



Universidade do Minho

Mestrado em Engenharia de Sistemas
Sistemas de Apoio à Decisão

Unidade Curricular de Armazéns de Dados - *Data Warehousing*

Ano Letivo de 2018/2019

Aluguer de Automóveis

Ana Margarida Rolim PG38332

Márcia Costa A67672

Célia Natália A67637

Daniel Sousa PG37112

Maio, 2019

AD-DW

Data de Recepção	
Responsável	
Avaliação	
Observações	

Aluguer de Automóveis

Margarida Rolim, PG38332

Márcia Costa, A67672

Célia Natália, A67637

Daniel Sousa, PG37112

Maio, 2019

Resumo

O presente relatório descreve todo o processo efetuado para cumprir o trabalho proposto de Sistemas de *Data Warehousing*.

Este processo surgiu da necessidade da empresa “OnRoad” que, após comprar a empresa “emViagem”, considerou importante implementar um sistema de suporte à decisão.

Primeiramente, serão descritas as motivações que levaram à implementação do armazém de dados, tal como os seus objetivos.

Posteriormente, são definidos os requisitos acordados com a empresa cliente. Apresenta-se, também, o modelo dimensional, alicerçado pelo método de quatro passos, proposto por Kimball e Ross, englobando todas as fases nele inerentes.

De seguida, foi elaborada não só a análise das duas fontes de dados, como também a modelação concetual do sistema de povoamento, descrevendo todo o processo ETL executado.

Em seguimento, procedeu-se à implementação física do modelo de *data warehouse* assim como a implementação e análise de todo o processo de povoamento.

Por fim, após a validação com o cliente, foi testado o modelo pedido, através de técnicas de análise de dados e *business intelligence*.

As conclusões e objetivos são a última área a ser abordada.

Área de Aplicação: Desenvolvimento de um sistema de Data Warehouse;

Palavras-chave: Data Warehouse, Fontes de Informação, Pentaho, Alugueres, Veículos, Clientes, Calendário;

Índice

Resumo	i
Índice	ii
Índice de Figuras	iv
Índice de Tabelas	v
1. Introdução	1
1.1 Contextualização	1
1.2 Apresentação do Caso de Estudo	2
1.3 Motivação e Objetivos	2
1.4 Viabilidade da Implementação do Sistema	3
1.5 Estrutura do Relatório	3
2. Planeamento e Gestão do Projeto	5
2.1 Definição da Identidade do Projeto	5
2.2 Identificação de Recursos	6
2.3 Definição de Medidas de Sucesso	6
3. Levantamento e Análise de Requisitos	8
3.1 Apresentação do Método de Aquisição de Requisitos	8
3.2 Requisitos de Descrição	8
3.3 Requisitos de Exploração	8
3.4 Requisitos de Controlo e Acesso	9
3.5 Revisão dos Requisitos com os Utilizadores	10
4. Modelação Dimensional do Sistema	11
4.1 Apresentação da Metodologia de Desenvolvimento	11
4.2 Esquematização da Matriz de Decisão	12
4.3 Definição e Caracterização do <i>Data Mart</i> e Grão	12
4.4 Definição e Caracterização das Dimensões	14
4.4.1 Dimensão Cliente	14
4.4.2 Dimensão Calendário	15
4.4.3 Dimensão Veículo	15
4.4.4 Dimensão Aluguer	15
4.5 Definição e Caracterização das Tabelas de Factos	16
4.6 Esquematização do Esquema Dimensional	19
4.7 Revisão do Esquema Dimensional Desenvolvido	19
5. Caracterização das Fontes de Informação	20
5.1 Identificação e Descrição das Fontes de Informação do Sistema	20
5.2 Análise dos Dados das Fontes	20
5.2.1 Fonte de Dados Relacional	20
5.2.2 Fonte de Dados Excel	22
5.3 Desenvolvimento do Esquema de Mapeamento de Dados	23

5.3.1 MySQL	23
5.3.2 Excel	24
5.4 Revisão do Esquema de Mapeamento de Dados	25
6. Modelação do Sistema de Povoamento	26
6.1 Apresentação e Descrição Sumária das Estruturas do Data Warehouse a Povoar e das Fontes de Dados Envolvidas	26
6.2 Esquematização conceptual e caracterização do sistema de povoamento	27
6.2.1 Extração	27
6.2.2. Transformação	28
6.2.3 Carregamento	29
7. Implementação do Sistema de Data Warehousing	31
7.1 Implementação Física do Data Warehouse	31
7.2 Implementação Física do Sistema de Povoamento	32
7.3 Análise da Execução do Sistema de Povoamento	32
8. Business Intelligence	37
9. Conclusões e Trabalho Futuro	44
Referências	45
Lista de Siglas e Acrónimos	46
Anexos	47
Script de criação de DW	47
I. Script de criação do data mart	48
II. Script de criação do data mart	51

Índice de Figuras

Figura 1- Esquema Dimensional	19
Figura 2: Esquema Geral	26
Figura 3: Extração dos dados da fonte de dados SQL	27
Figura 4: Extração dos dados da fonte de dados em Excel	28
Figura 5: Limpeza de nulos	28
Figura 6: Processo de transformação para a Dim Cliente	29
Figura 7: Processo de Transformação para a Dim Calendário	29
Figura 8: Processo de Povoamento do Data Warehouse	30
Figura 9: Esquema do <i>Data Warehouse</i>	31
Figura 10 -Esquema Carregamento Inicial	32
Figura 11: Carregamento em Sequência	33
Figura 12: Carregamento dados Clientes	33
Figura 13 - Carregamento dos dados Veiculo para data warehouse	34
Figura 14 - Carregamento de dados para a tabela DWCalendario	35
Figura 15 - Carregamento de dados para a tabela TFAluguer	36
Figura 16: Clientes que mais alugaram veículos	37
Figura 17: Clientes que investiram mais verbas na empresa	38
Figura 18 – Clientes que, em média, mais verbas investiram na empresa	39
Figura 19: Marcas de carros mais alugadas	40
Figura 20: Meses com mais procura	41
Figura 21: Modelo de carros mais alugados	41
Figura 22: Trimestre de determinado ano com maior frequência de alugueres	42
Figura 23 – Gráfico circular representativo das cidades com maior frequência de alugueres	43

Índice de Tabelas

Tabela 1 - Caracterização da Matriz de Decisão	12
Tabela 2- Caracterização das dimensões	14
Tabela 3 - Caracterização da TF-Aluguer	17
Tabela 4- Tabela com Dicionário dos dados	20
Tabela 5: Mapeamento de dados a partir da fonte SQL	23

1. Introdução

Este primeiro capítulo tem como objetivo apresentar uma breve introdução ao projeto a realizar. Sendo assim, é necessário definir o contexto no qual se desenvolve o caso de estudo, seguido da sua descrição. É também importante perceber os motivos que levaram à criação de um armazém de dados na empresa *OnRoad*, assim como os seus objetivos.

