



BI em Dados de Seguros de Vida

ALEXANDRA ANDREIA DA SILVA FERREIRA

novembro de 2022

BI em Dados de Seguros de Vida

Real Vida Seguros

Alexandra Andreia da Silva Ferreira

**Dissertação para obtenção do Grau de Mestre em
Engenharia Informática, Área de Especialização em
Sistemas de Informação e Conhecimento**

Orientador: Paulo Oliveira

Coorientador: Nelson Freire

Porto, outubro 2022

Resumo

A Real Vida Seguros, tal como o nome indica, trabalha no mercado de seguros de vida, e possui informação relativa à atividade a nível nacional. A empresa forneceu os dados para se elaborar esta dissertação e responsabiliza-se por monitorizar e armazenar dados referentes aos titulares e todo o processo de aquisição e um produto.

O conjunto destas informações torna-se útil para os profissionais, já que estes são responsáveis pelas melhorias relacionadas com a aquisição de um produto, portanto, o objetivo é fornecer a informação certa e sempre atualizada, antecipando algumas necessidades futuras.

Desenvolveu-se um conjunto de indicadores sobre titulares, sobre produtos, sobre propostas e apólices, de forma a retirar conclusões acerca dos clientes ao longo dos anos, os motivos que levam à aquisição de um produto, quais os tipos de produtos mais comuns, as faixas etárias que mais adquirem um seguro, onde se concentra o maior número de titulares, etc. É o conjunto destes indicadores que permite identificar causas até então desconhecidas.

Foi desenvolvido um sistema de informação de BI de forma a solucionar estas necessidades e onde a informação acerca das potenciais apólices e apólices efetivadas se encontra centralizada. Desta forma, o sistema encontra-se preparado para necessidades futuras sem ter de se fazer grandes alterações à estrutura do mesmo.

Desta forma, foi disponibilizado um conjunto de relatórios para que os profissionais da seguradora possam acompanhar os indicadores implementados, nos quais a informação pode ser filtrada e se encontra sob a forma de *dashboards*. Os profissionais da seguradora têm à sua disposição relatórios com os dados dos titulares, dos produtos, das propostas, entre outros, que respondem a alguns cenários identificados.

Por fim, como forma de validar o sistema desenvolvido, foram elaborados inquéritos de satisfação a profissionais de IT e da seguradora, onde se avalia a usabilidade e a utilidade do mesmo. Ambos os grupos demonstraram opiniões bastante positivas às afirmações que lhes foram colocadas acerca da aplicação desenvolvida. De acordo com os resultados obtidos, o desenvolvimento da aplicação foi conseguido segundo os objetivos a atingir e é de interesse utilizar este tipo de aplicação.

Palavras-chave: Dados de Seguros de Vida, *Business Intelligence*, indicadores, relatórios, *dashboards*

Abstract

Real Vida Seguros, as the name implies, works in the life insurance market, and has information on the activity at a national level. The company provided the data to prepare this dissertation and is responsible for monitoring and storing data regarding the holders and the entire acquisition process of a product.

All of this information becomes useful for professionals, as they are responsible for improvements related to the acquisition of a product, therefore, the objective is to provide the right information and always updated, anticipating some future needs.

A set of indicators was developed on holders, on products, on proposals and policies, in order to draw conclusions about customers over the years, the reasons that lead to the acquisition of a product, which are the most common types of products, the age groups that buy insurance the most, where the largest number of holders are concentrated, etc. It is the set of these indicators that makes it possible to identify hitherto unknown causes.

A BI information system was developed in order to solve these needs and where information about potential policies and effective policies is centralized. In this way, the system is prepared for future needs without having to make major changes to its structure.

In this way, a set of reports was made available so that the insurance company's professionals can monitor the implemented indicators, in which the information can be filtered and is in the form of dashboards. The insurer's professionals have at their disposal reports with data on the holders, products, proposals, among others, which respond to some identified scenarios.

Finally, as a way of validating the developed system, satisfaction surveys were carried out with IT professionals and the insurance company, in which the usability and usefulness of the system are evaluated. Both groups showed very positive opinions to the statements that were made to them about the developed application. According to the results obtained, the development of the application was achieved according to the objectives to be achieved and it is of interest to use this type of application.

Keywords: Life Insurance Data, Business Intelligence, indicators, reports, dashboards

Índice

1	Introdução	1
1.1	Contexto	1
1.2	Problema	1
1.3	Objetivos	2
1.4	Abordagem	3
1.5	Contribuições	3
1.6	Estrutura do Documento	4
2	Contexto e Estado da Arte	5
2.1	Área de Negócio	5
2.1.1	Domínio dos Seguros	5
2.1.2	Variáveis	7
2.1.3	Intervenientes	7
2.2	Business Intelligence	7
2.3	Armazém de Dados	8
2.3.1	Conceito	8
2.3.2	Data Mart	10
2.3.3	Arquiteturas de Armazéns de Dados	10
2.4	Modelação Dimensional	12
2.4.1	Tabela de Factos	12
2.4.2	Tabelas de Dimensão	12
2.4.3	Tipos de Modelos Dimensionais	13
2.4.4	Metodologia de Kimball	14
2.5	Processo de ETL	15
2.6	Processamento Analítico de Dados	15
2.7	Ferramentas	16
2.7.1	Ferramentas de ETL	16
2.7.2	Ferramentas de Análise de Dados	20
2.8	BI na Área dos Seguros	24
3	Análise de Valor	27
3.1	Definição de Análise de Valor e Terminologia	27
3.2	Fuzzy Front-End	27
3.2.1	Identificação da Oportunidade	28
3.2.2	Análise da Oportunidade	28
3.2.3	Geração de Ideias e Enriquecimento	28
3.2.4	Seleção de Ideias	29
3.2.5	Definição do Conceito	29
3.3	Análise de Valor	29

3.4	Proposta de Valor	30
3.5	Modelo de Negócio CANVAS.....	30
3.6	Cadeia de Valor de Porter	34
3.7	Processo de Análise Hierárquica (AHP)	36
3.7.1	Divisão Hierárquica	36
3.7.2	Comparação das Alternativas e Critérios	37
3.7.3	Definição de Prioridades	38
3.7.4	Consistência Lógica	38
3.7.5	Matriz de Comparação Paritária para cada Critério.....	39
3.7.6	Prioridade Composta para as Alternativas.....	41
4	Análise e Design da Solução Proposta	43
4.1	Requisitos	43
4.1.1	Funcionalidade	44
4.1.2	Usabilidade.....	47
4.1.3	Confiabilidade	47
4.1.4	Desempenho	47
4.1.5	Suportabilidade.....	47
4.1.6	Outros (+)	47
4.2	Arquitetura Proposta	48
4.2.1	Metodologia de Kimball	48
4.2.2	Fonte de Dados	49
4.2.3	Staging Area	50
4.2.4	Modelo Dimensional.....	53
4.2.5	Mapeamento de Dados entre os Sistemas.....	59
5	Implementação da Solução	61
5.1	Tecnologias Utilizadas.....	61
5.2	Processo de ETL	61
5.2.1	Operações de Extração.....	61
5.2.2	Operações de Transformação	62
5.2.3	Operações de Carregamento	73
5.3	Análises de Dados	77
6	Apresentação e Análise de Resultados	79
6.1	Indicadores.....	79
6.1.1	Informação Titulares.....	79
6.1.2	Informação Produtos.....	82
6.1.3	Informação Potenciais Apólices	83
6.1.4	Informação Apólices	86
7	Avaliação/Experimentação	89
7.1	Grandezas de Avaliação	89
7.2	Hipóteses.....	89
7.3	Metodologia de Avaliação.....	89

7.4	Resultado dos Inquéritos	90
7.4.1	Profissionais de IT	91
7.4.2	Profissionais da Seguradora	92
7.5	Comentários Finais	95
8	Conclusão	97
8.1	Objetivos Alcançados	97
8.2	Problemas/Limitações	98
8.3	Trabalho Futuro	98
Anexo I	104
Anexo II	107
Anexo III	117
Anexo IV	118

Lista de Figuras

Figura 1 – Exemplo da orientação de dados ao assunto (Inmon, 2005).....	8
Figura 2 – Exemplo de integração (Inmon, 2005).....	9
Figura 3 – Exemplo de variação temporal (Inmon, 2005).....	9
Figura 4 – Exemplo de não volatilidade (Inmon, 2005).....	9
Figura 5 – Arquitetura BUS (ZenTut, 2020).....	10
Figura 6 – Arquitetura CIF (ZenTut, 2020).....	11
Figura 7 – Modelo em Estrela (MSBI, 2017).....	13
Figura 8 – Modelo em Floco de Neve (MSBI, 2017).....	14
Figura 9 – Modelo em Galáxia (EDUCBA., 2020).....	14
Figura 10 – Processo de ETL (Tobin, D., 2020).....	15
Figura 11 – Quadrante Mágico das ferramentas de ETL (Gartner, 2020a).....	17
Figura 12 – Interface do SSIS (Chugugrace, 2019).....	18
Figura 13 – Interface do ODI (Oracle, 2020b).....	18
Figura 14 – Interface da SAS (SAS, 2020b).....	19
Figura 15 – Interface do Talend (Talend, 2020b).....	20
Figura 16 – Quadrante Mágico das ferramentas de análise de dados (Gartner, 2020b).....	21
Figura 17 – Interface da Qlik Sense (QlikTech, 2020).....	22
Figura 18 – Interface da Power BI (Microsoft, 2020b).....	22
Figura 19 – Interface da Tableau (TABLEAU, 2020b).....	23
Figura 20 – Interface da MicroStrategy (MicroStrategy,2020b).....	23
Figura 21 – Segmentos de clientes (Berlin S. D., 2012).....	30
Figura 22 – Proposta de valor (Berlin S. D., 2012).....	31
Figura 23 – Canais (Berlin S. D., 2012).....	31
Figura 24 – Relações com os Clientes (Berlin S. D., 2012).....	31
Figura 25 – Fluxos de Rendimento (Berlin S. D., 2012).....	32
Figura 26 – Recursos-Chave (Berlin S. D., 2012).....	32
Figura 27 – Atividades-Chave (Berlin S. D., 2012).....	32
Figura 28 – Parcerias-Chave (Berlin S. D., 2012).....	33
Figura 29 – Estrutura de Custos (Berlin S. D., 2012).....	33
Figura 30 – Modelo de Negócio CANVAS.....	34
Figura 31 – Cadeia de Valor de Porter (Serafim A., 2013).....	35
Figura 32 – Exemplo de uma hierarquia de critérios/objetivos (Vargas, R. V., 2010).....	36
Figura 33 – Hierarquia AHP.....	37
Figura 34 – Escala fundamental de Saaty (Níveis de importância de comparações).....	37
Figura 35 – Diagrama de casos de uso.....	44
Figura 36 – Arquitetura Proposta.....	48
Figura 37 – Excerto do Modelo da Base de Dados Operacional.....	50
Figura 38 – Modelo Dimensional.....	54
Figura 39 – Fase de extração na arquitetura do sistema.....	62
Figura 40 – Fase de transformação na arquitetura do sistema.....	63

Figura 41 – Processo de Transformação na tabela CodigosPostais	64
Figura 42 – Processo de Transformação na tabela Concelhos	64
Figura 43 – Processo de Transformação na tabela Distritos	65
Figura 44 – Linhas duplicadas nos CodigosPostais	65
Figura 45 – Remoção de linhas duplicadas nos CodigosPostais.....	66
Figura 46 – Processo de Transformação na tabela PROPET	67
Figura 47 – Redireccionamento de DQPs para a Tabela PROPET_DQP	68
Figura 48 – Exemplo de linhas com os atributos (PTSEXO, PTACTI, PTNACD, PTNATD e PTESTC) em falta	68
Figura 49 – Exemplo de linhas com código postal inválido.....	69
Figura 50 – Exemplo de linhas com o PTACTI e PT\$CAE inválidos	69
Figura 51 – Exemplo de linhas com a data de nascimento inválida.....	69
Figura 52 – Exemplo de linhas com a nacionalidade inválida	69
Figura 53 – Processo de Transformação na tabela PROPEO	70
Figura 54 – Processo de Transformação na tabela MODA00.....	72
Figura 55 – Fase de carregamento na arquitetura do sistema	74
Figura 56 – Carregamento da tabela DimDadosProposta.....	75
Figura 57 – Carregamento da tabela DimEstruturaComercial	75
Figura 58 – Carregamento da tabela FactProposta (1)	76
Figura 59 – Carregamento da tabela FactProposta (2)	76
Figura 60 – Carregamento da tabela FactProposta (3)	77
Figura 61 – Fase de análise de dados na arquitetura do sistema	77
Figura 62 – <i>Dashboard</i> de informação acerca dos titulares.....	79
Figura 63 – Gráfico Drill Down do Número de Titulares por Ano	80
Figura 64 – Relatório de informação acerca dos titulares	81
Figura 65 – <i>Dashboard</i> de informação acerca dos titulares.....	81
Figura 66 – <i>Dashboard</i> de informação acerca dos titulares.....	82
Figura 67 – <i>Dashboard</i> de informação acerca dos produtos	83
Figura 68 – <i>Dashboard</i> de informação acerca das propostas	84
Figura 69 – <i>Dashboard</i> de informação acerca das propostas	85
Figura 70 – Relatório de informação acerca das propostas.....	86
Figura 71 – <i>Dashboard</i> de informação acerca das apólices (capital total em caso de morte) ..	86
Figura 72 – <i>Dashboard</i> de informação acerca das apólices (capital total em caso de vida).....	87
Figura 73 – Outras Análises	88
Figura 74 – 1ª Questão do Inquérito de Satisfação (IT)	91
Figura 75 – 2ª Questão do Inquérito de Satisfação (IT)	91
Figura 76 – 3ª Questão do Inquérito de Satisfação (IT)	92
Figura 77 – 4ª Questão do Inquérito de Satisfação (IT)	92
Figura 78 – 1ª Questão do Inquérito de Satisfação (Profissionais da Seguradora).....	93
Figura 79 – 2ª Questão do Inquérito de Satisfação (Profissionais da Seguradora).....	93
Figura 80 – 3ª Questão do Inquérito de Satisfação (Profissionais da Seguradora).....	93
Figura 81 – 4ª Questão do Inquérito de Satisfação (Profissionais da Seguradora).....	94
Figura 82 – 5ª Questão do Inquérito de Satisfação (Profissionais da Seguradora).....	94

Figura 83 – 6ª Questão do Inquérito de Satisfação (Profissionais da Seguradora)	95
Figura 84 – Inquérito de satisfação (IT).....	117
Figura 85 – Inquérito de satisfação (profissionais da seguradora)	118

Lista de Tabelas

Tabela 1 – Comparação de Arquiteturas de DW (Breslin, M., 2004)	12
Tabela 2 – Comparação entre Ferramentas de ETL	20
Tabela 3 – Comparação entre Aplicações de Análise de Dados	24
Tabela 4 – Comparações Par a Par	38
Tabela 5 – Matriz de Comparação Normalizada e Pesos Estimados	38
Tabela 6 – Teste da Consciência	38
Tabela 7 – Valores de IR para matrizes quadradas de ordem n	39
Tabela 8 – Matriz de Alternativas Normalizada e Pesos Estimados (Custo).....	40
Tabela 9 – Teste da Consciência (Custo).....	40
Tabela 10 – Matriz de Alternativas Normalizada e Pesos Estimados (Tempo).....	40
Tabela 11 – Teste da Consciência (Tempo).....	40
Tabela 12 – Matriz de Alternativas Normalizada e Pesos Estimados (Fiabilidade)	41
Tabela 13 – Teste da Consciência (Fiabilidade).....	41
Tabela 14 – Matriz com Prioridades Globais	41
Tabela 15 – Tabela PROPET.....	51
Tabela 16 – Tabela PROPEO.....	51
Tabela 17 – Tabela PROPET_DQP	53
Tabela 18 – Nomenclaturas das tabelas	53
Tabela 19 – Tabela DimData	54
Tabela 20 – Tabela DimRegiao	55
Tabela 21 – Tabela DimProduto.....	55
Tabela 22 – Tabela DimTitular	56
Tabela 23 – Tabela DimEstruturaComercial.....	57
Tabela 24 – Tabela DimDadosProposta	57
Tabela 25 – Tabela FactProposta	58
Tabela 26 – Nomenclaturas das tabelas	59
Tabela 27 – Transformações na tabela PROPET	67
Tabela 28 – Transformações na tabela PROPEO	70
Tabela 29 – Transformações na tabela MODA00	72
Tabela 30 – Escala de Likert	90
Tabela 31 – Tabela MODA00.....	104
Tabela 32 – Tabela ESTRUTN	104
Tabela 33 – Tabela TBLLIN	105
Tabela 34 – Tabela CodigosPostais	105
Tabela 35 – Tabela Concelhos.....	105
Tabela 36 – Tabela Distritos.....	105
Tabela 37 – Mapeamento dos dados.....	107

1 Introdução

Este capítulo é dedicado à visão geral deste trabalho de mestrado ("BI em Dados de Seguros de Vida") e o contexto no qual está inserido. Começa por apresentar uma contextualização do mesmo, seguindo-se de uma descrição do problema. Descrevem-se os seus objetivos, a abordagem seguida, os contributos do trabalho e, finalmente, a estrutura do documento.

1.1 Contexto

Atualmente, dada a complexidade e nível de exigência requerido às organizações, importa cada vez mais ter um controlo sobre os fatores que condicionam o seu desempenho e o que o influencia.

A Real Vida Seguros, tal como o nome indica, trabalha no mercado de seguros de vida, possuindo informação relativa à atividade a nível nacional e responsabiliza-se por monitorizar toda essa informação. Esta organização está a acabar com a gestão dos contratos por papel para cada vez mais, passar a utilizar novas tecnologias, desde emissões *online* e gestão da área reservada a clientes e agentes, bem como os processamentos e gestão interna dos seguros. Sendo esta considerada uma empresa em constante evolução e onde a desvalorização dos indicadores de desempenho são uma ameaça frequente, exige análises cuidadosas e regulares. Além disso, as análises pedidas, como por exemplo, o número de Apólices emitidas do período, as receitas cobradas do período (por produto), a inclusão de indicadores de desempenho da Seguradora, tais como receita processada e indemnizações/resgates processados do período, entre outras, tem um custo diário para a organização.

Os sistemas de informação permitem que os dados sejam apresentados aos utilizadores de forma simples e útil, permitindo que a informação fique armazenada de forma centralizada e que possa ser acedida pelas diversas áreas da organização. Assim, o objetivo é fornecer a informação certa e sempre atualizada, antecipando algumas necessidades futuras.

Ao longo de vários anos, estas informações foram adquiridas de diversas fontes que originaram dados de má qualidade. A falta de exploração nos dados permitiu que muita informação essencial para a obtenção de conhecimento fosse descartada, impossibilitando novas descobertas.

Neste projeto irão ser analisados dados relativos aos Seguros de Vida para se poder retirar conclusões e reavaliar os dados atuais/estruturas de informação. Na análise a efetuar é essencial que os dados sejam de qualidade, estejam atualizados e que haja um histórico associado a cada cliente, para que este possa ser devidamente acompanhado.

Na génese deste trabalho encontra-se um *software* ("o cubo"), surgindo a necessidade de reestruturação do processo de extração de informação da ferramenta. Devido às diferentes áreas envolvidas e com distintas definições de conceitos pretende-se a consolidação da informação e existe ainda, a falta de variáveis importantes para a gestão do negócio na empresa em questão.

1.2 Problema

O *software* "o cubo" é antigo pelo que torna difícil a sua utilização e também a sua compreensão por parte dos seus utilizadores. Por outro lado, a base de dados subjacente não se revela adequada à utilização de uma ferramenta de processamento analítico de dados *online* (OLAP), uma vez que apresenta problemas ao nível do esquema dos dados e também ao nível das instâncias (i.e., dos valores). Esse tipo de ferramentas (OLAP) permitem a realização de análises de dados diversificadas seguindo uma filosofia *self-service* por parte do utilizador final.

Apesar de existir este *software*, não permite visualizar dados estatísticos automaticamente acerca dos seguros efetuados ao longo do tempo e leva um tempo excessivo na construção/disponibilização da informação.

Alguns exemplos de indicadores que deveriam estar disponíveis em tempo real seriam: se o número de seguros tem vindo a aumentar ou a diminuir; se existem seguros que foram anulados e quais os motivos; perceber o tipo de produto que é mais popular; perceber se houve aumentos de seguros em determinadas regiões e quais as faixas etárias que mais aderiram.

A necessidade de implementação do presente projeto derivou da identificação de:

- Decorrer de um processo manual a consolidação da informação para reporte comercial referente a um período, o que torna o processo moroso e com risco operacional elevado, uma vez que o processo depende de relatórios feitos pelo colaborador.
- Distintas definições dos conceitos, dependendo da área envolvida o que leva a desentendimentos ou confusões. Entre esses conceitos encontram-se os de:
 - Receita/indemnizações processadas
 - Receita cobrada
 - Prémios anualizados de produção
 - Objetivos comerciais
 - Segmentação de produtos/modalidades
 - Data de vigência
- Falta de variáveis importantes para a gestão do negócio. Entre essas variáveis, encontram-se:
 - Anulações/transferências/substituições
 - Comissionamento agentes/remuneração canais
 - Gestão de campanhas
 - Sinistros
 - Relatórios/gráficos/*outputs* (normalizados e customizáveis)

Em suma, existem problemas de má qualidade dos dados e má exploração dos dados existentes. Atualmente, os dados destes conceitos encontram-se no sistema ERP AS400, sendo que o cálculo e partilha semi-manual dos indicadores chave de desempenho (em inglês: *Key Performance Indicators*¹ - KPI) é um processo moroso e sujeito a erros e atrasos.

1.3 Objetivos

O principal objetivo proposto consiste no desenvolvimento de uma plataforma de *Business Intelligence* (BI).

A empresa pretende, recorrendo aos valores calculados dos chamados KPI da componente financeira, comercial e gestão/administração da empresa, adquirir benefícios tais como:

- Processo de extração e apresentação de informação automático;

¹ Key Performance Indicators são os indicadores críticos (principais) do progresso em direção ao resultado pretendido. Os KPIs fornecem um foco para a melhoria estratégica e operacional, criam uma base analítica para a tomada de decisões e ajudam a concentrar a atenção no que é mais importante (KPI.org., 2020).

- Redução do tempo despendido na elaboração do mapa diário;
- Mitigação dos erros operacionais;
- Acesso imediato e individual aos principais indicadores (comerciais e de gestão).

Pretende-se a reestruturação do processo de extração de informação da ferramenta “CUBO”, reavaliando os dados atuais/estruturas de informação. Esta passa por uma coordenação com a área comercial e financeira acerca dos dados a serem considerados, antecipando desde já, algumas necessidades futuras relacionadas com canais e políticas de segmentação e vendas cruzadas.

Existe a necessidade de reduzir o tempo necessário à construção dessa informação, melhorando a eficiência da consulta e envio da mesma.

Em concreto, com a realização deste trabalho de mestrado, pretende-se alcançar os seguintes objetivos:

- Definir o modelo conceptual de dados (modelo dimensional) que suportará o armazenamento dos dados no armazém de dados;
- Extrair os dados a partir da base de dados que suporta o sistema operacional;
- Aplicar técnicas de transformação e limpeza de dados, de modo a remover problemas de qualidade existentes nos dados, respetivamente, ao nível do esquema dos dados e ao nível das instâncias (i.e., dos valores);
- Proceder ao carregamento dos dados no armazém de dados, manipulando adequadamente o histórico de alterações que ocorram nos dados;
- Criar um cubo de dados que suporte a realização de análises de dados;
- Realizar análises diversificadas aos dados e produzir *dashboards* com os indicadores chave de desempenho relevantes para cada departamento.

1.4 Abordagem

De forma a abordar o problema indicado e dar resposta às suas necessidades, é seguido um conjunto de passos:

- Estudar a área de negócio, o conceito de BI e a arquitetura de um sistema de BI.
- Realizar uma pesquisa sobre as ferramentas de BI existentes no mercado e analisar as suas funcionalidades.
- Analisar e desenhar diferentes alternativas/soluções e escolher a melhor solução.
- Implementar a solução desenhada.
- Avaliar a solução desenhada.

1.5 Contribuições

O sistema a desenvolver deve permitir que a Real Vida Seguros, possa controlar e melhorar o desempenho da empresa, analisando ao longo do tempo os indicadores de eficiência de sustentabilidade relativos à sua atividade. Desta forma mais fácil e amigável de procurar, analisar e enviar as informações dos indicadores, os colaboradores ocuparão menos tempo com estas tarefas, podendo concentrar-se em tarefas com maior valor acrescentado. Além disso, e devido ao esforço no desenvolvimento de uma aplicação reutilizável e expansível, alguns dos componentes desenvolvidos podem ser usados noutros projetos que podem não estar relacionados a este. Isso beneficia a equipa de desenvolvimento e, conseqüentemente, a organização, em projetos futuros.

Assim, os profissionais conseguem ter o quotidiano mais facilitado, devido à simplificação e desmaterialização de processos até então usados. A forma como lhes é apresentada a informação dos clientes é um fator importante, pois o tempo de análise aos dados diminuiu e a tomada de decisão é conseguida mais rapidamente.

A normalização e consolidação da informação dos contratos de seguros irá:

- Melhorar a qualidade:
 - Base de dados dos contratos de seguros
 - Controlo antecipado dos objetivos
- Aumentar eficiência:
 - Automatização dos processos / libertação de recursos
 - Relatórios finais de acesso automático
- Reduzir custos:
 - Processos manuais (recursos e materiais)
- Antecipar necessidades de informação nos sistemas/base de dados:
 - Segmentos, canais, etc

1.6 Estrutura do Documento

Esta secção expõe a estrutura e organização deste documento, apresentando um resumo do conteúdo encontrado em cada capítulo, a fim de facilitar a sua leitura e consulta.

O capítulo atual – Introdução – pretende fornecer uma contextualização generalizada do projeto e problema a solucionar, bem como os seus objetivos e contributos do projeto.

No Capítulo 2, referente ao Contexto e Estado da Arte, é feito um breve estudo acerca da área de negócio nomeadamente, domínio dos seguros, variáveis e intervenientes que surgem nos seguros, sendo essencial para perceber os conceitos que permitem retirar conclusões através dos resultados obtidos ao longo da dissertação. Também é feito um estudo das arquiteturas e de diversas tecnologias existentes que poderiam ser usadas neste projeto, assim como, algumas comparações entre elas.

O Capítulo 3 é dedicado à análise de valor ao tema proposto, através de utilização de ferramentas que existem no mercado e que procedem a essa avaliação.

No Capítulo 4 é feita a avaliação da solução, nomeadamente a identificação das grandezas que serão utilizadas para avaliar o trabalho, que hipótese(s) se pretende(m) testar, qual a metodologia de avaliação e como se pretende testar essas hipóteses.

Finalmente, no quinto capítulo é feita a análise e *design* da solução proposta, detalhando os requisitos e a arquitetura proposta.

2 Contexto e Estado da Arte

Para a realização de um projeto é necessário efetuar um estudo, para que seja possível concluir onde se irá situar o produto que resultará da realização do projeto.

Embora este capítulo seja dedicado ao estado da arte, as secções seguintes iniciam-se com a apresentação dos conceitos essenciais relacionados com a área de negócio. Seguidamente, pode ser encontrada uma descrição das aplicações e das tecnologias utilizadas no processo de *Business Intelligence*, já existentes no mercado.

2.1 Área de Negócio

A presente secção tem por objetivo, contextualizar melhor o problema, o qual será alvo de estudo ao longo desta dissertação e apresentar o domínio dos seguros e conceitos relacionados com variáveis e intervenientes que se podem encontrar em seguros. É essencial perceber estes conceitos para se entender os resultados e retirar conclusões.

2.1.1 Domínio dos Seguros

Segundo a Autoridade de Supervisão de Seguros e Fundos de Pensões (ASF, 2015), o contrato de seguro é um acordo através do qual a entidade legalmente autorizada a exercer a atividade seguradora, e que é parte no contrato de seguro (segurador), assume a cobertura de determinada incerteza associada a um acontecimento futuro. Seja quanto à sua realização, ao momento em que ocorre ou aos danos dele decorrentes (risco), comprometendo-se a satisfazer as indemnizações ou a pagar o capital seguro em caso de ocorrência de sinistro, nos termos acordados. Em contrapartida, a pessoa ou entidade que celebra o seguro com a empresa de seguros (o tomador do seguro) fica obrigada a pagar ao segurador o prémio correspondente, ou seja, o custo do seguro. A prestação que ficou acordada no contrato pode ser efetuada à pessoa ou entidade no interesse do qual o seguro é celebrado (o segurado) ou de terceiro designado pelo tomador do seguro (o beneficiário) ou ainda a uma terceira pessoa ou entidade que tenha sofrido prejuízos que o segurado deva indemnizar (o terceiro lesado).

Ou seja, um seguro é um acordo entre uma pessoa ou entidade e uma companhia de seguros de modo que, se a pessoa ou entidade em particular incorrer num dano coberto no contrato de seguro, pode ser compensada total ou parcialmente. O segurado, em troca das coberturas oferecidas pela seguradora, paga um prémio à seguradora (Patrocínio, N., 2019).

A proposta de seguro é o documento através do qual o tomador do seguro expressa a sua vontade de celebrar um contrato de seguro e informa o segurador do risco que pretende segurar. É o primeiro passo para se celebrar um contrato de seguro.

Depois de receber a proposta preenchida e assinada pelo tomador do seguro, o segurador pode aceitar ou recusar o contrato de seguro. Se aceitar, emite a apólice de seguro, que é o documento que contém o que foi acordado pelas partes, nomeadamente as condições do contrato (condições gerais, especiais e particulares) celebrado entre o tomador do seguro e o segurador (ASF, 2021a).

O contrato de seguro é o conjunto de elementos, onde se expressa e quantifica a vontade entre o segurador e o tomador do seguro, contendo disposições de direitos e obrigações das partes (condições gerais, condições especiais e particulares) (Vida Seguros, R., 2021a).

Os seguros podem ser obrigatórios (quando a respetiva celebração é exigida por lei) ou facultativos (quando é opção do tomador do seguro celebrá-lo ou não) (ASF, 2015). Podem ser individuais ou de grupos. Os seguros individuais cobrem os riscos referentes a uma única pessoa, um agregado familiar

ou um conjunto de pessoas que vivam em economia comum. Os seguros de grupo são contratos através dos quais o segurador cobre riscos de um conjunto de pessoas ligadas ao tomador do seguro por uma relação distinta do seguro (por exemplo, seguro de saúde do conjunto dos trabalhadores de uma empresa).

Os seguros de grupo podem ser contributivos ou não contributivos. Serão contributivos se os segurados suportarem o pagamento de parte ou da totalidade do prémio e não contributivos se o prémio for totalmente suportado pelo tomador do seguro (por exemplo, seguro de saúde oferecido aos trabalhadores, que são os segurados, pela empresa em que trabalham, que é a tomadora do seguro) (ASF, 2015).

Os seguros são divididos em dois ramos: os chamados ramo “Não Vida” (artigo 123.º do DL n.º 94-B/98, de 17 de Abril) e ramo “Vida” (artigo 124.º do mesmo diploma) (Advogados, N., 2014).

Existem dois tipos de licenças para exploração no mercado, nomeadamente o ramo “Vida” e ramo “Não Vida”. A divisão é baseada nos aspetos ligados diretamente à atividade seguradora (Casimiro, N., 2011).

O artigo 9.º, n.º 1, do Decreto-Lei n.º 94-B/98, de 17 de Abril, define que a atividade de seguro direto e de resseguro do ramo “Vida” pode ser exercida cumulativamente apenas com a de seguro direto e resseguro dos ramos “Não Vida” referidos nos n.ºs 1) e 2) do artigo 123.º. Excecionalmente, o artigo 240.º, esta regra, ao estabelecer que as empresas de seguros que, à data da publicação daquele diploma, se encontrassem autorizadas a explorar cumulativamente em Portugal a atividade de seguros dos ramos “Não Vida” e a atividade de seguros do ramo “Vida”, podem continuar essa exploração cumulativa. Uma empresa de seguros autorizada aos ramos “Não vida”, mesmo que no passado tenha estado autorizada ao exercício da atividade seguradora também no ramo “Vida”, não poderá ser novamente autorizada a explorar ambos os ramos, por não se enquadrar na exceção prevista no artigo 240.º, do Decreto-Lei n.º 94-B/98 (ASF, 2021b).

Os seguros do ramo “Vida”, englobam os seguros de vida clássicos e os seguros financeiros e os seguros do ramo não vida ou também denominado de ramos reais englobam todos os seguros que têm objeto bens patrimoniais e também os seguros pessoais (excetuando o seguro de vida) (Casimiro, N., 2011).

No ramo “Vida” tem-se três grandes grupos ou modalidades de seguros: os seguros de vida, os seguros financeiros e os seguros de nupcialidade e natalidade (Advogados, N., 2014).

Um seguro de vida é um contrato pelo qual o Segurador se compromete a pagar o capital contratado em caso de morte do segurado (seguro em caso de morte) ou sobrevivência (seguro em caso de vida). Os seguros de vida podem também incluir coberturas complementares que garantam suporte financeiro em caso de invalidez da pessoa segura. Ao subscrever um seguro de vida, o segurado pode ainda beneficiar de apoio financeiro diário enquanto estiver hospitalizado (Vida Seguros, R., 2021b).

Existem, outros tipos de coberturas complementares, como por exemplo o de acidente ou de desemprego. No seguro de vida que cobre o risco de morte da pessoa segura (seguro em caso de morte), o segurador paga ao beneficiário o capital acordado, se a pessoa segura morrer durante o período fixado no contrato. No seguro de vida que cobre o risco de sobrevivência da pessoa segura (seguro em caso de vida), o segurador paga ao beneficiário o capital acordado, se a pessoa segura se encontrar viva no final do contrato. Estes seguros são usualmente utilizados para a constituição de uma poupança. Neste caso, o beneficiário pode ser a própria pessoa segura (ASF, 2015).

O ramo “Não Vida” engloba os seguros que têm objeto bens patrimoniais e também os seguros pessoais (exceto o seguro de vida). São exemplos do ramo “Não Vida”, os seguros de responsabilidade civil automóvel, acidentes de trabalho, seguro de doença, multiriscos habitação, multiriscos condomínio, seguro de caça, acidentes pessoais, seguro de viagem, entre outros (Casimiro, N., 2011).

A Real Vida Seguros, empresa na qual o presente trabalho está a ser realizado, opera no mercado de seguros de vida.

2.1.2 Variáveis

Define-se uma listagem de definições para os conceitos que são utilizados ao longo do trabalho:

- Receita processada no período n – Aquando da emissão da fatura-recibo considera-se o montante respetivo;
- Receita cobrada no período n – Montante pago no período n (não se refere à data da emissão do contrato de seguro);
- Prémios anualizados no período n – Prémio total para o período de 1 ano, independentemente do tipo de pagamento (mensal, trimestral, semestral);
- Prémio único – Prémio total entregue para o período integral de vigência da apólice;
- Data de vigência – Data de início ou data de emissão do contrato de seguro, a mais recente;
- Indemnizações processadas no período n – Valor da indemnização de sinistro aquando da abertura do mesmo;
- Segmentos de produtos – Listagem de produtos existentes sendo estes, Habitação, Previdência, Saúde, Acidentes Pessoais, Proteção ao Crédito e Financeiros;
- Estrutura comercial – Estrutura constituída por cinco níveis de hierarquia, dos quais 1º Nível Direção, 2º Nível Canal, 3º Nível Responsável Comercial, 4º Nível Designação tipo de Agente, 5º Nível Agente.

2.1.3 Intervenientes

- Tomador – Pessoa singular ou coletiva que celebra o contrato de seguro com o Segurador, sendo responsável pelo pagamento do prémio;
- Pessoa Segura – É a pessoa cuja vida, saúde ou integridade física se visa segurar com o contrato de seguro;
- Beneficiário – Pessoa singular ou coletiva a favor de quem se efetua o seguro de vida e reverterem as prestações do Segurador, devidas por um contrato de seguro ou operação de capitalização.

2.2 Business Intelligence

O termo *Business Intelligence* (BI) tem ganho relevância nas organizações que se querem manter competitivas (CCG, 2019). O BI é um termo abrangente que engloba os processos e métodos de extração, armazenamento e análise de dados das operações ou atividades para otimizar o desempenho dos negócios. O BI é usado de maneira integrada para criar uma visão abrangente da organização e ajudar a tomarem melhores decisões.

Os sistemas de BI têm feito parte da evolução das TI (tecnologia da informação) nas organizações e combinam a recolha, armazenamento e processamento de dados através de ferramentas analíticas com objetivo de apresentar informação complexa e competitiva aos gestores e decisores. Por isso, os sistemas de BI usam os dados presentes no seio das organizações para fornecer informação pertinente tendo em vista diferentes fins.

Os processos organizacionais geram uma quantidade significativa de dados e os sistemas de BI tiram partido dos mesmos, transformando os dados brutos existentes em informação útil. Essa informação irá suportar a tomada de decisão e promover o aumento de conhecimento dentro da organização. Os sistemas de BI baseados na análise de dados permitem a melhoria da estratégia organizacional.

Os sistemas de BI marcam a presença em vários setores de negócio e conseguem desenvolver estratégias e oportunidades de negócios, tendo em vista (a curto prazo) a vantagem competitiva no mercado e (a longo prazo) a estabilidade/sobrevivência da organização. Assim sendo, através do uso do BI, as organizações garantem a qualidade e promovem a criação e disseminação de conhecimento.

2.3 Armazém de Dados

O armazém de dados é um sistema utilizado para armazenar dados, de uma forma organizada, sendo considerado a base para o BI.

2.3.1 Conceito

O *data warehouse* (DW) ou armazém de dados, é um conjunto de dados orientado ao assunto, integrado, variável no tempo e não volátil usado no apoio à tomada de decisão e negócios (Inmon, 2005). O DW tipicamente contém dados corporativos no nível de granularidade mais elementar.

Na Figura 1 pode-se verificar como é feita a orientação dos dados ao assunto, sendo que os sistemas operacionais típicos são organizados em torno das aplicações da empresa. Por outro lado, o armazém de dados é organizado em torno dos principais assuntos/áreas da empresa, tais como clientes, pacientes, estudantes, vendas, entre outros. O DW foca-se na modelação e análise de dados para os decisores, não em operações diárias ou processamento de transações e fornece uma visão simples e concisa ao redor de assuntos específicos, excluindo dados não úteis no processo de suporte à decisão.

