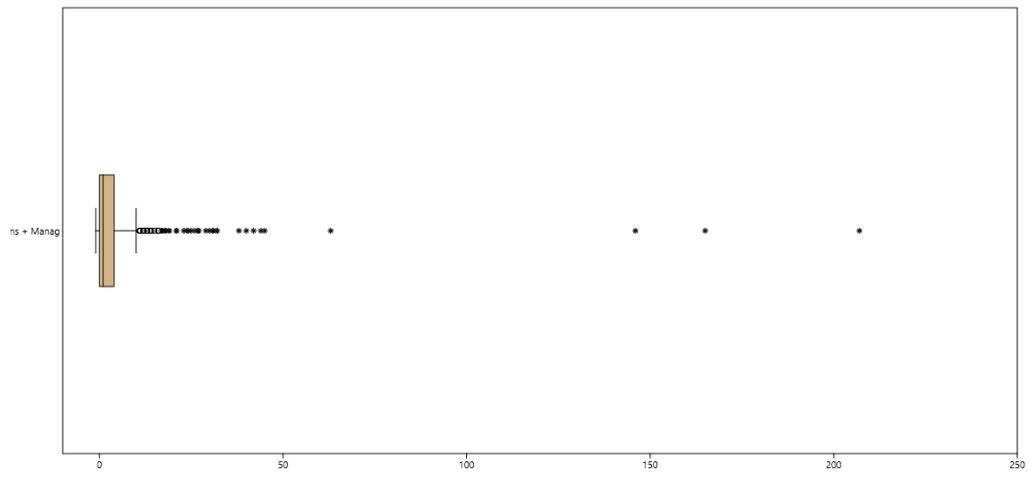


Figura 24 - Boxplot da variável *Total Admins+ Managers*

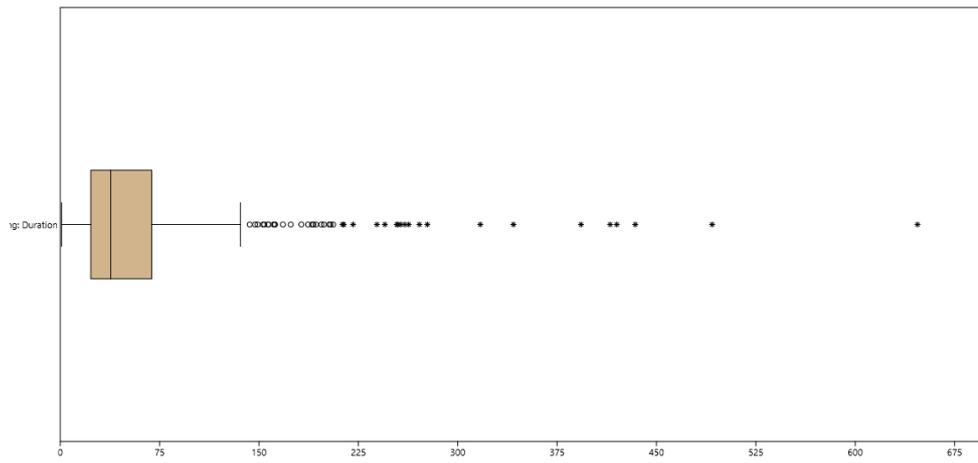


Figura 25 - Boxplot da variável *Onboarding Duration*

APÊNDICE II – CRIAÇÃO DAS TABELAS DO MODELO FÍSICO

```
-- MySQL Workbench Forward Engineering

SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0;
SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS,
FOREIGN_KEY_CHECKS=0;
SET @OLD_SQL_MODE=@@SQL_MODE,
SQL_MODE='ONLY_FULL_GROUP_BY,STRICT_TRANS_TABLES,NO_ZERO_IN
_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_ZERO,NO_ENGINE_SUB
STITUTION';

-- Schema mydb
-----
-- Schema esquema_estrela
-----
CREATE SCHEMA IF NOT EXISTS `esquema_estrela` DEFAULT
CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci ;
USE `esquema_estrela` ;
-----
-- Table `esquema_estrela`.`dimensao_area_negocio`
-----
CREATE TABLE IF NOT EXISTS
`esquema_estrela`.`dimensao_area_negocio` (
  `ID_Area_Negocio` INT NOT NULL,
  `Segment` VARCHAR(20) NULL DEFAULT NULL,
  `Industry` VARCHAR(33) NULL DEFAULT NULL,
  `Macro_Industry` VARCHAR(36) NULL DEFAULT NULL,
  `Deal_type` VARCHAR(9) NULL DEFAULT NULL,
  PRIMARY KEY (`ID_Area_Negocio`))
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8mb4
```

```

-----
-- Table `esquema_estrela`.`dimensao_local`
-----
CREATE          TABLE          IF          NOT          EXISTS
`esquema_estrela`.`dimensao_local` (
  `ID_Local` INT NOT NULL,
  `Cidade` VARCHAR(45) NULL DEFAULT NULL,
  `Distrito` VARCHAR(45) NULL DEFAULT NULL,
  `Regiao` VARCHAR(45) NULL DEFAULT NULL,
  `Pais` VARCHAR(45) NULL DEFAULT NULL,
  `Continente` VARCHAR(45) NULL DEFAULT NULL,
  PRIMARY KEY (`ID_Local`),
  INDEX `fk_ID_Local` (`ID_Local` ASC) VISIBLE)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8mb4
COLLATE = utf8mb4_0900_ai_ci;
-----