1.1 Contextualização

A *OnRoad* é uma empresa de aluguer de viaturas localizada no centro do Porto há 8 anos. O *corebusiness* da empresa são os alugueres de curta duração, apesar de gerar mais processos administrativos é o que gera mais lucro. O seu principal público alvo são os turistas e representantes de empresas em viagens de negócios.

A cidade do Porto foi eleita o melhor destino europeu pela *European Best Destinations* duas vezes consecutivas (2016 e 2017), isto levou a um aumento exponencial no número de turistas e atraiu ainda investimento de grandes empresas internacionais que decidiram construir instalações na cidade. Com isto, a carteira de clientes habituais e esporádicos da *OnRoad* aumentou de forma significativa.

A empresa *OnRoad* efetua todo o processo de aluguer de um veículo desde escolha do cliente em loja até à conclusão do contrato de aluguer e posterior assinatura do mesmo. Está ainda incluído um controlo minucioso sobre as datas previstas e efetivas de entrega do veículo. A empresa está dividida em vários sectores nomeadamente o sector financeiro, recursos humanos, compra de veículos a entidades externas e aluguer do mesmo ao consumidor final. Apenas serão alvo de estudo o sector de aluguer de veículos.

Neste documento, todo este processo será alvo de um estudo profundo, documentado através de uma pormenorizada análise de requisitos, para de acordo com a metodologia traçada, ser então desenvolvido um sistema que cumpra todos os objetivos definidos.

1.2 Apresentação do Caso de Estudo

A empresa *OnRoad* efetua todo o processo de aluguer de um veículo desde escolha do cliente em loja até à conclusão do contrato de aluguer e posterior assinatura do mesmo. Está ainda incluído um controlo minucioso sobre as datas previstas e efetivas de entrega do veículo. A empresa está dividida em vários sectores nomeadamente o sector financeiro, recursos humanos, compra de veículos a entidades externas e aluguer do mesmo ao consumidor final. Apenas serão alvo de estudo o sector de aluguer de veículos. Não será englobado no problema, a gestão dos funcionários que fazem parte da empresa. Assim sendo, os funcionários e o supervisor de secção de aluguer não serão abrangidos, tal como todos os outros trabalhadores de outras secções. Assim, ambos, não farão parte do sistema a desenvolver.

O processo de aluguer de veículos segue a sequência descrita de seguida: na primeira fase, um funcionário atende um determinado cliente, procurando satisfazer o pedido do mesmo. Um cliente seleciona um veículo que pretenda alugar. Após selecionar um dos veículos disponíveis na empresa, cabe ao respetivo funcionário tratar de todo o processo de aluguer.

O processo de aluguer requer que o mesmo tenha associado uma caução e um seguro, que para o caso são obrigatórios, isto é, todo o cliente que pretenda efetuar o aluguer de um veículo tem obrigatoriamente que pagar uma caução e assinar e pagar um seguro. Cada seguro tem associado uma descrição, de acordo com cada tipo de seguro. Seguidamente, é entregue ao cliente o veículo e respetiva chave.

Neste documento, todo este processo será alvo de um estudo profundo, documentado através de uma pormenorizada análise de requisitos, para de acordo com a metodologia traçada, ser então desenvolvido um sistema que cumpra todos os objetivos definidos.

Durante o processo de desenvolvimento, foram utilizadas duas bases de dados em ficheiros distintos, uma das quais num ficheiro Excel, proveniente da empresa emViagem e a restante no proveniente do MySQL da empresa OnRoad.

1.3 Motivação e Objetivos

Devido à diminuição do número de alugueres, o CEO da OnRoad decidiu implementar um data warehouse com o objetivo de centralizar e analisar os dados armazenados em duas fontes de dados distintas e, desse modo, contrariar o problema anteriormente mencionado.

Desta forma, com a implementação de um data warehouse, a OnRoad poderia melhorar o seu processo da tomada de decisão, apoiado no sistema de suporte à decisão elaborado.

A construção deste armazém de dados, tem a finalidade de aumentar o número de alugueres de veículos da empresa em questão. Para alcançar este objetivo, a empresa propõe-se não só a promover campanhas de fidelização de clientes regulares, como também a cativar novos clientes, através de uma análise e adaptação de preços de aluguer e de gama de automóveis disponíveis, de acordo com as preferências dos clientes. Para além destas motivações, a OnRoad compromete-se a, de acordo com média de dias que um veículo se encontra alugado, por cliente, atribuir prémios de excelência de cliente, refletidos numa promoção em alugueres futuros.

1.4 Viabilidade da Implementação do Sistema

A empresa *OnRoad* guarda os seus dados em duas fontes distintas, o que torna a análise dos mesmos mais complexa, dificultando o processo de tomada de decisão. Assim, para suportar esse processo, a construção de um armazém de dados torna-se fulcral para que os dados estejam integrados e acessíveis de forma coerente.

O *data warehouse* permite, também, à empresa, uma análise detalhada do tipo de clientes, permitindo um *profiling* dos mesmos. Desta forma, a *OnRoad* consegue adaptar a gama de produtos que oferece, de acordo com as preferências do cliente. Em seguimento, estas medidas permitem um maior controlo sobre a frota de veículos existente, evitando o investimento em veículos que não são atrativos para os seus clientes.

De acordo com o mencionado, uma análise de dados detalhada, fornece informação sobre o posicionamento da empresa no mercado, evidenciando os aspetos necessários a melhorar, com a finalidade em desencadear o processo de tomada de decisão suportado nos dados reais da empresa.

1.5 Estrutura do Relatório

Os capítulos deste projeto estão estruturados de forma a cumprir os passos conhecidos como corretos para a implementação de um *data warehouse*.