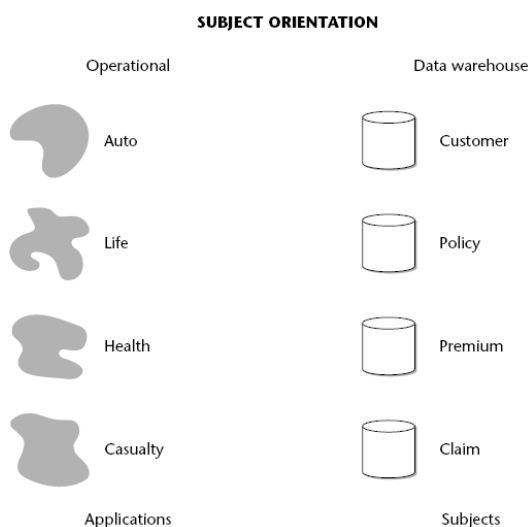


Figura 1 – Exemplo da orientação de dados ao assunto (Inmon, 2005)

O DW é construído integrando várias fontes de dados heterogêneas. Como este é “alimentado” por diferentes fontes, são aplicadas técnicas de limpeza e transformação dos dados, desde convenções de nomenclatura consistentes, formatos, medidas de atributos, entre outros (Figura 2).

2.3.2 Data Mart

Data marts são um subconjunto de dados do armazém de dados e são específicos para um departamento ou grupo de utilizadores contendo apenas os dados que são relevantes (Inmon, 2003). Os dados em cada *data mart* são geralmente para uma função específica, como análises da lucratividade do produto, análises de KPI (*Key Performance Indicators*), análises demográficas de clientes e assim por diante. Os *data marts* permitem uma resposta rápida a *queries*, uma vez que a informação neles contida é mais agregada e menos volumosa do que no DW.

2.3.3 Arquiteturas de Armazéns de Dados

A arquitetura do armazém de dados pode variar de acordo com as necessidades da empresa. Esta secção descreve duas arquiteturas finalizando com a comparação entre as mesmas.

2.3.3.1 Arquitetura BUS (Ralph Kimball)

A abordagem de Kimball propõe uma abordagem incremental ou *bottom-up* para a construção do DW. Esta abordagem começa por identificar quais são os principais processos de negócios que o DW necessita de responder. As principais fontes de dados são analisadas e documentadas, sendo utilizado o processo de ETL para trazer dados de todas as fontes e carregar numa área de preparação (*Data Staging Area*). Os dados são carregados num modelo dimensional que não é normalizado. O conceito fundamental da modelação dimensional é o esquema em estrela, no qual normalmente, existe uma tabela de factos rodeada por muitas tabelas de dimensão (Naeem, T., 2020).

Kimball propõe o conceito de "dimensões conformadas" em que as principais tabelas de dimensão são compartilhadas entre as diferentes tabelas de factos, ou seja, apenas serão construídas uma vez e usadas por todas as tabelas de factos (Group, K., 2020a). Isto garante que um conceito seja usado da mesma maneira entre os factos.

A Figura 5 mostra uma arquitetura típica de armazém de dados Kimball.

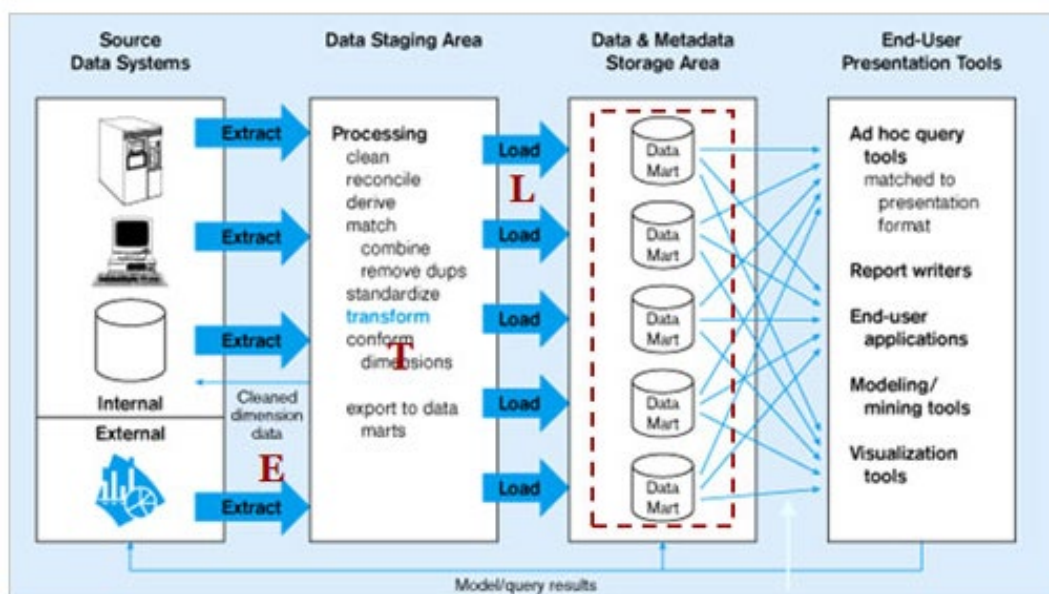


Figura 5 – Arquitetura BUS (ZenTut, 2020)

Esta arquitetura tem a vantagem de o processo de implementação ser decomposto em partes gerenciáveis, com cada parte focada nos processos de negócios específicos. Esta abordagem envolve soluções específicas para cada departamento, desenvolvidas de maneira mais independente e resulta em *data marts* isolados. Isso resulta em problema de coesão e interoperabilidade. Estes problemas são resolvidos pela arquitetura BUS, que aplica uma interface de barramento para melhorar a interação entre os *data marts*. Além disso, esta arquitetura é independente de tecnologia e base de dados.

2.3.3.2 Arquitetura CIF (Bill Inmon)

A abordagem de Inmon, CIF (*Corporate Information Factory*) propõe uma abordagem *top-down* para a construção do DW (Inmon, 2003). Como a redundância de dados é evitada o máximo possível, leva a uma identificação clara dos conceitos de negócios e evita anomalias de atualização de dados.

O DW é o componente mais importante do sistema e centraliza todos os dados vindos de vários sistemas operacionais, pelo que possui a informação normalizada.

Esse modelo normalizado torna o carregamento dos dados menos complexo, mas o uso dessa estrutura para consulta é difícil, pois envolve muitas tabelas e junções. Portanto, Inmon sugere a construção de *data marts* específicos, quando é necessário obter algum segmento do negócio. Todos os dados do armazém de dados são integrados e este é a única fonte de dados para os diferentes *data marts*. Isto garante que a integridade e a consistência dos dados sejam mantidas intactas em toda a organização (Figura 6).

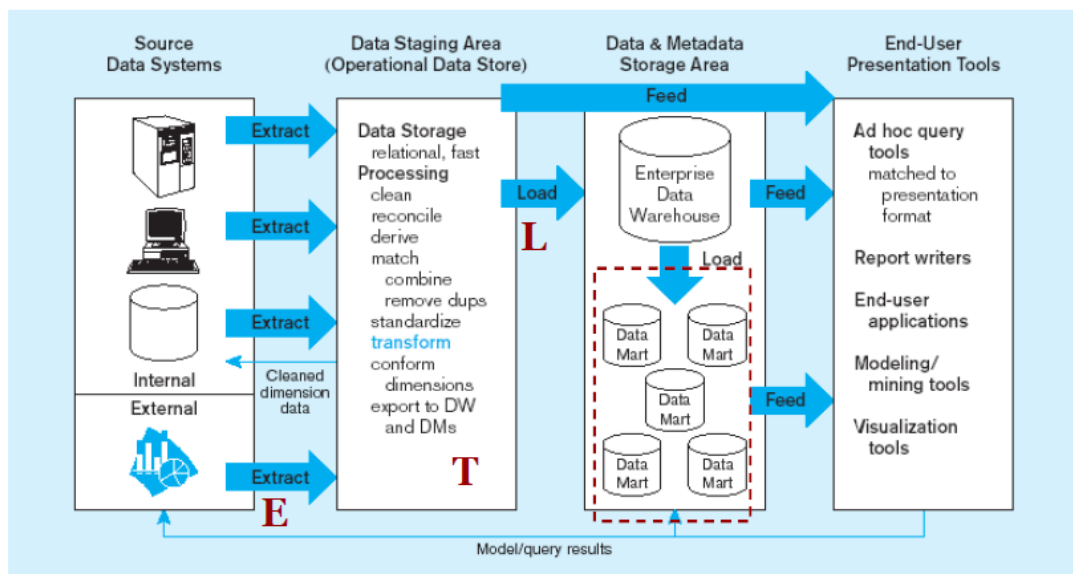


Figura 6 – Arquitetura CIF (ZenTut, 2020)

2.3.3.3 Comparação

A abordagem de Kimball difere da abordagem de Bill Inmon, sendo que ambas as abordagens são válidas, tendo em conta os objetivos traçados pelas mesmas. O acesso a dados da *Staging Area* na arquitetura BUS está completamente fora dos limites para os utilizadores e o acesso direto ao DW na arquitetura CIF é desencorajado (Inmon, 2003).

O modelo de Bill Inmon reconhece o DW apenas como sendo parte do sistema geral onde os *data marts* obtêm os seus dados. A informação é guardada no DW na 3ª forma normal. Enquanto o Kimball fornece uma arquitetura para integrar dados entre *data marts*. Portanto, essa união de *data marts* específicos de processos por si só, pode não ser suficiente para atender a todas as informações

necessárias para o suporte à decisão. Assim, Kimball considera o DW como sendo o conglomerado de todos os *data marts* da empresa. A informação é armazenada no modelo dimensional.

A abordagem de Inmon pode ser aplicada em organizações que dispõem de bastante tempo para a implementação do sistema de BI, enquanto que a abordagem de Kimball pode ser aplicada em organizações que pretendem demorar menos tempo na implementação e cujo foco é a otimização.

Na Tabela 1 pode-se comparar de forma resumida as vantagens e as desvantagens de cada um dos modelos de implementação.

Tabela 1 – Comparação de Arquiteturas de DW (Breslin, M., 2004)

	Vantagens	Desvantagens
Inmon	Engloba todos os SI (sistemas de informação) da empresa Construído de raiz Centralizado	Demorado Tecnicamente complexo
Kimball	Resultados rápidos Pouco complexo	Descentralizado

2.4 Modelação Dimensional

A modelação dimensional constitui a técnica usada para a construção do modelo de dados que suporta um armazém de dados.

2.4.1 Tabela de Factos

A tabela de factos é a tabela primária no modelo dimensional, sendo constituída por medidas numéricas do negócio que podem ser analisadas por meio de estatísticas (IBM, 2014a). Na sua génese é composta por uma chave primária, que é composta por chaves estrangeiras que se conectam a tabelas de dimensão.

2.4.2 Tabelas de Dimensão

As tabelas de dimensão definem os detalhes de cada transação. Na sua génese é composta por uma chave primária que se irá conectar à tabela de factos (IBM, 2014b).

2.4.2.1 Slowly Changing Dimension

Slowly Changing Dimensions (SCD), é uma técnica que permite classificar as atualizações dos atributos das dimensões de acordo com o tipo de mudança que se pretende que ocorra. Os tipos mais vulgares de SCD são: Tipo 1, Tipo 2 e Tipo 3 (McHugh, J., 2017).

O SCD de tipo 1 corresponde à alteração que não guarda nenhum histórico na dimensão. Trata-se do tipo mais simples, pois não existe controlo sobre a atualização de dados, apenas existe sobreposição de informação.

O SCD de tipo 2 é a técnica mais utilizada nas atualizações das dimensões. Com este tipo de SCD é adicionado um novo registo com as mudanças, preservando os dados anteriores.

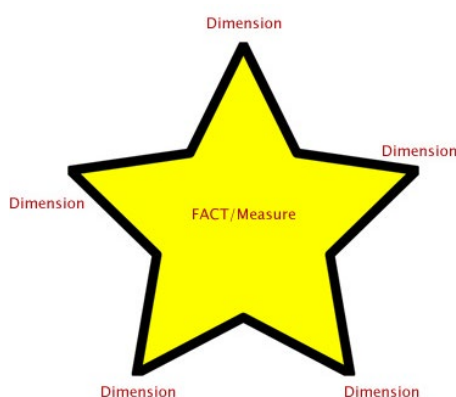
Por último, o SCD de tipo 3 permite manter as alterações no mesmo registo. Esta técnica possibilita a adição de uma nova coluna com os dados novos, mantendo também na antiga coluna o valor anterior.

2.4.3 Tipos de Modelos Dimensionais

O modelo dimensional contém as mesmas informações que o modelo normalizado, mas armazena os dados num formato que oferece a compreensão do utilizador, desempenho das consultas/análises e flexibilidade para a mudança.

O modelo dimensional é também conhecido por esquema em estrela, onde existe uma tabela central (tabela de factos), com várias junções conectadas às tabelas de dimensão (IBM, 2014c). Contudo existem três tipos de modelos dimensionais.

- Modelo em Estrela: constituído por uma tabela de factos central conectada a um conjunto de tabelas de dimensão, assemelhando-se a uma estrela. Relacionamento de muitos-para-um entre a tabela de factos e cada tabela de dimensão (Figura 7).



Star Schema

Figura 7 – Modelo em Estrela (MSBI, 2017)

- Modelo em Floco de Neve: refinamento do esquema em estrela. A tabela de factos encontra-se conectada a várias tabelas de dimensão, que podem ser conectadas a outras tabelas de dimensão por meio de um relacionamento muitos-para-um. Normalmente são normalizadas para a terceira forma normal. Cada tabela de dimensão representa exatamente um nível numa hierarquia (Figura 8).

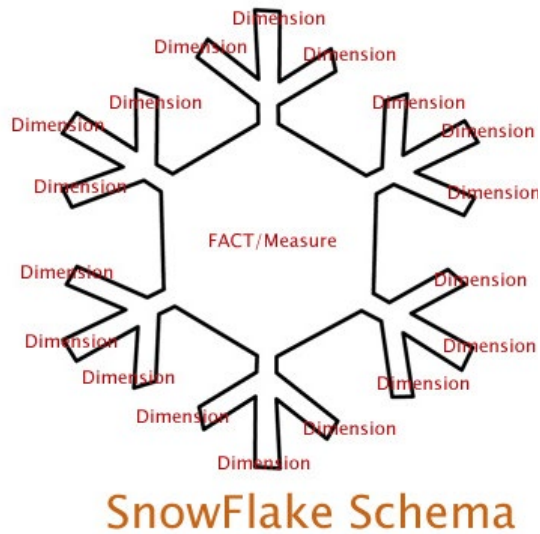


Figura 8 – Modelo em Floco de Neve (MSBI, 2017)

- Modelo em Galáxia: visa aproveitar os benefícios dos esquemas em estrela e dos esquemas de floco de neve sendo composta por várias tabelas de factos que partilham tabelas de dimensão, também conhecido como multi-estrela (Figura 9).

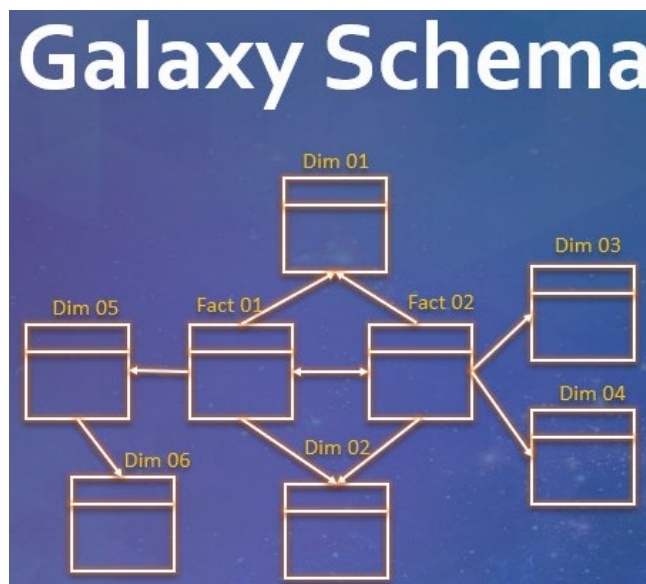


Figura 9 – Modelo em Galáxia (EDUCBA., 2020)

Os modelos dimensionais concentram-se no evento de todo o processo de negócio, dividindo os dados em medidas ou no contexto descritivo “quem, o quê, onde, quando, porquê e como” (Group, K., 2020b).

2.4.4 Metodologia de Kimball

Na metodologia de Kimball, existem quatro principais passos que devem ser seguidos durante o *design* do modelo dimensional, nomeadamente (Group, K., 2020c):

1. Selecionar a área de negócios: processo operacional suportado pelo sistema sobre o qual o *data mart* será desenvolvido;
2. Declarar o nível de granularidade: nível de detalhe dos dados a serem representados na tabela de factos;
3. Identificar as tabelas de dimensão: para além de se identificar as dimensões, envolve também descrever os atributos de cada dimensão;
4. Identificar os factos/medidas: definir quais as medidas que constituem a tabela de factos.

2.5 Processo de ETL

Uma das operações mais decisivas e que consomem mais tempo no ambiente DW/BI é a extração, transformação e carregamento de dados provenientes de vários sistemas operacionais, também chamada de ETL (*Extract, Transform and Load*) (Kimball R. and Ross M., 2013).

Extração, é a primeira etapa do processo, os dados são retirados do sistema de origem, geralmente de base de dados ou ficheiros internos ou externos à organização e destinam-se a uma área de preparação (*Staging Area*), onde são manipulados/transformados.

Depois de extraídos, existem inúmeras transformações, como limpar os dados (corrigir erros de ortografia, resolver conflitos de domínios, lidar com elementos ausentes, duplicados ou conflitos de regras de negócio) e garantir o formato dos dados, combinando dados de várias fontes. O processo de ETL adiciona valor aos dados com estas tarefas de limpeza e conformidade, uma vez que se garante a qualidade dos dados.

A etapa final do processo de ETL, é o carregamento de dados nos modelos dimensionais alvo da área de apresentação. A Figura 10 representa o processo de ETL.

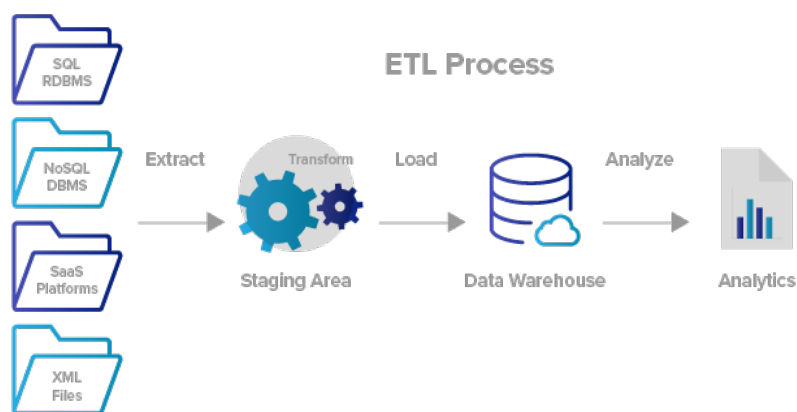


Figura 10 – Processo de ETL (Tobin, D., 2020)

2.6 Processamento Analítico de Dados

Processamento analítico *online* (OLAP) é uma tecnologia que organiza as bases de dados de grandes empresas e oferece suporte à análise complexa. Assim, o OLAP serve para agilizar a consulta de enormes quantidades de dados com grande rapidez de resposta. Uma das análises mais comuns é o filtro dos indicadores de negócio por categorias, por produto, por região, etc.

O OLAP é uma tecnologia de análise dinâmica e multidimensional dos dados. A informação analisada por ferramentas OLAP é proveniente de base de dados multidimensionais (DW), o que possibilita aos utilizadores combinar a informação de diferentes formas.

O cubo OLAP, elimina as limitações de bases de dados relacionais, fornecendo uma análise rápida de dados. Os cubos OLAP apresentam grandes quantidades de dados, permitindo também aos utilizadores o acesso pesquisável a qualquer informação. Os dados podem ser agregados, segmentados e repartidos conforme são necessários (Microsoft, 2019).

Existe um conjunto de operações que podem ser aplicadas sobre o cubo para se obter a informação pretendida, sendo as seguintes:

- *Slice & dice* – restringe os dados que estão a ser analisados a um subconjunto desses dados, ou seja, permite dividir a informação em partes menores;
- *Roll up (drill up)* – permite ao utilizador resumir a informação que está a analisar, ou seja, aumentar o grau de granularidade para diminuir o nível de detalhe da informação;
- *Drill down* – permite aumentar o nível de detalhe da informação que se está a visualizar, ou seja, diminuir o grau de granularidade.

2.7 Ferramentas

As ferramentas de *Business Intelligence* permitem recolher, analisar e converter dados em informações valiosas, sendo possível criar *dashboards* (representação gráfica da informação que se encontra em dois ou mais indicadores), relatórios (conjunto de análises de *dashboards*) e visualizações de dados para as organizações, sejam elas de pequeno, médio ou grande porte. Através deste tipo de ferramentas as organizações conseguem medir o seu desempenho e determinar estratégias para obter sucesso. As empresas obtêm informações acerca do seu crescimento, visualizam as tendências e comportamentos relativos aos seus clientes.

2.7.1 Ferramentas de ETL

A comparação entre ferramentas de ETL é efetuada, tendo em conta os critérios definidos no Quadrante Mágico 2020 pela Gartner (Gartner, 2020a). Neste Quadrante Mágico, a Gartner avalia os pontos fortes e fracos de 16 ferramentas de integração de dados que considera mais significativas no mercado e fornece aos leitores um gráfico (o Quadrante Mágico) apresentando as ferramentas com base na sua capacidade de execução e na sua integridade de visão. O gráfico está dividido em quatro quadrantes: concorrentes de nicho, desafiadores, visionários e líderes.

O quadrante mágico é aplicado na comparação das ferramentas de ETL existentes atualmente no mercado, como se pode ver na Figura 11.



Figura 11 – Quadrante Mágico das ferramentas de ETL (Gartner, 2020a)

Seguem-se alguns exemplos de ferramentas que podem ser utilizadas para a extração, transformação e integração dos dados.

2.7.1.1 Microsoft SQL Server Integration Services (SSIS)

A Microsoft oferece recursos de integração de dados via SQL Server Integration Services (SSIS), incluído na licença DBMS do SQL Server. Este também inclui integração de dados como parte do serviço de integração de dados em nuvem do Azure Data Factory (ADF) (Gartner, 2019).

O SSIS permite resolver problemas complexos através da cópia ou transferência de dados, carregamento de dados para DW, limpeza e mineração de dados e ainda, gerir objetos e dados do SQL Server (Figura 12). Inclui um conjunto de tarefas e transformações, ferramentas gráficas para criar pacotes e uma base de dados, onde se armazena, executa e gere pacotes (Chugugrace, 2018).

A Microsoft não relata uma contagem específica de clientes que utilizam as suas ferramentas de integração de dados.

A Gartner considera como forças desta ferramenta a velocidade de implementação, a facilidade de utilização e a capacidade de integração com outras ferramentas da Microsoft ou serviços do Azure. Outro ponto de referência é o facto de fornecer recursos de implantação híbrida, devido ao forte acoplamento para a integração local de dados com dados armazenados em vários serviços do Azure e repositórios de dados (Gartner, 2019).

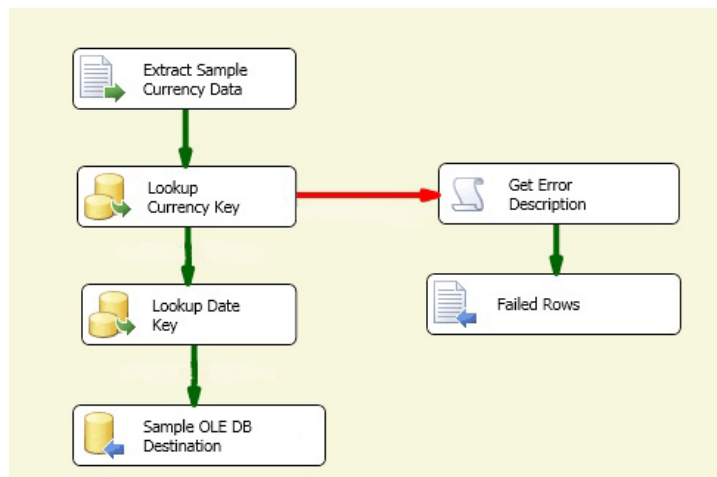


Figura 12 – Interface do SSIS (Chugugrace, 2019)

2.7.1.2 Oracle Data Integrator (ODI)

A Oracle oferece vários produtos de integração de dados, nomeadamente o Oracle Data Integration Platform Cloud Service (inclui o Oracle GoldenGate Cloud Service e Oracle Data Integrator Cloud Service), Oracle GoldenGate (para replicação de dados, sincronização e *streaming* de *big data*), Oracle Data Integrator (ODI - para integração em massa/lote), Oracle Big Data SQL (para cenários de virtualização de dados), Oracle Service Bus (para cenários de mensagens) e Oracle Integration Cloud Service (iPaaS). A base de clientes para este conjunto de produtos é superior a 12.000 organizações (Gartner, 2019).

O Oracle Data Integrator é uma plataforma de integração de dados abrangente que cobre todos os requisitos de integração de dados, ou seja, permite elevados volumes de dados com grande desempenho (Figura 13). O Oracle Data Integrator, baseia-se na sua arquitetura flexível e de alto desempenho, com suporte a *big data* e inclui interoperabilidade com o Oracle Warehouse Builder (OWB) para uma migração rápida e simples. Além disso, o ODI pode ser executado como uma solução única ou integrado com outras tecnologias e aplicações da Oracle (Oracle, 2020a).

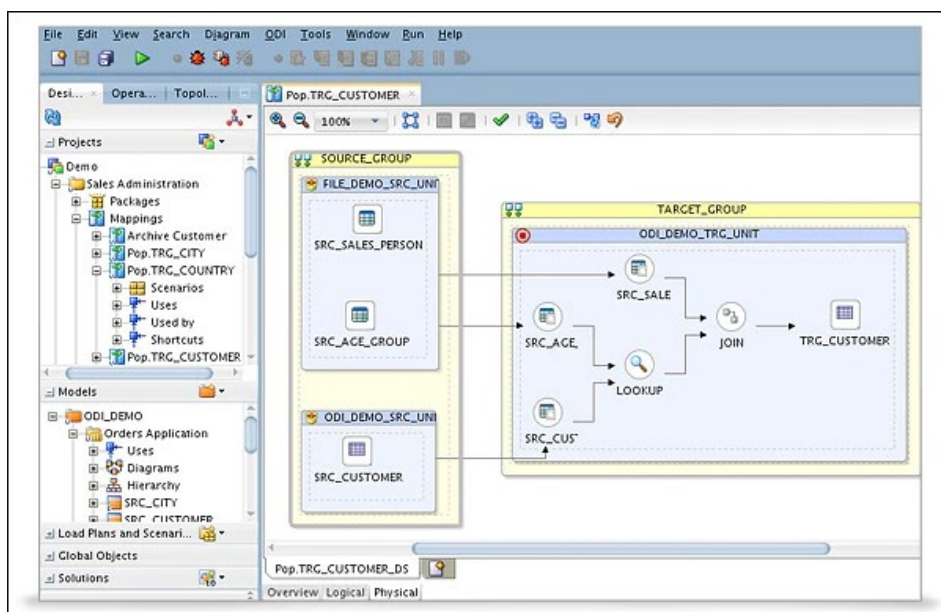


Figura 13 – Interface do ODI (Oracle, 2020b)

2.7.1.3 SAS Data Management

A SAS oferece vários produtos de integração de dados e a sua base de clientes é de 14.000 organizações (Gartner, 2019).

A SAS acredita que a sua vantagem no mercado é confirmada pela avaliação da Gartner, uma solução abrangente que transforma, integra, governa e protege dados (Figura 14), além de melhorar sua qualidade e confiabilidade gerais (SAS, 2020a).

Os seus pontos fortes passam pelos investimentos alinhados à visão de mercado. A SAS fez investimentos estratégicos de projetos de gestão de metadados em código aberto. De referir também, o alinhamento para operacionalizar a preparação de dados de autoatendimento (Gartner, 2019).

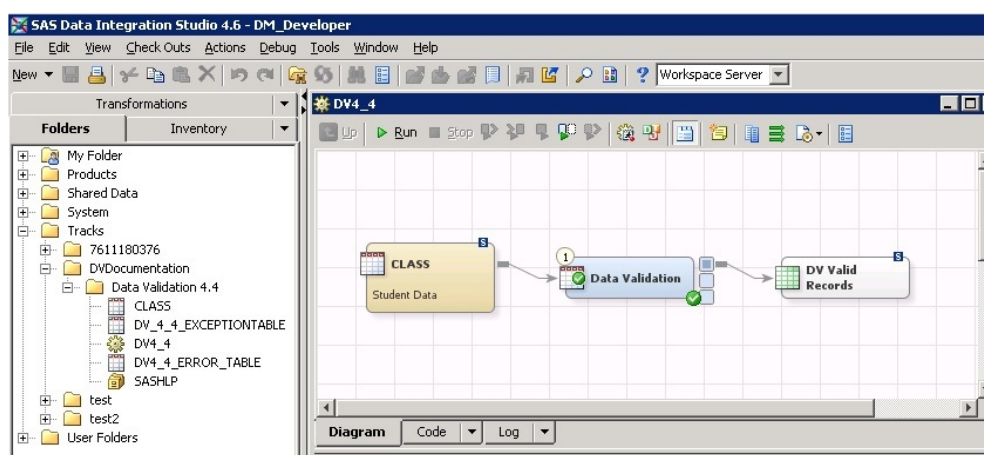


Figura 14 – Interface da SAS (SAS, 2020b)

2.7.1.4 Talend

Os clientes valorizam a abordagem unificada da Talend para solucionar a integração e a integridade dos dados, desafiando ambientes híbridos e de várias nuvens para uma gama completa de funções de TI e de negócios. A Talend acredita que a visão de uma abordagem unificada e abrangente de integração de dados (Figura 15) é o motivo pelo qual a Gartner posicionou a Talend no quadrante de líderes pelo quarto ano consecutivo (Talend, 2020a).

A Talend é uma ferramenta que inclui integração de dados, preparação de dados, catálogo de dados, administração de dados e gestão de API (em inglês: *Application Programming Interface*). Um dos pontos fortes é o seu compromisso de código aberto (Gartner, 2019).

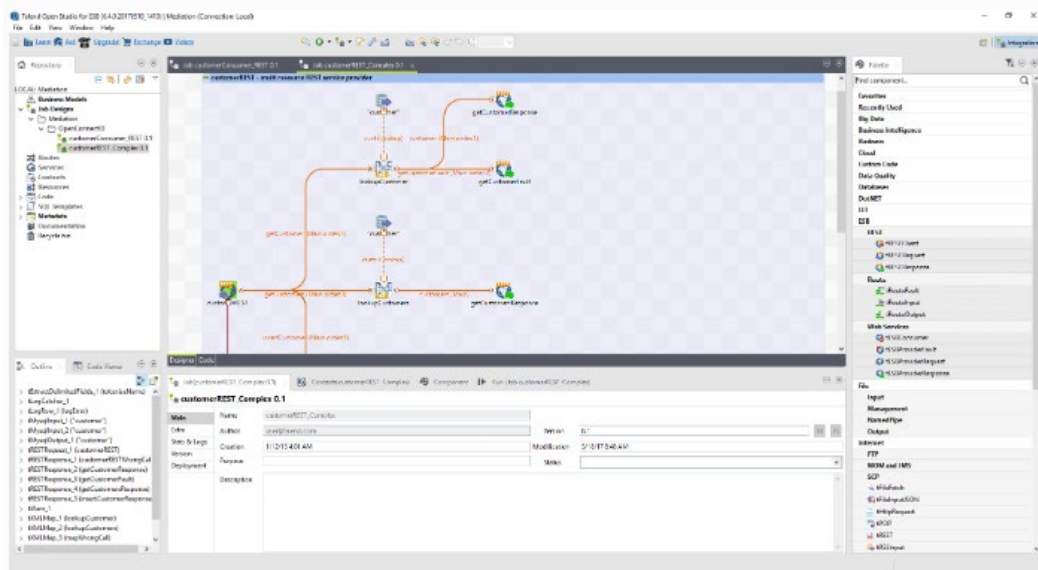


Figura 15 – Interface do Talend (Talend, 2020b)

2.7.1.5 Comparação entre Ferramentas de ETL

Na Tabela 2 observa-se uma pequena comparação entre as aplicações existentes.

Tabela 2 – Comparação entre Ferramentas de ETL

Ferramenta	Integração de Dados	Múltiplas Fontes	Consistência de Dados	Interoperabilidade
SSIS	Sim	Sim	Sim	Sim
ODI	Sim	Sim	Sim	Sim
SAS	Sim	Sim	Sim	Sim
Talend	Sim	Sim	Sim	Sim

Como se pode verificar pela comparação efetuada, todas as ferramentas selecionadas possuem as mesmas características quando aplicados os critérios propostos pela Gartner.

2.7.2 Ferramentas de Análise de Dados

A comparação entre as ferramentas de análise de dados é efetuada tendo por base o Quadrante de Gartner, tal como foi efetuada para as ferramentas de ETL.

Neste Quadrante Mágico, a Gartner avalia os pontos fortes e fracos de 22 ferramentas de análise de dados e BI que considera mais significativas no mercado e fornece aos leitores um gráfico (o Quadrante Mágico) (Gartner, 2020b).

O quadrante mágico é aplicado na comparação das ferramentas de análise de dados existentes atualmente no mercado, como se pode ver na Figura 16.



Figura 16 – Quadrante Mágico das ferramentas de análise de dados (Gartner, 2020b)

Aplicações como Qlik Sense, Power BI, Tableau, MicroStrategy ou Logi Analytics são soluções no mercado atual que permitem a análise dos KPI numa plataforma BI. Nas subsecções seguintes encontra-se a descrição das aplicações anteriormente referidas, dada a sua aplicabilidade ao projeto descrito.

2.7.2.1 Qlik Sense

O Qlik Sense é uma poderosa e eficiente plataforma de análise e monitorização de negócio. Esta inovadora ferramenta de BI permite gerir os dados eficazmente e tomar as decisões certas de forma mais rápida e inteligente (QlikTech, 2020).

Devido à sua simplicidade, qualquer pessoa pode consolidar, pesquisar, visualizar e analisar facilmente todos os dados existentes para obter uma perceção do negócio sem precedentes. O Qlik Sense é assim, uma plataforma de BI capaz de garantir uma gestão empresarial e uma visão única da informação.

A escalabilidade e a segurança são outras das suas principais características, assim como a possibilidade de ser utilizada a partir de diferentes tipos de *software* cliente (*browser, rich-client, mobile*) ou dispositivos (computadores, telemóveis) (Figura 17).

As suas principais vantagens são a redução do tempo de análise; garantia de retorno do investimento; aumento da satisfação e fidelização do cliente; autonomia do utilizador de negócio na gestão e análise de indicadores; pesquisa associativa que oferece respostas rápidas ao seu negócio; tecnologia *in memory* (em memória) que permite análises instantâneas dos dados em tempo real; acessibilidade total e independente do local; elevada escalabilidade; aumento da segurança e proteção de dados críticos.

Oferece análises simples através de indicadores, *dashboards* intuitivos, painel analítico e muitos outros componentes que respondem de forma dinâmica e em tempo real, sendo uma característica diferente de outras aplicações que possuem ferramentas de uso complexo, com necessidade de equipas de desenvolvimento e TI incluídas. Oferece também liberdade de tomada de decisões, podendo construir

os seus próprios indicadores de desempenho com informações que realmente sejam relevantes e podem visualizar dados, acompanhar a evolução dos indicadores e auxiliar a tomada de decisões.

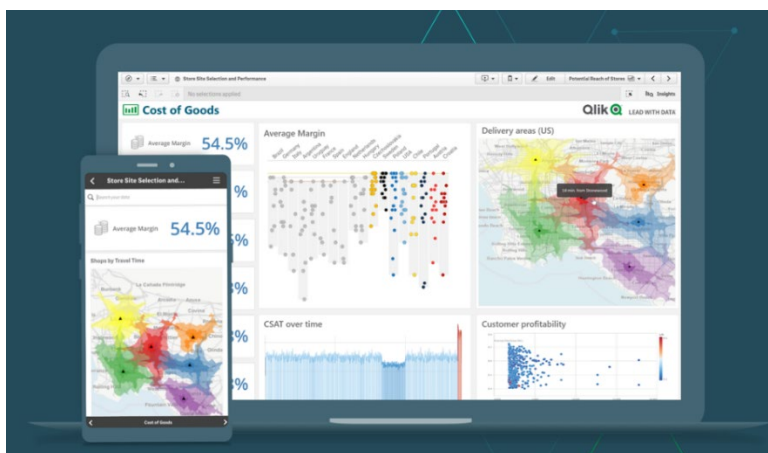


Figura 17 – Interface da Qlik Sense (QlikTech, 2020)

2.7.2.2 Power BI

O Power BI é uma solução de análise de negócios que permite visualizar dados e partilhar ideias na organização ou incorporá-los em aplicações ou *sites* (Microsoft, 2020a).

O Power BI faz a conexão a centenas de fontes de dados, simplifica a preparação de dados e conduz a análise ad hoc. Pode-se produzir relatórios e publicá-los para a organização consumir na Web e em dispositivos móveis, podendo criar painéis personalizados com uma visão única de 360 graus dos seus negócios (Figura 18). Esta ferramenta escala toda a empresa, com hierarquia e segurança embutidas.

O Microsoft Power BI oferece preparação de dados, descoberta de dados, painéis interativos e análises aumentadas por meio de um único produto (Gartner, 2019).

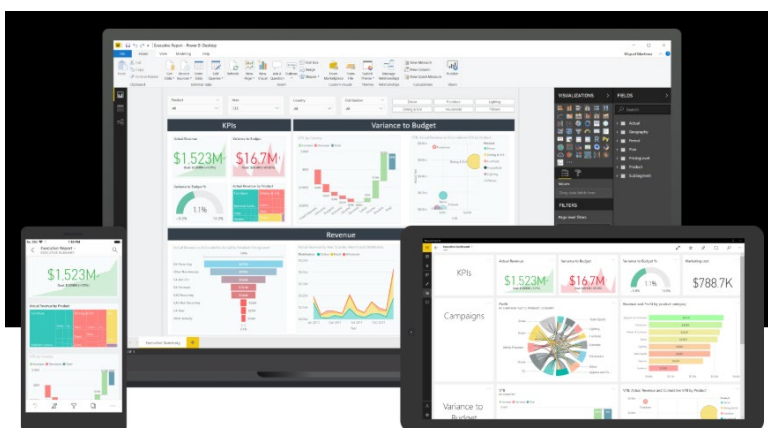


Figura 18 – Interface da Power BI (Microsoft, 2020b)

2.7.2.3 Tableau

O Tableau oferece uma experiência de exploração intuitiva e interativa baseada num visual que permite que os utilizadores de negócios e qualquer autor de conteúdo acedam, preparem, analisem e apresentem descobertas nos seus dados sem habilidades técnicas ou codificação (TABLEAU, 2020a). O Tableau oferece quatro produtos principais: o Tableau Desktop, o Tableau Prep, o Tableau Server e o

Tableau Online (na nuvem) (Figura 19). O Tableau comprometeu-se a oferecer a todos os elementos da empresa o poder de explorar os dados, que se tornou uma característica esperada de todas as plataformas modernas de análise e BI. Durante o ano passado, o Tableau forneceu vários recursos empresariais, sendo esta uma mudança na estratégia que atraiu os compradores de TI, como por exemplo, lançou fontes de dados certificados e recomendados para melhorar a gestão de grandes implantações; suporte a dados híbridos da nuvem; agendamento e alertas; um SDK e APIs aprimorados; e acrescentou colaboração.