-- Table `esquema_estrela`.`dimensao_tempo`
-----
CREATE          TABLE          IF          NOT          EXISTS
`esquema_estrela`.`dimensao_tempo` (
  `ID_tempo` INT NOT NULL,
  `db_date` DATE NOT NULL,
  `year` INT NOT NULL,
  `month` INT NOT NULL,
  `day` INT NOT NULL,
  `quarter` INT NOT NULL,
  `week` INT NOT NULL,
  `day_name` VARCHAR(9) NOT NULL,
  `month_name` VARCHAR(9) NOT NULL,
  PRIMARY KEY (`ID_tempo`),
  UNIQUE INDEX `td_ymd_idx` (`year` ASC, `month` ASC,
`day` ASC) VISIBLE,
  UNIQUE INDEX `td_dbdate_idx` (`db_date` ASC) VISIBLE)

```

```

ENGINE = MyISAM
DEFAULT CHARACTER SET = utf8mb4
COLLATE = utf8mb4_0900_ai_ci;
-----
-- Table `esquema_estrela`.`tabela_factos`
-----
CREATE TABLE IF NOT EXISTS `esquema_estrela`.`tabela_factos`
(
  `ID_Facto` INT NOT NULL AUTO_INCREMENT,
  `ID_Local` INT NOT NULL,
  `ID_Area_negocio` INT NOT NULL,
  `ID_Onboarding_kick_off_date` INT NOT NULL,
  `ID_Customer_Since` INT NOT NULL,
  `Duracao_Onboarding` VARCHAR(45) NULL DEFAULT NULL,
  `Onboarding_Data_esperada` VARCHAR(45) NULL DEFAULT NULL,
  `MRR` DECIMAL(6,2) NULL DEFAULT NULL,
  `Setup_Fee` DECIMAL(7) NULL DEFAULT NULL,
  `Total_Tecnicos` INT NULL DEFAULT NULL,
  `Total_Admins_Gestores` INT NULL DEFAULT NULL,
  `Fase` VARCHAR(58) NULL DEFAULT NULL,
  `Entidade` VARCHAR(45) NULL,
  PRIMARY KEY (`ID_Facto`),
  INDEX `fk_ID_Local_idx` (`ID_Local` ASC) VISIBLE,
  INDEX `fk_ID_Area_negocio_idx` (`ID_Area_negocio` ASC)
  VISIBLE,
  INDEX `fk_ID_Onboarding_kick_off_idx`
  (`ID_Onboarding_kick_off_date` ASC) VISIBLE,
  INDEX `fk_ID_Customer_Since_idx` (`ID_Customer_Since` ASC)
  VISIBLE,
  CONSTRAINT `fk_ID_Local`
  FOREIGN KEY (`ID_Local`)
  REFERENCES `esquema_estrela`.`dimensao_local` (`ID_Local`)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
  CONSTRAINT `fk_ID_Area_negocio`
  FOREIGN KEY (`ID_Area_negocio`)