Em primeiro lugar, o capítulo 2 apresenta não só um breve sumário do projeto a realizar com a descrição da equipa destacada para a realização do mesmo, como também a definição das medidas de sucesso para a implementação do *data warehouse*.

Seguidamente, o capítulo 3 expõe todos os requisitos definidos de acordo com o estabelecido com o cliente para a implementação do armazém de dados, tais como, requisitos de descrição, requisitos de exploração e requisitos de controlo e acesso.

No capítulo 4, é apresentado o modelo dimensional definido através do método dos quatro passos, proposto por *Kimball*. O primeiro passo consistiu na eleição da área de suporte à decisão a implementar, seguido da definição do detalhe dos factos (o grão) do processo selecionado. Posteriormente, procedeu-se à seleção das dimensões de análise sobre as quais se pretende analisar os factos e, por fim, definiram-se as medidas a integrar na estrutura de cada facto.

O capítulo 5, exhibe a análise das fontes de dados utilizadas, base de dados relacional e folha de cálculo Excel. Na parte final do capítulo mostra-se o esquema do mapeamento de dados das fontes para o *data warehouse*.

Ao longo do capítulo 6, encontra-se exposta a modelação concetual do sistema de povoamento de dados com a descrição de todo o processo ETL executado.

Relativamente ao capítulo 7, este evidencia a implementação física do *data warehouse*, assim como a implementação e análise de todo o sistema de povoamento do mesmo.

Em seguimento, no capítulo 8, são apresentadas as análises feitas de acordo com os dados existentes no armazém de dados, a partir das quais se verificou que o *data warehouse* implementado cumpria os requisitos definidos.

Por fim, descrevem-se no capítulo 9, as conclusões retiradas do trabalho realizado, assim como possíveis trabalhos futuros de modo a melhorar a implementação do *data warehouse* em questão.

2. Planeamento e Gestão do Projeto

2.1 Definição da Identidade do Projeto

O primeiro passo a efetuar no desenvolvimento de uma solução que terá por base um sistema de *Data Warehouse* é identificar o âmbito do projeto e justificar a sua necessidade. Após o desenvolvimento da base de dados que suporta as ações operacionais da empresa OnRoad, esta decidiu comprar uma empresa do mesmo ramo e assim aumentar o seu volume de clientes e negócios mais rapidamente.

A empresa adquirida emViagem, já detinha um sistema que suportava os seus processos internos e agora pretende-se migrar os dados para um único sistema. Cada uma das empresas irá então manter os seus sistemas atuais e independentes. Ao contrário da OnRoad, a emViagem armazena os seus dados operacionais em ficheiros Excel.

Para os administradores, agora das duas empresas, torna-se indispensável conseguir ter uma visão global, de forma a tomar as decisões mais acertadas para potencial o crescimento das duas marcas. Devido aos dois sistemas atuais, considera-se extremamente difícil conseguir ter esta visão global, uma vez que os dois sistemas foram desenhados para responderem às exigências operacionais, e não com a finalidade em responder a perguntas orientadas à tomada de decisão.

Imagine-se que um administrador pretende saber quais os clientes que mais veículos alugaram no mês de junho. Para obter esta informação, um administrador teria de cruzar informação sobre as duas fontes, que provavelmente terão identificadores diferentes para o mesmo cliente, para depois obter os resultados aglomerados. Os sistemas atuais teriam muita dificuldade em responder de forma clara às seguintes perguntas:

Quais os melhores clientes?

Quais as marcas de carros mais alugadas?

Quais os meses com mais procura?

As respostas às perguntas acima descritas são essenciais para que os administradores possam influenciar, de forma positiva, as vendas e a respetiva faturação das marcas. Com base da informação mencionada, surgiu a necessidade da

criação de um *Data Warehouse* denominado de *DWonRoad*. Neste sistema, a informação das duas fontes deverá ser homogeneizada permitindo responder de forma rápida às questões acima levantadas. Com base na resposta a estas questões, os agentes de decisão iram conseguir, de forma mais acertada, influenciar o rumo das marcas e aumentar o lucro da empresa. Este sistema deverá focar-se na faturação da empresa, tendo como vértices de análise os veículos alugados, os respetivos clientes e as datas em que o aluguer foi efetivado. Todo o desenvolvimento da *DWonRoad* deverá ter o âmbito agora descrito em conta.

O projeto em questão, traduz-se na composição de um *Data Warehouse*. Este é um repositório de dados integrados não voláteis, com a finalidade em suportar os processos de tomada de decisão dos gestores.

2.2 Identificação de Recursos

Para levar a cabo a execução deste projeto foi criada uma equipa de 4 elementos. Todas as tarefas necessárias para concluir a implementação do projeto deverá ser efetuada pela equipa. Os 4 elementos que formam a equipa são multifacetados e conseguiram entre eles realizar tarefas desde a análise de requisitos, modelação, implementação do povoamento do *Data Warehouse* e à validação do mesmo com o cliente.

Para além disso, cada elemento da equipa deverá ter ao seu dispor um computador com MySQL, Word, Pentaho Data Integration, Bizagi Modeler, bem como acesso aos sistemas de armazenamento da OnRoad e da emViagem.

Com o objetivo de promover um bom desenvolvimento do projeto, eliminando possíveis complicações, são requeridos recursos tais como, uma equipa de quatro pessoas capazes de realizar o projeto em questão, denominada de equipa de desenvolvimento. Em acréscimo, serão também necessárias pelo menos quatro máquinas, uma por cada elemento do grupo de trabalho.

2.3 Definição de Medidas de Sucesso

A implementação de um *Data Warehouse* é muitas vezes um processo complexo e com custos monetários elevados. É de extrema importância que se definam medidas de sucesso do projeto desenvolvido, estas deverão ser garantidas aquando da sua

finalização. Com a participação do cliente, foram definidas as seguintes medidas de sucesso:

O tempo necessário para a implementação, é no máximo o tempo estimado.

O *Data Warehouse* responde as perguntas dos tomadores de decisão.

O custo do projeto não ultrapassa o valor estimado.

3. Levantamento e Análise de Requisitos

3.1 Apresentação do Método de Aquisição de Requisitos

Para compreender que requisitos o *data warehouse* deveria conter, analisaram-se as características essenciais para a empresa em questão, relacionadas com os seus clientes, com os seus veículos, assim como, de que tipo de informações a OnRoad necessita de obter a partir deste sistema. Desta forma, definiu-se as restrições e as características que o armazém deveria conter para o bom funcionamento do mesmo, ao longo de nove anos.