Figura 19 – Interface da Tableau (TABLEAU, 2020b)

2.7.2.4 MicroStrategy

Com a mais ampla variedade de objetos reutilizáveis, combinada com uma arquitetura de memória altamente escalável, a MicroStrategy facilita o fornecimento de análises personalizadas para milhares de utilizadores em toda a empresa, com apenas uma pequena equipa (MicroStrategy, 2020a). A experiência de exploração visual interativa totalmente equipada da MicroStrategy é combinada com os melhores recursos corporativos e de relatórios corporativos fornecidos numa única plataforma e fluxo de trabalho totalmente integrados (Figura 20). Por esta razão torna-se a mais adequada para relatórios de grande escala e implantações de procura de dados regulada para conjuntos de dados grandes e complexos do que a maioria das outras ofertas. A MicroStrategy foi criado para permitir que as organizações implantem rapidamente aplicações analíticas e de segurança sofisticada em escala, sendo especialmente adequada para fornecer aplicações de alto desempenho e atender às demandas de *business intelligence* de todos os utilizadores e de todas as organizações.

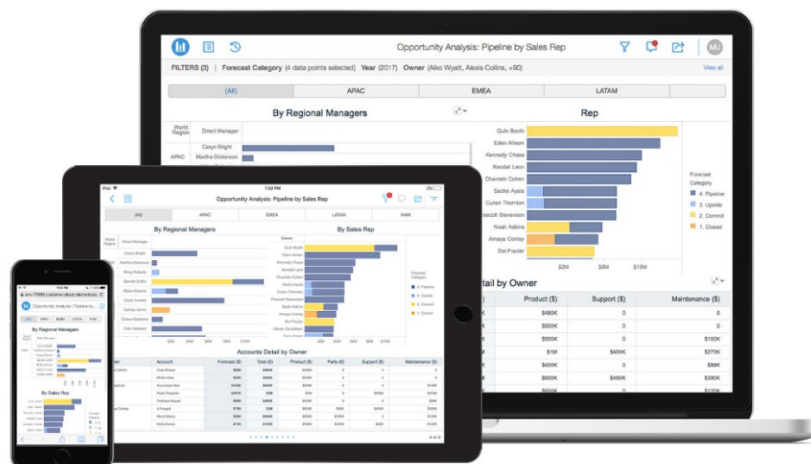


Figura 20 – Interface da MicroStrategy (MicroStrategy, 2020b)

2.7.2.5 Comparação entre Aplicações de Análise de Dados

Na Tabela 3 observa-se uma pequena comparação entre as aplicações existentes.

Tabela 3 – Comparação entre Aplicações de Análise de Dados

Aplicação	Interação com base de dados	Permite Autenticação	Diferentes tipos de clientes	Preço
Qlik Sense	Sim	Sim	Web e mobile	<ul style="list-style-type: none">• Qlik Sense - Grátis• Qlik Sense Business - €27,5/mês/utilizador• Qlik Sense Enterprise: Professional - €64,21/mês/utilizador• Analyzer - €36,69/mês/utilizador
Power BI	Sim	Sim	Web e mobile	<ul style="list-style-type: none">• Power BI - Grátis• Power BI Pro - €8,40/mês/utilizador• Power BI Enterprise - €4212,30/mês/recurso
Tableau	Sim	--	Web e mobile	<ul style="list-style-type: none">• Tableau - Grátis• Tableau Creator - €64,21/mês/utilizador• Tableau Explorer - €32,10/mês/utilizador• Tableau Viewer - €11,00/mês/utilizador
MicroStrategy	Sim	--	Web e mobile	<ul style="list-style-type: none">• MicroStrategy - Grátis

Como se pode verificar pela comparação efetuada, todas as ferramentas selecionadas possuem uma versão grátis para testar a ferramenta. A Power BI e a Qlik Sense são as ferramentas que mais se assemelham.

2.8 BI na Área dos Seguros

A área de seguros é obrigada a trabalhar com um grande volume de dados, uma vez que engloba muitas informações a respeito dos clientes, das apólices, de valores, comissões, propostas enviadas, entre outras. São muitos dados que, para serem aproveitados, necessitam de ser separados por categoria e, depois, interpretados (Tecnologia, F., 2018). A partir dessa interpretação torna-se possível verificar os pontos fortes e fracos da seguradora: “onde estamos a falhar? É na fidelização do cliente ou no envio da proposta? O que podemos melhorar para fidelizá-los?”, por isso, é vital o acompanhamento das mudanças para que possam ter utilidade no processo de tomada de decisões.

Torna-se necessário um fator estratégico que possa gerar uma vantagem competitiva proporcionando informação privilegiada para responder aos problemas do negócio.

As seguradoras são uma área que tem investido num sistema de BI, no sentido de ganhar uma capacidade de resposta mais rápida às suas necessidades de negócio, melhor desempenho operacional, capacidade de resposta e decisão (Computerworld, 2014).

A introdução de soluções tecnológicas permitiu que os sistemas tradicionais fossem agregados, para que fosse mais fácil a gestão de dados para auxiliar na visualização de históricos e retroceder operações,

descartando alterações feitas (*Roll-back*); controlo de acesso de dados; parâmetros de risco flexíveis; perfil de valor e risco; cálculos de seguro dinâmicos (TI, K., 2015).

Integrar as diversas aplicações de seguros num único sistema, torna-se vantajoso pois consegue-se aceder a diversos *scripts* e pacotes de lógica de negócios, modelos de cálculo, indicadores de desempenho (KPI), soluções em forma de gráficos e relatórios.

Os sistemas de BI permitem melhorar o processo de tomada de decisões, agilizar processos internos e que as soluções para todas as áreas sejam cuidadosamente projetadas com um foco na agilidade e no resultado. Algumas operações por áreas são as seguintes:

- Marketing:
 - Análise de vendas: Trata-se de um processo periódico de investigação dos resultados do setor, em busca de *insights* valiosos para as próximas estratégias;
 - *Cross selling* (ou venda cruzada): É uma técnica de vendas utilizada para convencer o cliente a comprar mais produtos do que havia pensado inicialmente;
 - Desempenho da seguradora: Definir metas de futuros negócios, monitorizar o progresso, avaliar a eficácia e alinhar objetivos e táticas;
- Subscrição:
 - Análise de rentabilidade: indica se o negócio está a gerar lucro;
 - Taxa de sinistralidade: razão expressa em percentagem, entre os custos com os serviços e o prémio pago;
 - Frequências de sinistro: o número de ocorrências de sinistro;
 - Parâmetros de risco: critérios/métodos/fatores que influenciam o risco associado;
- Sinistros:
 - Participações de sinistro: ocorrência de um acontecimento/sinistro;
 - Pagamentos: tipo de pagamento (débito direto, multibanco, entre outros);

A i2S Brokers desenvolveu um módulo específico de – BI para seguros – que lhe permite tomar decisões baseadas numa forte “inteligência de negócio” através de previsões do que está para acontecer e possa responder adequadamente, duma forma rápida e eficaz às alterações do mercado (Brokers, i2S., 2020).

3 Análise de Valor

Para a realização de um projeto é necessário efetuar um estudo da análise de valor do mesmo. Assim, este capítulo, refere-se a uma metodologia utilizada para a identificação de funções e o relacionamento com os custos para realizar cada função, visando identificar e eliminar custos desnecessários.

3.1 Definição de Análise de Valor e Terminologia

Lawrence Miles definiu a Análise de Valor (AV) como sendo "uma abordagem organizada e criativa que tem como objetivo a identificação eficiente do custo desnecessário, isto é, custo que não fornece qualidade, nem uso, nem aparência, nem características do cliente" (Lawrence D. M., 1961).

A AV é também definida como aplicação de técnicas reconhecidas para identificar as funções de um produto ou serviço e fornecer essas funções ao menor custo possível, atendendo às necessidades do cliente. Por isso, pode-se afirmar que a técnica de Análise de Valor suporta atividades de redução de custo, relacionando o custo dos componentes às suas contribuições de função.

Diferentes clientes interpretarão o valor de um produto de maneiras diferentes. O valor pode ser expresso como maximização da função do produto em relação ao seu custo:

Valor = (desempenho + capacidade) / custo = Função / custo

A Análise de Valor define uma "função básica" como qualquer coisa que faça o produto funcionar ou vender, sendo que uma função básica não pode mudar. Funções secundárias, também chamadas de "funções de suporte", podem ser modificadas ou eliminadas para reduzir o custo do produto.

3.2 Fuzzy Front-End

O processo de negócio e de inovação sugerido por Peter Koen, pode ser dividido em três áreas: o *Front-End* difuso (FFE), o processo de desenvolvimento de novos produtos (NPD) e a comercialização (Koen P. A., 2002).

Cada vez mais a atenção das empresas está focada nas atividades de *front-end* que precedem a um processo formal e estruturado, a fim de aumentar o valor, a quantidade e a probabilidade de sucesso dos conceitos de alto lucro que entram no desenvolvimento e na comercialização do produto.

O novo conceito de desenvolvimento (NCD) divide o *front-end* em três partes: o motor, a roda e a jante. O motor consiste nos elementos principais - atributos organizacionais, como visão, estratégia, recursos, cultura, equipas e colaboração. A roda, a parte interna do modelo, compreende cinco elementos de atividade:

- 1) Identificação da oportunidade;
- 2) Análise da oportunidade;
- 3) Geração de ideias e enriquecimento;
- 4) Seleção de ideias;
- 5) Definição do conceito.

O terceiro elemento do modelo NCD, a jante, consiste no ambiente externo, fatores que influenciam o mecanismo e os elementos de atividade.

Os cinco elementos de atividade são aplicados ao contexto do problema em questão, sendo que os resultados obtidos são apresentados nas secções seguintes.

3.2.1 Identificação da Oportunidade

A oportunidade surge da necessidade de reestruturação do processo de extração de informação da ferramenta “CUBO”. Devido às diferentes áreas envolvidas e com distintas definições de conceitos pretende-se a consolidação da informação e existe ainda, a falta de variáveis importantes para a gestão do negócio na empresa em questão, pelo que surgiu a oportunidade de colmatar este problema.

A oportunidade passa pelo desenvolvimento de um sistema de BI, sendo este um fator estratégico que pode gerar uma vantagem competitiva proporcionando informação privilegiada para responder aos problemas do negócio.

3.2.2 Análise da Oportunidade

A análise da oportunidade determina como a oportunidade se encaixa no mercado da empresa e pontos fortes, lacunas e ameaças da tecnologia. Assim, como já foi referido, as seguradoras são uma área que tem investido num sistema de BI, no sentido de ganhar uma capacidade de resposta mais rápida às suas necessidades de negócio, melhor desempenho operacional, capacidade de resposta e decisão, sendo imprescindível o desenvolvimento do mesmo para esta organização.

O sistema de BI:

- Reduz os custos;
- Melhora a eficiência e a produtividade do negócio;
- Melhora o processo de tomada de decisão, baseando-a em evidências;
- Facilita o acesso e o compartilhamento de informações que darão suporte à gestão;
- Permite análises com navegação rápida;
- Entrega resultados corretos, no tempo certo, para a pessoa exata;
- Minimiza erros e riscos;
- Fideliza clientes/Obtém informações valiosas sobre o comportamento de clientes;
- Dá maior competitividade à empresa;
- Facilita a identificação de oportunidades;
- Melhora as rotinas internas;
- Permite o reconhecimento do mercado.

Contudo, existem custos associados ao desenvolvimento do sistema de BI, bem como as fontes de dados estão desatualizadas e propícias a erros, não menosprezando o facto de a tecnologia estar em constante evolução.

3.2.3 Geração de Ideias e Enriquecimento

O sistema de *Business Intelligence* auxilia a organização a estruturar um plano estratégico para construir vantagens competitivas sustentáveis que agreguem valor ao negócio e às soluções oferecidas aos clientes. O suporte à tomada de decisão é oferecido pela facilidade de visualização de análises e relatórios resultantes dos processos. Isto acontece uma vez que se tem acesso ao panorama geral da

empresa. Assim, pode-se decidir de forma pensada, estratégica e inteligente em questão de minutos, poupando tempo e esforços. Surge então, a ideia de desenvolver um sistema de BI para guiar o negócio da empresa estrategicamente.

As ferramentas de BI auxiliam na identificação de tendências de consumo e mudanças no comportamento dos clientes. Com isto, auxiliam na detecção de oportunidades de negócios, no desenvolvimento e lançamento de novos produtos, na ampliação de receitas e na conquista de novos clientes e mercados. Consequentemente, conseguem ganhar competitividade.

3.2.4 Seleção de Ideias

A ideia selecionada passa pelo desenvolvimento de um sistema que permite obter a informação detalhada, de forma resumida e de fácil interpretação, uma vez que permite a geração de tabelas, gráficos e relatórios.

3.2.5 Definição do Conceito

Além da informação surgir de forma resumida, fácil de interpretar, é possível obter indicadores sob a forma de *dashboards* e relatórios. Para além disto, áreas distintas têm especificidades diferentes que devem ser levadas em consideração na construção de modelos, para que as análises tenham relevância para a organização.

3.3 Análise de Valor

“O valor foi definido em diferentes contextos teóricos como necessidade, desejo, interesse, padrão/critério, crenças, atitudes e preferências” (Nicola et al., 2012).

“A criação de valor é fundamental para qualquer negócio, e qualquer atividade de negócio é sobre trocar bens ou serviços tangíveis e/ou intangíveis e ter o seu valor aceite e recompensado por clientes, dentro ou fora da empresa ou rede colaborativa” (Nicola et al., 2012).

Zeithmal (1988, p14), sugeriu que "o valor percebido é a avaliação geral do consumidor sobre a utilidade de um produto com base nas percepções do que é recebido e do que é dado".

Lindgreen e Wynstra (2005) dizem que o valor percebido pelo cliente, significa algo diferente do valor percebido pelo utilizador, por exemplo, o produtor é menos sensível ao preço, enquanto o consumidor é mais sensível à qualidade do produto.

Valor para o cliente (VC) é qualquer percepção pessoal de vantagem resultante da associação do cliente com a oferta de uma organização e pode ocorrer como redução no sacrifício; presença de benefício (percebidos como atributos ou resultados); resultante de qualquer combinação ponderada de sacrifício e benefício (determinado e expresso, racional ou intuitivo); ou uma agregação, ao longo do tempo, de um ou de todos eles (Woodall T., 2003).

O DW é uma das ferramentas de BI mais importantes numa organização. Transforma a enorme quantidade de dados gerados de várias fontes num formato fácil de entender, oferece suporte à decisão e permite o desenvolvimento de análises, relatórios e outras funções de BI.

Devido ao seu funcionamento, os dados são extraídos regularmente de diferentes sistemas de origem usando o *software* de ETL. Os dados são limpos, formatados, resumidos e transferidos para os servidores. A instalação de um DW requer servidores dedicados que executam um sistema de gestão de bases de dados (DBMS), implicando custos adicionais (Technologies, F., 2014).

Assim, o valor da solução passa pela validade e a utilidade do sistema de apoio à decisão, e consequentemente, pela oferta de maximizar a qualidade de dados, na medida em que os benefícios superem os custos de implementação. Esta ferramenta será utilizada pelos colaboradores da empresa, para que estes possam aceder aos dados relevantes de uma forma mais rápida e com maior confiança nos dados e assim, tomar decisões ponderadas no menor espaço de tempo.

3.4 Proposta de Valor

Segundo Alexander Osterwalder "proposta de valor é entendida como o conjunto de produtos e serviços que criam valor para um segmento específico de clientes" (Osterwalder et al., 2014). Nesta perspetiva uma proposta de valor é uma afirmação através da qual se pretende demonstrar a correspondência entre as necessidades dos consumidores e os benefícios que resultam de utilizar o produto.

Uma proposta de valor pode ser construída na perspetiva do produto, na perspetiva do negócio e na perspetiva do investidor.

Deve-se começar por criar uma afirmação que descreva o problema que se identificou existir ao nível do mercado e ao qual o produto irá dar resposta.

Assim, a Real Vida Seguros, com o intuito de promover melhores decisões empresariais em toda a organização, pretende desenvolver um sistema de BI, uma ferramenta que permite que qualquer pessoa de qualquer nível da organização tome decisões com confiança, através de análises atualizadas e de informação de qualidade. Com isto, os colaboradores conseguem obter a informação de que necessitam num único local, o que oferece tudo o que precisam para aceder, visualizar e analisar dados. As análises rápidas e flexíveis ajudam a maximizar o impacto dos dados, traçando uma base para um futuro impulsionado por estes.

3.5 Modelo de Negócio CANVAS

O Modelo de Negócio Canvas foi proposto por Alexander Osterwalder. "Um modelo de negócio CANVAS descreve a lógica de como uma organização cria, proporciona e obtém valor" (Osterwalder A., Pigneur Y., 2010).

Este modelo descreve nove segmentos que formam os alicerces do modelo de negócios numa única página. Segue-se uma descrição dos nove segmentos os quais devem ser preenchidos pela seguinte ordem:

- Segmentos de clientes – vários grupos de pessoas ou organizações que a empresa visa atingir e servir (Figura 21);

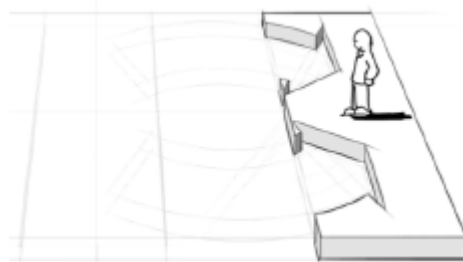


Figura 21 – Segmentos de clientes (Berlin S. D., 2012)

- Proposta de valor – conjunto de produtos e serviços que criam valor para um segmento de clientes específico. O importante não é o produto ou serviço em si, mas o valor percebido pelo cliente (Figura 22);

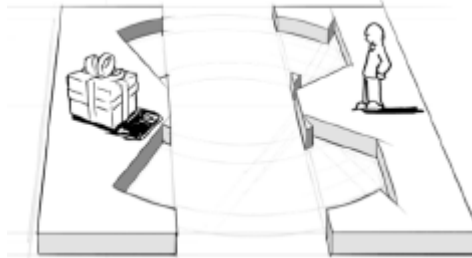


Figura 22 – Proposta de valor (Berlin S. D., 2012)

- Canais – como a empresa comunica e tenta influenciar os seus segmentos de clientes, para lhes proporcionar uma proposta de valor. Os canais podem ser de comunicação, distribuição, venda e apoio técnico (Figura 23);

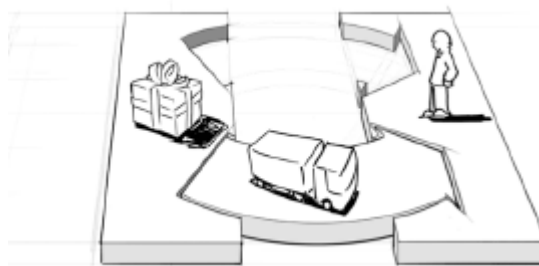


Figura 23 – Canais (Berlin S. D., 2012)

- Relações com os Clientes – tipo de relações que uma empresa estabelece com os seus segmentos de clientes (Figura 24);



Figura 24 – Relações com os Clientes (Berlin S. D., 2012)

- Fluxos de Rendimento – ou fontes de receita, representa o dinheiro que uma empresa gera a partir de cada segmento de clientes (Figura 25);

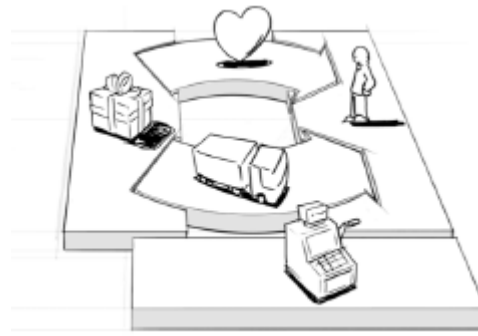


Figura 25 – Fluxos de Rendimento (Berlin S. D., 2012)

- Recursos-Chave – ativos mais importantes para o funcionamento do modelo de negócio. Permitem que uma empresa crie e ofereça a proposta de valor, mantenha as relações com segmentos de clientes e consiga rendimentos e a vantagem competitiva (Figura 26);

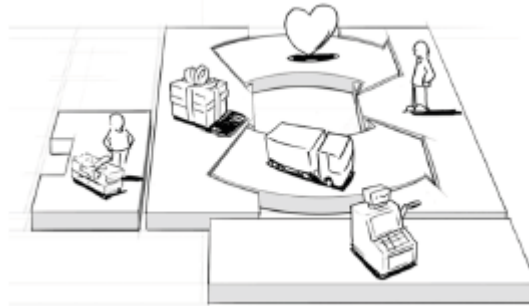


Figura 26 – Recursos-Chave (Berlin S. D., 2012)

- Atividades-Chave – atividades mais importantes, críticas ou essenciais, que a empresa deve fazer para conseguir que o modelo de negócio funcione corretamente (Figura 27);

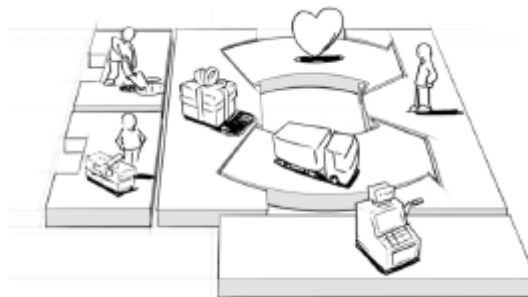


Figura 27 – Atividades-Chave (Berlin S. D., 2012)

- Parcerias-Chave – rede de fornecedores e de parceiros que fazem o modelo de negócio funcionar (Figura 28);

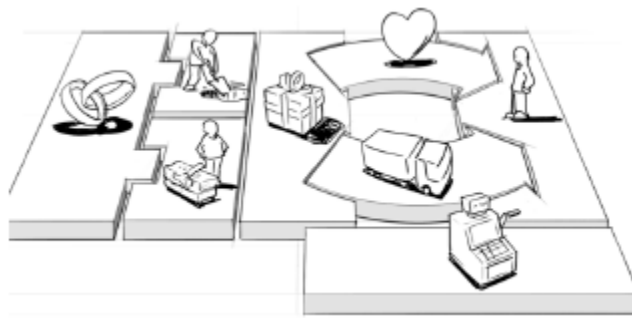


Figura 28 – Parcerias-Chave (Berlin S. D., 2012)

- Estrutura de Custos – principais custos envolvidos na operação de um modelo de negócios (Figura 29).

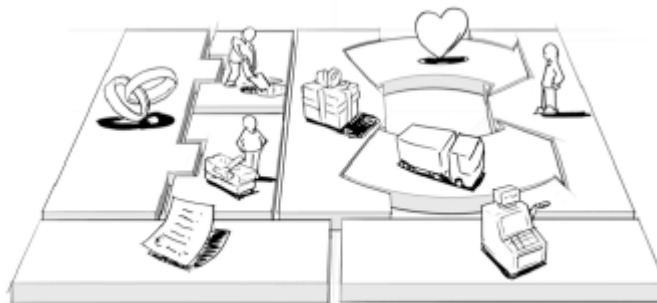


Figura 29 – Estrutura de Custos (Berlin S. D., 2012)

Na Figura 30 é possível verificar o modelo de negócio de Canvas utilizado para descrever a ideia de negócio que se pretende desenvolver.

Business Model Canvas - BI em Dados de Seguros de Vida

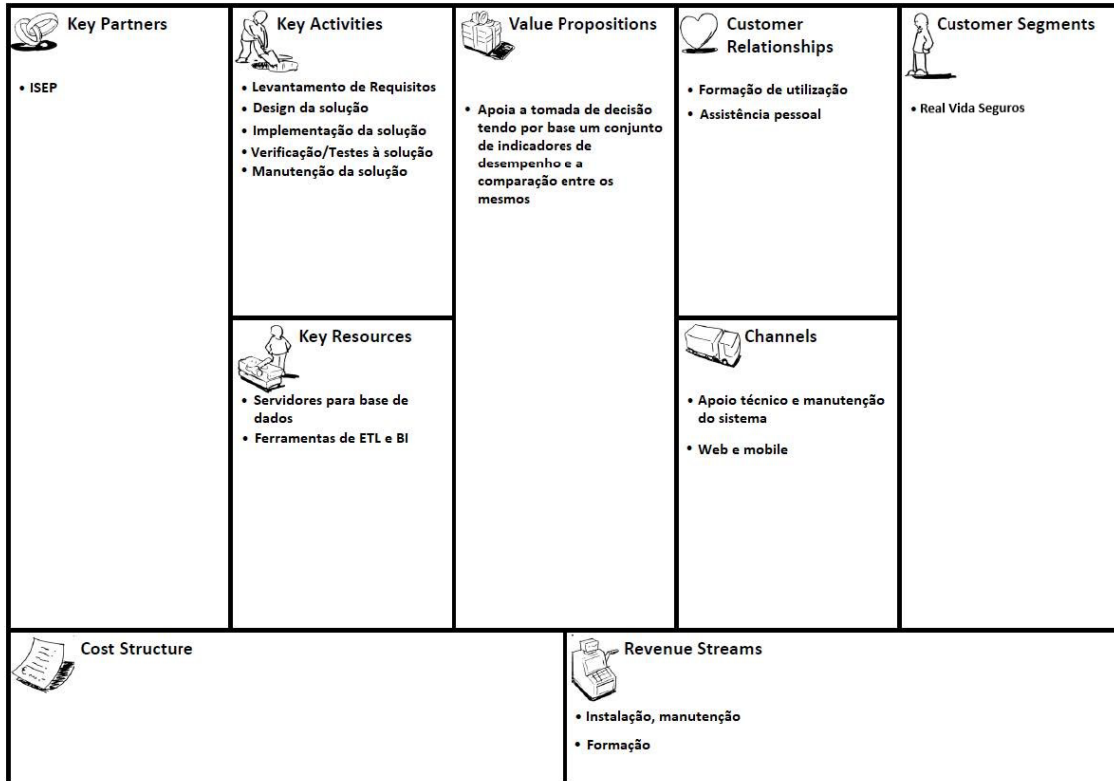


Figura 30 – Modelo de Negócio CANVAS

3.6 Cadeia de Valor de Porter

A cadeia de valor pode ser definida como um conjunto de atividades e operações, de modo a criar valor desde as matérias primas até ao produto ou serviço final.

Sistema de atividades interdependentes conectadas por elos, enquadrando as atividades de valor da empresa em categorias, distribuídas em: atividades primárias e atividades de apoio (ou suporte). Existem dois tipos possíveis de vantagem competitiva (liderança de custos ou diferenciação) em cada etapa da cadeia de valor (Porter, 1989).

Conforme a Figura 31, as estruturas das atividades da cadeia de valor são subdivididas em dois conjuntos: as atividades de apoio e as atividades primárias.



Figura 31 – Cadeia de Valor de Porter (Serafim A., 2013)

O conjunto das atividades primárias é representado por:

- Logística de Entrada – são as atividades relacionadas com a recepção, armazenamento e distribuição das entradas. Relativamente ao projeto em questão as entradas são todos os registos de dados relativos a seguros de vida.
- Operações – são as atividades que transformam entradas em saídas. Neste caso, os dados iniciais são transformados em informações úteis para o utilizador final.
- Logística de Saída – são as atividades que entregam o produto ou serviço ao cliente. Nesta circunstância, não existe entrega de um produto físico, sendo que o que se pode considerar é a entrega do sistema de BI.
- Marketing e Vendas – são os processos que diferenciam os produtos da concorrência persuadindo os clientes a comprar os mesmos. A criação de tutoriais que mostram o modo de utilização do sistema, assim como a facilidade de utilização podem ser aplicados neste contexto.
- Serviços – são as atividades relacionadas com a manutenção do valor do produto ou serviço após a compra. Neste contexto, a existência de um serviço de *helpdesk* associada ao produto é essencial, uma vez que estes são os responsáveis pela instalação, correção e ajustes do produto. Bem como, o desenvolvimento de tutoriais para aprendizagem do mesmo.

O conjunto das atividades de apoio é constituído por:

- Aquisição – obter os recursos necessários para operar. Neste cenário, é necessário adquirir servidores para alojar bases de dados e *software* para a análise dos dados.
- Desenvolvimento tecnológico – atividades relacionadas com a gestão e processamento de informações. Minimizar os custos de tecnologia da informação, manter-se atualizado com os avanços tecnológicos e manter a excelência técnica são fontes de criação de valor. Por isso, através da utilização de tecnologias e ferramentas recentes no mercado e que contribuem para inovar os produtos, é possível alcançar este objetivo.
- Gestão de Recursos Humanos – forma como uma organização recruta, contrata, treina, motiva, recompensa e retém os funcionários. Existe um conjunto de ações que incentivam os colaboradores a permanecerem fiéis à organização nomeadamente: promoções, bom ambiente de trabalho e flexibilidade horária.
- Infraestrutura – sistemas de suporte de uma empresa e as funções que lhe permitem manter as operações diárias. Para tal, o sistema de BI deverá permitir a gestão da organização por diferentes departamentos (e.g.: Financeiro, Recursos Humanos, Comercial).

Finda a análise acima referida, é evidenciado o potencial que o tema proposto tem na criação de valor no negócio.

3.7 Processo de Análise Hierárquica (AHP)

O Processo de Análise Hierárquica (*Analytic Hierarchy Process – AHP*) foi desenvolvido por Tomas L. Saaty em meados da década de 1970 e é uma técnica de análise de decisão e solução de problemas complexos, envolvendo múltiplos critérios (Vargas, R. V., 2010).

A aplicação do AHP começa com a decomposição de um problema numa hierarquia de critérios, sejam eles qualitativos ou quantitativos, facilitando, assim, sua compreensão e avaliação (Figura 32).

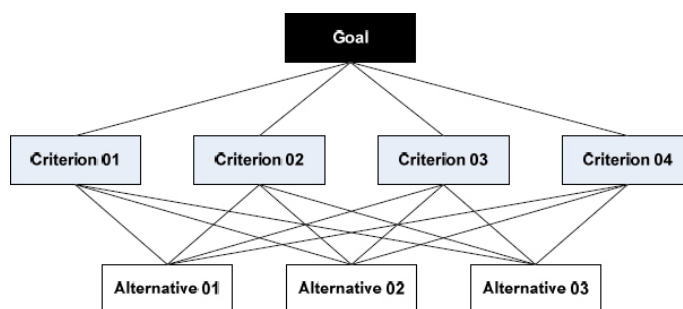


Figura 32 – Exemplo de uma hierarquia de critérios/objetivos (Vargas, R. V., 2010)

O problema é decomposto em fatores, que por sua vez podem ainda ser decompostos até ao nível mais baixo, de forma a estabelecer relações e sintetizar a informação, comparando aspetos de cada alternativa.

O processo AHP pode-se dividir nas etapas que são enunciadas nas secções a seguir.

3.7.1 Divisão Hierárquica

Para o desenvolvimento da resolução do problema, de acordo com o método utilizado, a primeira etapa consiste em decompor o problema numa hierarquia, sendo que o primeiro nível da hierarquia corresponde ao problema, o segundo aos critérios e o terceiro às alternativas. No presente estudo, a hierarquia de decisão foi estruturada como mostra a Figura 33.

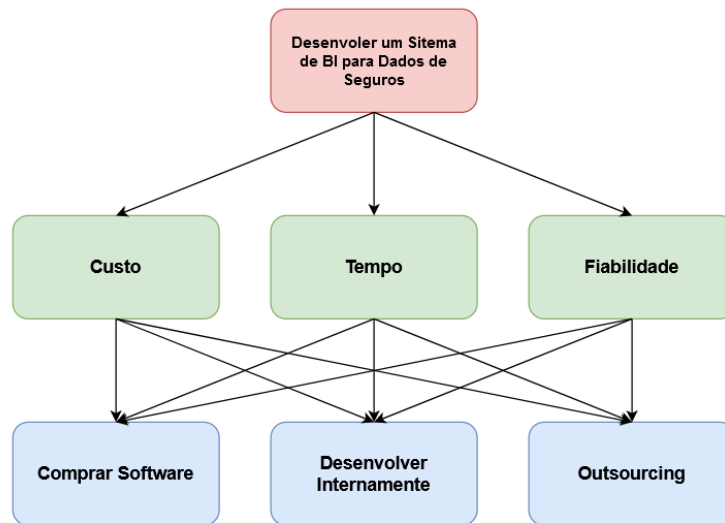


Figura 33 – Hierarquia AHP

3.7.2 Comparação das Alternativas e Critérios

A segunda fase consiste em estabelecer prioridades entre os elementos para cada nível da hierarquia, por meio de uma matriz de comparação.

Para tal é necessário determinar uma escala de valores que não deve exceder um total de nove fatores, a fim de se manter a matriz consistente. Existe uma escala-padrão que foi sugerida por Saaty, a Escala Fundamental (Figura 34) com o objetivo de avaliar de 1 a 9 os critérios e as alternativas no processo de decisão. Estes valores numéricos possuem uma definição qualitativa.

Nível de importância	Definição	Explicação
1	Igual importância	As duas atividades contribuem igualmente para o objetivo
3	Fraca importância	A experiência e o julgamento favorecem levemente uma atividade em relação à outra
5	Forte importância	A experiência e o julgamento favorecem fortemente uma atividade em relação à outra
7	Muito forte importância	Uma atividade é muito fortemente favorecida em relação a outra
9	Importância absoluta	A evidência favorece uma atividade em relação a outra com o mais alto grau de certeza
2,4,6,8	Valores intermediários	Quando se procura uma condição de compromisso entre duas definições

Figura 34 – Escala fundamental de Saaty (Níveis de importância de comparações)

3.7.3 Definição de Prioridades

Nesta fase são atribuídas prioridades aos critérios e alternativas apresentados na “Divisão Hierárquica”.

Para tal, estabelece-se primeiro prioridades para os critérios e depois, estabelecesse prioridades para cada alternativa em relação ao critério em causa.

Seguidamente normaliza-se os valores da matriz de comparações, que tem por objetivo igualar todos os critérios a uma mesma unidade, para isto cada valor da matriz é dividido pelo total da sua respetiva coluna (Tabela 4).

Tabela 4 – Comparações Par a Par

Critérios	Custo	Tempo	Fiabilidade
Custo	1	1/3	1/8
Tempo	3	1	1/5
Fiabilidade	8	5	1
Soma	12	19/3	53/40

Divide-se cada elemento da matriz de comparação pelo total da coluna respetiva. Finalmente obtém-se o vetor de prioridades, que tem por objetivo identificar a ordem de importância de cada critério, para isto é calculado a média aritmética dos valores de cada linha da matriz normalizada obtida no item anterior (Tabela 5).

Tabela 5 – Matriz de Comparação Normalizada e Pesos Estimados

Critérios	Custo	Tempo	Fiabilidade	Pesos
Custo	0.0833	0.0526	0.0943	0.0768
Tempo	0.2500	0.1579	0.1509	0.1863
Fiabilidade	0.6667	0.7895	0.7547	0.7370
	1	1	1	1

3.7.4 Consistência Lógica

Esta fase passa por calcular a Razão de Consistência (RC) para medir o quanto os julgamentos foram consistentes em relação a grandes amostras de juízos completamente aleatórios.

As avaliações do método AHP são baseadas no pressuposto de que o decisor é racional, isto é, se A é preferido a B e B é preferível a C, então A é preferido a C.

Se o RC é superior a 0,1 os julgamentos não são confiáveis porque estão demasiado perto para o conforto de aleatoriedade, neste caso os resultados obtidos não apresentam valores consistentes.

Para calcular a RC é necessário primeiro obter o valor de λ_{max} (Tabela 6) que representa o maior valor próprio da matriz, obtido a partir da seguinte equação: $Ax = \lambda_{max} x$.

Tabela 6 – Teste da Consciência

Critérios	Custo	Tempo	Fiabilidade	Pesos	
Custo	1	1/3	1/8	0.0768	3.0088
Tempo	3	1	1/5	0.1863	3.0276
Fiabilidade	8	5	1	0.7370	3.0972
λ_{max}					3.0445

Uma vez calculado λ_{max} , deve-se calcular o Índice de Consistência (IC) para seguidamente calcular a RC. O cálculo do índice de consistência é dado pela seguinte fórmula:

$$IC = \frac{\lambda_{max} - n}{n - 1}$$

Onde IC é o índice de consistência e o n é o número de critérios avaliados.

$$IC = \frac{3.0445 - 3}{3 - 1} = 0.022264$$

A RC é obtida pela fórmula:

$$RC = \frac{IC}{IR}$$

Onde o IR é o índice aleatório referente a um grande número de comparações par a par efetuadas. Este é um índice aleatório calculado para matrizes quadradas de ordem n pelo Laboratório Nacional de Oak Ridge, nos EUA. A Tabela 7 define os valores de IR em função do número de critérios:

Tabela 7 – Valores de IR para matrizes quadradas de ordem n

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.48	1.56	1.57	1.59

Cada um dos números desta tabela é a média de ICs derivada de uma amostra de matrizes recíprocas selecionadas aleatoriamente usando a escala AHP.

Um RC de 10% ou menos implica que o ajuste é pequeno em comparação com os valores atuais das entradas. Um RC de 90% significaria que os julgamentos são praticamente emparelhados aleatoriamente e são completamente não confiáveis.

Assim sendo, o resultado de RC é:

$$RC = \frac{0.022264}{0.58} = 0.038386$$

3.7.5 Matriz de Comparação Paritária para cada Critério

Após definidas as prioridades dos critérios segue-se a construção da matriz de comparação paritária para cada critério, considerando cada uma das alternativas selecionadas. Assim sendo, todos os procedimentos para a construção da matriz de comparação e para a determinação da prioridade relativa de cada critério devem ser feitos novamente, observando agora a importância relativa de cada uma das alternativas que compõem a estrutura hierárquica do problema em questão.

As tabelas seguintes (Tabela 8 e Tabela 9) ilustram as matrizes de alternativas normalizada e pesos estimados do critério custo, bem como o teste da consciência.

Tabela 8 – Matriz de Alternativas Normalizada e Pesos Estimados (Custo)

Custo	Comprar Software	Outsourcing	Desenvolver Internamente	Pesos
Comprar Software	1/9	0	1/8	0.1066
Outsourcing	2/9	0.1818	0.1765	0.1935
Desenvolver Internamente	2/3	5/7	5/7	0.6999

Tabela 9 – Teste da Consciência (Custo)

Custo	Comprar Software	Outsourcing	Desenvolver Internamente	Pesos	
Comprar Software	1	1/2	1/6	0.1066	3.002788
Outsourcing	2	1	1/4	0.1935	3.005629
Desenvolver Internamente	6	4	1	0.6999	3.019242
λ_{\max}					3.00922

Finda a análise, chegou-se aos valores de IC = 0.00461 e RC = 0.007948, para o critério de custo.