```

```

REFERENCES `esquema_estrela`.`dimensao_area_negocio`
(`ID_Area_Negocio`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
CONSTRAINT `fk_ID_Onboarding_kick_off`
    FOREIGN KEY (`ID_Onboarding_kick_off_date`)
    REFERENCES `esquema_estrela`.`dimensao_tempo` (`ID_tempo`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
CONSTRAINT `fk_ID_Customer_Since`
    FOREIGN KEY (`ID_Customer_Since`)
    REFERENCES `esquema_estrela`.`dimensao_tempo` (`ID_tempo`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8mb4
COLLATE = utf8mb4_0900_ai_ci;
-- -----
-- Table `esquema_estrela`.`temporaria`
-- -----
CREATE TABLE IF NOT EXISTS `esquema_estrela`.`temporaria` (
    `Duracao_Onboarding` VARCHAR(5) NULL DEFAULT NULL,
    `Onboarding_Data_esperada` VARCHAR(24) NULL DEFAULT NULL,
    `MRR` DECIMAL(19,14) NULL DEFAULT NULL,
    `Setup_Fee` DECIMAL(7,2) NULL DEFAULT NULL,
    `Total_Tecnicos` INT NULL DEFAULT NULL,
    `Total_Admins_Gestores` INT NULL DEFAULT NULL,
    `Fase` VARCHAR(58) NULL DEFAULT NULL,
    `Onboarding_Kick_off_date` DATETIME NULL DEFAULT NULL,
    `Customer_Since` DATETIME NULL DEFAULT NULL,
    `Cidade` VARCHAR(58) NULL DEFAULT NULL,
    `Deal_market` VARCHAR(5) NULL DEFAULT NULL,

```

```

`Pais` VARCHAR(17) NULL DEFAULT NULL,
  `Entidade` VARCHAR(12) NULL DEFAULT NULL,
  `Segment` VARCHAR(20) NULL DEFAULT NULL,
  `Industry` VARCHAR(33) NULL DEFAULT NULL,
  `Macro_Industry` VARCHAR(36) NULL DEFAULT NULL,
  `Deal_type` VARCHAR(9) NULL DEFAULT NULL)
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8mb4
COLLATE = utf8mb4_0900_ai_ci;

USE `esquema_estrela` ;

-- -----
-- Placeholder table for view `esquema_estrela`.`vw_area_negocio`
-- -----

CREATE TABLE IF NOT EXISTS `esquema_estrela`.`vw_area_negocio`
(`Duracao_Onboarding` INT, `Onboarding_Data_esperada` INT, `MRR` INT, `Setup_Fee` INT,
`Total_Tecnicos` INT, `Total_Admins_Gestores` INT,
`Onboarding_Kick_off_date` INT, `Customer_Since` INT,
`Deal_market` INT, `Pais` INT, `Entidade` INT, `Segment` INT,
`Industry` INT, `Macro_Industry` INT, `Deal_type` INT,
`ID_Area_Negocio` INT);

-- -----
-- Placeholder table for view `esquema_estrela`.`vw_local`
-- -----

CREATE TABLE IF NOT EXISTS `esquema_estrela`.`vw_local`
(`Duracao_Onboarding` INT, `Onboarding_Data_esperada` INT,
`MRR` INT, `Setup_Fee` INT, `Total_Tecnicos` INT,
`Total_Admins_Gestores` INT, `Onboarding_Kick_off_date` INT,
`Customer_Since` INT, `Deal_market` INT, `Pais` INT,
`Entidade` INT, `Segment` INT, `Industry` INT,
`Macro_Industry` INT, `Deal_type` INT, `ID_Local` INT,
`Cidade` INT);

-- -----
-- Placeholder table for view `esquema_estrela`.`vw_tempo`
-- -----