3.2 Requisitos de Descrição

- Para a análise de um aluguer é necessário considerar informação sobre os clientes, veículos e a data de aluguer.
- Todos os alugueres são identificados inequivocamente.
- A data deve conter informações relativas ao dia, ao mês, trimestre e ano.
- Cada cliente é identificado com um identificador único, com um nome, um NIF, a carta de condução, o email, o telefone, a data de nascimento, a cidade e o país de residência.
- Cada veículo tem um identificador único. É identificado também pela matrícula, o seu modelo, a sua marca, e o seu número de quilómetros.
- Um veículo pode ser alugado pelo mesmo cliente no mesmo dia, mais do que uma vez.

3.3 Requisitos de Exploração

- Averiguar quais os veículos mais alugados pelos clientes da *OnRoad*.
- Identificar o tipo de clientes de acordo com o seu capital investido na empresa.
- Identificar os períodos do ano com mais ou menos alugueres de veículos.
- Relacionar o preço de aluguer de veículos com a sua frequência de aluguer.

- Identificar a marca ou modelo de veículo preferida pelos clientes da empresa.
- Estabelecer relações entre as características do cliente, tais como a cidade ou país onde reside, com a sua frequência de aluguer de veículos.
- Identificar os clientes com o maior número de alugueres.

3.4 Requisitos de Controlo e Acesso

Um armazém de dados, é utilizado para armazenar informações relativas às atividades de uma organização numa base de dados.

De realçar que, todos os utilizadores devem aceder aos seus próprios dados sem que, de alguma forma, não seja possível que esses mesmos utilizadores acedam a dados de outros. Esta realidade é cada vez mais relevante na atualidade, na medida em que as leis de proteção de dados estão cada vez mais atentas a situações que possam comprometer questões relacionadas com a privacidade e proteção de dados pessoais. Como proteção de dados, entende-se a possibilidade de cada utilizador determinar de forma autónoma, a utilização que é feita dos seus próprios dados. Para tal, é de extrema importância garantir formas capazes de monitorizar e limitar o acesso a vários dados, por parte do utilizador.

Foram identificados dois tipos de utilizadores, o administrador e o cliente.

- **Administrador**

Profissional responsável por gerir, instalar, configurar, atualizar e monitorizar a base de dados ou sistemas de base de dados. O administrador tem as permissões necessárias para resolver questões relacionadas com a informação guardada na base de dados. Assim sendo, o administrador pode alterar, remover, inserir e consultar toda a informação da base de dados.

- **Cliente**

O cliente pode fazer todas as operações que dizem respeito aos seus dados, exceção feita à remoção completa do seu registo.

3.5 Revisão dos Requisitos com os Utilizadores

Para garantir que os requisitos definidos são suficientes para suprir as necessidades da OnRoad, realizaram-se diversas reuniões com a empresa, para validar o *data warehouse* definido.

4. Modelação Dimensional do Sistema

4.1 Apresentação da Metodologia de Desenvolvimento

Uma das formas mais usuais de fazer o desenvolvimento de um esquema dimensional é através da utilização do método dos “4 passos”, de *Kimball*, que pressupõe o desenvolvimento do sistema de data *warehousing* tipicamente de baixo para cima. Os quatro passos do método são:

1. Seleção da **área** de suporte à decisão a implementar.
2. Definição do detalhe dos factos (**o grão**) do processo seleccionado.
3. Seleção das **dimensões de análise** sobre as quais se pretende analisar os factos.
4. Definição das **medidas** a integrar na estrutura de cada facto.

Depois de definidos e caracterizados, cada passo é necessário dialogar com o cliente, a fim de se verificar o desenvolvimento do *Data Warehouse*.

4.2 Esquematização da Matriz de Decisão

Tabela 1 - Caracterização da Matriz de Decisão

Caracterização do <i>Data Mart</i> “Comercial”	
Identificação: Comercial	
Descrição geral: Informação para dar suporte à tomada de decisão na área de alugueres de veículos, relativa à empresa “ <i>OnRoad</i> ”, providenciando elementos de dados selecionados acerca de alugueres efetuados, com motivação à gestão e controlo das mesmas, para ações comerciais e análise de vendas, isto é, de alugueres.	
Estrutura base:	
Tabela de Factos >>>	TF-Aluguer
<<< Dimensões	
Calendário	√
Cliente	√
Veículo	√
Número de Dimensões	3
Tipo	Transacional
Periodicidade	Diária
Descrição	Transações comerciais de alugueres de veículos.
Utilidade estratégica	Avaliação das características mais apreciadas nos alugueres. Incentivar o aluguer de automóveis. Definição de ações promocionais. Profiling de clientes.
Utilizadores	Administradores.
Observações:	
Nada a assinalar.	

4.3 Definição e Caracterização do *Data Mart* e Grão

Depois de analisar a área de negócio de suporte à decisão a implementar, foi necessário definir o grão do processo selecionado, que o define: “O aluguer de um veículo, numa determinada data, solicitado por um cliente, consoante determinado preço de aluguer.” A construção do armazém de dados, tem a finalidade de aumentar o número de alugueres de veículos da empresa. Como tal, a tabela de factos correspondente a alugueres será a única necessária para o caso em estudo.

Na **dimensão Calendário** é possível efetuar relacionamentos com a data de um aluguer e permanecerá inalterada até ao momento em que se esgote a sua validade. A **dimensão Cliente** permite relacionar os clientes com os respetivos alugueres, e, por fim a **dimensão Veículo** com os diferentes veículos existentes na empresa. Pressupõem-se que as dimensões serão atualizadas diariamente. Relativamente à dimensão Cliente, crescerá de acordo com o número de novos clientes diários ou alterações a um cliente existente., assim como a dimensão Veículo que também crescerá conforme aquisição de novos modelos, marcas, etc.