As tabelas seguintes (Tabela 10 e Tabela 11) ilustram as matrizes de alternativas normalizada e pesos estimados do critério tempo, bem como o teste da consciência.

Tabela 10 – Matriz de Alternativas Normalizada e Pesos Estimados (Tempo)

Tempo	Comprar Software	Outsourcing	Desenvolver Internamente	Pesos
Comprar Software	0	0	1/7	0.0754
Outsourcing	0.6000	0.1957	0.1724	0.3227
Desenvolver Internamente	1/3	7/9	2/3	0.6019

Tabela 11 – Teste da Consciência (Tempo)

Tempo	Comprar Software	Outsourcing	Desenvolver Internamente	Pesos	
Comprar Software	1	1/9	1/5	0.0754	3.070729
Outsourcing	9	1	1/4	0.3227	3.570518
Desenvolver Internamente	5	4	1	0.6019	3.771353
λ_{\max}					3.470867

Finda a análise, chegou-se aos valores de IC = 0.235433 e RC = 0.40592, para o critério de tempo.

As tabelas seguintes (Tabela 12 e Tabela 13) ilustram as matrizes de alternativas normalizada e pesos estimados do critério fiabilidade, bem como o teste da consciência.

Tabela 12 – Matriz de Alternativas Normalizada e Pesos Estimados (Fiabilidade)

Fiabilidade	Comprar Software	Outsourcing	Desenvolver Internamente	Pesos
Comprar Software	0	0	1/9	0.0612
Outsourcing	1/2	1/7	1/8	0.2658
Desenvolver Internamente	2/5	5/6	3/4	0.6731
	1	1	1	1

Tabela 13 – Teste da Consciência (Fiabilidade)

Fiabilidade	Comprar Software	Outsourcing	Desenvolver Internamente	Pesos	
Comprar Software	1	1/9	1/7	0.0612	3.054273
Outsourcing	9	1	1/6	0.2658	3.493862
Desenvolver Internamente	7	6	1	0.6731	4.005537
λ_{\max}					3.517891

Finda a análise, chegou-se aos valores de IC = 0.258945 e RC = 0.446458, para o critério de fiabilidade.

3.7.6 Prioridade Composta para as Alternativas

Nesta última etapa, obtemos as prioridades compostas das alternativas, multiplicando os valores anteriores e os das prioridades relativas, obtidos no início do método, conforme a Tabela 14.

Tabela 14 – Matriz com Prioridades Globais

	Custo	Tempo	Fiabilidade	Sí'	
Comprar Software	0.1066	0.0754	0.0612	0.0768	0.067329
Outsourcing	0.1935	0.3227	0.2658	0.1863	0.27085
Desenvolver Internamente	0.6999	0.6019	0.6731	0.7370	0.661921
	1	1	1	1	

A alternativa selecionada é a de desenvolver internamente.

4 Análise e Design da Solução Proposta

Este capítulo é dedicado à análise e *design* da solução proposta. Nomeadamente, identificação dos atores e requisitos do sistema e ainda à apresentação da arquitetura proposta.

Para além disto, é apresentada a fonte de dados utilizada e os diversos componentes que foram necessários para a elaboração da arquitetura.

4.1 Requisitos

Nesta secção são sistematizados e analisados os objetivos do negócio em termos de requisitos do projeto de *software*. Os requisitos são capacidades e condições às quais o sistema (e o projeto) deve satisfazer e dividem-se em duas categorias, os requisitos funcionais, no qual está inserida a funcionalidade, e os requisitos não funcionais, no qual estão inseridos a usabilidade, confiabilidade, desempenho e suportabilidade, entre outros.

As categorias FURPS são de dois tipos diferentes: Funcional (F) e não funcional (URPS). Essas categorias podem ser usadas tanto como requisitos do produto como na avaliação da qualidade do produto.

O modelo FURPS originalmente apresentado por Robert Grady (Eeles, P., 2005) classifica os requisitos segundo cinco categorias, sendo o seu significado o seguinte:

- Funcionalidade (*Functionality*) - Especifica as funcionalidades próprias do domínio (sob a forma de casos de uso), e também especifica requisitos funcionais arquiteturalmente significativos no sistema, nomeadamente: Auditoria, Reporte, Interoperabilidade e Segurança.
- Usabilidade (*Usability*) - Avalia a interface com o utilizador. Possui diversas subcategorias, entre elas: estética e *design*; consistência e padrões.
- Confiabilidade (*Reliability*) - Refere-se a integridade, conformidade e interoperabilidade do *software*. Os requisitos a serem considerados são: frequência e gravidade de falha, possibilidade de recuperação, possibilidade de previsão, exatidão e tempo médio entre falhas.
- Desempenho (*Performance*) - Avalia os requisitos de desempenho do *software*, nomeadamente: tempo de resposta, consumo de memória, utilização da CPU, capacidade de carga e disponibilidade da aplicação.
- Suportabilidade (*Supportability*) - Os requisitos de suportabilidade agrupam várias características, como: testabilidade, adaptabilidade, manutenibilidade, compatibilidade, configurabilidade, instabilidade e escalabilidade.

Posteriormente o modelo FURPS foi atualizado, passando a ser conhecido por FURPS+ (Coepd., 2014). A adição do sinal "+", representa as restrições impostas ao sistema, inclusive:

- Requisitos de *design* (padrões de *design*, bibliotecas de classes, etc.)
- Requisitos de implementação (ferramentas de desenvolvimento, linguagens de programação, políticas de integridade de base de dados, limites de recursos, sistema operativo, etc.)
- Requisitos de interface/interação (especifica ou restringe as funcionalidades inerentes às interações do sistema com o exterior)
- Requisitos físicos (especifica uma restrição física imposta no *hardware* do sistema)

4.1.1 Funcionalidade

Relativamente ao projeto em questão, os requisitos funcionais foram capturados em forma de caso de uso. No caso particular desta aplicação, consideraram-se os seguintes casos de uso:

- UC1 – Analisar valores totais e quantidade de propostas
- UC2 – Analisar valores totais e quantidade de apólices
- UC3 – Analisar o número de apólices potenciais e efetivas
- UC4 – Analisar o valor total do capital das propostas e das apólices
- UC5 – Analisar os valores dos prémios potenciais e efetivos
- UC6 – Analisar o número de apólices potenciais e efetivas emitidas e anuladas
- UC7 – Analisar os valores totais do capital potencial e efetivo
- UC8 – Analisar clientes potenciais e efetivos
- UC9 – Analisar os principais devedores

Os atores identificados foram o gestor comercial, o gestor financeiro e o mediador, que necessitam de consultar o sistema para efetuar a tomada de decisão com base nos dados apresentados (Figura 35).

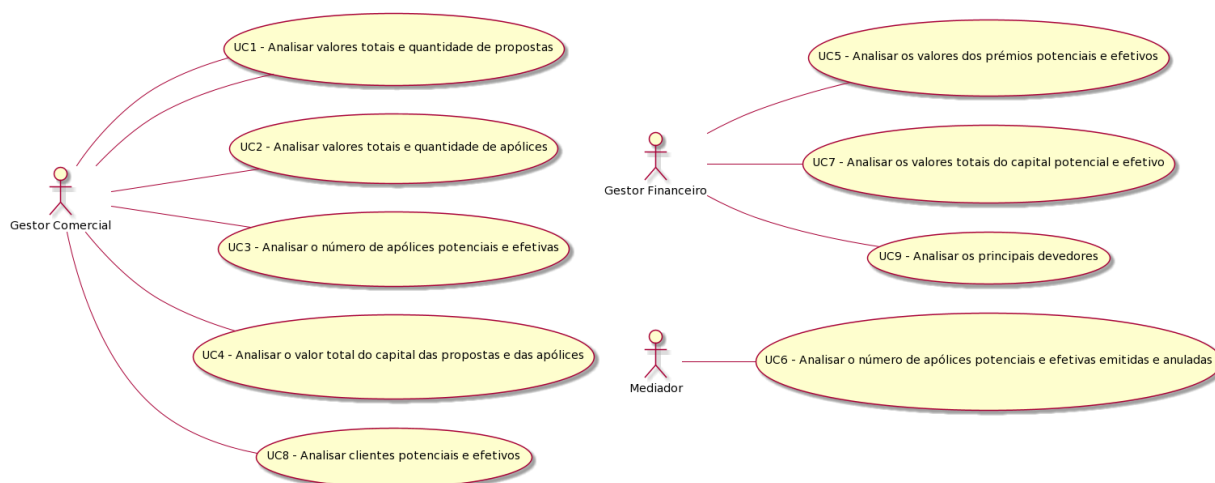


Figura 35 – Diagrama de casos de uso

4.1.1.1 UC1 – Analisar valores totais e quantidade de propostas

Este caso de uso decompõe-se nas seguintes *user stories*:

- US1 – Como gestor comercial, pretendo analisar os valores totais e respetivas quantidades efetuadas em cada mês de um dado ano, detalhados por categoria do produto, com possibilidade de análise detalhada (i.e., drill down) ao nível individual do produto.
- US2 – Como gestor comercial, pretendo analisar os valores totais e respetivas quantidades adquiridas num dado ano, dos produtos cujo tomador do seguro é a pessoa segura, detalhados pelo distrito, com possibilidade de análise detalhada (i.e., drill down) ao nível do próprio distrito.
- US3 – Como gestor comercial, pretendo analisar os valores totais e respetivas quantidades referentes aos produtos adquiridos num semestre de um dado ano, detalhados por agente e por zona geográfica do cliente.

- US4 – Como gestor comercial, pretendo analisar os valores totais das propostas e respetivas quantidades, efetuadas num quadrimestre de um dado ano, detalhados pela zona geográfica do cliente e por produto, com possibilidade de análise agregada (i.e., roll up) ao nível da categoria do produto.
- US5 – Como gestor comercial, pretendo analisar os valores totais das propostas efetuadas durante a primavera e verão de um dado ano, detalhados por país e género do cliente.
- US6 – Como gestor comercial, pretendo analisar os valores totais das propostas efetuadas no primeiro trimestre de um dado ano, detalhados pelo género do cliente e pela categoria do produto.

4.1.1.2 UC2 – Analisar valores totais e quantidade de apólices

Este caso de uso decompõe-se nas seguintes *user stories*:

- US7 – Como gestor comercial, pretendo analisar os valores totais e respetivas quantidades efetuadas em cada mês de um dado ano, detalhados por categoria do produto, com possibilidade de análise detalhada (i.e., drill down) ao nível individual do produto.
- US8 – Como gestor comercial, pretendo analisar os valores totais e respetivas quantidades adquiridas num dado ano, dos produtos cujo tomador do seguro é a pessoa segura, detalhados pelo distrito, com possibilidade de análise detalhada (i.e., drill down) ao nível do próprio distrito.
- US9 – Como gestor comercial, pretendo analisar os valores totais e respetivas quantidades referentes aos produtos adquiridos num semestre de um dado ano, detalhados por agente e por zona geográfica do cliente.
- US10 – Como gestor comercial, pretendo analisar os valores totais das apólices e respetivas quantidades, efetuadas num quadrimestre de um dado ano, detalhados pela zona geográfica do cliente e por produto, com possibilidade de análise agregada (i.e., roll up) ao nível da categoria do produto.
- US11 – Como gestor comercial, pretendo analisar os valores totais das apólices efetuadas durante a primavera e verão de um dado ano, detalhados por país e género do cliente.
- US12 – Como gestor comercial, pretendo analisar os valores totais das apólices efetuadas no primeiro trimestre de um dado ano, detalhados pelo género do cliente e pela categoria do produto.

4.1.1.3 UC3 – Analisar o número de apólices potenciais e efetivas

Este caso de uso decompõe-se nas seguintes *user stories*:

- US13 – Como gestor comercial, pretendo analisar o número de potenciais² apólices.
- US14 – Como gestor comercial, pretendo analisar o número de potenciais apólices por idade atuarial; por género; por NIF (pessoa coletiva/individual); por estado civil; por classe de risco (profissão); por cúmulo de risco (capitais acumulados das várias apólices da pessoa segura); por localização; por produto; por situação.
- US15 – Como gestor comercial, pretendo analisar o número de potenciais apólices do ano corrente comparativamente ao ano anterior.

² Potencial uma vez que, uma proposta pode ou não se transformar numa apólice.

- US16 – Como gestor comercial, pretendo analisar o número de apólices por situação; por idade atuarial; por género; por NIF (pessoa coletiva/individual); por estado civil; por classe de risco (profissão); por cúmulo de risco (capitais acumulados das várias apólices da pessoa segura); por localização; por produto; por situação e produto; por estrutura comercial.

4.1.1.4 UC4 – Analisar o valor total do capital das propostas e das apólices

Este caso de uso decompõe-se nas seguintes *user stories*:

- US17 – Como gestor comercial, pretendo analisar os valores totais dos capitais referentes às potenciais aquisições realizadas numa dada semana de um dado ano, detalhados por distrito, com possibilidade de análise detalhada (i.e., drill down) ao nível do local.
- US18 – Como gestor comercial, pretendo analisar os valores totais dos capitais referentes às potenciais aquisições por cliente, com possibilidade de análise agregada (i.e., roll up) ao nível do país do cliente, e por mês de um dado ano, com possibilidade de análise agregada (i.e., roll up) ao nível do trimestre, semestre ou ano, unicamente para os produtos que pertençam às categorias “Vida Pleno” e “Habitação”.
- US19 – Como gestor comercial, pretendo analisar os valores totais dos capitais referentes às aquisições realizadas numa dada semana de um dado ano, detalhados por distrito, com possibilidade de análise detalhada (i.e., drill down) ao nível do local.
- US20 – Como gestor comercial, pretendo analisar os valores totais dos capitais referentes às aquisições por cliente, com possibilidade de análise agregada (i.e., roll up) ao nível do país do cliente, e por mês de um dado ano, com possibilidade de análise agregada (i.e., roll up) ao nível do trimestre, semestre ou ano, unicamente para os produtos que pertençam às categorias “Vida Pleno” e “Habitação”.

4.1.1.5 UC5 – Analisar os valores dos prémios potenciais e efetivos

Este caso de uso decompõe-se nas seguintes *user stories*:

- US21 – Como gestor financeiro, pretendo analisar os valores dos potenciais prémios
- US22 – Como gestor financeiro, pretendo analisar os valores dos prémios anualizados/únicos por produto; por situação

4.1.1.6 UC6 – Analisar o número de apólices potenciais e efetivas emitidas e anuladas

Este caso de uso decompõe-se nas seguintes *user stories*:

- US23 – Como mediador, pretendo analisar o número de potenciais apólices emitidas e anuladas (com o motivo da anulação) por agente, num determinado período
- US24 – Como mediador, pretendo analisar o número de apólices emitidas e anuladas (com o motivo da anulação) por agente, num determinado período

4.1.1.7 UC7 – Analisar os valores totais do capital potencial e efetivo

Este caso de uso decompõe-se nas seguintes *user stories*:

- US25 – Como gestor financeiro, pretendo analisar os valores totais do capital proposto durante um mês de um dado ano, detalhados pelo distrito e por categoria de produto
- US26 – Como gestor financeiro, pretendo analisar os valores totais do capital durante um mês de um dado ano, detalhados pelo distrito e por categoria de produto

4.1.1.8 UC8 – Analisar clientes potenciais e efetivos

Este caso de uso decompõe-se nas seguintes *user stories*:

- US27 – Como gestor comercial, pretendo analisar quais serão os potenciais clientes
- US28 – Como gestor comercial, pretendo analisar quais serão os principais clientes (dimensão da carteira do cliente, em prémio)

4.1.1.9 UC9 – Analisar os principais devedores

Este caso de uso decompõe-se na seguinte *user storie*:

- US29 – Como gestor financeiro, pretendo analisar quais serão os principais devedores (prémio em cobrança)

4.1.2 Usabilidade

O requisito de usabilidade passa pela facilidade de aprendizagem e de utilização do sistema.

4.1.3 Confiabilidade

Os requisitos de confiabilidade relacionam-se com a integridade dos dados, ou seja, as alterações que são efetuadas nas fontes de dados devem refletir-se na solução final.

4.1.4 Desempenho

A realização das consultas à informação não deve demorar mais do que 30 segundos.

4.1.5 Suportabilidade

O sistema deve ser expansível e permitir a integração de novos serviços e/ou módulos.

4.1.6 Outros (+)

Os outros requisitos (+) são:

- Adoção do processo de desenvolvimento de *software* iterativo e incremental.
- Adoção de boas práticas de design.
- Os dados devem ser armazenados no nível mais baixo de granularidade, de forma a suportar consultas *ad hoc* por parte dos utilizadores.

4.2 Arquitetura Proposta

Escolher e definir a arquitetura do armazém de dados é um dos fatores chave de sucesso de um projeto nesta área. Uma arquitetura inadequada pode levar a problemas como falta de escalabilidade, de desempenho e da veracidade dos dados.

Conforme analisado na subsecção 2.3.3, as arquiteturas mais relevantes são as arquiteturas de Kimball (subsecção 2.3.3.1) e Inmon (subsecção 2.3.3.2), sendo que a solução mais realista é a abordagem defendida por Ralph Kimball. Esta arquitetura procede à construção de *data marts* independentes para cada área de negócio e integram-se de forma a ter um DW mais abrangente. No entanto, esta abordagem não é adequada se os *data marts* estiverem isolados e não puderem ser ligados entre eles. Como analisado, esta arquitetura propõe o conceito de "dimensões conformadas" que consiste em garantir que todos os *data marts* tenham dimensões e factos comuns entre eles em conformidade, ou seja, tenham dimensões que possam ser compartilhadas entre factos nos vários *data marts*.

Esta abordagem torna-se vantajosa, uma vez que caso seja necessário adicionar novas áreas de negócio à solução no futuro, a criação de novos *data marts* não comprometem a arquitetura existente. Para além disso, caso se pretenda alterar a estrutura do sistema, quer para a adição/remoção de funcionalidades, quer para a sua manutenção, consegue-se fazê-lo rapidamente e sem grandes problemas.

A arquitetura proposta é apresentada na Figura 36.

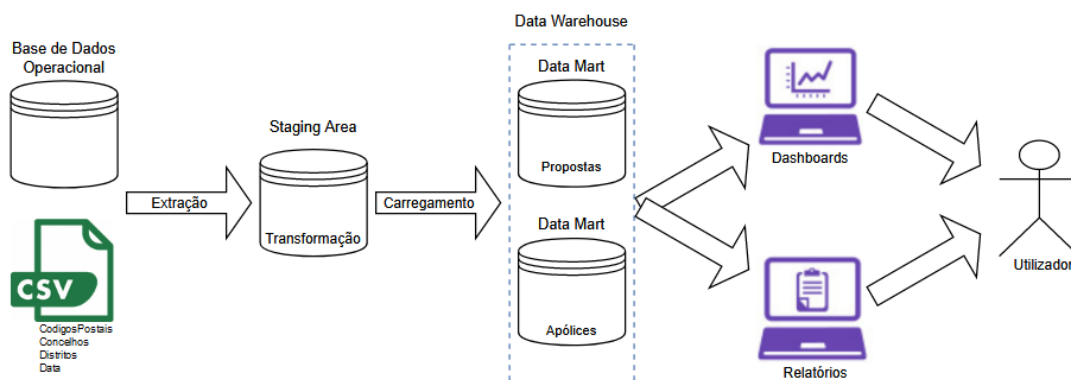


Figura 36 – Arquitetura Proposta

Os dados de origem necessitam sofrer transformações, por isso estes são extraídos para uma base de dados intermédia onde ocorrem os eventos necessários. Após os dados passarem pelo processo de limpeza e transformação são carregados num DW, que por sua vez possui dois *data marts* com a informação das Propostas e Apólices. As ferramentas de BI acedem diretamente ao *data mart* e só depois é que a informação é transmitida ao utilizador.

O sistema deve ser concebido de forma a contemplar dois níveis diferentes de carregamento de dados:

Base de dados Operacional >> *Staging Area* e *Staging Area* >> *Data Warehouse/Data Marts*.

4.2.1 Metodologia de Kimball

Segundo a metodologia de Kimball, existem quatro principais decisões a serem tomadas durante o *design* do modelo dimensional:

1. Selecionar a área de negócios

2. Declarar o nível de granularidade
3. Identificar as dimensões
4. Identificar os factos/medidas

Na área dos seguros de vida existem três processos de negócio essenciais:

- Apólices
- Recibos (prémios/estornos)
- Sinistros/Indemnizações/Resgates

Um armazém de dados é composto por um ou mais *data marts*, no entanto, no presente contexto e através da análise do problema identificou-se que a **área de negócios** a considerar são as apólices, por esta razão será desenvolvido um *data mart* para as mesmas e um *data mart* para as propostas, uma vez que é interessante analisar quais são as potenciais apólices.

Através do requisito não funcional enunciado anteriormente (4.1.6), surgiu o **nível de granularidade** a considerar, sendo este o mais elementar, ou seja, cada linha da tabela de factos reflete o resultado de um seguro de um determinado produto, num determinado momento, realizado por um dado titular.

Após análise de dados, foram identificadas as **dimensões** que dão resposta às necessidades dos utilizadores, sendo estas:

- Data
- Região
- Produto
- Titular
- Estrutura Comercial

As tabelas de factos dos diferentes *data marts* são as Apólices e as Propostas conforme o *data mart* em questão. As **medidas** são, o prémio total, o prémio pago e o capital.

4.2.2 Fonte de Dados

Os dados dos seguros de vida disponibilizados do sistema operacional, contêm informações acerca de regiões, produtos, estrutura comercial, titulares, apólices e propostas. Segue-se um excerto do modelo relacional da base de dados operacional na Figura 37 **Error! Reference source not found..**

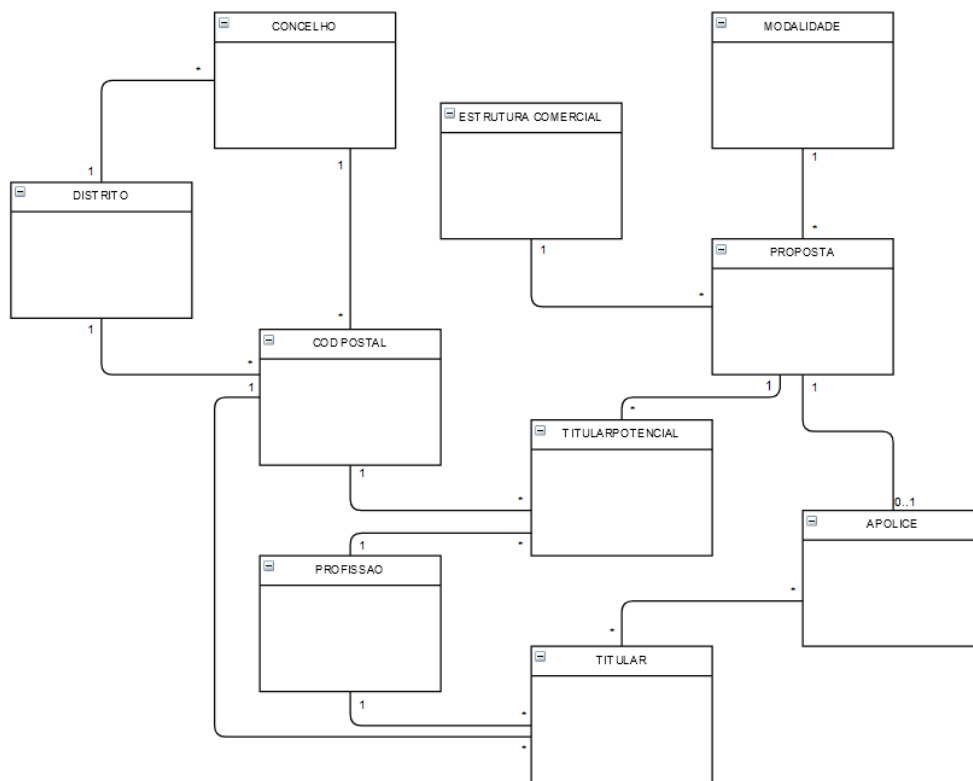


Figura 37 – Excerto do Modelo da Base de Dados Operacional

A base de dados fornecida continha uma amostra de dados que abrange informação durante um período, nomeadamente os anos de 2015 a 2020.

Devido à confidencialidade exigida pela empresa não se entrou em mais detalhe acerca da base de dados, sendo que apenas se desenhou um esquema das entidades de forma mais genérica.

4.2.3 Staging Area

Após a análise da área de negócios da Real Vida Seguros e da base de dados fornecida, foi feito um esboço do que poderia ser a estrutura de dados da *Staging Area*, e quais as tabelas a criar e os respetivos atributos/valores a alocar nas mesmas.

Assim, com o intuito de não perder informação, todas as tabelas existentes na base de dados fonte, ou seja, as tabelas do sistema operacional, foram consideradas e criadas na *Staging Area*.

Outra consideração foi o facto de alguns atributos, que se considerou importantes para futuras análises, na *Staging Area* serem obrigatórios, para possuir os dados necessários a carregar depois no *Data Mart*.

Segue-se uma ilustração das tabelas e respetivos atributos a serem criadas na *Staging Area*.

A tabela PROPET (apresentada na Tabela 15) refere-se aos potenciais clientes. Nesta tabela são guardadas informações acerca dos clientes a que se efetuou uma proposta de seguro.

Tabela 15 – Tabela PROPET

Tabela PROPET			
Atributo/Campo	Descrição	Tipo de Dados	Tamanho
PT\$PRO	Proposta	numeric	(8,0)
PTTIPO	Tipo de Cliente	nvarchar	(1)
PT\$COD	Código de Titular	numeric	(6,0)
PTNVAL	Tipo de Titular	nvarchar	(1)
PTSEXO	Género	nvarchar	(1)
PT\$CTT	Código Postal 4	numeric	(4,0)
PT\$CT3	Código Postal 3	numeric	(3,0)
PTACTI	Atividade	nvarchar	(1)
PT\$CAE	Cae	numeric	(6, 0)
PT\$PRF	Profissão	numeric	(3, 0)
PTDNAN	Ano Data Nascimento	numeric	(4, 0)
PTDNME	Mês Data Nascimento	numeric	(2, 0)
PTDNDI	Dia Data Nascimento	numeric	(2, 0)
PTUSER	Utilizador de Criação	nvarchar	(10)
PTDATA	Data de Criação	date	-
PTNACD	Nacionalidade	nvarchar	(20)
PTNATD	Naturalidade	nvarchar	(20)
PTESTC	Estado Civil	nvarchar	(1)

A tabela PROPE0 (apresentada na Tabela 16) refere-se às propostas de potenciais apólices. Nesta tabela são guardadas informações acerca da potencial compra de seguro.

Tabela 16 – Tabela PROPE0

Tabela PROPE0			
Atributo/Campo	Descrição	Tipo de Dados	Tamanho
PO\$PRO	Proposta	numeric	(8,0)
POTIPO	Tipo de Criação	nvarchar	(1)
PO\$MOD	Modalidade	numeric	(2,0)
PO\$VER	Versão	numeric	(2,0)
PO\$AG1	Agente Angariador	numeric	(6,0)
PO\$AG2	Agente	numeric	(6,0)
PONIV4	Designação Tipo Agente	numeric	(6,0)
PONIV3	Responsável Comercial	numeric	(6,0)
PONIV2	Canal	numeric	(6,0)
PONIV1	Direção	numeric	(6,0)
PO\$APL	Apólice	numeric	(6,0)
PONAPO	Numero de Apólice	numeric	(6,0)
POPRT0	Premio Total	numeric	(11,2)
PO\$FRA	Fracionamento	nvarchar	(1)
POPPAG	Premio Pago	numeric	(11,2)
POCAPT	Capital	numeric	(13,2)
PO\$TIT	Código de Titular	numeric	(6,0)

Tabela PROPE0			
Atributo/Campo	Descrição	Tipo de Dados	Tamanho
POTITP	Tipo de Titular	nvarchar	(1)
PO\$1CB	Código da 1ª Cabeça	numeric	(6,0)
POP1TP	Tipo 1ª Pessoa Segura	nvarchar	(1)
PO\$2CB	Código da 2ª Cabeça	numeric	(6,0)
POP2TP	Tipo 2ª Pessoa Segura	nvarchar	(1)
POHERD	Herdeiros Legais	nvarchar	(1)
PODATA	Data	date	-
PODEFT	Data de Efeito	date	-
PODSIT	Data de Situação	date	-
POSITU	Situação	numeric	(1,0)
POUSER	Utilizador	nvarchar	(10)
PODURA	Duração	numeric	(2,0)
POTOSE	Tomador Seguro é Beneficiário	nvarchar	(1)
POINF2	Forma Pagamento	nvarchar	(1)
POFPRS	Forma Pagamento Recibos Seguintes	nvarchar	(1)

Segue-se uma breve explicação das restantes tabelas a criar na *Staging Area*, sendo que as mesmas se encontram em anexo (Anexo I).

A tabela MODA00 (apresentada na Tabela 31) refere-se à tabela de modalidades do seguro. É nesta tabela que é definido o produto e coberturas do seguro.

A tabela ESTRUTN (apresentada na Tabela 32) refere-se à estrutura comercial. Normalmente o seguro é efetuado por um agente, o qual pertence a uma estrutura de cinco níveis (1º Nível – Direção; 2º Nível – Canal; 3º Nível – Responsável Comercial; 4º Nível – Designação tipo de Agente; 5º Nível – Agente).

A tabela TBLLIN (apresentada na Tabela 33) é referente às profissões. A profissão tem um risco associado o que influencia o valor do seguro.

Os CTT disponibilizam ficheiros na sua central de dados com informação acerca das moradas (Carvalho, A.I., Lafuente, R., 2021). Esta informação é interessante para se analisar uma vez que proporciona uma visão das regiões onde há mais seguros. Assim sendo, descarregou-se essa informação e criaram-se as seguintes tabelas: a tabela CodigosPostais (apresentada na Tabela 34) possui informação acerca do código postal; a tabela Concelhos (apresentada na Tabela 35) possui informação acerca dos concelhos; a tabela Distritos (apresentada na Tabela 36) possui informação acerca dos distritos.

Nesta análise surgiram também as designadas tabelas DQP, que servem para armazenar os problemas de qualidade detetados nos dados. Vulgarmente, estes DQP são denominados de erros, e nestas tabelas podem ser armazenados os valores que se encontrem a “null” e que não podiam estar assim, ou seja, existem dados em falta.

Durante as operações de transformação e limpeza de dados, tem de se identificar e corrigir os erros nos dados, como o caso dos erros ortográficos, violações do domínio, valores em falta, valores duplicados, e violações de regras de negócio. De notar que, embora a grande maioria dos DQP possa ser automaticamente detetada, tem de ser manualmente corrigida.

Assim, através do processo de limpeza de dados surgiram tabelas de DQP e respetivos atributos criados na *Staging Area*, sendo exemplo disso a Tabela 17 (PROPET_DQP).

Tabela 17 – Tabela PROPET_DQP

Tabela PROPET_DQP			
Atributo/Campo	Descrição	Tipo de Dados	Tamanho
PT\$PRO	Proposta	numeric	(8,0)
PTTIPO	Tipo de Cliente	nvarchar	(1)
PT\$COD	Código de Titular	numeric	(6,0)
PTNVAL	Tipo de Titular	nvarchar	(1)
PTSEXO	Sexo	nvarchar	(1)
PT\$CTT	Código Postal 4	numeric	(4,0)
PT\$CT3	Código Postal 3	numeric	(3,0)
PTACTI	Atividade	nvarchar	(1)
PT\$CAE	Cae	numeric	(6, 0)
PT\$PRF	Profissão	numeric	(3, 0)
PTDNAN	Ano Data Nascimento	numeric	(4, 0)
PTDNME	Mês Data Nascimento	numeric	(2, 0)
PTDNDI	Dia Data Nascimento	numeric	(2, 0)
PTUSER	Utilizador de Criação	nvarchar	(10)
PTDATA	Data de Criação	date	-
PTNACD	Nacionalidade	nvarchar	(20)
PTNATD	Naturalidade	nvarchar	(20)
PTESTC	Estado Civil	nvarchar	(1)
DQP	Descrição do Erro	nvarchar	(100)

4.2.4 Modelo Dimensional

Um modelo dimensional caracteriza-se por uma técnica usada em DW e difere da modelação relacional. O conceito foi desenvolvido por Ralph Kimball, sendo que existem dois pilares, as tabelas de dimensão e a(s) tabela(s) de factos. As tabelas de dimensão fornecem a base para analisar os dados na tabela de factos.

O objetivo é conseguir obter a informação do DW mais facilmente. Nos modelos relacionais, os dados são normalizados para que não haja redundância na informação, enquanto os modelos dimensionais apenas organizam os dados para se obter mais facilmente a informação, conseguir transmitir a mesma ao utilizador final e também por questões de *performance*.

Após identificar as necessidades/requisitos de análise de dados anteriormente apresentados definiu-se o modelo dimensional e aplicaram-se as transformações necessárias ao mesmo.

Definiram-se as nomenclaturas para as tabelas do *data mart* apresentadas na Tabela 18:

Tabela 18 – Nomenclaturas das tabelas

Tipo de tabelas	Nomenclatura
Nomenclatura das tabelas de dimensões	Dim<Nome da tabela fonte>
Nomenclatura das tabelas de factos	Fact<Nome da tabela fonte>

Das necessidades de análise de dados surgiram as tabelas dimensão *DimRegiao*, *DimProduto*, *DimTitular* e *DimEstruturaComercial*. Adicionalmente acrescentou-se a tabela de dimensão *DimData*.

Esta tabela foi criada com o intuito de ter os dados no espaço tempo, mais elementar e descritivo possível.

Surgiu também as tabelas de factos *FactProposta* e *FactApolice*. Segue-se uma ilustração das tabelas a serem criadas no *data mart* (Figura 38).

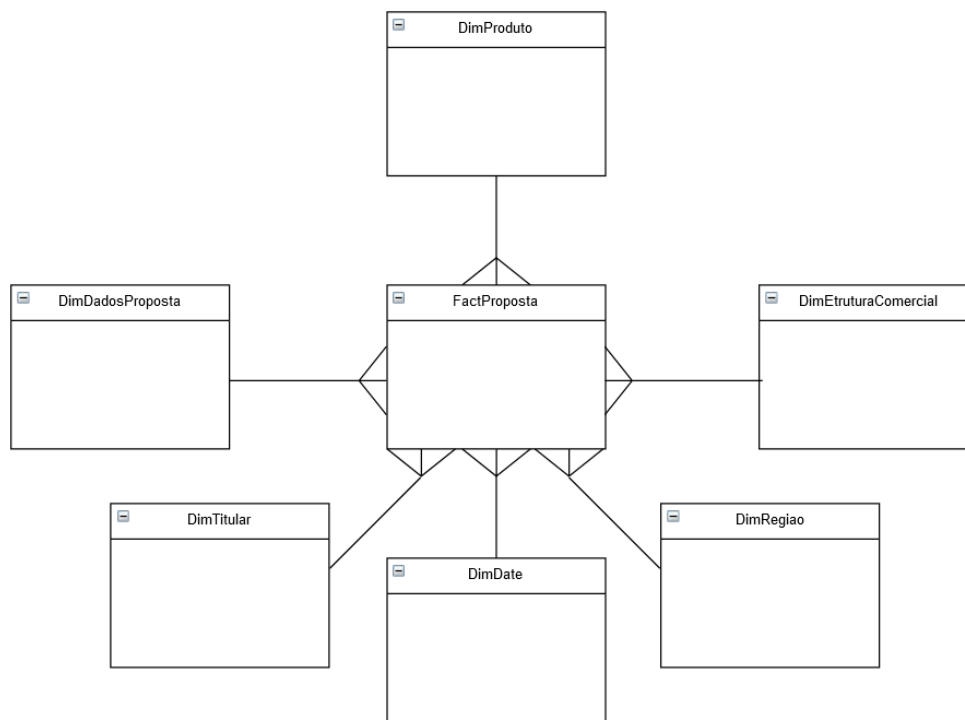


Figura 38 – Modelo Dimensional

4.2.4.1 Tabela DimData

Uma das primeiras dimensões desenvolvidas foi a *DimData*, necessária em qualquer *DW*, uma vez que permite contextualizar os factos no tempo (granularidade diária), possibilitando analisar o processo numa perspetiva temporal.

Este negócio em específico não traz grandes benefícios analisar a informação no tempo (granularidade hora/minuto/segundo), por isso o nível de granularidade restringiu-se apenas a esta dimensão (Tabela 19).

Tabela 19 – Tabela DimData

Tabela DimData		
Atributo/Campo	Tipo de Dados	Tamanho
DataKey (PK)	int	-
DataCompleta	datetime	-
Ano	int	-
Semestre	tinyint	-
Trimestre	tinyint	-
Mes	tinyint	-
NomeMes	nvarchar	10
Semana	tinyint	-
NumeroAno	int	-

Tabela DimData		
Atributo/Campo	Tipo de Dados	Tamanho
NumeroMes	tinyint	-
NumeroSemana	tinyint	-
DiaDaSemana	nvarchar	10
FimDeSemana	nvarchar	3

4.2.4.2 Tabela DimRegiao

A dimensão *DimRegiao* permite analisar os valores das métricas por concelho ou distrito de um país. O código postal é o nível mais elementar e é onde é guardada informação acerca da morada do titular e da morada onde uma proposta ou apólice foi registada.

A Tabela 20 (*DimRegiao*), foi criada a partir das tabelas CodigosPostais, Concelhos e Distritos do sistema operacional.

Para além dos atributos provenientes das tabelas enumeradas acrescentou-se o atributo RegiaoKey.

De notar ainda que os atributos identificados com “(PK-SO)” correspondem à PK do sistema operacional.

Tabela 20 – Tabela DimRegiao

Tabela DimRegiao		
Atributo/Campo	Tipo de Dados	Tamanho
RegiaoKey (PK)	int	-
CodigoPostal4 (PK-SO)	numeric	(4,0)
CodigoPostal3 (PK-SO)	numeric	(3,0)
FreguesiaDesc	nvarchar	(50)
ConcelhoDesc	nvarchar	(50)
DistritoDesc	nvarchar	(50)

4.2.4.3 Tabela DimProduto

A dimensão *DimProduto* descreve o seguro e coberturas do seguro. Esta tabela foi criada com base na tabela MODA00 do sistema operacional (Tabela 21).

Para além dos atributos provenientes da tabela enumerada, acrescentaram-se os atributos ProdutoKey (constitui a chave de substituição), EffectiveDate, ExpiredDate e IsCurrent (permitem armazenar as alterações que ocorram, através do mecanismo de Slow Changing Dimension – Tipo 2).