```

```

CREATE TABLE IF NOT EXISTS `esquema_estrela`.`vw_tempo`
(`Duracao_Onboarding` INT, `Onboarding_Data_esperada` INT,
`MRR` INT, `Setup_Fee` INT, `Total_Tecnicos` INT,
`Total_Admins_Gestores` INT, `Onboarding_Kick_off_date`
INT, `Customer_Since` INT, `Deal_market` INT, `Pais` INT,
`Entidade` INT, `Segment` INT, `Industry` INT,
`Macro_Industry` INT, `Deal_type` INT, `ID_Customer_Since`
INT, `ID_Onboarding_Kick_off_date` INT);
-- -----
-- procedure fill_date_dimension
-- -----
DELIMITER $$
USE `esquema_estrela`$$
CREATE DEFINER=`root`@`localhost` PROCEDURE
`fill_date_dimension`(IN startdate DATE,IN stopdate DATE)
BEGIN
    DECLARE currentdate DATE;
    SET currentdate = startdate;
    WHILE currentdate <= stopdate DO
        INSERT INTO dimensao_tempo VALUES (
            YEAR(currentdate)*10000+MONTH(currentdate)*100
+ DAY(currentdate),
            currentdate,
            YEAR(currentdate),
            MONTH(currentdate),
            DAY(currentdate),
            QUARTER(currentdate),
            WEEKOFYEAR(currentdate),
            DATE_FORMAT(currentdate,'%W'),
            DATE_FORMAT(currentdate,'%M')
        );
        SET currentdate = ADDDATE(currentdate,INTERVAL 1
DAY);
    END WHILE;
END$$

```

```

DELIMITER ;

-- View `esquema_estrela`.`vw_area_negocio`--
DROP TABLE IF EXISTS `esquema_estrela`.`vw_area_negocio`;
USE `esquema_estrela`;
CREATE          OR          REPLACE          ALGORITHM=UNDEFINED
DEFINER=`root`@`localhost` SQL SECURITY DEFINER VIEW
`esquema_estrela`.`vw_area_negocio` AS select
`t`.`Duracao_Onboarding` AS
`Duracao_Onboarding`,`t`.`Onboarding_Data_esperada` AS
`Onboarding_Data_esperada`,`t`.`MRR` AS
`MRR`,`t`.`Setup_Fee` AS `Setup_Fee`,`t`.`Total_Tecnicos`
AS `Total_Tecnicos`,`t`.`Total_Admins_Gestores` AS
`Total_Admins_Gestores`,`t`.`Onboarding_Kick_off_date` AS
`Onboarding_Kick_off_date`,`t`.`Customer_Since` AS
`Customer_Since`,`t`.`Deal_market` AS
`Deal_market`,`t`.`Pais` AS `Pais`,`t`.`Entidade` AS
`Entidade`,`t`.`Segment` AS `Segment`,`t`.`Industry` AS
`Industry`,`t`.`Macro_Industry` AS
`Macro_Industry`,`t`.`Deal_type` AS
`Deal_type`,`an`.`ID_Area_Negocio` AS `ID_Area_Negocio`
from (`esquema_estrela`.`temporaria` `t` left join
`esquema_estrela`.`dimensao_area_negocio` `an`
on((`t`.`Segment` = `an`.`Segment`))) where
((`t`.`Segment` = `an`.`Segment`) and (`t`.`Industry` =
`an`.`Industry`) and (`an`.`Macro_Industry` =
`t`.`Macro_Industry`) and (`t`.`Deal_type` =
`an`.`Deal_type`));

-- View `esquema_estrela`.`vw_local`--
DROP TABLE IF EXISTS `esquema_estrela`.`vw_local`;
USE `esquema_estrela`;
CREATE          OR          REPLACE          ALGORITHM=UNDEFINED
DEFINER=`root`@`localhost` SQL SECURITY DEFINER VIEW
`esquema_estrela`.`vw_local` AS select
`t`.`Duracao_Onboarding` AS
`Duracao_Onboarding`,`t`.`Onboarding_Data_esperada` AS
`Onboarding_Data_esperada`,`t`.`MRR` AS
`MRR`,`t`.`Setup_Fee` AS `Setup_Fee`,`t`.`Total_Tecnicos`
AS `Total_Tecnicos`,`t`.`Total_Admins_Gestores` AS