4.4 Definição e Caracterização das Dimensões

Tabela 2- Caracterização das dimensões

Dimensões do Data Mart “Comercial”			
Nr	Identificação	Descrição	Esquema (tipo)
1	Cliente	Identificação e caracterização dos clientes da empresa “OnRoad”.	Dim-Cliente (com variação, com criação de novos registos na tabela base)
2	Veículo	Identificação e caracterização dos veículos da empresa “OnRoad”.	Dim-Veículo (com variação, com criação de novos registos na tabela base)
3	Calendário	Corresponde à dimensão temporal. Acolhe todos os atributos que sustentem análises temporais, como a data, mês, trimestre e ano. Regista a data de um aluguer.	Dim-Calendário (Normal)

4.4.1 Dimensão Cliente

Caracteriza-se pela dimensão que guarda os dados dos clientes que já alugaram um veículo, tendo como atributos:

- Nome
- Telemóvel
- Data de nascimento
- Cidade
- País

Dos atributos que caracterizam a dimensão cliente, foram considerados dois atributos agregados:

- Cidade
- País

4.4.2 Dimensão Calendário

Dimensão responsável por guardar a data em que cada veículo foi alugado, tendo como atributos:

- Id da data
- Data (“dia-mês-ano”),
- Mês da data (de 1 a 12), trimestre (“trim-ano”)
- Ano da data

Destes, há o interesse e possibilidade em agrupar por:

- Data
- Mês do Ano
- Trimestre
- Ano

4.4.3 Dimensão Veículo

Dimensão que caracteriza cada veículo que a “OnRoad” possui. Cada veículo é caracterizado através dos atributos:

- Id do veículo
- Matrícula
- Número de quilómetros
- Modelo
- Marca

Será possível agregar conforme:

- Modelo
- Marca

4.4.4 Dimensão Aluguer

A dimensão aluguer é uma dimensão degenerada. Esta não possuiu qualquer atributo que a caracterize, para além da que integra a chave da tabela de factos. Assim, esta apenas serve para garantir que um cliente, num determinado dia, poderá alugar o

mesmo veículo mais de uma vez. Na prática, esta dimensão não origina uma tabela de suporte autónoma.

4.5 Definição e Caracterização das Tabelas de Factos

O tipo de uma tabela de factos releva algumas características bastante pertinentes de uma tabela de factos, nomeadamente a forma como é povoada, o tipo de elementos de dados que irá acolher, a forma como os seus dados serão extraídos das fontes de informação, a cadência do seu povoamento, entre outras. A tabela de factos deve ser simples, por forma a simplificar a sua leitura e também as operações sobre a mesma. A tabela de factos representada a seguir refere-se a um aluguer de um veículo. Nesse aluguer será guardada informação relativa ao preço de aluguer. Existem três formas de caracterizar o tipo de uma tabela factos, nomeadamente: transacional, instantâneo e acumulativo. Neste caso trata-se de uma tabela de factos transacional uma vez que, representa um evento que ocorreu num momento instantâneo de tempo.

Tabela 3 - Caracterização da TF-Aluguer

Caracterização da tabela de factos				
Identificação		TF-Aluguer		
Descrição		Tabela que acolhe os vários registos de cada um dos alugueres de cada um dos clientes.		
Data mart		Comercial		
Tipo		Transacional		
Utilidade Estratégica		Avaliação das características mais apreciadas nos alugueres. Incentivar ao aluguer de veículos. Definição de ações promocionais. <i>Profiling</i> de clientes. Melhoria de base de negociação com cadeias de automóveis.		
Povoamento		Realizado diariamente entre as 00:00h até às 8:00h.		
Dimensão Inicial		64 KB		
Crescimento		25% ano		
Período de dados		Desde o início do ano de 2016. As restantes informações ficarão em arquivos.		
Atributos				
Dimensões				
Nr	Identificação	Chave	Descrição	Exemplo
1	Id_Cliente	S	Chave de substituição para um cliente da <i>OnRoad</i> .	1
2	Id_Veiculo	S	Chave de substituição para um veículo da <i>OnRoad</i> .	1
3	Id_Data	S	Código da data referente à data de um aluguer.	1
Medidas				
Nr	Identificação	Domínio	Descrição	Exemplo
1	PreçoAluguer	Decimal(8,2)	Valor total do aluguer efetuado.	2

Perfis de Utilização
Administrador da base de dados.
Observações:
Nada a assinalar.

4.6 Esquematização do Esquema Dimensional

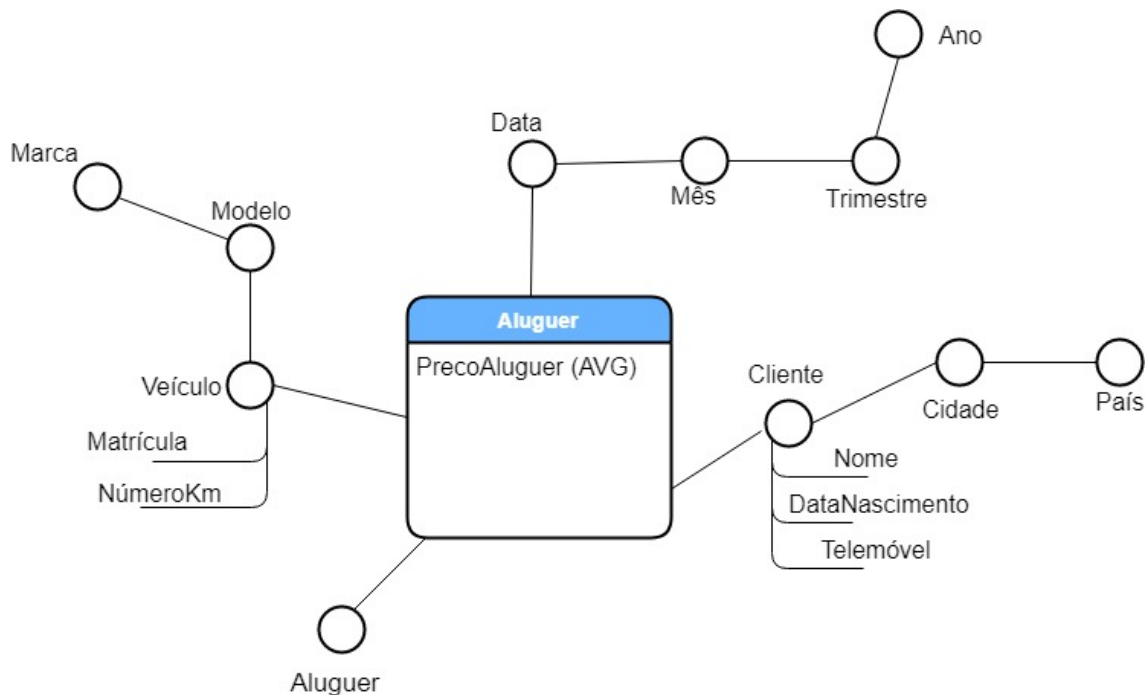


Figura 1- Esquema Dimensional

4.7 Revisão do Esquema Dimensional Desenvolvido

Em seguimento do trabalho desenvolvido, foi apresentado ao cliente, o modelo dimensional elaborado. Após discussão de algumas temáticas relacionadas com o modelo e com questões subjacentes ao mesmo, este foi aprovado com sucesso.