Tabela 21 – Tabela DimProduto

Tabela DimProduto		
Atributo/Campo	Tipo de Dados	Tamanho
ProdutoKey (PK)	int	-
CodigoModalidade (PK-SO)	numeric	(2,0)
Versao (PK-SO)	numeric	(2,0)
Grupo	numeric	(2,0)
DesignacaoAbreviada	nvarchar	(8)
Descricao	nvarchar	(30)
TipoSeguro	nvarchar	(29)

Tabela DimProduto		
Atributo/Campo	Tipo de Dados	Tamanho
PessoaSegura	nvarchar	(21)
DuracaoContrato	nvarchar	(22)
TipoContrato	nvarchar	(7)
TipoModalidade	nvarchar	(5)
DataInicioVersaoKey	int	-
DataFimVersaoKey	int	-
UtilizadorAlteracao	nvarchar	(10)
DataAlteracaoKey	int	-
EffectiveDate	int	-
ExpiredDate	int	-
IsCurrent	nvarchar	(3)

4.2.4.4 Tabela DimTitular

A dimensão DimTitular armazena informação acerca dos titulares das propostas e apólices (Tabela 22). Esta dimensão é necessária para que seja possível compreender e analisar o comportamento de certos segmentos de titulares. Esta tabela foi criada a partir das tabelas PROPET e TBLIN do sistema operacional.

Para além dos atributos provenientes da tabela enumerada, acrescentaram-se os atributos TitularKey, EffectiveDate, ExpiredDate e IsCurrent.

Tabela 22 – Tabela DimTitular

Tabela DimTitular		
Atributo/Campo	Tipo de Dados	Tamanho
TitularKey (PK)	int	-
Proposta (PK-SO)	numeric	(8,0)
TipoCliente (PK-SO)	nvarchar	(24)
CodigoTitular	numeric	(6,0)
TipoTitular	nvarchar	(9)
Sexo	nvarchar	(9)
CodigoPostalKey	int	-
Atividade	nvarchar	(10)
Cae	numeric	(6, 0)
Profissao	nvarchar	(60)
DataNascimentoKey	int	-
UtilizadorCriacao	nvarchar	(10)
DataCriacaoKey	int	-
Nacionalidade	nvarchar	(20)
Naturalidade	nvarchar	(20)
EstadoCivil	nvarchar	(14)
EffectiveDate	int	-
ExpiredDate	int	-
IsCurrent	nvarchar	(3)

4.2.4.5 Tabela DimEstruturaComercial

A dimensão DimEstruturaComercial indica quem foram os intervenientes que produziram a proposta ou apólice, desde o agente produtor (nível mais elementar possível de granularidade) até à direção (nível mais grosso).

A Tabela 23 (DimEstruturaComercial) foi criada com base na tabela ESTRUTN do sistema operacional.

Para além dos atributos provenientes da tabela enumerada, acrescentaram-se os atributos EstruturaComKey, EffectiveDate, ExpiredDate e IsCurrent.

Tabela 23 – Tabela DimEstruturaComercial

Tabela DimEstruturaComercial		
Atributo/Campo	Tipo de Dados	Tamanho
EstruturaComKey (PK)	int	-
Ano (PK-SO)	decimal	(4,0)
Mes (PK-SO)	decimal	(2,0)
Direcao	decimal	(6,0)
NomeDirecao	nvarchar	(42)
Canal	decimal	(6,0)
NomeCanal	nvarchar	(42)
ResponsavelComercial	decimal	(6,0)
DesignacaoTipoAgente	decimal	(6,0)
Agente (PK-SO)	decimal	(6,0)
EffectiveDate	int	-
ExpiredDate	int	-
IsCurrent	nvarchar	(3)

4.2.4.6 Tabela DimDadosProposta

A Tabela 24 (DimDadosProposta) foi criada a partir da tabela PROPEO do sistema operacional.

Para além dos atributos provenientes da tabela enumerada, acrescentou-se os atributos EffectiveDate, ExpiredDate e IsCurrent.

Tabela 24 – Tabela DimDadosProposta

Tabela DimDadosProposta		
Atributo/Campo	Tipo de Dados	Tamanho
PropostaKey (PK)	int	-
Proposta	numeric	(8,0)
TipoCriacao	nvarchar	(9)
Apolice	numeric	(6,0)
NumeroApolice	numeric	(6,0)
Fracionamento	nvarchar	(10)
TipoTitular	nvarchar	(9)
Tipo1PSegura	nvarchar	(9)
Tipo2PSegura	nvarchar	(9)
HerdeirosLegais	nvarchar	(3)
Situacao	nvarchar	(24)

Tabela DimDadosProposta		
Atributo/Campo	Tipo de Dados	Tamanho
Utilizador	nvarchar	(10)
TomadorSeguroBeneficiario	nvarchar	(3)
FormaPagamento	nvarchar	(18)
FormaPagamentoRecibosSeguintes	nvarchar	(18)
EffectiveDate	int	-
ExpiredDate	int	-
IsCurrent	nvarchar	(3)

4.2.4.7 Tabela FactProposta

A Tabela 25 (FactProposta), foi criada com base nas tabelas MODA00, PROPET, ESTRUTN e PROPE0 do sistema operacional.

Para além dos atributos provenientes das tabelas enumeradas, acrescentaram-se os atributos TitularKey, ProdutoKey, EstruturaComKey e DataKey, os quais representam a chave primária.

Tabela 25 – Tabela FactProposta

Tabela FactProposta		
Atributo/Campo	Tipo de Dados	Tamanho
TitularKey (PK)	int	-
ProdutoKey (PK)	int	-
EstruturaComKey (PK)	int	-
DataKey (PK)	int	-
Proposta	numeric	(8,0)
PremioTotal	numeric	(11,2)
PremioPago	numeric	(11,2)
Capital	numeric	(13,2)
Cabeca1TitularKey	int	-
Cabeca2TitularKey	int	-
DataEfeitoKey	int	-
DataSituacaoKey	int	-
Duracao	numeric	(2,0)

4.2.5 Mapeamento de Dados entre os Sistemas

De modo a obter uma melhor perceção das transformações, definiu-se as seguintes abreviações (Tabela 26):

Tabela 26 – Nomenclaturas das tabelas

Tipo de Tabela	Nome da Base de Dados
D – Tabelas de dimensão	S. A. – Staging Area
F – Tabelas de Factos	D. W. – Data Warehouse

O mapeamento dos dados que foi feito encontra-se representado na tabela em anexo (Anexo II).

5 Implementação da Solução

Neste capítulo descreve-se a implementação da solução. Encontra-se dividido em várias secções, onde se descreve os principais aspetos da aplicação.

De forma a facilitar um melhor entendimento do trabalho realizado, serão utilizados alguns excertos de código, bem como imagens da solução desenvolvida. Deste modo, pretende-se proporcionar uma abordagem visual e uma linguagem padrão, permitindo uma melhor compreensão da informação e dos conceitos abordados.

Assim sendo, é descrito todo o trabalho realizado em cada um dos componentes da arquitetura. Numa primeira secção são apresentadas as tecnologias que foram utilizadas na implementação da solução e, nas secções seguintes, são descritas as etapas do processo de ETL durante o desenvolvimento da solução. Estas etapas refletem as operações de extração que foram necessárias, as operações de transformação aos dados, incluindo as operações de limpeza e, por fim, o carregamento dos dados. Para além disso, são ainda mencionadas as formas de análises de dados a que os colaboradores têm acesso.

5.1 Tecnologias Utilizadas

A escolha das ferramentas foi efetuada com base nas características apresentadas na Secção 2.7 e também tendo em conta que, tanto o ISEP como a Real Vida Seguros, possuem acordo com a Microsoft, para que as ferramentas sejam utilizadas gratuitamente.

Para implementar a solução, foram eleitas tecnologias Microsoft, nomeadamente SQL Server para criar o modelo e suportar o armazenamento de dados; Microsoft Integration Services para efetuar o processo de ETL (extração, transformação e carregamento) e integrar facilmente com o Office e o SQL Server; por fim a ferramenta Microsoft Power BI para realizar as análises de dados.

5.2 Processo de ETL

Na Secção 2.5 é descrito o processo de ETL. Em suma, este processo consiste na extração de dados da(s) fonte(s), transformação desses dados tendo em conta as regras de negócio e, por fim, carregamento dos dados no repositório de destino, para poderem ser analisados.

Na Figura 36 é representada a arquitetura que caracteriza a solução final, as subsecções seguintes descrevem cada uma das etapas que constituem o processo de ETL.

5.2.1 Operações de Extração

Extração, é a primeira etapa do processo, os dados são extraídos do sistema de origem para uma área de preparação (Staging Area), permanecendo nas tabelas dessa estrutura temporariamente. A Figura 39 representa o processo de extração da arquitetura proposta para a solução.

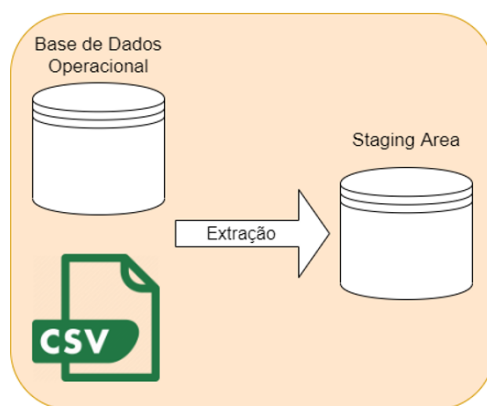


Figura 39 – Fase de extração na arquitetura do sistema

A *Staging Area* foi implementada na forma de tabelas, usando pequenos scripts para a sua criação, posteriormente foram inseridos os dados nas tabelas criadas, dados esses que estão presentes na base de dados do sistema operacional, referente à parte que suporta a gestão das apólices dos clientes da Real Vida Seguros.

A extração convém ser efetuada da forma mais rápida possível para não sobrecarregar o sistema operacional, uma vez que no futuro pode ter implicações no tempo que há disponível para extrair os dados, por esta razão a extração foi feita sem qualquer tipo de transformação, para não tornar o processo moroso.

Especificando, a partir da base de dados operacional foi desenvolvida uma base de dados idêntica restrita apenas aos dados que servirão de fonte ao sistema de análise. Esta não é mais do que um excerto da base de dados operacional, carregada diariamente com a informação atualizada (de fecho de dia), sem qualquer tipo de transformação de dados. Apenas se extraem/replicam os dados relevantes para as análises de dados. Esta estrutura não armazena quaisquer dados históricos, ficando essa função entregue ao armazém de dados, que é alimentado por esta estrutura. Nesse sentido, a *Staging Area* é apagada antes de efetuar o carregamento de dados diário, garantindo a integridade da informação diária, eliminando a provável existência de dados desatualizados.

As restrições PK-FK não se encontram implementadas, porque na realidade há relacionamentos entre as tabelas (até porque estas são um reflexo do sistema operacional). No fundo, há relacionamentos só que não estão implementados por questões de performance/desempenho.

5.2.2 Operações de Transformação

Nesta fase é adicionado valor aos dados com as tarefas de limpeza e transformação, uma vez que se garante a qualidade dos dados. A Figura 40 representa o processo de transformação da arquitetura proposta para a solução.

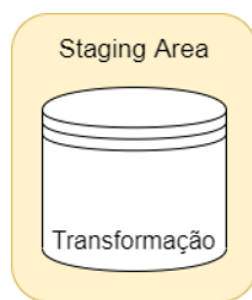


Figura 40 – Fase de transformação na arquitetura do sistema

Nas tabelas foi necessário aplicar regras, proceder-se à remoção de duplicados e até ao envio de dados incoerentes para tabelas DQP que armazenam os problemas de qualidade dos dados, para que a organização possa validar e corrigir os mesmos junto dos tomadores do seguro e mais tarde possam então ser carregados para o sistema.

Posto isto, e de forma a não existir redundância na informação que se pretende disponibilizar, foram efetuadas etapas de transformação, incluindo operações de limpeza nestas tabelas.

Seguidamente são apresentadas as transformações e operações de limpeza que se realizaram sobre os dados das tabelas na *Staging Area*.

5.2.2.1 Tabelas de CodigosPostais, Concelhos e Distritos

A Figura 41 representa o processo de transformação realizado na tabela "CodigosPostais". Nesta tabela houve necessidade de realizar algumas transformações, sendo mencionadas tais operações de seguida.

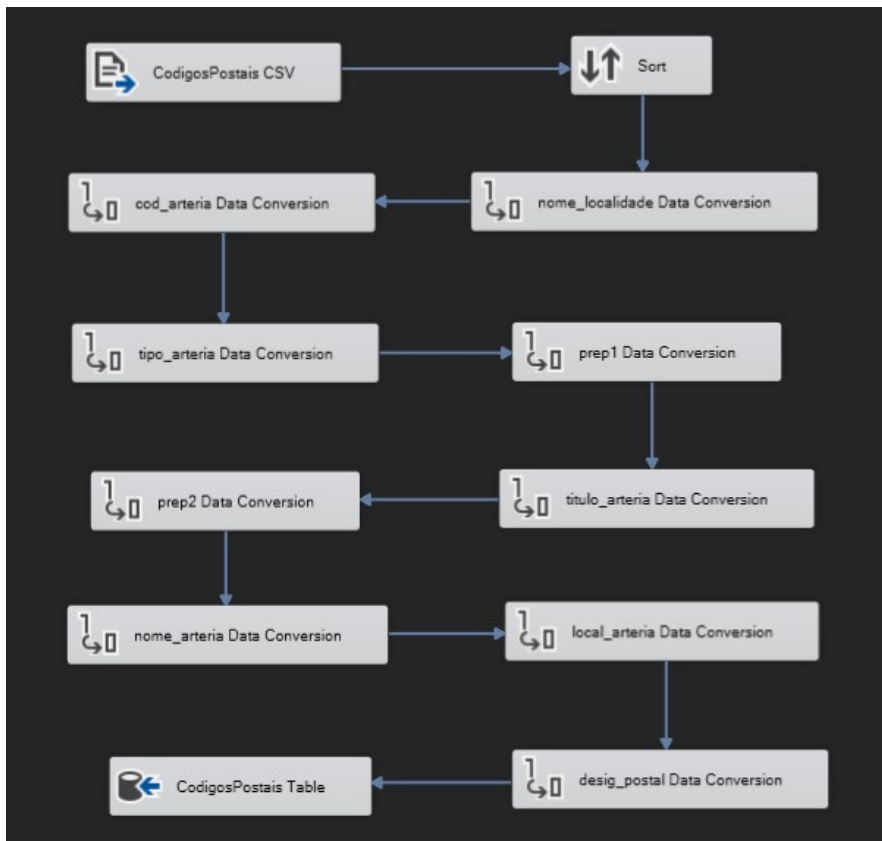


Figura 41 – Processo de Transformação na tabela CodigosPostais

Houve necessidade de realizar algumas conversões relativamente aos tipos de dados, para que os mesmos se aproximassem o mais possível do repositório de destino e evitar que problemas futuros de carregamento possam existir. É exemplo disso, a alteração do tipo de dados para os campos visíveis na figura anterior, uma vez que na base de dados original estes se encontravam como *string* e foram convertidos para o tipo *Unicode string*.

A Figura 42 representa o processo de transformação realizado na tabela “Concelhos”. Nesta tabela houve necessidade de realizar uma transformação, sendo mencionada de seguida.

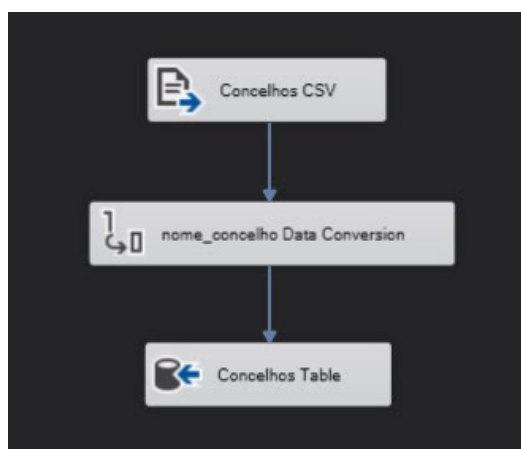


Figura 42 – Processo de Transformação na tabela Concelhos

Tal como na tabela de CodigosPostais, houve a necessidade de realizar a conversão relativamente ao tipo de dados. Para tal, procedeu-se à conversão de *string* para o tipo *Unicode string* no nome do concelho.

A Figura 43 representa o processo de transformação realizado na tabela “Distritos”. Nesta tabela houve necessidade de realizar uma transformação, sendo mencionada de seguida.

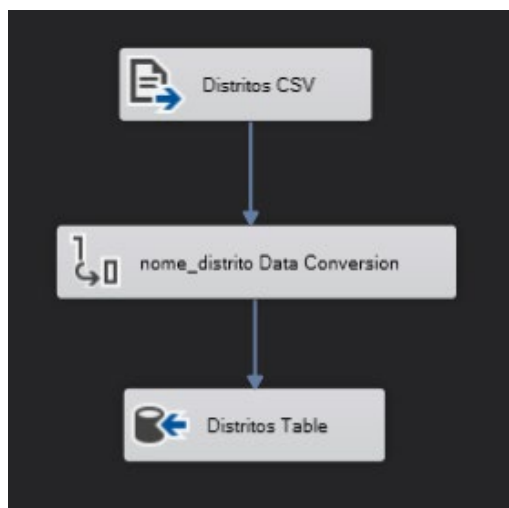


Figura 43 – Processo de Transformação na tabela Distritos

Igualmente como a tabela Concelhos, houve a necessidade de realizar a conversão relativamente ao tipo de dados. Para tal, procedeu-se à conversão de *string* para o tipo *Unicode string* no nome do distrito.

Pela análise da tabela de CodigosPostais verificou-se que existem diferentes endereços com o mesmo código postal (Figura 44).

cod_distrito	cod_concelho	cod_localidade	nome_localidade	cod_arteria	tipo_arteria	prep1	nome_arteria	num_cod_postal	ext_cod_postal	desig_postal
16	9	49015	Afife	100916	Caminho	das	Presas	4900	12	AFIFE
16	9	49015	Afife	10916	Caminho	de	Monte Agudo	4900	12	AFIFE
16	9	49015	Afife	110916	Caminho	do	Loureiro	4900	12	AFIFE
16	9	49015	Afife	120916	Caminho	da	Lapa	4900	12	AFIFE
16	9	49015	Afife	130916	Estrada		Florestal	4900	12	AFIFE
16	9	49015	Afife	140916	Caminho	dos	Cortiços	4900	12	AFIFE
16	9	49015	Afife	150916	Caminho	do	Viso	4900	12	AFIFE
16	9	49015	Afife	160916	Largo		Tomás Fernandes Pinto	4900	12	AFIFE
16	9	49015	Afife	170916	Largo	de	Cabanas	4900	12	AFIFE
16	9	49015	Afife	180916	Caminho	da	Tomenga	4900	12	AFIFE
16	9	49015	Afife	190916	Caminho	da	Revolta	4900	12	AFIFE
16	9	49015	Afife	200916	Caminho	do	Agro	4900	12	AFIFE
16	9	49015	Afife	20916	Urbanização	de	Bouça Cabrita	4900	12	AFIFE
16	9	49015	Afife	220916	Caminho	do	Calvário	4900	12	AFIFE
16	9	49015	Afife	230916	Caminho	do	Telheiro	4900	12	AFIFE
16	9	49015	Afife	240916	Caminho	de	Quelhães	4900	12	AFIFE
16	9	49015	Afife	250916	Caminho		Fonte Gatenha	4900	12	AFIFE
16	9	49015	Afife	260916	Caminho	do	Tojeiro	4900	12	AFIFE
16	9	49015	Afife	270916	Caminho	dos	Penedos	4900	12	AFIFE
16	9	49015	Afife	280916	Caminho	da	Ribela	4900	12	AFIFE
16	9	49015	Afife	290916	Caminho	do	Pincho	4900	12	AFIFE
16	9	49015	Afife	30916	Largo	da	Fonte da Gamosa	4900	12	AFIFE

Figura 44 – Linhas duplicadas nos CodigosPostais

Como para a implementação desta solução apenas é necessário distinguir as freguesias, concelhos e distritos, pode-se remover a duplicação destes resultados. Através da Figura 45 é possível verificar que no componente “Sort 1” foi aplicada a transformação ao lado, em que se ordena os resultados pelas

chaves que se pretendem e selecionando a opção “Remove rows with duplicate sort values” remove-se as duplicações acima mencionadas.

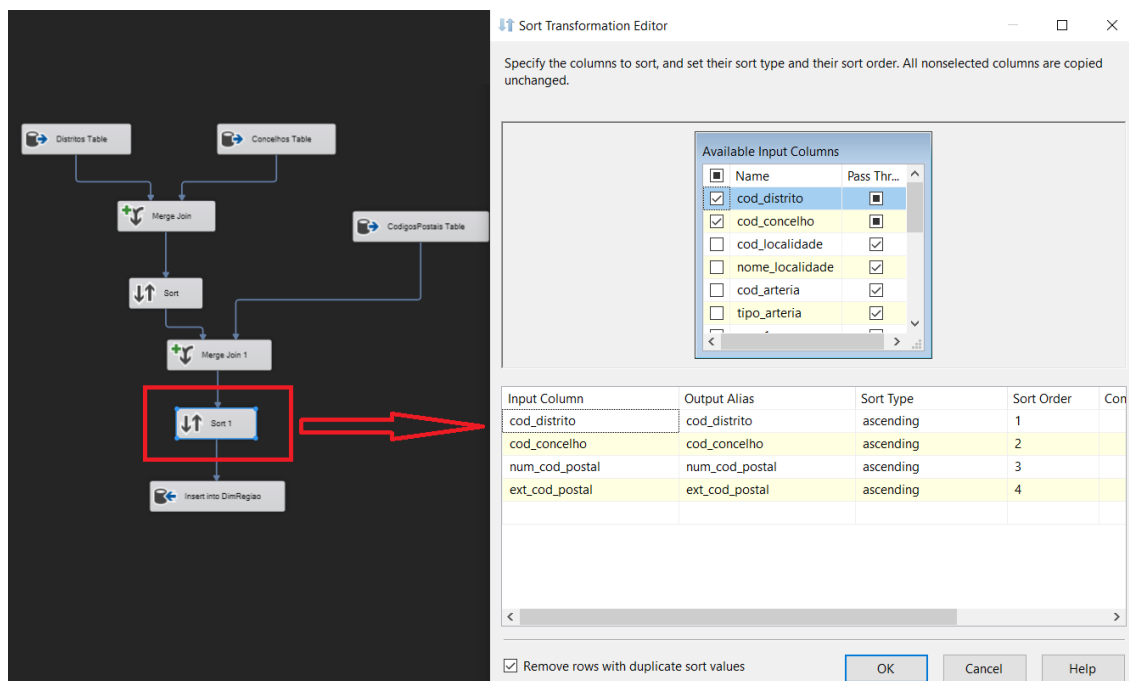


Figura 45 – Remoção de linhas duplicadas nos CodigosPostais

5.2.2.2 Tabela PROPET

A Figura 46 representa o processo de transformação realizado na tabela “PROPET”. Nesta tabela houve necessidade de realizar algumas transformações, sendo mencionadas tais operações de seguida.

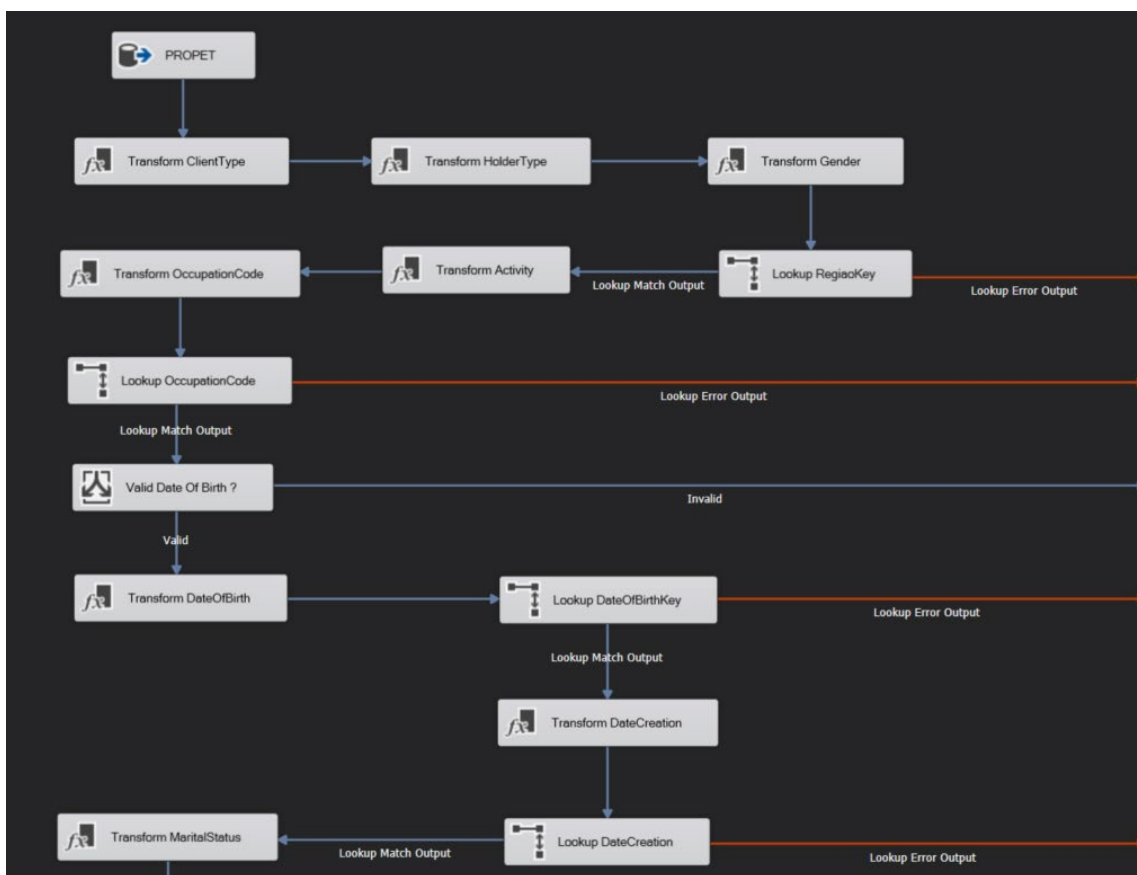


Figura 46 – Processo de Transformação na tabela PROPET

Na Tabela 27 houve necessidade de realizar algumas conversões relativamente aos tipos de dados. Segue-se as transformações efetuadas.

Tabela 27 – Transformações na tabela PROPET

Transformação	Origem			Destino		
	Tabela PROPET			DimTitular		
	Atributo	Tipo Dados	Valores	Atributo	Tipo Dados	Valores
<i>Transform ClientType</i>	PTTIPO	<i>nvarchar(1)</i>	T 1 2 " "	TipoCliente	<i>nvarchar(24)</i>	Titular Primeira Cabeça Segunda Cabeça Nenhum
<i>Transform HolderType</i>	PTNVAL	<i>nvarchar(1)</i>	N A I " "	TipoTitular	<i>nvarchar(9)</i>	Novo Alterar Importado Nenhum
<i>Transform Gender</i>	PTSEXO	<i>nvarchar(1)</i>	F M	Sexo	<i>nvarchar(9)</i>	Feminino Masculino
<i>Transform Activity</i>	PTACTI	<i>nvarchar(1)</i>	E I " "	Atividade	<i>nvarchar(10)</i>	Empresa Individual Nenhum

Transformação	Origem			Destino		
	Tabela PROPET			DimTitular		
	Atributo	Tipo Dados	Valores	Atributo	Tipo Dados	Valores
Transform OccupationCode	PT\$PRF	numeric(3,0)		Profissao	nvarchar(60)	
Transform DateOfBirth	PTDNAN	numeric(4,0)		DataNascimentoKey	date	
	PTDNME	numeric(2,0)				
	PTDNDI	numeric(2,0)				
Transform DateCreation	PTDATA	date		DataCriacaoKey	date	
Transform MaritalStatus	PTESTC	nvarchar(1)	V S C D U "	EstadoCivil	nvarchar(14)	Viúvo Solteiro Casado Divorciado União de Facto Nenhum

Aquando da fase de carregamento de dados, as transformações que não forem possíveis efetuar são redirecionadas para a tabela PROPET_DQP (Figura 47).

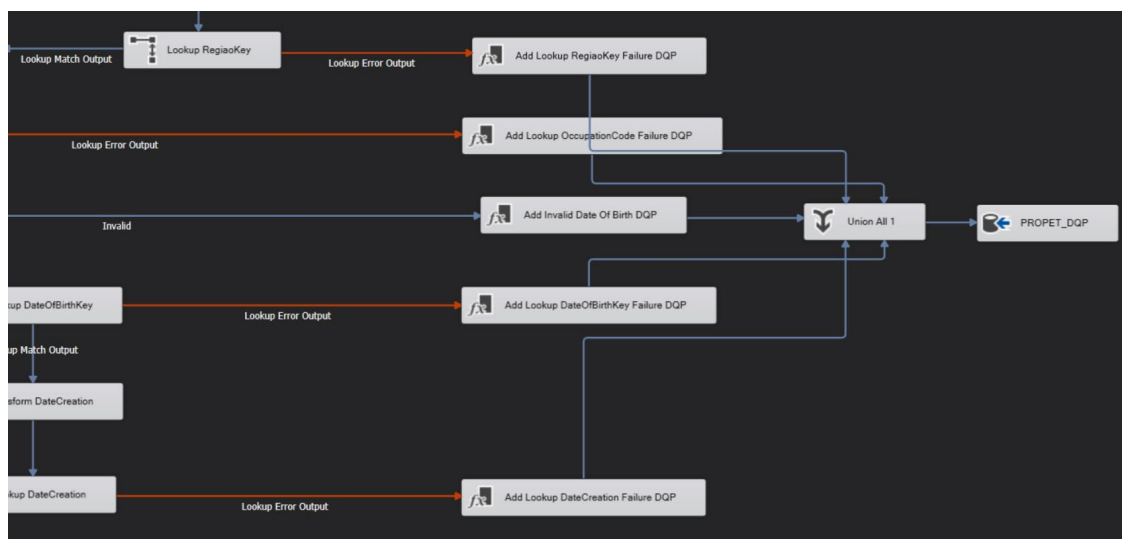


Figura 47 – Redirecionamento de DQPs para a Tabela PROPET_DQP

Pela análise da tabela PROPET_DQP verificou-se que existem valores em falta nos atributos PTSEXO (sexo), PTACTI (atividade), PTNACD (nacionalidade), PTNATD (naturalidade) e PTESTC (estado civil) (Figura 48).

	PTSPRO	PTTIPO	PTSODD	PTNVAL	PTSEXO	PTSCTT	PTSCT3	PTACTI	PTSCAE	PTS\$PRF	PTDNAN	PTDNME	PTDNDI	PTUSER	PTDATA	PTNACD	PTNATD	PTESTC
1	30709	1	445546	N		4450	111		0	0	1990	1	1	T0000003	2019-03-20			
2	30709	T	16102	N		4450	111		0	0	1990	1	1	T0000003	2019-03-20			
3	30717	1	0	N		4050	276		0	0	1980	6	11	T0000004	2019-03-20			
4	30717	T	0	N		4050	276		0	0	1980	6	11	T0000004	2019-03-20			
5	30718	1	445565	N		4450	123		0	0	1980	1	1	T0000005	2019-03-20			

Figura 48 – Exemplo de linhas com os atributos (PTSEXO, PTACTI, PTNACD, PTNATD e PTESTC) em falta

Existe uma violação de domínio nos atributos PT\$CTT e PT\$CT3, uma vez que o código postal 9999-999 é o valor de defeito da empresa não existindo esse código postal (Figura 49).

	PT\$PRO	PTTIPO	PT\$COD	PTNVAL	PTSEXO	PT\$CTT	PT\$CT3	PTACTI	PT\$CAE	PT\$PRF	PTDNAN	PTDNME	PTDNDI	PTUSER	PTDATA	PTNACD	PTNATD	PTESTC
1	1038	1	175469	A	M	9999	999	I	0	0	0	0	0	W00946501	2015-02-06	***	***	
2	1038	T	0	N	M	9999	999	I	0	0	0	0	0	W00946501	2015-02-06	***	***	
3	1039	1	0	N	F	9999	999	I	0	0	0	0	0	W00946501	2015-02-06	***	***	
4	1039	T	0	N	F	9999	999	I	0	0	0	0	0	W00946501	2015-02-06	***	***	
5	1040	1	0	N	F	9999	999	I	0	0	0	0	0	W00946501	2015-02-06	***	***	

Figura 49 – Exemplo de linhas com código postal inválido

Existe uma violação de regra de negócio nos atributos PTACTI e PT\$CAE, uma vez que quando o tipo de atividade é “E” (empresa) está em falta o CAE e quando o tipo de atividade é “I” (individual) está a mais o CAE (Figura 50).

	PT\$PRO	PTTIPO	PT\$COD	PTNVAL	PTSEXO	PT\$CTT	PT\$CT3	PTACTI	PT\$CAE	PT\$PRF	PTDNAN	PTDNME	PTDNDI	PTUSER	PTDATA	PTNACD	PTNATD	PTESTC
1	30758	1	134519	A	M	2790	379	I	11	195	1961	8	11	U295642	2019-03-21			
2	30758	T	295642	A	M	2790	379	I	11	195	1961	8	11	U295642	2019-03-21			
3	65086	T	74840	N		7860	207	E	0	0	1	1	1	W02622401	2020-09-16	Portuguesa		
4	68819	T	79570	N		7800	751	E	0	0	1	1	1	W02622401	2020-10-28	Portuguesa		

Figura 50 – Exemplo de linhas com o PTACTI e PT\$CAE inválidos

Existe uma violação de domínio no atributo PT\$PRF (profissão), uma vez que o valor 0 é o valor por omissão da empresa (Figura 50).

Existe uma violação de domínio nos atributos PTDNAN, PTDNME e PTDNDI, uma vez que existem valores a 0-0-0 e 1-1-1 que são valores de defeito de datas vazias ou datas por defeito (Figura 51).

	PT\$PRO	PTTIPO	PT\$COD	PTNVAL	PTSEXO	PT\$CTT	PT\$CT3	PTACTI	PT\$CAE	PT\$PRF	PTDNAN	PTDNME	PTDNDI	PTUSER	PTDATA	PTNACD	PTNATD	PTESTC
1	1144	1	0	N	M	9999	999		0	0	0	0	0	W00935101	2015-10-13	***	***	
2	1144	T	0	N	M	9999	999	I	0	0	0	0	0	W00935101	2015-10-13	***	***	
3	1156	T	695619	N		2970	0	E	94995	0	1	1	1	W01702601	2015-10-30			
4	1162	T	689101	I		4520	112	E	15201	0	1	1	1	W00938701	2015-11-12			

Figura 51 – Exemplo de linhas com a data de nascimento inválida

O atributo PTNACD é violado, uma vez que existem duplicados aproximados resultantes da existência de erros ortográficos, ou seja, existem erros ortográficos, como por exemplo “Alemã” vs “Alema”, “POTRUGUESA”, “poryuguesa”, “PORT|UGUES”, “Portugfues”, “portugursa”, “Portuguêsa”, entre outros e existe ainda uma violação do domínio nos valores “***” (Figura 52).

	PTNACD	PTNACD
1	Portugues	P0ORTUGUES
2	***	P0rtuguesa
3	*Portugues	Partugues
4	.Portuguesa	PARTUGUESA
5	Alema	PÇORTUGUES
6	Alemã	POÇRTUGUES
7	ALEMANHA	POERTUGUES
8	Alemão	Poetuguesa
9	Oprtuguesa	Poftuguesa
10	ORTUGUESA	poortugues
11	OURTUGUESA	POPRTUGUES
12	P+ORTUGUES	Poretugues

Figura 52 – Exemplo de linhas com a nacionalidade inválida

Existe também uma violação de domínio no atributo PTNATD, pois existem valores “***”.

5.2.2.3 Tabela PROPE0

A Figura 53 representa o processo de transformação realizado na tabela “PROPE0”. Nesta tabela houve necessidade de realizar algumas transformações, sendo mencionadas tais operações de seguida.

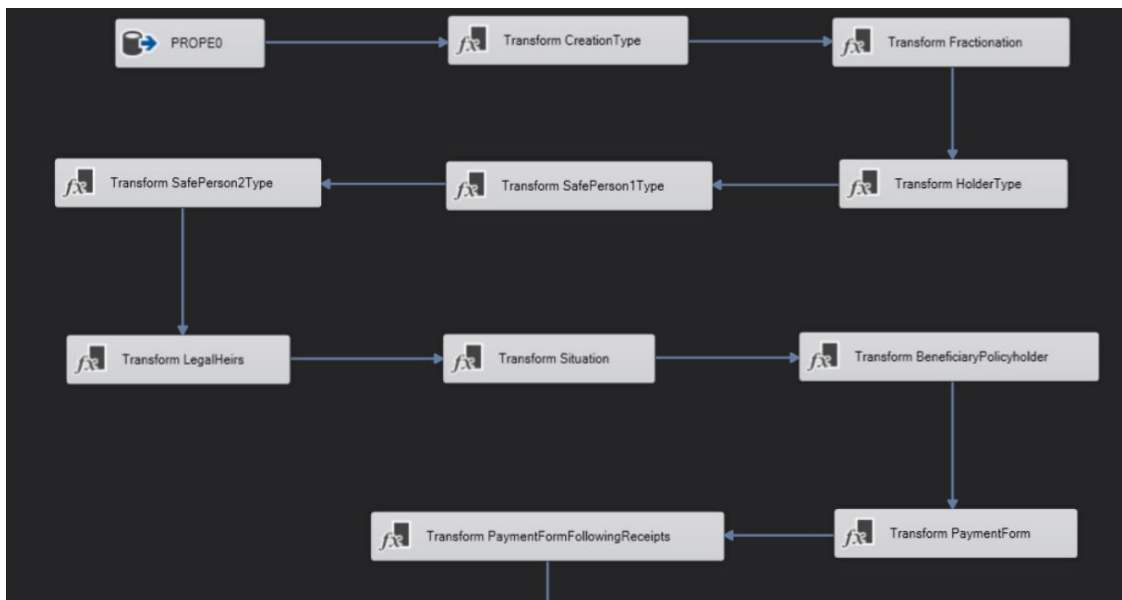


Figura 53 – Processo de Transformação na tabela PROPE0

Na Tabela 28 houve necessidade de realizar algumas conversões relativamente aos tipos de dados (e.g, de nvarchar(1) para nvarchar(9)), mas também transformações ao nível dos próprios valores (e.g., de "N" para "Novo").

Segue-se as transformações efetuadas.