```

```

AS `Deal_market`,`t`.`Pais` AS `Pais`,`t`.`Entidade` AS
`Entidade`,`t`.`Segment` AS `Segment`,`t`.`Industry` AS
`Industry`,`t`.`Macro_Industry` AS
`Macro_Industry`,`t`.`Deal_type` AS `Deal_type`,`l`.`ID_Local`
AS `ID_Local`,`l`.`Cidade` AS `Cidade` from
(`esquema_estrela`.`temporaria` `t` left join
`esquema_estrela`.`dimensao_local` `l` on((`t`.`Cidade` =
`l`.`Cidade`)));
-- View `esquema_estrela`.`vw_tempo`--
DROP TABLE IF EXISTS `esquema_estrela`.`vw_tempo`;
USE `esquema_estrela`;
CREATE OR REPLACE ALGORITHM=UNDEFINED
DEFINER=`root`@`localhost` SQL SECURITY DEFINER VIEW
`esquema_estrela`.`vw_tempo` AS
select
`t`.`Duracao_Onboarding` AS
`Duracao_Onboarding`,`t`.`Onboarding_Data_esperada` AS
`Onboarding_Data_esperada`,`t`.`MRR` AS `MRR`,`t`.`Setup_Fee`
AS `Setup_Fee`,`t`.`Total_Tecnicos` AS
`Total_Tecnicos`,`t`.`Total_Admins_Gestores` AS
`Total_Admins_Gestores`,`t`.`Onboarding_Kick_off_date` AS
`Onboarding_Kick_off_date`,`t`.`Customer_Since` AS
`Customer_Since`,`t`.`Deal_market` AS `Deal_market`,`t`.`Pais`
AS `Pais`,`t`.`Entidade` AS `Entidade`,`t`.`Segment` AS
`Segment`,`t`.`Industry` AS `Industry`,`t`.`Macro_Industry` AS
`Macro_Industry`,`t`.`Deal_type` AS
`Deal_type`,`tp`.`ID_tempo` AS
`ID_Customer_Since`,`dtp`.`ID_tempo` AS
`ID_Onboarding_Kick_off_date` from
((`esquema_estrela`.`temporaria` `t` left join
`esquema_estrela`.`dimensao_tempo` `tp`
on((cast(`t`.`Customer_Since` as date) = `tp`.`db_date`)))
left join `esquema_estrela`.`dimensao_tempo` `dtp`
on((cast(`t`.`Onboarding_Kick_off_date` as date) =
`dtp`.`db_date`)));

SET SQL_MODE=@OLD_SQL_MODE;
SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS;
SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS;

```

APÊNDICE III – VIEW TEMPO E VIEW ÁREA DE NEGÓCIO

```
-----  
-- view_tempo --  
-----  
  
CREATE VIEW esquema_estrela.vw_tempo as  
SELECT T.Duracao_Onboarding,  
T.Onboarding_Data_esperada,  
T.MRR,  
T.Setup_Fee,  
T.Total_Tecnicos,  
T.Total_Admins_Gestores,  
T.Onboarding_Kick_off_date,  
T.Customer_Since,  
T.Deal_market,  
T.Pais,  
T.Entidade,  
T.Segment,  
T.Industry,  
T.Macro_Industry,  
T.Deal_type,      TP.ID_tempo      as      ID_Customer_Since,  
DTP.ID_tempo      as      ID_Onboarding_Kick_off_date      from  
temporaria T  
LEFT JOIN dimensao_tempo TP ON  
CAST (T.Customer_Since AS DATE) = TP.db_date  
LEFT JOIN dimensao_tempo DTP ON  
CAST(T.Onboarding_Kick_off_date      AS      DATE)      =  
DTP.db_date;
```

```

-----
-- view_area_negocio --
-----

Create view esquema_estrela.vw_area_negocio as
Select  T.Duracao_Onboarding,
T.Onboarding_Data_esperada,
T.MRR,
T.Setup_Fee,
T.Total_Tecnicos,
T.Total_Admins_Gestores,
T.Onboarding_Kick_off_date,
T.Customer_Since,
T.Deal_market,
T.Pais,
T.Entidade,
T.Segment,
T.Industry,
T.Macro_Industry,
T.Deal_type, AN.ID_Area_Negocio from temporaria T
LEFT JOIN dimensao_area_negocio AN ON
T.Segment=AN.Segment
WHERE T.Segment=AN.Segment and T.Industry=AN.Industry
and AN.Macro_Industry= T.Macro_Industry and
T.Deal_type=AN.Deal_type

```