Em conversa com o cliente, o mesmo observou que as necessidades de obtenção de informação relativas aos veículos que compõem os alugueres, bem como os seus atributos, se encontravam conforme o proposto.

Desta forma, foi então iniciado o processo de ETL.

5. Caracterização das Fontes de Informação

Para que o povoamento do *data warehouse* seja efetuado corretamente é necessário caracterizar as fontes de dados da OnRoad. Desta forma, sabemos exatamente que tipo de informação existe e, que transformações serão necessárias para carregar os dados pretendidos, devidamente adaptados, no armazém de dados desenvolvido.

5.1 Identificação e Descrição das Fontes de Informação do Sistema

A OnRoad possui, como já referido anteriormente, duas fontes de dados. A principal, que contém a maioria da informação, é relacional, desenvolvida no MySQL. A segunda fonte de dados é não relacional, mas os dados guardados na mesma são exatamente do mesmo tipo que a informação existente na base de dados relacional. Desta forma, é possível recolher, nestas duas fontes, todas as informações relativas aos alugueres efetuados pela empresa.

5.2 Análise dos Dados das Fontes

5.2.1 Fonte de Dados Relacional

A principal fonte de dados é uma base de dados relacional desenvolvida em MySQL. Esta fonte tem informação relativa aos alugueres, assim como de todos os clientes da empresa, inclusive a cidade e o país de origem dos mesmos. Para além destes dados, a OnRoad guardou, nesta base de dados, informação relativa aos veículos, aos seguros efetuados e, por fim, a informação necessária sobre os funcionários da empresa. A Tabela 4 apresenta uma descrição de todos os atributos existentes na fonte relacional.

Tabela 4- Tabela com Dicionário dos dados

Entidade	Atributo	Descrição	Tipo de dados e tamanho	Nulo	Multi-valor
	Id_Cliente	Código que identifica um cliente	INT	N	N

Cliente	Nome	Nome do Cliente	VARCHAR(80)	S	N
	Data de nascimento	Data de nascimento do cliente	DATE	N	N
	Telemovel	Número de telemóvel do cliente	INT(9)	S	N
	Email	Email do cliente	VARCHAR(45)	S	N
	Rua	Rua do cliente	VARCHAR(100)	S	N
	Carta de condução	Bolleano que esclarece se o cliente é ou não portador de carta de condução	BOOLEAN	N	N
	Nif	Número de contribuinte do cliente	INT(5)	S	N
Funcionário	Id_funcionario	Código que identifica um Funcionário	INT	N	N
	nome	Nome do Cliente	VARCHAR(80)	N	N
	salario	Salário corresponde ao valor que o funcionário é remunerado.	DECIMAL(8,2)	N	N
	Data_contratação	Data correspondente ao dia em que o funcionário foi contratado.	DATE	N	N
	Data_nascimento	Data correspondente ao dia de nascimento do funcionário.	DATE	N	N
	email	Email do funcionário	VARCHAR(45)	N	N
	telémovel	Número de telemóvel do funcionário.	INT	N	N
	rua	Nome da rua onde o funcionário reside.	VARCHAR(75)	S	N
Aluguer	Id_aluguer	Código que identifica um Aluguer	INT	N	N
	caução	Valor que cliente paga como caução do aluguer.	DECIMAL(8,2)	N	N
	Data_real_entrega	Data e hora em que o cliente entregou o automóvel.	DATETIME	S	N
	Kms_percorridos	Número de quilómetros percorridos durante o aluguer.	DECIMAL(8,2)	S	N
	Data_aluguer	Data em que o cliente assinou contrato de aluguer.	DATE	S	N
	Data_prevista_entrega	Data estipulada para entrega do automóvel.	DATE	S	N
	Data_prevista_levantamento	Data estipulada para o cliente levantar o veículo.	DATE	S	N

Veiculo	Id_veiculo	Código que identifica um Veículo	INT	N	N
	Taxa_desvalorizaç ao	Valor associado a cada veículo alugado.	DECIMAL(8,3)	S	N
	Nº_kms	Número de quilómetros que o veículo já possui na data de levantamento.	DECIMAL(8,2)	S	N
	matrícula	Matrícula que identifica o veículo.	VARCHAR(20)	N	N
	Preço_em_novo	Preço que o veículo custou à empresa <i>OnRoad</i> .	DECIMAL(8,2)	S	N
	Marca	Marca do veículo.	VARCHAR(50)	S	N
	Modelo	Modelo do veículo.	VARCHAR(50)	S	N
	Ano_compra	Data relativa ao ano em que o veículo foi comprado.	DATE	S	N
Seguro	Id_seguro	Código que identifica um Seguro	INT	N	N
	Preço_seguro	Valor que custa o seguro.	DECIMAL(8,2)	S	N
	Data_validade	Data de validade do seguro relativamente ao seguro.	DATE	S	N
Cidade	Id_cidade	Código que identifica uma Cidade	INT	N	N
	designação	Nome da cidade à qual pertence o cliente e o funcionário.	TEXT	S	N
País	Id_país	Código que identifica um País	INT	N	N
	designação	Nome do país à qual pertence o cliente e o funcionário.	TEXT	S	N

5.2.2 Fonte de Dados Excel

A segunda fonte de dados disponibilizada corresponde a um ficheiro Excel da empresa emViagem, que contém informação relativa a alugueres efetuados pela mesma. Como tal, em qualquer um dos campos que compõem esta segunda fonte, poderão haver erros de escrita. Assim, é necessário fazer uma revisão de todo o ficheiro e verificar que, a título de exemplo, a marca de um carro não é escrita de maneira diferente em dois casos distintos. Este ficheiro apresenta a informação abaixo mencionada.