Tabela 28 – Transformações na tabela PROPE0

Transformação	Origem			Destino		
	Tabela PROPE0			DimDadosProposta		
	Atributo	Tipo Dados	Valores	Atributo	Tipo Dados	Valores
<i>Transform CreationType</i>	POTIPO		N	TipoCriacao		Novo
<i>Transform HolderType</i>	POTITP	nvarchar(1)	A	TipoTitular	nvarchar(9)	Alterar
<i>Transform SafePerson1Type</i>	POP1TP		I	Tipo1PSegura		Importado
<i>Transform SafePerson2Type</i>	POP2TP		“ ”	Tipo2PSegura		Nenhum
<i>Transform Fractionation</i>	PO\$FRA	nvarchar(1)	M S T A U “ ”	Fracionamento	nvarchar(10)	Mensal Semestral Trimestral Anual Único Nenhum

Transformação	Origem			Destino		
	Tabela PROPEO			DimDadosProposta		
	Atributo	Tipo Dados	Valores	Atributo	Tipo Dados	Valores
<i>Transform LegalHeirs</i>	POHERD	<i>nvarchar(1)</i>	S	HerdeirosLegais	<i>nvarchar(3)</i>	Sim
<i>Transform BeneficiaryPolicy holder</i>	POTOSE		N	TomadorSeguroBeneficiario		Não
<i>Transform Situation</i>	POSITU	<i>numeric(1,0)</i>	-1 0 1 2 3 4 5 6 7 8 9 -6 -7 -8 -9 " "	Situacao	<i>nvarchar(24)</i>	Em Validação Criada Pendente Pagamento Submetida Área Técnica Pago Por Cheque Pago SIBS Apólice Criada Arquivada Sem Efeito Apólice Condicionada Anulada Pendente de Registo IBAN Pendente de Data Inicio Pendente de Autorização Recusada Todos
<i>Transform PaymentForm</i>	POINF2	<i>nvarchar(1)</i>	C M D	FormaPagamento	<i>nvarchar(18)</i>	Cheque Multibanco Débito Direto SEPA
<i>Transform PaymentForm FollowingReceipts</i>	POFPRS		T A N " "	FormaPagamento		Tesouraria Mediador Nenhum Vazio

5.2.2.4 Tabela MODA00

A Figura 54 representa o processo de transformação realizado na tabela "MODA00". Nesta tabela houve necessidade de realizar algumas transformações, sendo mencionadas tais operações de seguida.

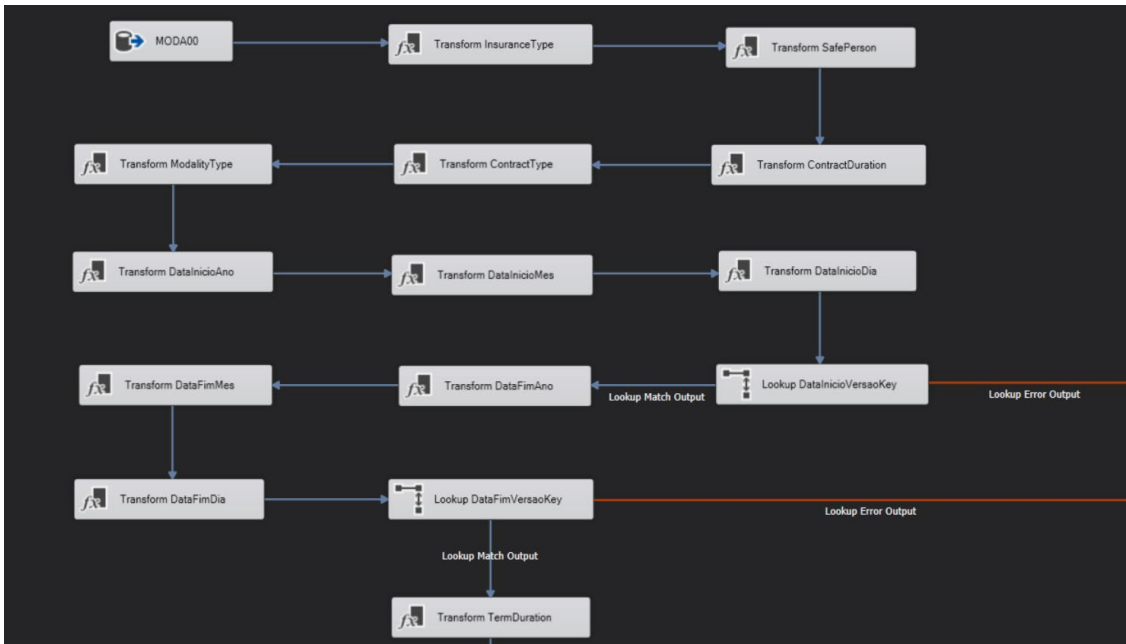


Figura 54 – Processo de Transformação na tabela MODA00

Na Tabela 29 houve necessidade de realizar algumas conversões relativamente aos tipos de dados, mas também transformações ao nível dos próprios valores. Segue-se as transformações efetuadas.

Tabela 29 – Transformações na tabela MODA00

Transformação	Origem			Destino		
	Tabela MODA00			DimProduto		
	Atributo	Tipo Dados	Valores	Atributo	Tipo Dados	Valores
<i>Transform InsuranceType</i>	MOTSEG	<i>nvarchar(1)</i>	S P F R O K C M E A U D I	TipoSeguro	<i>nvarchar(29)</i>	SAE PPR FINANCEIRO RISCO OURO REAL PPI UNIT-LINKED (FUNDO UNICO) CLASSICO MISTO RENDAS PPA PPR UNIDADES DE CONTA SEG SAUDE ACIDENTES PESSOAIS
<i>Transform SafePerson</i>	MOPSEG	<i>nvarchar(1)</i>	1 2 A F	PessoaSegura	<i>nvarchar(21)</i>	1 Cabeça (Individual) 2 Cabeça (Individual) Grupo Aberto Grupo Fechado

Transformação	Origem			Destino		
	Tabela MODA00			DimProduto		
	Atributo	Tipo Dados	Valores	Atributo	Tipo Dados	Valores
<i>Transform ContractDuration</i>	MODURC	<i>nvarchar(1)</i>	V T R I C	DuracaoContra	<i>nvarchar(22)</i>	Vitalício Temporário Temp. Anual Renovável Temp. Prazo Inf. 1 Ano Compl. Reforma
<i>Transform ContractType</i>	MOTCCR	<i>nvarchar(1)</i>	C R	TipoContrato	<i>nvarchar(7)</i>	Capital Renda
<i>Transform ModalityType</i>	MOVIMO	<i>nvarchar(1)</i>	V M A	TipoModalidad	<i>nvarchar(15)</i>	Vida Morte Ambos
<i>Transform DataInicioAno</i>	MOAVIN	<i>numeric(4, 0)</i>		DataInicioVersaoKey	<i>int</i>	
<i>Transform DataInicioMes</i>	MOMVIN	<i>numeric(2, 0)</i>				
<i>Transform DataInicioDia</i>	MODVIN	<i>numeric(2, 0)</i>				
<i>Transform DataFimAno</i>	MOAVFI	<i>numeric(4, 0)</i>		DataFimVersaoKey	<i>int</i>	
<i>Transform DataFimMes</i>	MOMVFI	<i>numeric(2, 0)</i>				
<i>Transform DataFimDia</i>	MODVFI	<i>numeric(2, 0)</i>				

5.2.3 Operações de Carregamento

A etapa final do processo ETL é o carregamento de dados nos modelos dimensionais alvo da área de apresentação, ou seja, nos Data Mart. A Figura 55 representa o processo de carregamento da arquitetura proposta para a solução.

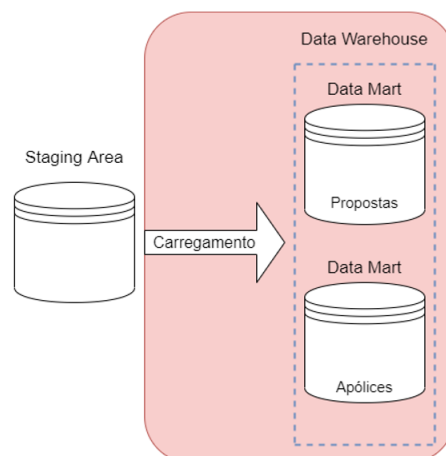


Figura 55 – Fase de carregamento na arquitetura do sistema

Primeiramente são carregadas as dimensões e de seguida é executado o carregamento da tabela de factos, uma vez que esta depende da chave das dimensões.

Aquando do carregamento das dimensões é necessário validar se os atributos devem sofrer alterações, quer seja para manter o histórico dos dados, quer seja para atualizar a informação (manter sempre a informação mais atual). Na solução implementada, os atributos das dimensões apenas sofreram alterações de Tipo 1 e de Tipo 2 (Anexo II).

Assim, foram inseridos três atributos nos scripts do *data mart* para guardar as datas: *“IsCurrent”*, com o intuito de perceber de imediato qual a informação mais atual para um determinado registo e *“EffectiveDate”* e *“ExpiredDate”*, sendo que a finalidade destes dois atributos é caracterizar o período em que uma determinada alteração foi válida (entre uma data início (*EffectiveDate*) e uma data fim (*ExpiredDate*)). O registo em vigor é também aquele que apenas tem data início (além do *IsCurrent* a *“Yes”*).

A Figura 56 corresponde ao processo implementado para se efetuar o carregamento da dimensão *“DimDadosProposta”*, para esse efeito foi utilizado o componente de *slowly changing dimension*, disponível num *Integration Services Project* criado no Visual Studio. O carregamento das tabelas *“DimTitular”*, *“DimProduto”* e *“DimDadosApolice”*, foi realizado de forma semelhante.



Figura 56 – Carregamento da tabela DimDadosProposta

O carregamento das tabelas “DimRegiao” e “DimEstruturaComercial” foi realizado de forma direta (Figura 57).

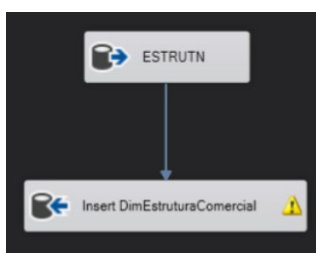


Figura 57 – Carregamento da tabela DimEstruturaComercial

No caso da existência de dados incorretos são reencaminhados para as tabelas DQP, que após a realização das correções necessárias nesses dados, numa 2ª iteração de execução, esses dados (corrigidos) são carregados com sucesso no DW.

Posto isto, findo o carregamento das dimensões, é necessário proceder-se ao carregamento da tabela de factos. As figuras seguintes (Figura 58, Figura 59 e Figura 60) corresponde ao processo implementado para o carregamento da tabela de factos.

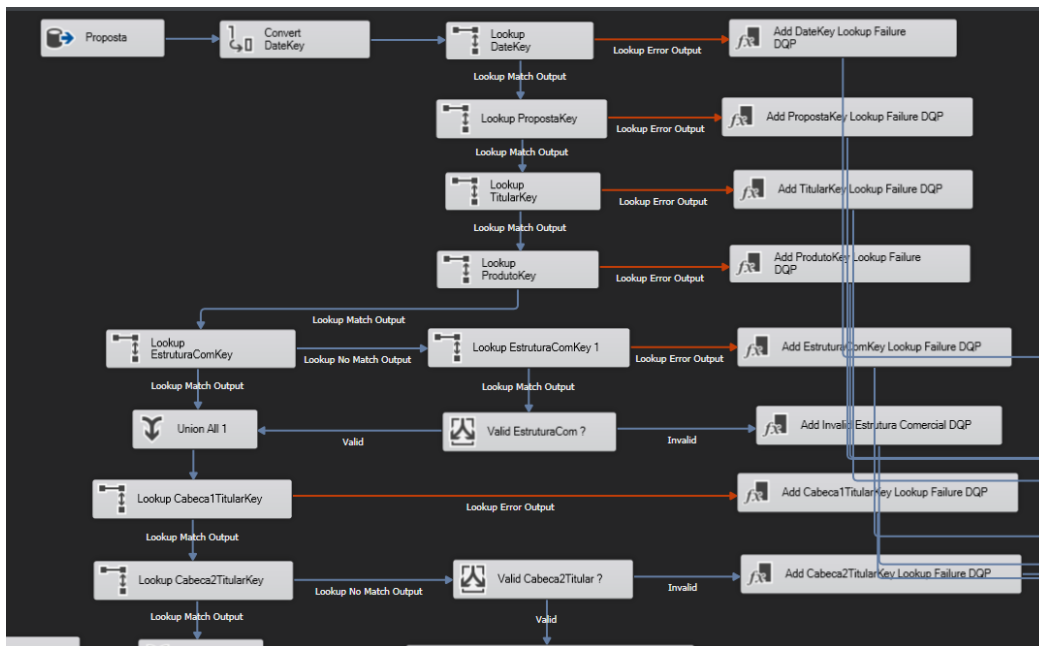


Figura 58 – Carregamento da tabela FactProposta (1)

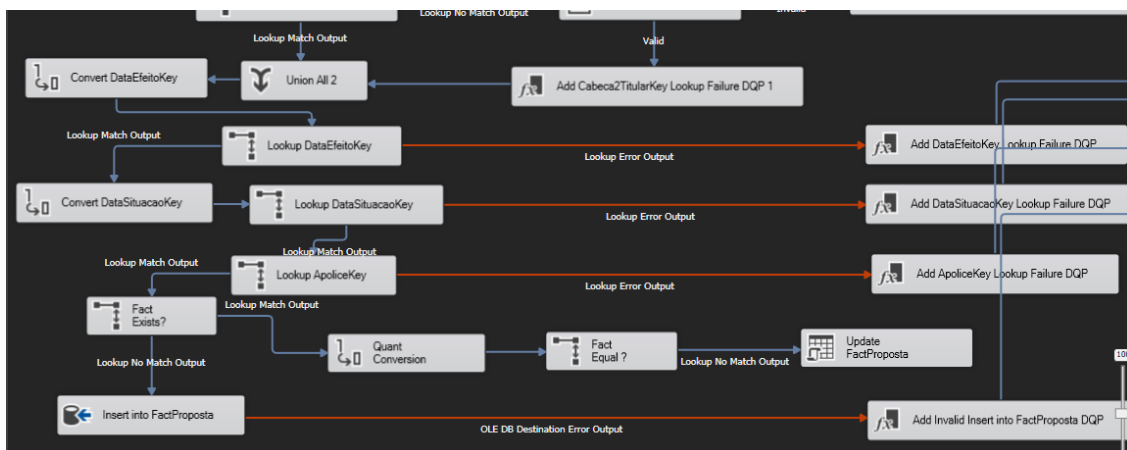


Figura 59 – Carregamento da tabela FactProposta (2)

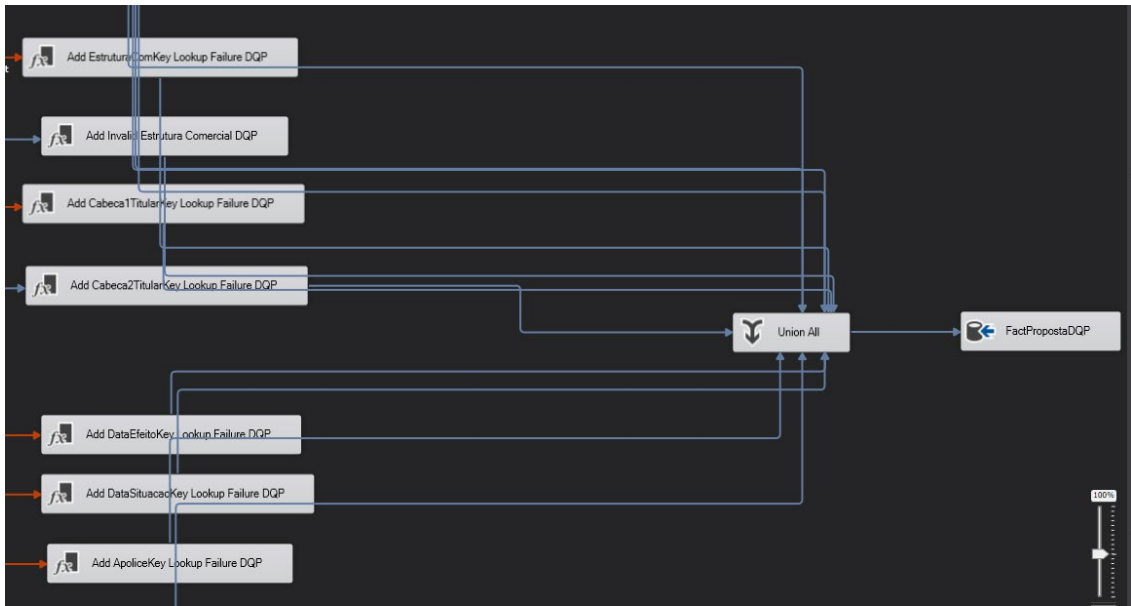


Figura 60 – Carregamento da tabela FactProposta (3)

5.3 Análises de Dados

A análise de dados é a última fase apresentada na arquitetura proposta do sistema. A Figura 61 representa a contextualização da mesma.

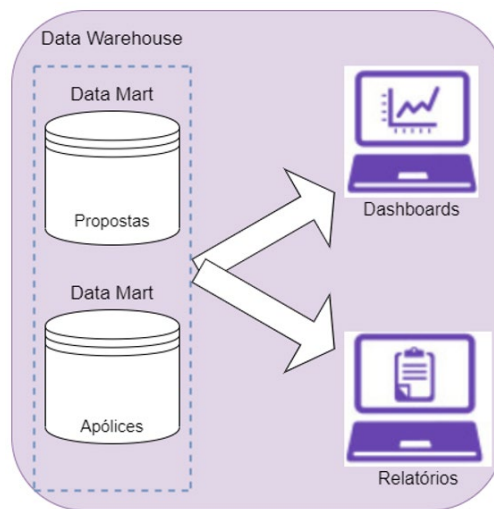


Figura 61 – Fase de análise de dados na arquitetura do sistema

Através da ferramenta Power BI é possível a visualização dos dados que se encontram nos Data Marts. Para tal, é necessário que a ferramenta aceda aos dados que se encontram no DW.

Pretende-se realizar análises diversificadas aos dados, algumas das quais serão apresentadas no capítulo seguinte, que correspondem às US definidas na Secção 4.1.1., e produzir dashboards com os indicadores chave de desempenho relevantes, de forma que o utilizador final consiga tirar conclusões de acordo com os filtros que aplica e ainda ser possível executar relatórios.

6 Apresentação e Análise de Resultados

Neste capítulo são exibidos exemplos de análises efetuadas com a solução implementada. Os *dashboards* desenvolvidos, encontram-se de acordo com o estudo que foi feito ao longo desta dissertação, evidenciando assim, toda a informação que se considera relevante para o utilizador final.

6.1 Indicadores

Existe um conjunto de indicadores que são alvo de estudo ao longo desta dissertação e que são exibidos ao longo de um conjunto de relatórios. Destacam-se indicadores sobre titulares, sobre produtos, sobre propostas e apólices. O conjunto desses indicadores possibilita averiguar números e motivos que levam a conclusões sobre aquisição dos seguros de vida.

As análises implementadas têm como utilizador final os profissionais da seguradora, pelo que estas devem apresentarem-se visualmente de fácil interpretação, de forma a conseguir-se retirar conclusões rapidamente e com pouco esforço.

6.1.1 Informação Titulares

A Figura 62 representa o primeiro *dashboard* implementado relativamente à informação de titulares, onde os dados podem ser obtidos durante os anos de 2015 a 2020.

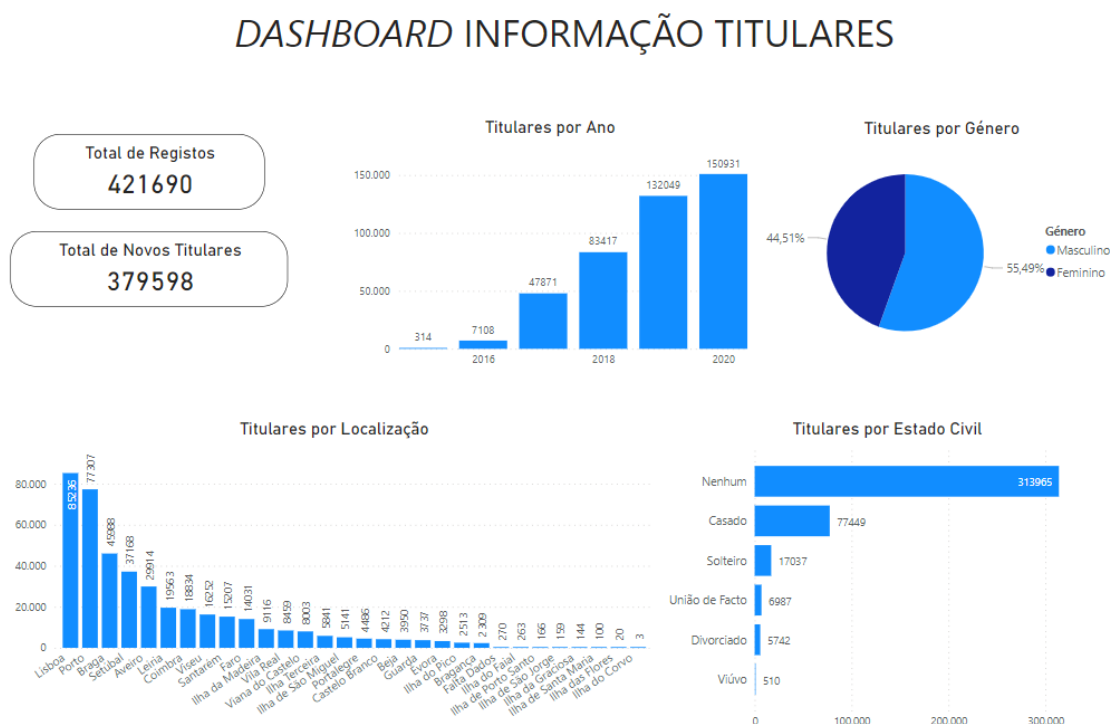


Figura 62 – *Dashboard* de informação acerca dos titulares

De acordo com o mesmo, pode-se verificar que durante os cinco anos foram efetuados 421690 registos, dos quais se contabilizam 379598 como sendo novos titulares.

Relativamente ao número de titulares por ano, pode-se visualizar que no ano de 2015 se contabilizaram 314 titulares, no ano de 2016 houve 7108 titulares, no ano de 2017 houve 47871 titulares, no ano de 2018 houve 83417, no ano de 2019 houve 132049 e por fim no ano de 2020 houve 150931 titulares, respetivamente houve um crescimento de 95.58%, 85.15%, 42.61%, 36.83% e 12.51% em comparação com o ano anterior. Este gráfico tem a particularidade de se conseguir fazer *drill down*, para se obter a informação de forma mais detalhada. Através do *drill down* o valor do número de titulares é conseguido para o trimestre, mês e até mesmo ao dia correspondente a cada um dos anos. A Figura 63 representa o *drill down* que se efetuou no ano de 2020, quarto trimestre, no mês de novembro, dando os valores dos titulares por dia.

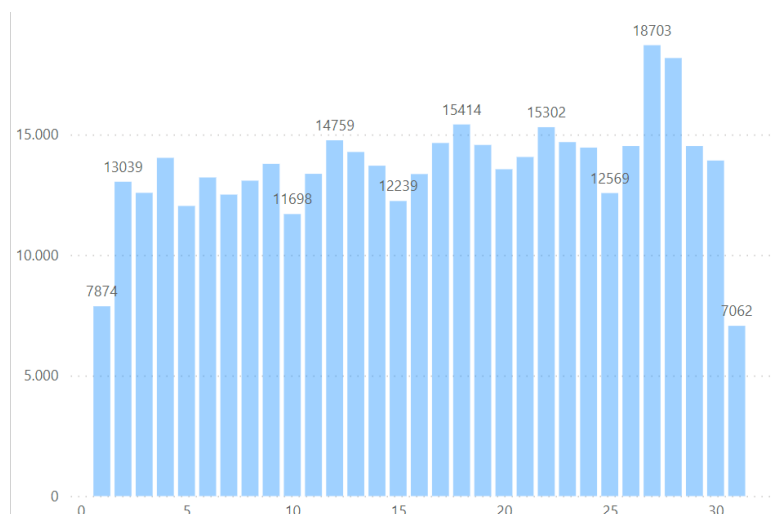


Figura 63 – Gráfico Drill Down do Número de Titulares por Ano

Ainda no *dashboard*, também se pode visualizar a percentagem de titulares por género, sendo que o 55,49% das pessoas são do sexo masculino, enquanto o sexo feminino possui uma percentagem de 44,51%.

É de notar no gráfico “Titulares por Localização” que as duas grandes potencias de titulares são em Lisboa e Porto, consideradas as maiores cidades de Portugal.

Por fim, pode-se visualizar o número de titulares por estado civil, neste gráfico é evidente a falta desta informação, uma vez que o estado civil que mais se destaca é o “Nenhum”. Por esta razão os profissionais saberão que terão de questionar e preencher essa informação. Assim sendo, os três estados civis que mais se destacam em seguida são: “Casado” com 77449 titulares, “Solteiro” com 17037 titulares e por fim, “União de Facto” com 6987 titulares. Por último, os estados civis que possuem menos titulares são “Divorciado” e “Viúvo” com 5742 e 510 titulares, respetivamente.

A Figura 64 representa um segundo relatório da informação acerca dos titulares. Saber o número de titulares com determinada profissão é considerado um fator importante de acordo com o contexto atual. Existem profissões que por sofrerem um elevado nível de exposição são consideradas de risco, dado estarem mais propensas a determinados aspetos perigosos. Assim sendo obter tal indicação pode ser útil, já que o profissional de seguros pode identificar circunstâncias que impeçam de imediato a obtenção de um certo seguro. Através do gráfico é possível verificar que a maioria dos titulares tem profissões como “Administrativo”, “Professor”, “Reformado”, “Empresário”, entre outras.

Para além da profissão, este relatório possui o número de titulares pela respetiva nacionalidade. Como era de esperar a nacionalidade com mais titulares é a “Portuguesa”, seguindo-se a categoria sem nacionalidade (“Nenhum”), que mais tarde terá de se resolver.

RELATÓRIO INFORMAÇÃO TITULARES

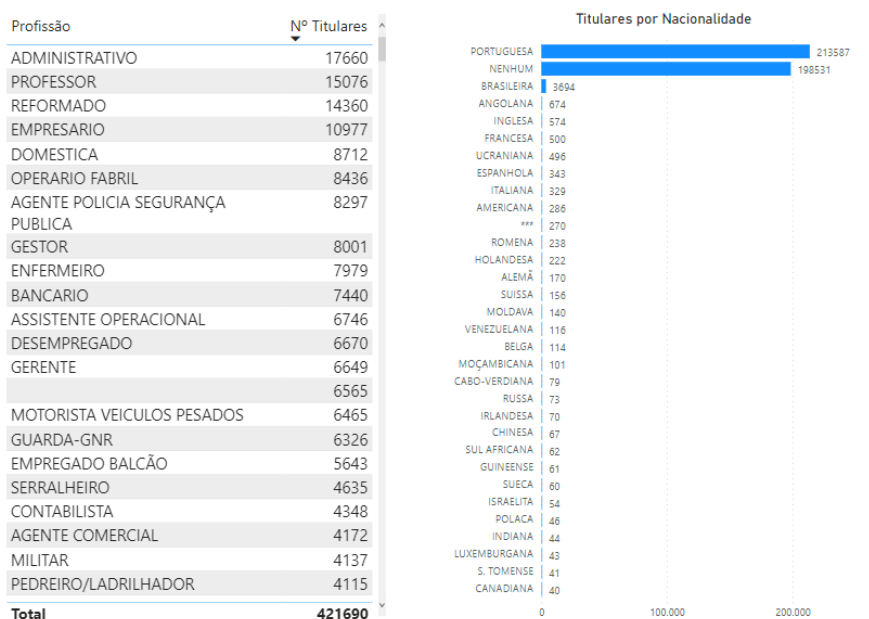


Figura 64 – Relatório de informação acerca dos titulares

Como já referido na análise do primeiro *dashboard*, as duas grandes potencias de titulares são Lisboa e Porto. No *dashboard* seguinte (Figura 65) é possível analisar estes dados de diferentes formas, tanto como por um mapa, por uma tabela ou por um mapa de formas, características visíveis no tamanho dos círculos, a quantidade de titulares ou cores mais escuras com mais titulares, respetivamente.

DASHBOARD INFORMAÇÃO TITULARES

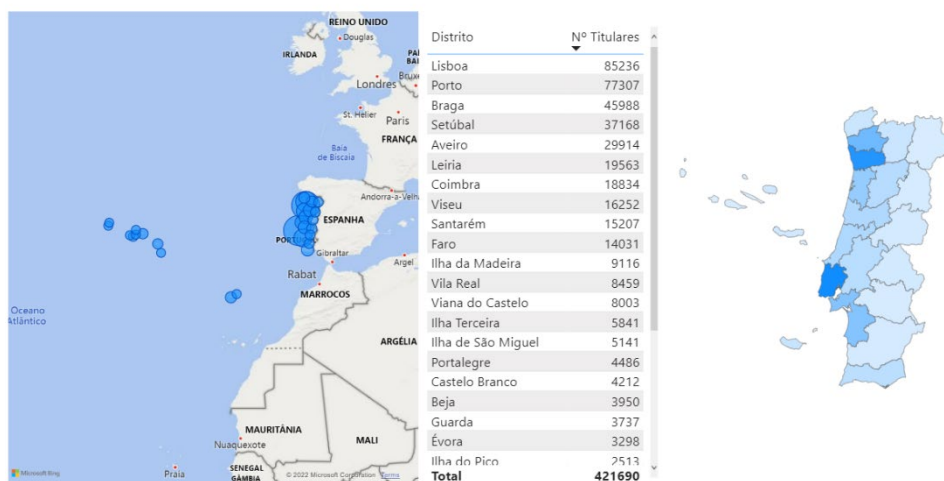


Figura 65 – Dashboard de informação acerca dos titulares

Neste *dashboard* (Figura 66), é evidenciado o número de registos de titulares que efetuaram uma proposta.

Apesar de já ser possível um cliente fazer um seguro sozinho sem ser por intermédio de um agente, normalmente são os agentes que vão ao encontro dos clientes com o objetivo de conseguirem efetuar a venda de um seguro.

Nos cinco anos verifica-se que houve um total de 410295 registos durante a semana e 11395 ao fim de semana.

Os registos efetuados durante a semana são praticamente idênticos sendo que o dia da semana com mais registos é a quarta-feira. No entanto é evidente a desproporcionalidade dos registos durante a semana versus fim-de-semana. Por outro lado, os registos por trimestre e semestre praticamente são idênticos.

Quanto aos registos mensais é evidente que há menos registos nas férias de verão (julho, agosto e setembro). Isto provavelmente deve-se ao facto de a empresa privilegiar o tempo de menor atividade para marcação de férias dos colaboradores.

É de realçar que toda esta informação pode ser obtida ao ano, mês e até mesmo ao nível do dia da semana, seleccionando os devidos filtros.

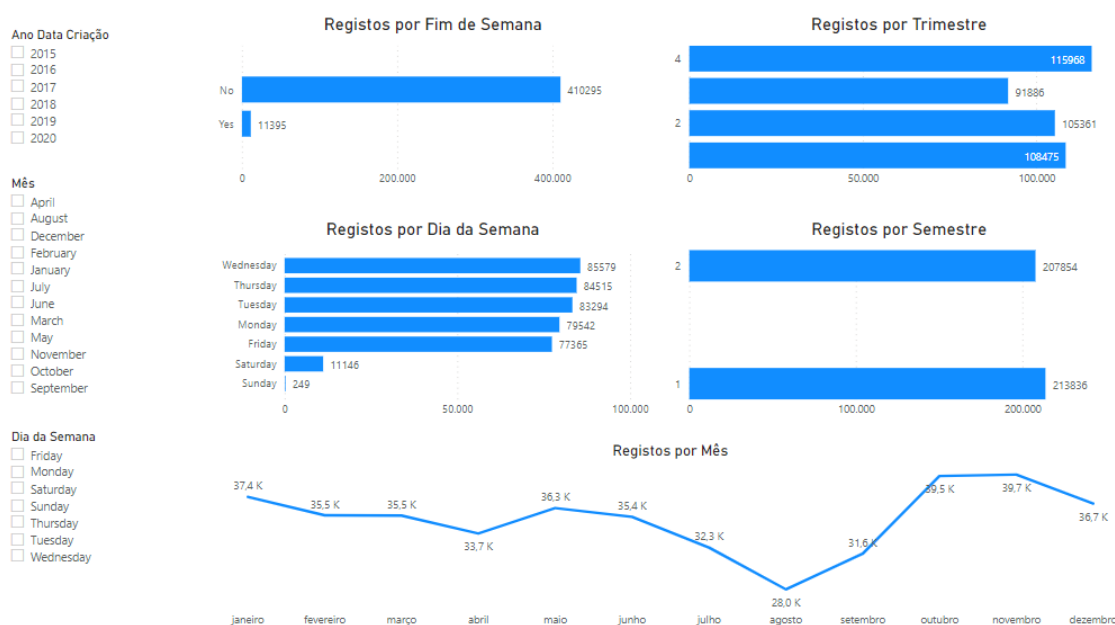


Figura 66 – Dashboard de informação acerca dos titulares

6.1.2 Informação Produtos

A Figura 67 representa o primeiro dashboard implementado acerca da informação dos produtos.

RELATÓRIO INFORMAÇÃO PRODUTOS

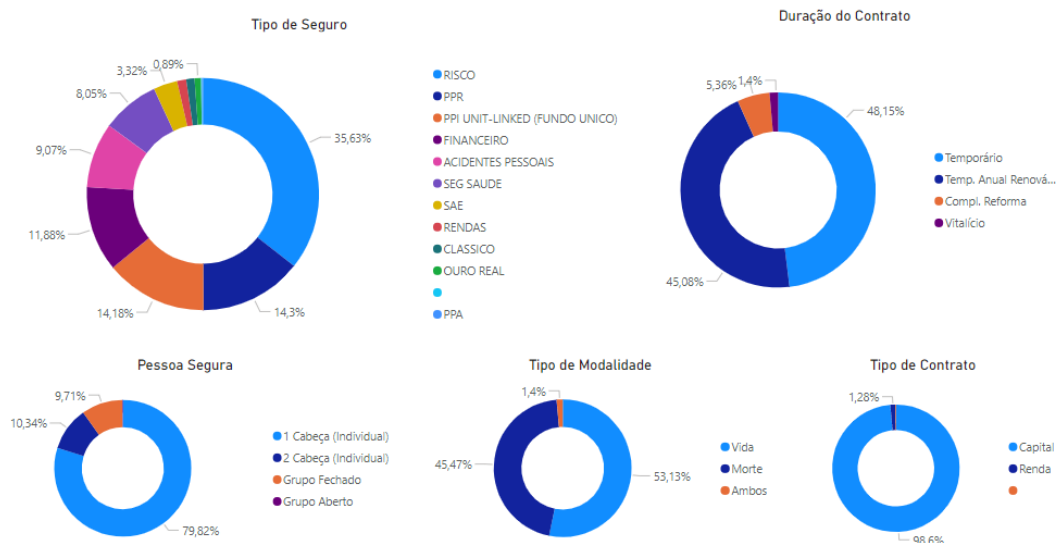


Figura 67 – Dashboard de informação acerca dos produtos

Verificou-se que a maior fatia do tipo de seguro corresponde aos produtos de risco e a menor fatia corresponde ao produto PPA – Planos de Poupança em Ações. Por sua vez, na duração do contrato, as grandes vertentes são o “Temporário” com 48.15% dos produtos e o “Temporário Anual Renovável” com 45.08%.

Tal como era de esperar, uma vez que a maioria dos seguros têm pelo menos um segurado, 79.82% dos produtos tem apenas primeira pessoa segura, seguindo-se de 10.34% dos seguros com segunda pessoa segura.

A maioria dos produtos são da modalidade de vida ou de morte, existindo apenas 11 produtos com ambas as modalidades.

O tipo de contrato da maioria dos produtos é o capital com cerca de 98.60% dos casos.

6.1.3 Informação Potenciais Apólices

Quanto à informação acerca das potenciais apólices, que representa o contrato para a venda de um produto, a Figura 68 representa alguns dados que compõem esse contrato.

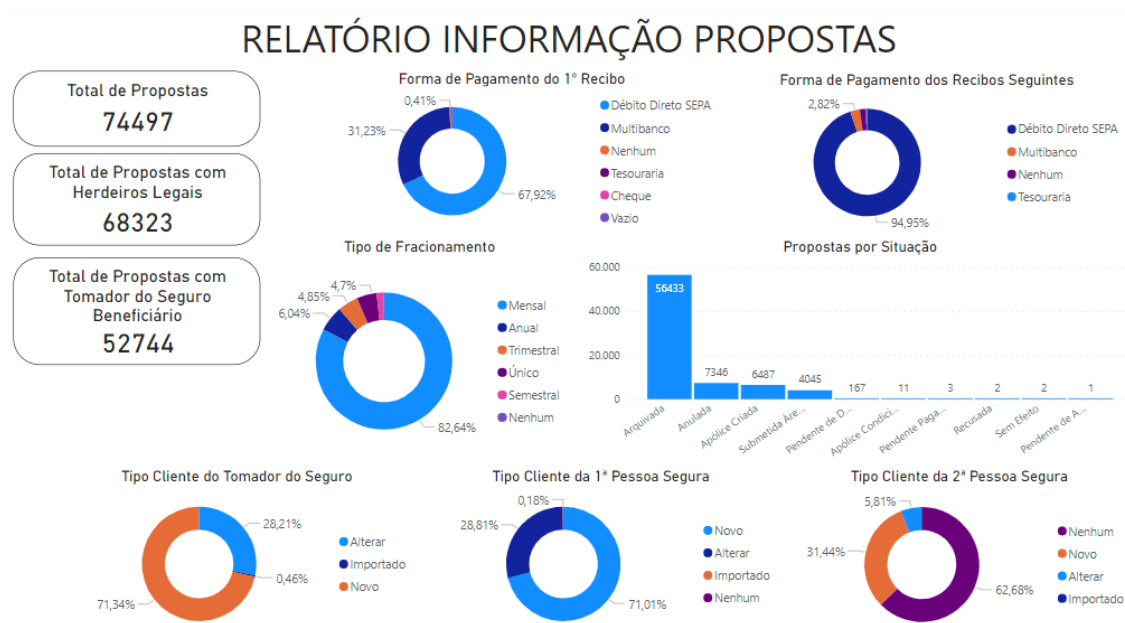


Figura 68 – *Dashboard* de informação acerca das propostas

Na Real Vida Seguros, durante os anos de 2015 a 2020 foram registadas um total de 74497 propostas (US13 que se encontra na subsecção 4.1.1.3). Existe um total de 68323 propostas em que os herdeiros legais são beneficiários e 52744 propostas em que o tomador do seguro é beneficiário.

Quanto à forma de pagamento, o débito direto, foi escolhido maioritariamente tanto no pagamento do primeiro recibo (67.92%) como nos seguintes (94.95%), uma vez que permite assegurar pagamentos recorrentes, garantindo que na data contratada é feito o pagamento, sem atrasos nem preocupações.

No gráfico acerca do tipo de fracionamento, é notável que o fracionamento mensal se destaca, seguindo-se do fracionamento anual, trimestral e único, com proporções semelhantes entre eles.

Cerca de 56433 propostas encontram-se arquivadas, 6487 propostas emitiram apólice e 4045 foram submetidas à área técnica para análise. Por outro lado, 7346 propostas foram anuladas, que é um indício menos favorável. Este gráfico responde ao US14 (subsecção 4.1.1.3) pelo detalhe da situação.

Verificou-se que quase 30% dos tomadores do seguro e da primeira pessoa segura fizeram outro seguro na empresa (tipo cliente alterado). Já na segunda pessoa segura o que prevalece é o nenhum uma vez que o seguro não possui segunda pessoa. No entanto 30% dos casos são novos clientes e 5% são pessoas que já continham um seguro.

A *dashboard* seguinte é muito focado no valor de capital proposto (Figura 69). Este *dashboard* pode ser filtrado por data ou por distrito, sendo possível analisar o capital versus a quantidade de vendas em cada mês, assim é possível verificar que no mês de novembro houve um capital total de 422209,4K³ e também é o mês com mais seguros propostos 6,6K. Apesar de abril e maio contarem com o mesmo número de vendas (4K), há uma diferença de capital com cerca de 32600000K entre eles. Deste também se pode retirar uma análise respondendo ao US21 (subsecção 4.1.1.5) pelo detalhe do capital em cada mês em função do número de vendas.

O capital proposto é cerca de 3771,64M⁴, sendo que o produto com mais capital é o Real Seguro Vida Habitação de 2 cabeças, ou seja, duas pessoas seguradas. Como é possível analisar no gráfico “Capital Total por Produto”, com cerca de 2111,24M tem-se o produto Real Seguro Vida Habitação de 2 cabeças

³ A unidade “K” significa milhares.

⁴ A unidade “M” significa milhões.

segundo-se o produto Real Seguro Vida Habitação de 1 cabeça e o Real Seguro Vida Pleno de 1 cabeça. O produto com menos capital é o Real Vida Super Rendimento.

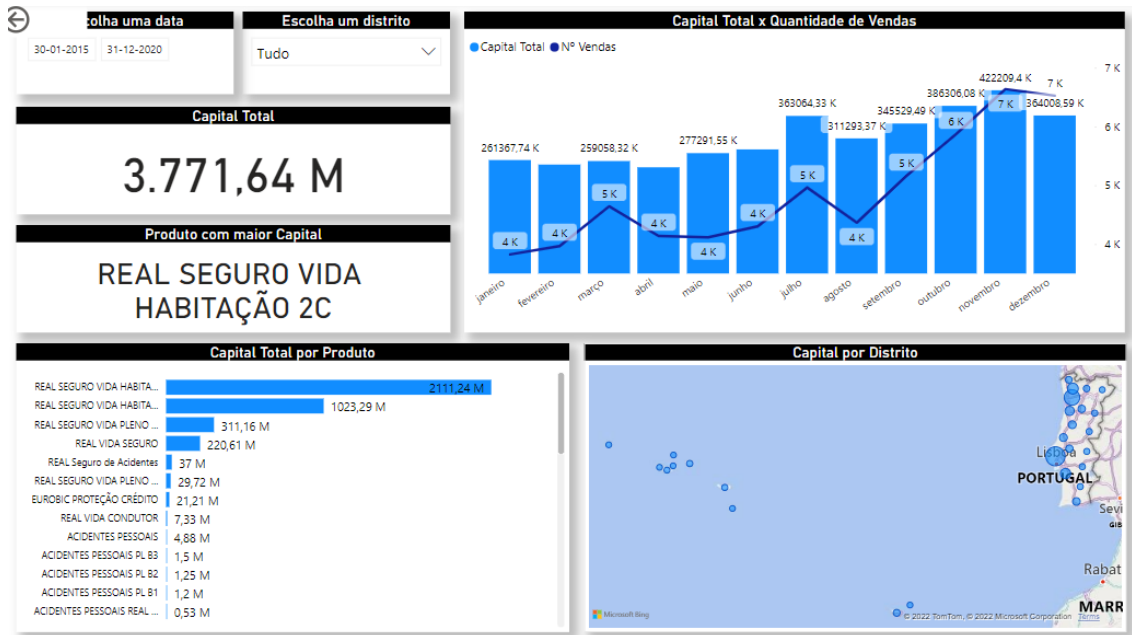


Figura 69 – Dashboard de informação acerca das propostas

Este gráfico, “Capital Total por Produto”, responde ao US22 (subsecção 4.1.1.5) pelo detalhe do produto.

Quanto ao capital por distrito o grande foco continua a ser as grandes cidades de Portugal, ou seja, Lisboa e Porto.

Na Figura 70 é notável um crescimento de capital proposto ao longo dos anos dos diferentes meses, sendo que o ano de 2020 teve um aumento significativo, sendo que este aumento pode dever-se ao facto de grande parte dos contratos ser feito online, ou por se entrar num ano de pandemia (COVID-19) e os clientes querem precaver-se subscrevendo seguros.

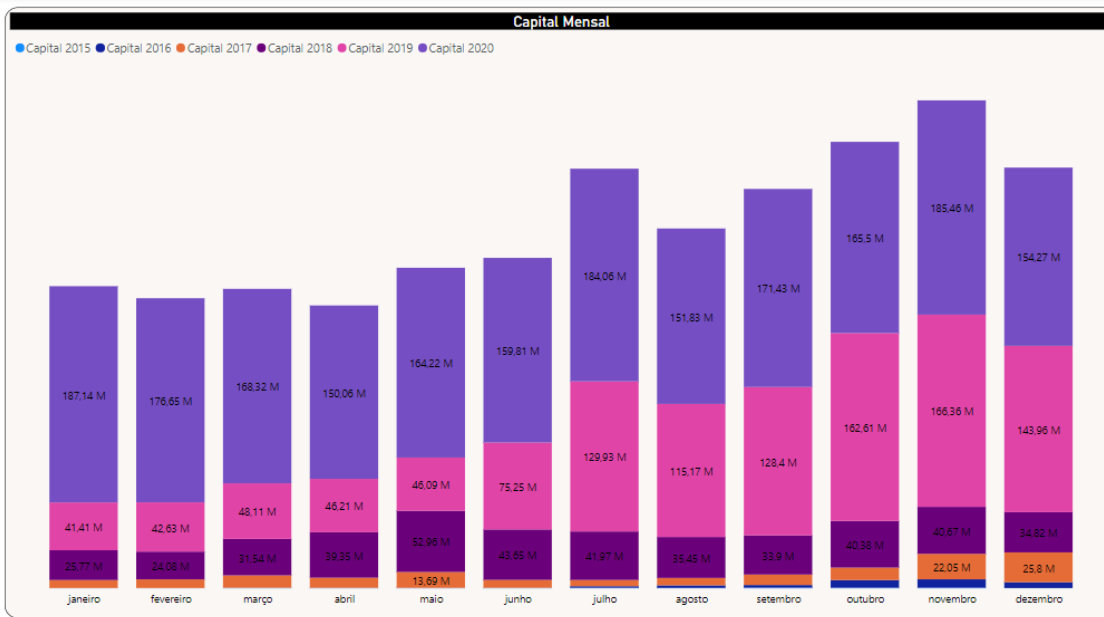


Figura 70 – Relatório de informação acerca das propostas

6.1.4 Informação Apólices

Acerca do resultado das análises aos capitais de morte e de vida nas Apólices foram criados dois *dashboards*, em que a informação pode ser obtida durante um período ou por distrito. Posto isto, foram obtidos dois *dashboards* que se encontram representadas na Figura 71 e Figura 72.

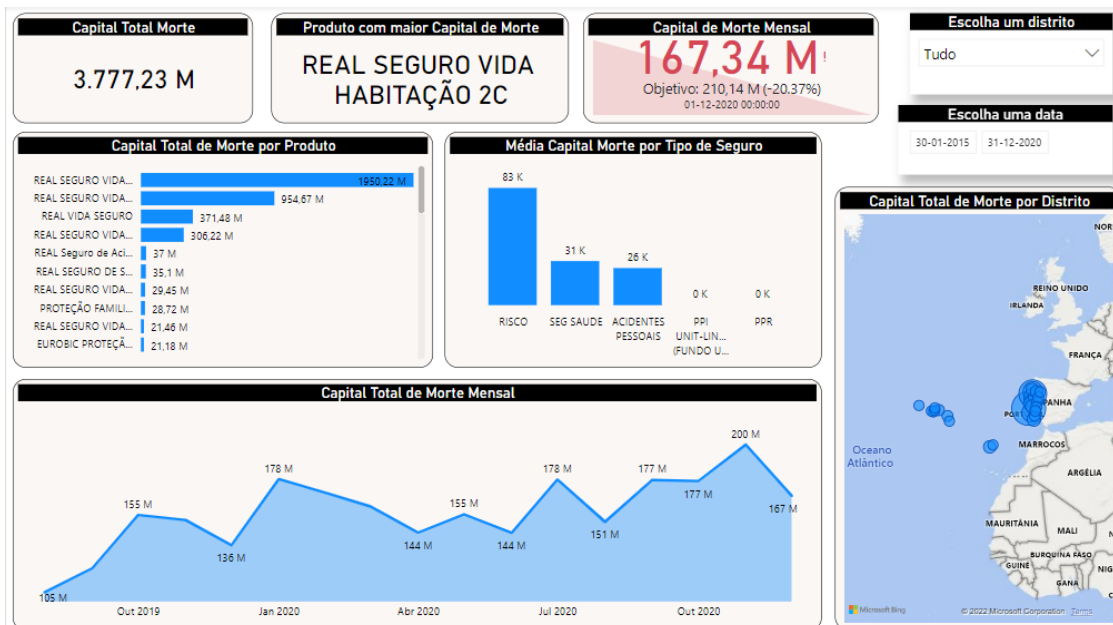


Figura 71 – Dashboard de informação acerca das apólices (capital total em caso de morte)

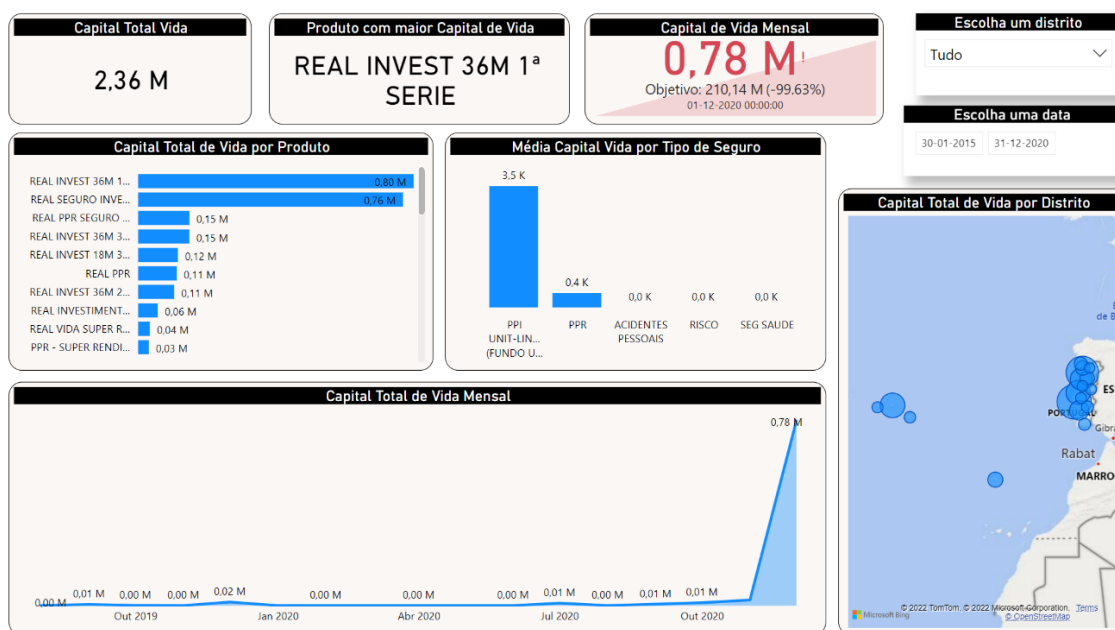


Figura 72 – Dashboard de informação acerca das apólices (capital total em caso de vida)

O primeiro dashboard representa o capital total em caso de morte e o segundo representa o capital total em caso de vida, sendo que para ambos os cenários foram feitos o mesmo tipo de análises.

A maior fonte de capital está nos seguros em caso de morte que conta com cerca de 3777,23M enquanto os seguros em caso de vida apenas contam com 2,36M.

O produto com maior capital em caso de morte é o Real Seguro Vida Habitação de 2 cabeças, sendo possível verificar através do gráfico “Capital Total de Morte por Produto” que os seguros de risco são que tem maior capital deste tipo. Por sua vez, produto com maior capital em caso de vida é o Real Invest 36M 1ª Série, sendo possível verificar que os seguros de fundo de investimento são que tem maior capital deste tipo.

Através da análise do capital por produtos, o tipo de seguro com a média de capital mais elevada em caso de morte é o seguro de risco, enquanto que a média mais elevada em caso de vida é o PPI – Unit-Linked (fundo de investimento). Em ambos os casos os tipos de seguros com maior capital é o inverso do outro, ou seja, no seguro em caso de morte o que prevalece é o tipo risco e o que menos prevalece é o tipo PPI. Já por outro lado, no seguro em caso de vida o que prevalece é o tipo PPI e o que menos prevalece é o tipo risco.

Quanto aos gráficos de KPI (“Capital de Morte Mensal” e “Capital de Vida Mensal”) se estiverem com uma cor verde significa que atingiram o objetivo, se estiverem com uma cor vermelha significa que não atingiram o objetivo, sendo que ambos os gráficos mostram o valor que obtiveram, o esperado e a data. Assim sendo, o objetivo definido foi haver um crescimento de 10% em comparação com o mês anterior. No último mês de registos (dezembro de 2020) não se conseguiu atingir esses objetivos tanto num cenário como no outro.

Na Figura 73 encontram-se outros exemplos de análises, mais precisamente para dar resposta aos user stories US15 (subsecção 4.1.1.3), US27 e US28 (subsecção 4.1.1.8) identificados.

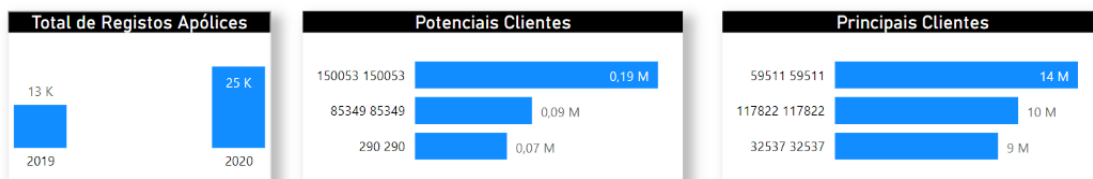


Figura 73 – Outras Análises

Assim sendo, no primeiro gráfico verifica-se que no ano 2020 houve cerca de 25K registos e apenas 13K comparativamente com o ano anterior.

No segundo e terceiro gráfico encontra-se representado o top 3 dos potenciais clientes e dos principais clientes pelo capital, em que o valor da esquerda é o código do titular seguindo-se do valor do capital.

De notar que apenas são apresentadas algumas *dashboards* por motivos de espaço ocupado.

7 Avaliação/Experimentação

Este capítulo é dedicado à descrição das experiências e avaliação realizadas à solução preconizada. Nomeadamente, identificar que grandezas são utilizadas para avaliar o trabalho, que hipótese(s) se pretende(m) testar para suportar os resultados do trabalho, qual a metodologia de avaliação e como se pretende testar essas hipóteses.

Assim sendo, tendo por base o objetivo de avaliar o sistema de BI dedicado à análise de dados de seguros de vida, recorreu-se a um conjunto de grandezas para avaliar a viabilidade da solução implementada. A solução foi avaliada através de inquéritos de satisfação, onde requisitos como a usabilidade e desempenho foram analisados.

7.1 Grandezas de Avaliação

Existem vastas grandezas de avaliação, como por exemplo, tempo, memória, exatidão, satisfação do utilizador, entre outras. Neste projeto as grandezas que se pretende avaliar são a exatidão e a satisfação do utilizador. A exatidão uma vez que o sistema de BI providencia análises de acordo com a realidade, sendo estas uma base para o processo de apoio à decisão, espera-se que os dados sejam exatos. Por outro lado, como será o colaborador a usar o sistema, pretende-se saber a satisfação do mesmo.

7.2 Hipóteses

Os colaboradores pretendem procurar informações que possam ser úteis para o processo de *Business Intelligence* (BI). Por isso, a integração dos dados é um dos momentos mais importantes do BI, pois é nesta fase que a informação é colocada no mesmo formato, de forma a transmitir ao utilizador final a informação útil e de confiança, sem comprometer a tomada de decisão. Além disso, neste momento as ferramentas eliminam registos de baixa qualidade ou que não sejam úteis para a análise, auxiliando a empresa a ter melhores resultados.

O desempenho na obtenção de indicadores é outra hipótese a validar, uma vez que importa entregar os resultados corretos, no tempo certo, para a pessoa certa.

Finalmente, a usabilidade do sistema é outra hipótese a considerar, uma vez que é importante que o acesso e o compartilhamento de informações que darão suporte à decisão sejam processos intuitivos e fáceis, sendo passível que um utilizador com pouco conhecimento na área consiga interagir e analisar os dados.

7.3 Metodologia de Avaliação

A solução desenvolvida, foi avaliada com recurso a dois inquéritos de satisfação, realizados a dois grupos de profissionais distintos: profissionais de IT⁵ e profissionais da seguradora.

O objetivo consistiu em saber a opinião técnica dos profissionais de IT sobre alguns pontos da aplicação e a opinião dos profissionais da seguradora acerca da forma como os dados lhes são apresentados.

⁵ *Information technology.*

Desta forma, os inquéritos de satisfação visam retratar pontos de vista diferentes em relação à aplicação.

O grupo de profissionais de IT que responderam ao inquérito, possuem vários anos de experiência em sistemas de informação e já tiveram oportunidade de trabalhar na área dos seguros. Desta forma, conseguem avaliar o sistema e perceber se foram alcançadas as necessidades dos sistemas das seguradoras.

Apesar desta dissertação se tratar de um trabalho académico, torna-se importante a opinião dos profissionais da seguradora já que estes são os utilizadores aos quais a aplicação se destina, caso a mesma fosse disponibilizada em ambiente real. Melhor do que ninguém, os profissionais da seguradora conseguem ter uma maior capacidade analítica acerca da utilidade deste tipo de aplicação, dado as circunstâncias que enfrentam no dia a dia.

Ambos os inquéritos foram disponibilizados e realizados online, dado ser uma forma rápida de fazer chegar os mesmos aos profissionais e também de tornar intuitiva a forma como a informação é apresentada. Para além disso, é uma forma de aproveitar o potencial de ferramentas online que permitem de imediato obter a avaliação acerca das respostas obtidas.

Em ambos os inquéritos de satisfação, foi utilizado um conjunto de afirmações as quais foram classificadas, segundo a escala Likert, para especificar o nível de concordância com as mesmas.

A escala de Likert mede a satisfação do utilizador segundo níveis de classificação. Na solução implementada foram aplicados cinco níveis de satisfação, de forma a ser possível medir o comportamento dos utilizadores relativamente à concordância com as afirmações, a avaliação do serviço prestado e a frequência de utilização. Nos inquéritos de satisfação foi utilizada a escala que se encontra na Tabela 30.

Tabela 30 – Escala de Likert

Escala	Descrição
1	Discordo Totalmente
2	Discordo
3	Neutro
4	Concordo
5	Concordo Totalmente

O inquérito de satisfação realizado aos profissionais de IT pode visualizar-se no Anexo III, enquanto que o inquérito realizado aos Profissionais da Seguradora, acerca da avaliação da solução é apresentado no Anexo IV.

7.4 Resultado dos Inquéritos

Os inquéritos são usados para recolher uma série de dados homogéneos que podem ser usados para efeitos estatísticos, assim com o intuito de avaliar os resultados obtidos nos inquéritos, recorre-se a testes estatísticos.

A escala Likert é um tipo de escala de classificação utilizada para medir atitudes ou opiniões. Com esta escala, os inquiridos são solicitados a classificar os itens num nível de acordo.

Nos inquéritos é definido um conjunto de afirmações as quais possuem diferentes classificações, segundo a escala Likert, para especificar o nível de concordância com as mesmas.

Nos inquéritos elaborados foram aplicados cinco níveis de classificação, pretendendo-se medir o comportamento dos utilizadores relativamente à concordância com as afirmações, à avaliação do serviço prestado e à frequência de utilização.

Os inquéritos de satisfação realizados contaram com a participação de 20 pessoas, os quais são analisados nos pontos seguintes.

7.4.1 Profissionais de IT

O inquérito de satisfação realizado ao grupo de profissionais de IT contou com a participação de 5 pessoas - de três funções diferentes, uma de desenvolvimento, dois analistas de dados (responsável pelo AS400 da empresa) e dois de infra-estruturas - as quais tinham de responder a um conjunto de quatro questões. De acordo com as opiniões obtidas, em geral, conseguiu-se resultados bastante positivos, os quais podem ser visualizados graficamente.

Segundo os resultados obtidos, na primeira questão (Figura 74) houve 3 participantes que seleccionaram o nível de concordância 4 (Concordo) e outros 2 que seleccionaram o nível 3 (Neutro) e o nível 5 (Concordo Totalmente). Posto isto, a maioria está de acordo de que a arquitetura do sistema é adequada ao negócio.

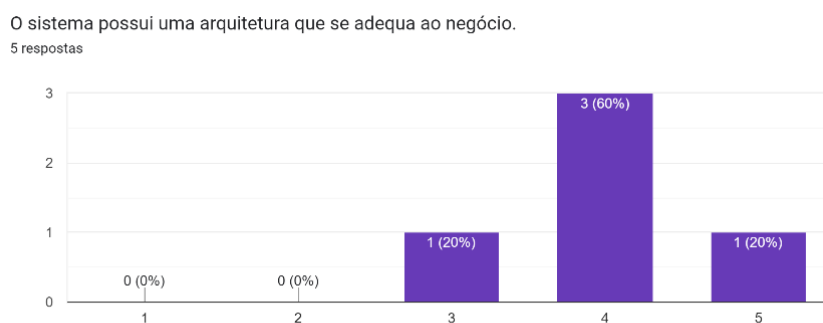


Figura 74 – 1ª Questão do Inquérito de Satisfação (IT)

Na segunda questão (Figura 75), 3 participantes responderam 5 (Concordo Totalmente), refletindo a opinião da maioria, 1 participante respondeu 3 (Neutro) e outro participante respondeu 4 (Concordo). Assim sendo, verifica-se que a maioria concorda que a solução encontra-se preparada para necessidades futuras, nomeadamente para mostrar novos indicadores.

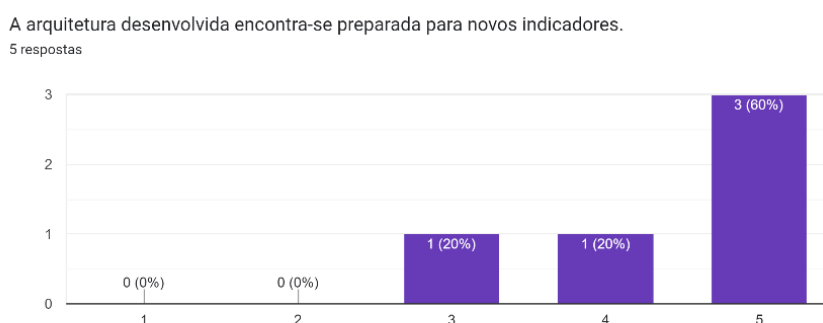


Figura 75 – 2ª Questão do Inquérito de Satisfação (IT)

Na terceira questão (Figura 76), 4 participantes mostraram concordância, pelo que responderam 4 (Concordo) e 1 respondeu 5 (Concordo Totalmente). Com isto, também estão de acordo que o sistema está construído de forma a integrar facilmente com outras aplicações.

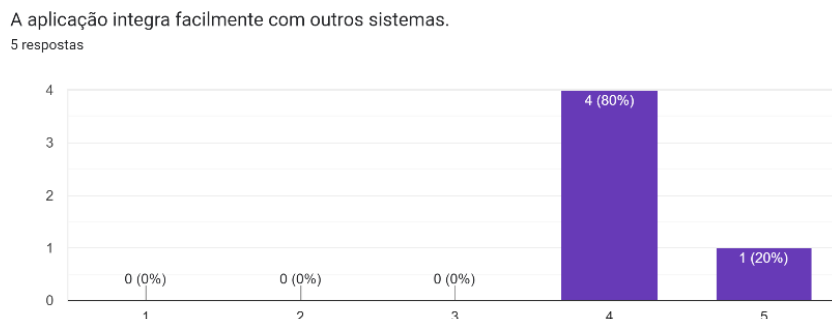


Figura 76 – 3ª Questão do Inquérito de Satisfação (IT)

Por último, e quarta questão (Figura 77), 3 participantes responderam 5 (Concordo Totalmente), refletindo assim a maioria e 2 participantes responderam 4 (Concordo). Esta questão evidencia que o tempo de resposta da aplicação é adequado.

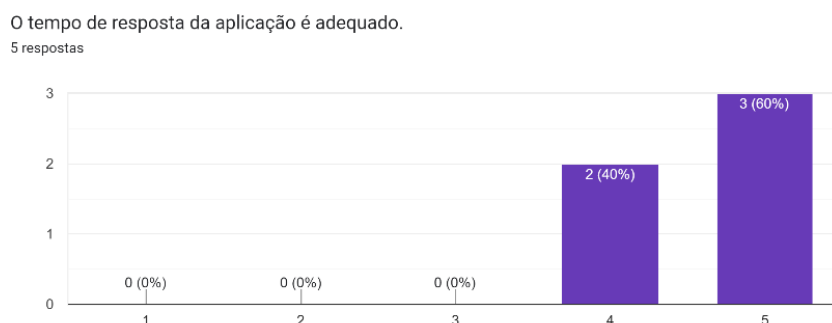


Figura 77 – 4ª Questão do Inquérito de Satisfação (IT)

7.4.2 Profissionais da Seguradora

O inquérito de satisfação realizado ao grupo dos profissionais da seguradora contou com o contributo de 15 participantes, sendo este grupo representado por diversas áreas distintas da empresa. Este grupo de inquiridos também se mostrou satisfeito com o resultado da aplicação. De seguida pode-se consultar os resultados obtidos graficamente para cada uma das questões efetuadas.

De acordo com a análise dos resultados, na primeira questão (Figura 78), 8 participantes classificaram a afirmação como 5 (Concordo Totalmente), enquanto 6 participantes escolheram o nível 4 (Concordo) e ainda 1 escolheu o nível 3 (Neutro). Desta forma conclui-se que a maioria considerou que a aplicação é útil e intuitiva.

A aplicação mostra-se útil e intuitiva.

15 respostas

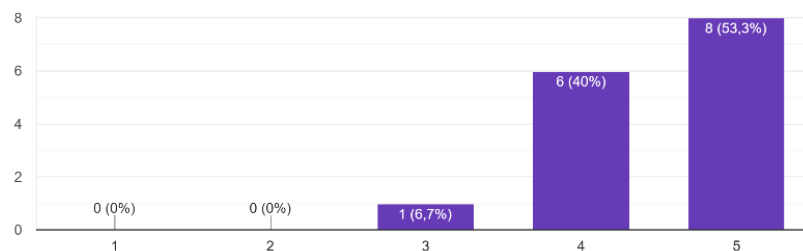


Figura 78 – 1ª Questão do Inquérito de Satisfação (Profissionais da Seguradora)

Na segunda questão (Figura 79), 8 participantes escolheram 4 (Concordo) e 7 escolheram o nível 5 (Concordo Totalmente). Desta forma, os inquiridos concordaram que os filtros se adequam ao tema em estudo.

Os filtros adequam-se ao tema em estudo.

15 respostas

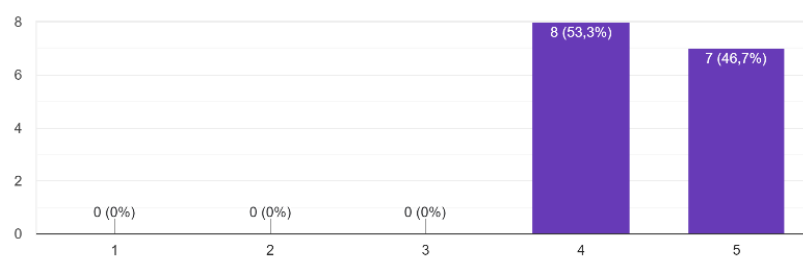


Figura 79 – 2ª Questão do Inquérito de Satisfação (Profissionais da Seguradora)

Na terceira questão (Figura 80), 2 participantes selecionaram o nível 3 (Neutro), 6 escolheram o nível 4 (Concordo) e os restantes 7 escolheram nível 5 (Concordo Totalmente). Assim, e mais uma vez, os inquiridos, concordam que a forma como a informação foi disponibilizada é uma mais valia, torna-se de fácil leitura e análise.

A informação encontra-se disponibilizada de forma a facilitar a leitura e a análise dos dados.

15 respostas

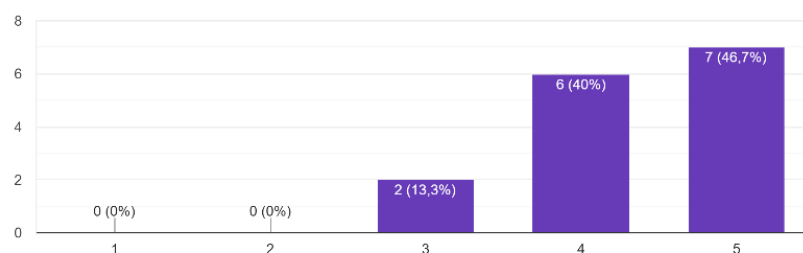


Figura 80 – 3ª Questão do Inquérito de Satisfação (Profissionais da Seguradora)

Na quarta questão (Figura 81), 2 participantes selecionaram o nível 3 (Neutro), 5 escolheram o nível 4 (Concordo) e os restantes 8 escolheram nível 5 (Concordo Totalmente). Assim, os inquiridos, concordam que a forma como a informação foi disponibilizada é preferível quando comparada a relatórios estáticos.

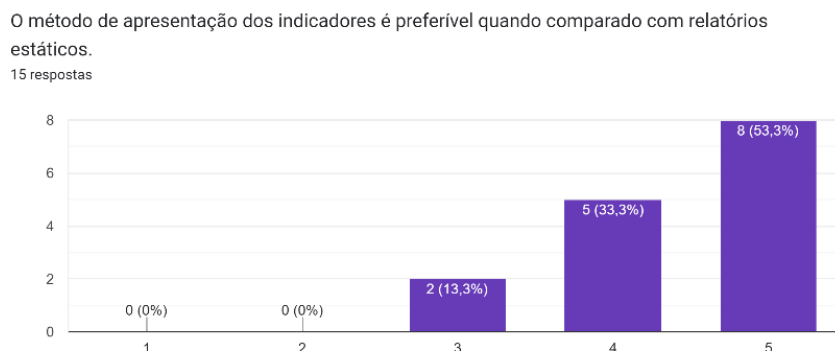


Figura 81 – 4ª Questão do Inquérito de Satisfação (Profissionais da Seguradora)

Na quinta questão (Figura 82), 9 participantes selecionaram o nível 5 (Concordo Totalmente), 1 participante selecionou o nível 3 (Neutro) e ainda houve 5 que selecionaram o nível 4 (Concordo). Desta forma, a maior parte dos inquiridos são da opinião de que a aplicação constitui uma boa ferramenta para exploração de dados.

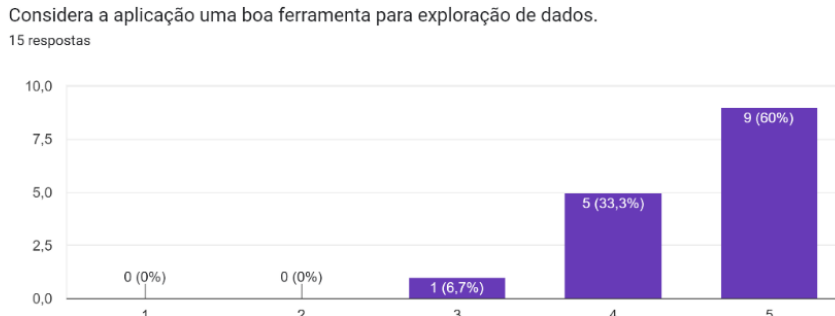


Figura 82 – 5ª Questão do Inquérito de Satisfação (Profissionais da Seguradora)

Na última, e sexta questão (Figura 83), 9 dos participantes selecionaram o nível 4 (Concordo) e 6 selecionaram o nível 5 (Concordo Totalmente). Assim, segundo os inquiridos a aplicação desenvolvida cumpriu com os objetivos a atingir.

Considera que a solução desenvolvida foi conseguida de acordo com os objectivos a atingir.

15 respostas

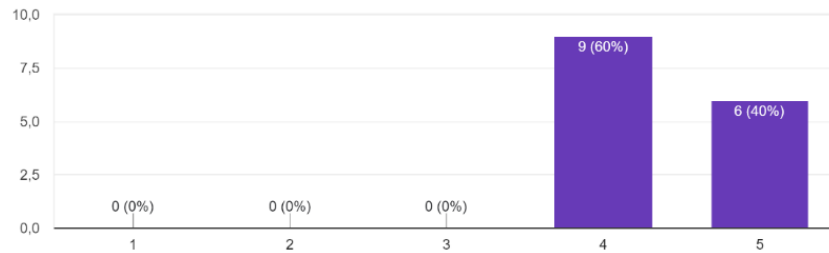


Figura 83 – 6ª Questão do Inquérito de Satisfação (Profissionais da Seguradora)

7.5 Comentários Finais

Os inquéritos de satisfação permitem avaliar a solução segundo dois grupos diferentes de profissionais. Um dos grupos, os profissionais de IT, avaliaram a solução de acordo com questões mais técnicas enquanto que o outro grupo, os profissionais da seguradora, avaliaram a solução na perspetiva de utilizador final.

Com base nas respostas obtidas, e ainda que a amostra tenha sido bastante pequena, conclui-se que o grau de satisfação foi elevado, refletindo-se através de resultados bastante positivos.

Os resultados, evidenciaram que a solução tem interesse e poderia ser útil, caso fosse disponibilizada em ambiente real.

Ainda assim, os inquéritos de satisfação deveriam ter sido disponibilizados a um maior número de pessoas, no entanto, considera-se que o facto de o sistema não ter sido disponibilizado num ambiente real é uma das causas para tal.

8 Conclusão

Neste capítulo são descritas todas as conclusões obtidas ao longo desta dissertação. Para além dos objetivos alcançados, são mencionados os problemas/limitações que existiram ao longo da mesma e por fim, é apresentado o trabalho futuro que ainda se poderia desenvolver, tendo por base o contexto desta dissertação.

8.1 Objetivos Alcançados

Esta dissertação teve como objetivo o desenvolvimento de um sistema de BI que permitisse a gestão de informação acerca de Dados de Seguros de Vida e que fosse capaz de fornecer toda a informação necessária aos profissionais da seguradora, de forma a apoiar no processo de tomada de decisão relativamente ao tema em questão.

O caso de estudo, parte de uma instituição de seguros que se responsabiliza por toda a gestão e monitorização de informação acerca dos Dados de Seguros de Vida e que tem vindo a mostrar grande importância a nível Nacional.

Assim sendo, na secção 1.3, definiu-se um conjunto de objetivos que serão detalhados de seguida.

Relativamente ao primeiro objetivo, “Definir o modelo conceptual de dados (modelo dimensional) que suportará o armazenamento dos dados no armazém de dados”, foi alcançado tal como descrito na secção 4.2.4.

Quanto ao segundo objetivo, “Extrair os dados a partir da base de dados que suporta o sistema operacional”, o sistema de BI possui uma arquitetura extensível, de forma a estar preparado para situações futuras que possam advir (Ex: outras fontes de dados diferentes de Access), disponibilizando toda a informação segundo um único padrão e capaz de armazenar histórico, desta forma, o objetivo em causa também foi alcançado.

Em relação ao terceiro objetivo, “Aplicar técnicas de transformação e limpeza de dados, de modo a remover problemas de qualidade existentes nos dados, respetivamente, ao nível do esquema dos dados e ao nível das instâncias (i.e., dos valores)”, foi alcançado uma vez que se aplicaram técnicas de transformação e limpeza de dados, de modo a remover problemas de qualidade existentes nos dados (secção 5.2) e procedeu-se ao carregamento dos dados no armazém de dados, sendo este o quarto objetivo, “Proceder ao carregamento dos dados no armazém de dados, manipulando adequadamente o histórico de alterações que ocorram nos dados”.

Quanto ao quinto objetivo, “Criar um cubo de dados que suporte a realização de análises de dados”, foi alcançado através da ferramenta de Power BI que serve exatamente para esse propósito.

Por último, em relação ao sexto objetivo, “Realizar análises diversificadas aos dados e produzir dashboards com os indicadores chave de desempenho relevantes para cada departamento”, foi também alcançado com a execução de *dashboards* e relatórios acerca de titulares, de produtos, de potenciais apólices realizadas e de apólices.

Através dos *dashboards* e relatórios implementados e a forma como estes são apresentados consegue-se obter conclusões de imediato, que caso contrário seria impossível.

Por fim, como forma de validar se os requisitos funcionais mencionados nesta dissertação foram cumpridos, foram realizados inquéritos de satisfação a uma amostra de profissionais de IT e profissionais da seguradora.

Os inquéritos disponibilizados visam obter a opinião relativamente à usabilidade do sistema, ao grau de satisfação relativo ao tempo de resposta, ao cumprimento dos requisitos e à importância dos indicadores implementados.

Através dos inquéritos de satisfação efetuados aos profissionais de IT e aos profissionais da seguradora, ambos os grupos mostram opiniões bastante positivas, que refletem que os objetivos a atingir com a aplicação foram conseguidos.

8.2 Problemas/Limitações

Um dos problemas que se verificou logo à partida nesta dissertação são os dados. A informação disponibilizada não era atual, pelo que o sistema teria mais interesse caso o Data Mart pudesse ser abastecido diretamente a partir do sistema operacional da organização, de forma a se poder obter os dados mais atualizados e não só referentes a um determinado período temporal.

Para além disso, como a dissertação é em âmbito académico, houve a necessidade de desenvolver um capítulo que servisse de auxílio para entender algumas informações ao nível das regras de negócio sobre os Dados de Seguros de Vida. Para além disso, houve dificuldade em encontrar profissionais da seguradora que pudessem analisar a aplicação com base nas *dashboards* desenvolvidas e responder aos questionários, uma vez que apesar de serem da área, não estão habituados a ver/analisar este tipo de ferramentas.