5.3 Desenvolvimento do Esquema de Mapeamento de Dados

Após a análise das fontes de informação existentes na OnRoad, procedeu-se ao mapeamento de dados. Assim, através do esquema apresentado, é possível verificar a origem e o destino dos dados, assim como, todas as transformações efetuadas de modo a que informação esteja da forma pretendida para o posterior carregamento no *data warehouse* desenvolvido.

5.3.1 MySQL

Tabela 5: Mapeamento de dados a partir da fonte SQL

Tabela Origem	Atributo Origem	Tipo de dados e tamanho	Transformação	Tabela Destino	Atributo Destino	Tipo de dados e tamanho
Cliente	Id_Cliente	INT	Direto	Dim-Cliente	IdCliente	INT
	Nome	VARCHAR(80)	Direto	Dim-Cliente	Nome	VARCHAR(45)
	Data de nascimento	DATE	Direto	Dim-Cliente	DataNascimento	DATE
	Telemovel	INT(9)	Direto	Dim-Cliente	Telemovel	INT(9)
Cidade	designação	TEXT	Juntar as tabelas Cliente e Cidade para retirar a designação	Dim-Cliente	Cidade	VARCHAR(45)
País	designação	TEXT	Juntar as tabelas Cliente e País para retirar a designação	Dim-Cliente	País	VARCHAR(45)
Veiculo	Id_veiculo	INT	Direto	Dim-Veiculo	IdVeiculo	INT
	Nº_kms	DECIMAL(8,2)	Direto	Dim-Veiculo	NrKms	DECIMAL(9)
	matrícula	VARCHAR(20)	Direto	Dim-Veiculo	Matricula	VARCHAR(45)
	Marca	VARCHAR(50)	Direto	Dim-Veiculo	Marca	VARCHAR(20)

	Modelo	VARCHAR(50)	Direto	Dim-Veiculo	Modelo	VARCHAR(20)
Aluguer	DataAluguer	DATE	Direto	Dim-Calendarario	Data	DATE
	DataAluguer	DATE	Aplicar a função <i>MonthName</i> e concatenar com o ano	Dim-Calendarario	Mes	VARCHAR(25)
	DataAluguer	DATE	Aplicar a função <i>Quarter</i> e concatenar com o ano	Dim-Calendarario	Trimestre	VARCHAR(25)
	DataAluguer	DATE	Aplicar a função <i>Year</i>	Dim-Calendarario	Ano	INT
	PreçoAluguer	DECIMAL(8,2)	Direto	TF-Alugueres	PrecoAluguer	DECIMAL(8,2)
	DataAluguer	DATE	Direto	TF-Alugueres	DataAluguer	DATE
	Cliente	INT	Direto	TF-Alugueres	Cliente	INT
	Veiculo	INT	Direto	TF-Alugueres	Veiculo	INT

5.3.2 Excel

Folha Origem	Atributo Origem	Tipo de dados e tamanho	Transformação	Tabela Destino	Atributo Destino	Tipo de dados e tamanho
Dados Cliente	IdCliente	INT	Direto	Dim-Cliente	IdCliente	INT
	Primeiro Nome	VARCHAR	Concatenar o Primeiro Nome com o Último Nome	Dim-Cliente	Nome	VARCHAR(45)
	Ultimo Nome	VARCHAR		Dim-Cliente		
	Data de nascimento	DATE	Direto	Dim-Cliente	DataNascimento	DATE
	Telemovel	INT	Direto	Dim-Cliente	Telemovel	INT(9)
	Cidade	TEXT	Direto	Dim-Cliente	Cidade	VARCHAR(45)
	País	TEXT	Direto	Dim-Cliente	Pais	VARCHAR(45)
Dados Veiculo	IdVeiculo	INT	Direto	Dim-Veiculo	IdVeiculo	INT

	Nrkms	DECIMAL(8,2)	Direto	Dim-Veiculo	NrKms	DECIMAL(9)
	matrícula	VARCHAR(20)	Direto	Dim-Veiculo	Matricula	VARCHAR(45)
	Marca	VARCHAR(50)	Direto	Dim-Veiculo	Marca	VARCHAR(20)
	Modelo	VARCHAR(50)	Direto	Dim-Veiculo	Modelo	VARCHAR(20)
Dados Aluguer	DataAluguer	DATE	Direto	Dim-Calendarario	Data	DATE
	DataAluguer	DATE	Aplicar a função <i>MonthName</i> e concatenar com o ano	Dim-Calendarario	Mes	VARCHAR(25)
	DataAluguer	DATE	Aplicar a função <i>Quarter</i> e concatenar com o ano	Dim-Calendarario	Trimestre	VARCHAR(25)
	DataAluguer	DATE	Aplicar a função <i>Year</i>	Dim-Calendarario	Ano	INT
	PreçoAluguer	DECIMAL(8,2)	Direto	TF-Alugueres	PrecoAluguer	DECIMAL(8,2)
	DataAluguer	DATE	Direto	TF-Alugueres	DataAluguer	DATE
	Cliente	INT	Direto	TF-Alugueres	Cliente	INT
	Veiculo	INT	Direto	TF-Alugueres	Veiculo	INT

5.4 Revisão do Esquema de Mapeamento de Dados

O esquema de mapeamento de dados do *data warehouse* foi revisto e aprovado pela equipa destacada para a execução do mesmo.

6. Modelação do Sistema de Povoamento

Tendo-se definido o modelo dimensional do *data warehouse* e caracterizado as duas fontes de informação, enumeram-se, neste capítulo, os detalhes de todo o sistema de povoamento, englobando o processo de extração de dados das mesmas, as transformações efetuadas aos mesmos, finalizando com o carregamento da informação tratada para o sistema de suporte à decisão.

O processo de ETL é elaborado com efeito de onda.

6.1 Apresentação e Descrição Sumária das Estruturas do Data Warehouse a Povoar e das Fontes de Dados Envolvidas

Tal como mencionado anteriormente, a empresa OnRoad, ao adquirir a empresa emViagem, teve de lidar com dados provenientes de mais uma fonte. Posto isto, é necessário analisar os dados dos Clientes, dos Veículos e dos Alugueres, de acordo com o apresentado na figura 2.