8.3 Trabalho Futuro

Os sistemas de informação contribuem com melhorias para solucionar problemas que têm grande impacto no quotidiano das pessoas.

Na área dos seguros, estes são fundamentais porque procuram dar resposta imediata a problemas que até então requerem grandes esforços para solucionar.

Nesta dissertação ainda existe um conjunto de trabalhos que podem ser realizados, para completar o trabalho até então desenvolvido. Destaca-se a adição de novos indicadores relativamente ao tema em estudo, “Dados de Seguros de Vida”, de forma a justificar os motivos que levam os titulares a adquirir certos produtos, perceber quais os produtos que vale a pena investir e publicitar mais e quais as alturas do ano em que são necessários mais produtos e os respetivos motivos.

Ficaria também para trabalho futuro, adaptar o sistema da empresa para este novo módulo, desenvolvido no âmbito deste trabalho, de forma a ser possível criar um novo data Mart com estes dados e desenvolver novos indicadores relativamente à área dos seguros de vida.

Referências

Advogados, N. (2014)	<i>O Ramo “Vida”</i> . Disponível em: https://www.guiadosseguros.pt/o-ramo-vida.html?fbclid=IwAR0o7SONa3ooP3POdcS-X3-nXCEoYMS4-IOKESp93o1nCLDeg2-o83CNHPs . Acedido em: janeiro 2021.
ASF. (2015)	<i>Guia de Seguros e Fundos de Pensões</i> . Disponível em: https://www.asf.com.pt/NR/exeres/DC513DEB-5E88-436A-8301-9D1B8E0C27D2,frameless.htm?NRMODE=Published&fbclid=IwAR00nENadUz6KPJOKPkiS6vvPk9wMYL_OXNnbt8JafGstk7xImPdzzbzL04 . Acedido em: janeiro 2021.
ASF. (2021a)	<i>Apoio ao Consumidor</i> . Disponível em: https://www.asf.com.pt/NR/exeres/9A06CE14-CA54-4FDB-8FC9-DA5939D30769.htm?fbclid=IwAR0B9HUOS_OKC76c3q9sCESvrlFB7BFzLIXG6phDO6YMs0Xq7thH0Cgy0b8 . Acedido em: janeiro 2021.
ASF. (2021b)	<i>Exploração de ramos Vida e Não Vida</i> . Disponível em: https://www.asf.com.pt/NR/exeres/C12B6BE5-2A94-4873-BE64-63565558FA6F.htm?fbclid=IwAR2UKlfJMuNBiK95GPEUINI_GcFe_6oj0a0fvCROeddC0FrxiGhdCny1B4 . Acedido em: janeiro 2021.
Berlin, S. D. (2012)	<i>Visual Business Modeling Service Design Drinks Berlin</i> . Disponível em: https://www.slideshare.net/ServiceDesignBerlin/visual-business-modeling-service-design-drinks-berlin . Acedido em: janeiro 2020.
Breslin, M. (2004)	<i>Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models</i> . Disponível em: http://www.tdwi.org/research/display.aspx?ID=6991 . Acedido em: janeiro 2020.
Brokers, i2S. (2020)	<i>BI para Seguros</i> . Disponível em: https://www.i2sbrokers.pt/solution/?id=6 . Acedido em: janeiro 2020.
Carvalho, A.I., Lafuente, R. (2021)	<i>Códigos postais — Central de Dados</i> . Disponível em: http://centraldedados.pt/codigos_postais/ . Acedido em: fevereiro 2021.
Casimiro, N. (2011)	<i>Tipos de seguros. Ramo vida e ramo não vida</i> . Disponível em: https://investidor.pt/tipos-de-seguros-ramo-vida-e-ramo-nao-vida/ . Acedido em: janeiro 2021.
CCG. (2019)	<i>Business Intelligence: Um bilhete de identidade das empresas?</i> . Disponível em: http://www.ccg.pt/business-intelligence-empresas/ . Acedido em: janeiro 2020.
Chugugrace. (2018)	<i>SQL Server Integration Services</i> . Disponível em: https://docs.microsoft.com/en-us/sql/integration-services/sql-server-integration-services . Acedido em: janeiro 2020.
Chugugrace. (2019)	<i>Lesson 4-5: Test the Lesson 4 package</i> . Disponível em: https://docs.microsoft.com/en-us/sql/integration-services/lesson-4-5-testing-the-lesson-4-tutorial-package . Acedido em: janeiro 2020.
Coepd. (2014)	<i>What is FURPS+?</i> . Disponível em: https://businessanalystraininghyderabad.wordpress.com/2014/08/05/what-is-furps/ . Acedido em: janeiro 2021.
Computerworld. (2014)	<i>Lusitania Seguros renova sistema de BI</i> . Disponível em: https://www.computerworld.com.pt/2014/01/29/lusitania-seguros-renova-sistema-de-bi/ . Acedido em: janeiro 2020.

EDUCBA. (2020)	<i>Galaxy Schema</i> . Disponível em: https://www.educba.com/galaxy-schema/ . Acedido em: janeiro 2020.
Eeles, P. (2005)	<i>Capturing Architectural Requirements</i> . Disponível em: http://www.ibm.com/developerworks/rational/library/4706.html . Acedido em: março 2020.
IBM. (2014a)	<i>The fact table</i> . Disponível em: www.ibm.com/support/knowledgecenter/sk/ssgu8g_11.50.0/com.ibm.whse.doc/ids_ddi_345.htm . Acedido em: janeiro 2020.
IBM. (2014b)	<i>Dimension tables</i> . Disponível em: www.ibm.com/support/knowledgecenter/sk/ssgu8g_11.50.0/com.ibm.whse.doc/ids_ddi_348.htm . Acedido em: janeiro 2020.
IBM. (2014c)	<i>Concepts of dimensional data modeling</i> . Disponível em: www.ibm.com/support/knowledgecenter/sk/ssgu8g_11.50.0/com.ibm.whse.doc/ids_ddi_350.htm . Acedido em: janeiro 2020.
Inmon. (2005)	<i>Building the Data Warehouse</i> . W. H. Inmon Wiley (4th Edition), 2005.
Inmon. (2003)	<i>Mastering Data Warehouse Design: Relational and Dimensional Techniques</i> . Claudia Imhoff, Nicholas Gallemmo, Jonathan G. Geiger Wiley, 2003.
Gartner. (2019)	<i>Magic Quadrant for Data Integration Tools</i> . Disponível em: https://b2bsalescafe.files.wordpress.com/2019/09/gartner-magic-quadrant-for-data-integration-tools-august-2019.pdf . Acedido em: janeiro 2020.
Gartner. (2020a)	<i>2020 Gartner Magic Quadrant for Data Integration Tools</i> . Disponível em: https://www.bmc.com/blogs/gartner-magic-quadrant-data-integration-tools/ . Acedido em: março 2021.
Gartner. (2020b)	<i>Gartner Magic Quadrant for Analytics and Business Intelligence Platforms</i> . Disponível em: https://www.gartner.com/en/documents/2019/2/magic-quadrant-for-analytics-and-business-intelligence-3980852 . Acedido em: março 2021.
Group, K. (2020a)	<i>Enterprise Data Warehouse Bus Architecture</i> . Disponível em: https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/kimball-data-warehouse-bus-architecture/ . Acedido em: janeiro 2020.
Group, K. (2020b)	<i>Star Schemas and OLAP Cubes</i> . Disponível em: https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/star-schema-olap-cube/ . Acedido em: janeiro 2020.
Group, K. (2020c)	<i>Four-Step Dimensional Design Process</i> . Disponível em: https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/four-4-step-design-process/ . Acedido em: janeiro 2020.
Kimball R. and Ross M. (2013)	<i>The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modelling</i> . Ralph Kimball, Margy Ross, Wiley (3rd Edition), 2013.
Koen P. A. (2002)	<i>Fuzzy Front End: Effective Methods, Tools, and Techniques</i> . Em: Bellviav, P., Griffin, A. and Somemeyer, S. (eds.) - The PDMA Tool-Book for New Product Development, p. 5–36. issn: 1876-2859. doi: 10.1021/ed072p378.

KPI.org. (2020)	<i>What is a Key Performance Indicator (KPI)?</i> . Disponível em: https://kpi.org/KPI-Basics . Acedido em: janeiro 2020.
Lawrence D. M. (1961)	<i>Techniques of Value Analysis and Engineering</i> . Lawrence D. Miles (1 st Edition), 1961.
McHugh, J. (2017)	<i>Data Warehouse Design Techniques - Slowly Changing Dimensions</i> . Disponível em: https://www.nuwavesolutions.com/slowly-changing-dimensions/ . Acedido em: março 2021.
Microsoft. (2019)	<i>Descrição geral dos cubos OLAP do Service Manager para análise avançada</i> . Disponível em: https://docs.microsoft.com/pt-pt/system-center/scsm/olap-cubes-overview?view=sc-sm-2019 . Acedido em: janeiro 2020.
Microsoft. (2020a)	<i>What is Power BI?</i> . Disponível em: https://powerbi.microsoft.com/en-us/what-is-power-bi/ . Acedido em: janeiro 2020.
Microsoft. (2020b)	<i>Why Power BI</i> . Disponível em: https://powerbi.microsoft.com/en-us/why-power-bi/ . Acedido em: janeiro 2020.
MicroStrategy. (2020a)	<i>A platform to power the Intelligent Enterprise</i> . Disponível em: https://www.microstrategy.com/us/product/enterprise-platform . Acedido em: janeiro 2020.
MicroStrategy. (2020b)	<i>The Intelligent Enterprise starts with you</i> . Disponível em: https://www.microstrategy.com/us/get-started/desktop . Acedido em: janeiro 2020.
MSBI. (2017)	<i>SSIS Dimensions, Fact Table, Star Schema and Snowflake - Theory Part 1</i> . Disponível em: http://www.learnmsbitutorials.net/ssis-dimension-fact-table-snowflake-star-schema-part1.php . Acedido em: janeiro 2020.
Naeem, T. (2020)	<i>Data Warehouse Concepts: Kimball vs. Inmon Approach</i> . Disponível em: https://www.astera.com/type/blog/data-warehouse-concepts/ . Acedido em: março 2021.
Nicola et al. (2012)	Susana Nicola, Eduarda Pinto Ferreira, and J. J. Pinto Ferreira, International Journal of Information Technology & Decision Making 2012 11:03, p. 661-703
Oracle. (2020a)	<i>Oracle Data Integrator</i> . Disponível em: https://www.oracle.com/middleware/technologies/data-integrator.html . Acedido em: janeiro 2020.
Oracle. (2020b)	<i>High-Performance Bulk Data Movement and Data Transformation</i> . Disponível em: https://www.oracle.com/in/middleware/data-integration/enterprise-edition/ . Acedido em: janeiro 2020.
Osterwalder A, Pigneur Y. (2010)	<i>Business Model Generation</i> . Disponível em: http://www.businessmodelgeneraJon.com/ . Acedido em: janeiro 2020.
Osterwalder et al. (2014)	Osterwalder, Alexander; Pigneur, Yves; Bernarda, Greg; Smith, Alan. Trad. Bruno Alexander, <i>Value Proposition Design: Como construir propostas de valor inovadoras</i> .
Patrocínio, N. (2019)	<i>O que é um seguro? Que tipos de seguro existem?</i> . Disponível em: https://www.rankia.pt/seguros/o-que-e-um-seguro-que-tipos-de-seguro-existem/ . Acedido em: janeiro 2021.

Porter. (1989)	<i>Vantagem competitiva: criando e sustentando um desempenho superior.</i> Tradução de Elizabeth Maria de Pinho Braga. Rio de Janeiro: Elsevier, 1989.
QlikTech. (2020)	<i>Data Analytics Platform.</i> Disponível em: https://www.qlik.com/us/products/qlik-sense . Acedido em: janeiro 2020.
SAS. (2020a)	<i>Gartner positions SAS as a Leader in the 2019 Magic Quadrant for Data Integration Tools.</i> Disponível em: https://www.sas.com/pt_pt/news/analyst-viewpoints/gartner-magic-quadrant-data-integration-tools.html . Acedido em: janeiro 2020.
SAS. (2020b)	<i>Support / Samples & SAS Notes.</i> Disponível em: http://support.sas.com/kb/53/323.html . Acedido em: janeiro 2020.
Serafim, A. (2013)	<i>O Modelo de Cadeia de Valor de Michael Porter.</i> Disponível em: https://www.portal-gestao.com/artigos/6991-o-modelo-de-cadeia-de-valor-de-michael-porter.html . Acedido em: janeiro 2020.
TABLEAU. (2020a)	<i>Products.</i> Disponível em: https://www.tableau.com/products#video . Acedido em: janeiro 2020.
TABLEAU. (2020b)	<i>Self-service analytics in the cloud with Tableau Online.</i> Disponível em: https://www.tableau.com/products/cloud-bi . Acedido em: janeiro 2020.
Talend. (2020a)	<i>2019 Gartner Magic Quadrant for Data Integration Tools.</i> Disponível em: https://www.talend.com/resources/2019-gartner-magic-quadrant-data-integration-tools/ . Acedido em: janeiro 2020.
Talend. (2020b)	<i>Open Studio for ESB.</i> Disponível em: https://www.talend.com/products/application-integration/esb-open-studio/ . Acedido em: janeiro 2020.
Technologies, F. (2014)	<i>Top 3 Data Warehousing Benefits for Companies.</i> Disponível em: http://www.fountainheadtechnologies.com/newsroom/56-top-3-data-warehousing-benefits-for-companies . Acedido em: janeiro 2020.
Tecnologia, F. (2018)	<i>Business Intelligence (BI): Gerencie sua corretora de seguros com alguns cliques.</i> Disponível em: https://solucoesparacoretoras.quiver.net.br/business-intelligence-bi/ . Acedido em: janeiro 2020.
TI, K. (2015)	<i>Business Intelligence para Seguradoras.</i> Disponível em: http://kalti.com.br/business-intelligence-cockpit/seguradoras/ . Acedido em: janeiro 2020.
Tobin, D. (2020)	<i>ETL & Data Warehousing Explained: ETL Tool Basics.</i> Disponível em: https://www.xplenty.com/blog/etl-data-warehousing-explained-etl-tool-basics/?fbclid=IwAR1dCe0VvLc2ZAf07-19ndsnBH6OdjPuknWTAcVIL2GZdm4p7BYgRg1psb0/ . Acedido em: março 2021.
Vargas, R. V. (2010)	<i>Using the analytic hierarchy process (ahp) to select and prioritize projects in a portfolio.</i> Paper presented at PMI® Global Congress 2010—North America, Washington, DC. Newtown Square, PA: Project Management Institute. Disponível em: https://www.pmi.org/learning/library/analytic-hierarchy-process-prioritize-projects-6608 . Acedido em: março 2021.
Vida Seguros, R. (2021a)	<i>Glossário.</i> Disponível em: https://www.realvidaseguros.pt/glossario?fbclid=IwAR3ozJR3_d-cNlaLQ-qrB6FA5e-J6k87pDwFZGUXyo9SmiGtwG6Fwp6n3dc . Acedido em: janeiro

	2021.
Vida Seguros, R. (2021b)	<i>Porque devo fazer um Seguro de Vida?</i> Disponível em: https://www.realvidaseguros.pt/particulares/seguros-de-vida?fbclid=IwAR1v6WpQXyV8u-daxgXyGeZPCIYO7QS-yZSNZaFPPjTO92_LdrWi1AJV_gs . Acedido em: janeiro 2021.
Woodall, T. (2003)	<i>Conceptualising 'Value for the Customer': An Attributional, Structural and Dispositional Analysis</i> . Academy of Marketing Science Review. 12. Disponível em: https://www.researchgate.net/publication/228576532_Conceptualising_'Value_for_the_Customer'_An_Attributional_Structural_and_Dispositional_Analysis . Acedido em: janeiro 2020.
Wynstra. (2005)	<i>The concept of perceived value</i> . Disponível em: https://www.researchgate.net/publication/237937875_The_concept_of_perceived_value_A_systematic_review_of_the_research . Acedido em: janeiro 2020.
Zeithmal. (1988, p14)	<i>The concept of perceived value</i> . Disponível em: https://www.researchgate.net/publication/237937875_The_concept_of_perceived_value_A_systematic_review_of_the_research . Acedido em: janeiro 2020.
ZenTut. (2020)	<i>Data Mart</i> . Disponível em: https://www.zentut.com/data-warehouse/data-mart/ . Acedido em: janeiro 2020.

Anexo I

Neste anexo encontra-se o esquema de dados das restantes tabelas a criar na *Staging Area*.

Tabela 31 – Tabela MODA00

Tabela MODA00			
Atributo/Campo	Descrição	Tipo de Dados	Tamanho
MO\$MOD	Código da Modalidade	numeric	(2,0)
MONVER	Versão	numeric	(2,0)
MOGRUP	Grupo	numeric	(2,0)
MODESI	Designação Abreviada	nvarchar	(8)
MODESC	Descrição	nvarchar	(30)
MOTSEG	Tipo de Seguro	nvarchar	(1)
MOPSEG	Pessoa Segura	nvarchar	(1)
MODURC	Duração do Contrato	nvarchar	(1)
MOTCCR	Tipo de Contrato	nvarchar	(1)
MOVIMO	Tipo de Modalidade	nvarchar	(1)
MOAVIN	Ano Início da Versão	numeric	(4,0)
MOMVIN	Mês Início da Versão	numeric	(2,0)
MODVIN	Dia Início da Versão	numeric	(2,0)
MOAVFI	Ano Fim da Versão	numeric	(4,0)
MOMVFI	Mês Fim da Versão	numeric	(2,0)
MODVFI	Dia Fim da Versão	numeric	(2,0)

Tabela 32 – Tabela ESTRUTN

Tabela ESTRUTN			
Atributo/Campo	Descrição	Tipo de Dados	Tamanho
ENANOR	Ano	decimal	(4,0)
ENMESR	Mês	decimal	(2,0)
ENNIV1	Direção	decimal	(6,0)
ENNOM1	Nome Direção	nvarchar	(50)
ENNIV2	Canal	decimal	(6,0)
ENNOM2	Nome Canal	nvarchar	(50)
ENNIV3	Responsável Comercial	decimal	(6,0)
ENNIV4	Designação Tipo Agente	decimal	(6,0)
ENNIV5	Agente	decimal	(6,0)

Tabela 33 – Tabela TBLIN

Tabela TBLIN			
Atributo/Campo	Descrição	Tipo de Dados	Tamanho
TLCODE	Código Informação	numeric	(3,0)
TLELEM	Informação	nvarchar	(3)
TLDESC	Descrição	nvarchar	(60)

Tabela 34 – Tabela CodigosPostais

Tabela CodigosPostais			
Atributo/Campo	Descrição	Tipo de Dados	Tamanho
cod_distrito	Código do distrito	numeric	(2,0)
cod_concelho	Código do concelho	numeric	(2,0)
cod_localidade	Código da localidade	numeric	(6,0)
nome_localidade	Nome da localidade	nvarchar	(50)
cod_arteria	Código da artéria	nvarchar	(20)
tipo_arteria	Tipo de artéria (Rua, Praça, etc)	nvarchar	(30)
prep1	Primeira preposição	nvarchar	10
titulo_arteria	Título (Doutor, Eng.º, Professor, etc)	nvarchar	30
prep2	Segunda preposição	nvarchar	10
nome_arteria	Designação da artéria	nvarchar	100
local_arteria	Local da artéria	nvarchar	100
num_cod_postal	Número do código postal	numeric	(4,0)
ext_cod_postal	Extensão do número do código postal	numeric	(3,0)
desig_postal	Designação postal	nvarchar	(50)

Tabela 35 – Tabela Concelhos

Tabela Concelhos			
Atributo/Campo	Descrição	Tipo de Dados	Tamanho
cod_distrito	Código do distrito	numeric	(2,0)
cod_concelho	Código do concelho	numeric	(2,0)
nome_concelho	Nome do concelho	nvarchar	(50)

Tabela 36 – Tabela Distritos

Tabela Distritos			
Atributo/Campo	Descrição	Tipo de Dados	Tamanho
cod_distrito	Código do distrito	numeric	(2,0)
nome_distrito	Nome do distrito	nvarchar	(50)

Anexo II

Neste anexo encontram-se os restantes mapeamentos de dados entre as tabelas da *Staging Area* com as da tabelas de dimensão e factos.

Tabela 37 – Mapeamento dos dados

Destino					Origem				Transformação
Nome da Tabela	Nome da Coluna	Tipo de Dados	Tipo de Tabela	Tipo de SCD	Nome da Base de Dados	Nome da Tabela	Nome da Coluna	Tipo de Dados	
DimRegiao	RegiaoKey (PK)	int	D	N/A				int	Surrogate key
DimRegiao	CodigoPostal4	numeric(4,0)	D	1	S. A.	CodigosPostais	G7\$CTT	numeric(4,0)	PK no sistema operacional
DimRegiao	CodigoPostal3	numeric(3,0)	D	1	S. A.	CodigosPostais	G7\$CT3	numeric(3,0)	PK no sistema operacional
Destino					Origem				Transformação

Nome da Tabela	Nome da Coluna	Tipo de Dados	Tipo de Tabela	Tipo de SCD	Nome da Base de Dados	Nome da Tabela	Nome da Coluna	Tipo de Dados	
DimRegiao	FreguesiaDesc	nvarchar(50)	D	1	S. A.	CodigosPostais	desig_postal	nvarchar(50)	SELECT desig_postal FROM CodigosPostais
DimRegiao	ConcelhoDesc	nvarchar(50)	D	1	S. A.	Concelhos	nome_concelho	nvarchar(50)	SELECT nome_concelho FROM Concelhos
DimRegiao	DistritoDesc	nvarchar(50)	D	1	S. A.	Distritos	nome_distrito	nvarchar(50)	SELECT nome_distrito FROM Distritos
FactProposta	TitularKey (PK)	int	F	N/A				int	Foreign key (DimTitular)
FactProposta	ProdutoKey (PK)	int	F	N/A				int	Foreign key (DimProduto)
FactProposta	EstruturaComKey (PK)	int	F	N/A				int	Foreign key (DimEstruturaComercial)
FactProposta	DataKey (PK)	int	F	N/A				int	Foreign key (DimData)
FactProposta	Proposta	numeric(8,0)	F	N/A	S. A.	PROPE0	PO\$PRO	numeric(8,0)	SELECT PO\$PRO FROM PROPE0
FactProposta	PremioTotal	numeric(11,2)	F	N/A	S. A.	PROPE0	POPRT0	numeric(11,2)	SELECT POPRTO FROM PROPE0
FactProposta	PremioPago	numeric(11,2)	F	N/A	S. A.	PROPE0	POPPAG	numeric(11,2)	SELECT POPPAG FROM PROPE0
FactProposta	Capital	numeric(13,2)	F	N/A	S. A.	PROPE0	POCAPT	numeric(13,2)	SELECT POCAPT FROM PROPE0
Destino					Origem				Transformação

Nome da Tabela	Nome da Coluna	Tipo de Dados	Tipo de Tabela	Tipo de SCD	Nome da Base de Dados	Nome da Tabela	Nome da Coluna	Tipo de Dados	
FactProposta	Cabeca1TitularKey	int	F	N/A					Foreign key (DimTitular)
FactProposta	Cabeca2TitularKey	int	F	N/A					Foreign key (DimTitular)
FactProposta	DataEfeitoKey	int	F	N/A					Foreign key (DimData)
FactProposta	DataSituacaoKey	int	F	N/A					Foreign key (DimData)
FactProposta	Duracao	numeric(2,0)	F	N/A	S. A.	PROPEO	PODURA	numeric(2,0)	SELECT PODURA FROM PROPEO

Destino					Origem				Transformação
Nome da Tabela	Nome da Coluna	Tipo de Dados	Tipo de Tabela	Tipo de SCD	Nome da Base de Dados	Nome da Tabela	Nome da Coluna	Tipo de Dados	
DimProduto	ProdutoKey (PK)	int	D	N/A				int	Surrogate key
DimProduto	CodigoModalidade	numeric(2,0)	D	1	S. A.	MODA00	M0\$MOD	numeric(2,0)	PK para o produto na Staging Area

DimProduto	Versao	numeric(2,0)	D	1	S. A.	MODA00	MONVER	numeric(2,0)	PK para o produto na Staging Area
DimProduto	Grupo	numeric(2,0)	D	1	S. A.	MODA00	M0GRUP	numeric(2,0)	SELECT M0GRUP FROM MODA00
DimProduto	DesignacaoAbreviada	nvarchar(8)	D	1	S. A.	MODA00	M0DESI	nvarchar(8)	SELECT M0DESI FROM MODA00
DimProduto	Descricao	nvarchar(30)	D	1	S. A.	MODA00	M0DESC	nvarchar(30)	SELECT M0DESC FROM MODA00
DimProduto	TipoSeguro	nvarchar(29)	D	2	S. A.	MODA00	M0TSEG	nvarchar(1)	SELECT M0TSEG FROM MODA00
DimProduto	PessoaSegura	nvarchar(21)	D	2	S. A.	MODA00	M0PSEG	nvarchar(1)	SELECT M0PSEG FROM MODA00
Destino					Origem				Transformação
Nome da Tabela	Nome da Coluna	Tipo de Dados	Tipo de Tabela	Tipo de SCD	Nome da Base de Dados	Nome da Tabela	Nome da Coluna	Tipo de Dados	
DimProduto	DuracaoContrato	nvarchar(22)	D	1	S. A.	MODA00	M0DURC	nvarchar(1)	SELECT M0DURC FROM MODA00
DimProduto	TipoContrato	nvarchar(7)	D	2	S. A.	MODA00	M0TCCR	nvarchar(1)	SELECT M0TCCR FROM MODA00
DimProduto	TipoModalidade	nvarchar(5)	D	2	S. A.	MODA00	M0VIMO	nvarchar(1)	SELECT M0VIMO FROM MODA00
DimProduto	DataInicioVersaoKey	int	D	1	S. A.	MODA00	M0AVIN M0MVIN M0DVIN	numeric(4,0) numeric(2,0) numeric(2,0)	SELECT M0AVIN, M0MVIN, M0DVIN FROM MODA00

DimProduto	DataFimVersaoKey	int	D	1	S. A.	MODA00	MOAVFI MOMVFI MODVFI	numeric(4,0) numeric(2,0) numeric(2,0)	SELECT MOAVFI, MOMVFI, MODVFI FROM MODA00
DimProduto	EffectiveDate	int	D	-					
DimProduto	ExpiredDate	int	D	-					
DimProduto	IsCurrent	nvarchar(3)	D	-					
DimTitular	TitularKey (PK)	int	D	N/A				int	Surrogate key
DimTitular	Proposta	numeric(8,0)	D	1	S. A.	PROPET	PT\$PRO	numeric(8,0)	PK para as propostas na Staging Area
Destino					Origem				Transformação
Nome da Tabela	Nome da Coluna	Tipo de Dados	Tipo de Tabela	Tipo de SCD	Nome da Base de Dados	Nome da Tabela	Nome da Coluna	Tipo de Dados	
DimTitular	TipoCliente	nvarchar(24)	D	2	S. A.	PROPET	PTTIPO	nvarchar(1)	PK para o titular na Staging Area
DimTitular	CodigoTitular	numeric(6,0)	D	1	S. A.	PROPET	PT\$COD	numeric(6,0)	SELECT PT\$COD FROM PROPET
DimTitular	TipoTitular	nvarchar(9)	D	2	S. A.	PROPET	PTNVAL	nvarchar(1)	SELECT PTNVAL FROM PROPET
DimTitular	Sexo	nvarchar(9)	D	1	S. A.	PROPET	PTSEXO	nvarchar(1)	SELECT PTSEXO FROM PROPET

DimTitular	RegiaoKey	int	D	1	S. A.	PROPET	PT\$CTT PT\$CT3	numeric(4,0) numeric(3,0)	SELECT PT\$CTT, PT\$CT3 FROM PROPET
DimTitular	Atividade	nvarchar(10)	D	1	S. A.	PROPET	PTACTI	nvarchar(1)	SELECT PTACTI FROM PROPET
DimTitular	Cae	numeric(6,0)	D	1	S. A.	PROPET	PT\$CAE	numeric(6,0)	SELECT PT\$CAE FROM PROPET
DimTitular	Profissao	nvarchar(60)	D	2	S. A.	PROPET	PT\$PRF	numeric(3,0)	SELECT PT\$PRF FROM PROPET
DimTitular	DataNascimentoKey	int	D	1	S. A.	PROPET	PTDNAN PTDNME PTDNDI	numeric(4,0) numeric(2,0) numeric(2,0)	SELECT PTDNAN, PTDNME, PTDNDI FROM PROPET
DimTitular	UtilizadorCriacao	nvarchar(10)	D	1	S. A.	PROPET	PTUSER	nvarchar(10)	SELECT PTUSER FROM PROPET
Destino					Origem				Transformação
Nome da Tabela	Nome da Coluna	Tipo de Dados	Tipo de Tabela	Tipo de SCD	Nome da Base de Dados	Nome da Tabela	Nome da Coluna	Tipo de Dados	
DimTitular	DataCriacaoKey	int	D	1	S. A.	PROPET	PTDATA	date	SELECT PTDATA FROM PROPET
DimTitular	Nacionalidade	nvarchar(20)	D	1	S. A.	PROPET	PTNACD	nvarchar(20)	SELECT PTNACD FROM PROPET
DimTitular	Naturalidade	nvarchar(20)	D	1	S. A.	PROPET	PTNATD	nvarchar(20)	SELECT PTNATD FROM PROPET
DimTitular	EstadoCivil	nvarchar(14)	D	2	S. A.	PROPET	PTESTC	nvarchar(1)	SELECT PTESTC FROM PROPET

DimTitular	EffectiveDate	int	D	-						
DimTitular	ExpiredDate	int	D	-						
DimTitular	IsCurrent	nvarchar(3)	D	-						
DimEstruturaComercial	EstruturaComKey (PK)	int	D	N/A				int	Surrogate key	
DimEstruturaComercial	Ano	decimal(4,0)	D	1	S. A.	ESTRUTN	ENANOR	decimal(4,0)	PK para a estrutura comercial na Staging Area	
DimEstruturaComercial	Mes	decimal(2,0)	D	1	S. A.	ESTRUTN	ENMESR	decimal(2,0)	SELECT ENMESR FROM ESTRUTN	
DimEstruturaComercial	Direcao	decimal(6,0)	D	1	S. A.	ESTRUTN	ENNIV1	decimal(6,0)	SELECT ENNIV1 FROM ESTRUTN	
Destino					Origem				Transformação	
Nome da Tabela	Nome da Coluna	Tipo de Dados	Tipo de Tabela	Tipo de SCD	Nome da Base de Dados	Nome da Tabela	Nome da Coluna	Tipo de Dados		
DimEstruturaComercial	NomeDirecao	nvarchar(42)	D	1	S. A.	ESTRUTN	ENNOM1	nvarchar(42)	SELECT ENNOM1 FROM ESTRUTN	
DimEstruturaComercial	Canal	decimal(6,0)	D	1	S. A.	ESTRUTN	ENNIV2	decimal(6,0)	SELECT ENNIV2 FROM ESTRUTN	
DimEstruturaComercial	NomeCanal	nvarchar(42)	D	1	S. A.	ESTRUTN	ENNOM2	nvarchar(42)	SELECT ENNOM2 FROM ESTRUTN	
DimEstruturaComercial	ResponsavelComercial	decimal(6,0)	D	1	S. A.	ESTRUTN	ENNIV3	decimal(6,0)	SELECT ENNIV3 FROM ESTRUTN	

DimEstruturaComercial	DesignacaoTipoAgente	decimal(6,0)	D	1	S. A.	ESTRUTN	ENNIV4	decimal(6,0)	SELECT ENNIV4 FROM ESTRUTN	
DimEstruturaComercial	Agente	decimal(6,0)	D	1	S. A.	ESTRUTN	ENNIV5	decimal(6,0)	SELECT ENNIV5 FROM ESTRUTN	
DimData	DataKey (PK)	int	D	N/A				int	Surrogate key	
DimData	DataCompleta	datetime	D	1						
DimData	Ano	int	D	1						
DimData	Semestre	tinyint	D	1						
DimData	Trimestre	tinyint	D	1						
Destino					Origem				Transformação	
Nome da Tabela	Nome da Coluna	Tipo de Dados	Tipo de Tabela	Tipo de SCD	Nome da Base de Dados	Nome da Tabela	Nome da Coluna	Tipo de Dados		
DimData	Mes	tinyint	D	1						
DimData	NomeMes	nvarchar(10)	D	1						
DimData	Semana	tinyint	D	1						
DimData	NumeroAno	int	D	1						

DimData	NumeroMes	tinyint	D	1					
DimData	NumeroSemana	tinyint	D	1					
DimData	DiaDaSemana	nvarchar(10)	D	1					
DimData	FimDeSemana	nvarchar(3)	D	1					
DimDadosProposta	PropostaKey (PK)	int	D	N/A				int	Surrogate key
DimDadosProposta	Proposta	numeric(8,0)	D	1	S. A.	PROPE0	PO\$PRO	numeric(8,0)	PK para a proposta na Staging Area
DimDadosProposta	TipoCriacao	nvarchar(9)	D	2	S. A.	PROPE0	POTIPO	nvarchar(1)	SELECT POTIPO FROM PROPE0
Destino					Origem				Transformação
Nome da Tabela	Nome da Coluna	Tipo de Dados	Tipo de Tabela	Tipo de SCD	Nome da Base de Dados	Nome da Tabela	Nome da Coluna	Tipo de Dados	
DimDadosProposta	Apolice	numeric(6,0)	D	1	S. A.	PROPE0	PO\$APL	numeric(6,0)	SELECT PO\$APL FROM PROPE0
DimDadosProposta	NumeroApolice	numeric(6,0)	D	1	S. A.	PROPE0	PONAPO	numeric(6,0)	SELECT PONAPO FROM PROPE0
DimDadosProposta	Fracionamento	nvarchar(10)	D	2	S. A.	PROPE0	PO\$FRA	nvarchar(1)	SELECT PO\$FRA FROM PROPE0
DimDadosProposta	TipoTitular	nvarchar(9)	D	2	S. A.	PROPE0	POTITP	nvarchar(1)	SELECT POTITP FROM PROPE0

DimDadosProposta	Tipo1PSegura	nvarchar(9)	D	2	S. A.	PROPE0	POP1TP	nvarchar(1)	SELECT POP1TP FROM PROPE0
DimDadosProposta	Tipo2PSegura	nvarchar(9)	D	2	S. A.	PROPE0	POP2TP	nvarchar(1)	SELECT POP2TP FROM PROPE0
DimDadosProposta	HerdeirosLegais	nvarchar(3)	D	2	S. A.	PROPE0	POHERD	nvarchar(1)	SELECT POHERD FROM PROPE0
DimDadosProposta	Situacao	nvarchar(24)	D	2	S. A.	PROPE0	POSITU	numeric(1,0)	SELECT POSITU FROM PROPE0
DimDadosProposta	Utilizador	nvarchar(10)	D	1	S. A.	PROPE0	POUSER	nvarchar(10)	SELECT POUSER FROM PROPE0
DimDadosProposta	TomadorSeguroBeneficiario	nvarchar(3)	D	2	S. A.	PROPE0	POTOSE	nvarchar(1)	SELECT POTOSE FROM PROPE0
DimDadosProposta	FormaPagamento	nvarchar(18)	D	2	S. A.	PROPE0	POINF2	nvarchar(1)	SELECT POINF2 FROM PROPE0
Destino					Origem				Transformação
Nome da Tabela	Nome da Coluna	Tipo de Dados	Tipo de Tabela	Tipo de SCD	Nome da Base de Dados	Nome da Tabela	Nome da Coluna	Tipo de Dados	
DimDadosProposta	FormaPagamentoRecibosSeguintes	nvarchar(18)	D	2	S. A.	PROPE0	POFPRS	nvarchar(1)	SELECT POFPRS FROM PROPE0
DimDadosProposta	EffectiveDate	int	D	-					
DimDadosProposta	ExpiredDate	int	D	-					
DimDadosProposta	IsCurrent	nvarchar(3)	D	-					

Anexo III

Neste anexo encontra-se o inquérito de satisfação aos profissionais de IT.

Inquérito de Satisfação (IT)

Foi desenvolvido um sistema de *Business Intelligence* para ser aplicado em Dados de Seguros de Vida e que visa auxiliar profissionais da seguradora. Como forma de avaliar a solução desenvolvida, são apresentados alguns pontos que devem ser avaliados de acordo com uma escala de 1 a 5, onde o 1 representa "Discordo Totalmente", o 2 representa "Discordo", o 3 representa "Não Concordo nem Discordo", o 4 representa "Concordo" e o 5 representa "Concordo Totalmente".

Concordo Totalmente".

O sistema possui uma arquitetura que se adequa ao negócio. *

1 2 3 4 5

Discordo Totalmente Concordo Totalmente

A arquitetura desenvolvida encontra-se preparada para novos indicadores. *

1 2 3 4 5

Discordo Totalmente Concordo Totalmente

A aplicação integra facilmente com outros sistemas. *

1 2 3 4 5

Discordo Totalmente Concordo Totalmente

O tempo de resposta da aplicação é adequado. *

1 2 3 4 5

Discordo Totalmente Concordo Totalmente

Figura 84 – Inquérito de satisfação (IT)

Anexo IV

Neste anexo encontra-se o inquérito de satisfação aos profissionais da seguradora.

Inquérito de Satisfação (Profissionais da Seguradora)

Foi desenvolvido um sistema de *Business Intelligence* para ser aplicado em Dados de Seguros de Vida e que visa auxiliar profissionais da seguradora. Como forma de avaliar a solução desenvolvida, são apresentados alguns pontos que devem ser avaliados de acordo com uma escala de 1 a 5, onde o 1 representa 'Discordo Totalmente', o 2 representa 'Discordo', o 3 representa 'Não Concordo nem Discordo', o 4 representa 'Concordo' e o 5 representa 'Concordo Totalmente'.

Concordo Totalmente*

A aplicação mostra-se útil e intuitiva. *

1 2 3 4 5

Discordo Totalmente Concordo Totalmente

Os filtros adequam-se ao tema em estudo. *

1 2 3 4 5

Discordo Totalmente Concordo Totalmente

A informação encontra-se disponibilizada de forma a facilitar a leitura e a análise * dos dados.

1 2 3 4 5

Discordo Totalmente Concordo Totalmente

O método de apresentação dos indicadores é preferível quando comparado com * relatórios estáticos.

1 2 3 4 5

Discordo Totalmente Concordo Totalmente

Considera a aplicação uma boa ferramenta para exploração de dados. *

1 2 3 4 5

Discordo Totalmente Concordo Totalmente

Considera que a solução desenvolvida foi conseguida de acordo com os * objectivos a atingir.

1 2 3 4 5

Discordo Totalmente Concordo Totalmente

Figura 85 – Inquérito de satisfação (profissionais da seguradora)