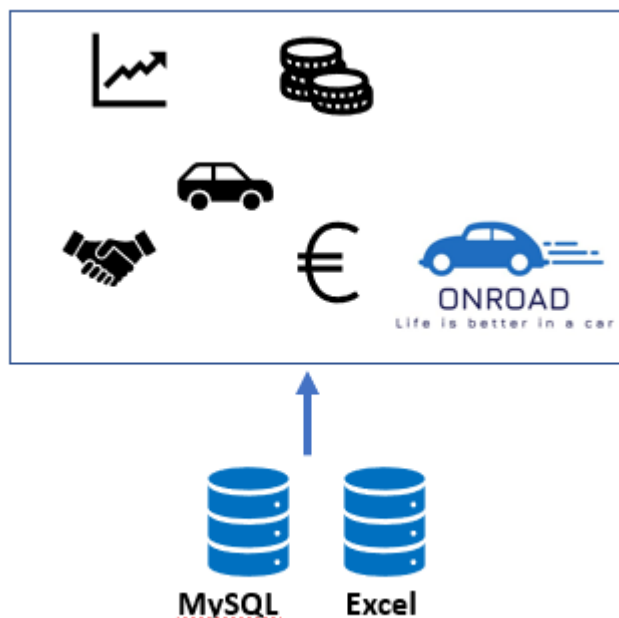


Figura 2: Esquema Geral

6.2 Esquematização conceptual e caracterização do sistema de povoamento

Neste subcapítulo, não só serão definidas e especificadas as transformações necessárias a para cada um dos passos do processo ETL elaborado, como também a modelação conceptual de povoamento em BPMN.

6.2.1 Extração

A primeira etapa do processo ETL é a extração dos dados das diversas fontes. Nesse sentido, extraíram-se os dados relativos às duas fontes de informação existentes, Base de Dados Relacional e Fonte de Dados Excel.

A fonte de dados relacional, na qual foram extraídos todos os dados pretendidos das tabelas Cliente, onde se substituiu os identificadores de cidade e país pelas respectivas designações, Veículo e Alugueres, de acordo com o exposto da figura 3.

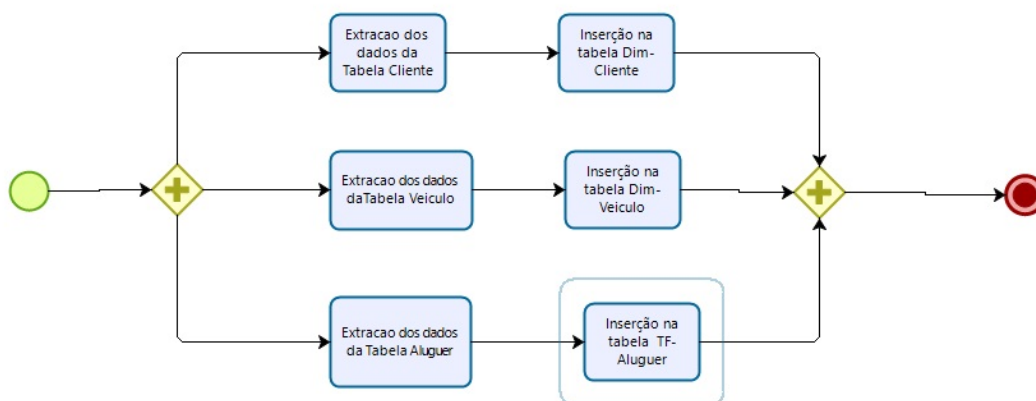


Figura 3: Extração dos dados da fonte de dados SQL

Da fonte dados Excel, foram extraídos os dados pretendidos das folhas DadosCliente, DadosVeículos, DadosAluguer e DadosCalendário conforme o ilustrado na figura 4.

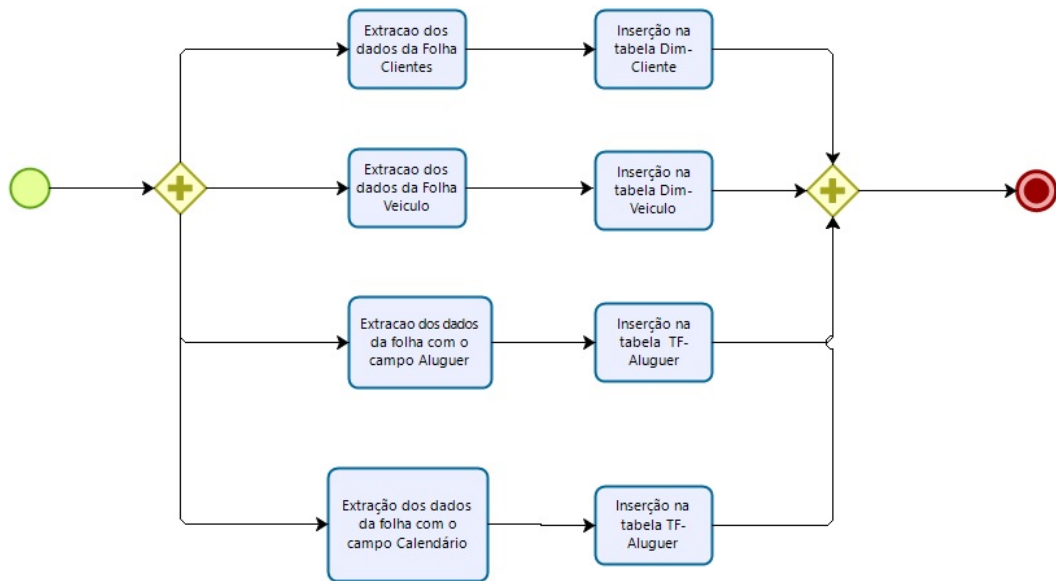


Figura 4: Extração dos dados da fonte de dados em Excel

6.2.2. Transformação

No processo ETL, o passo de transformação é composto por vários processos. Os abordados pela equipa de desenvolvimento do *data warehouse* denomina-se por limpeza.

A fase de limpeza de um processo ETL tem como objetivo limpar os dados nulos da base de dados relacional, nomeadamente nos campos Cidade e Telemóvel, que serão substituídos por 'Desconhecido' e '999999999', respetivamente. A figura 5 evidencia essa mesma situação, exemplificando o caso do atributo Cidade.

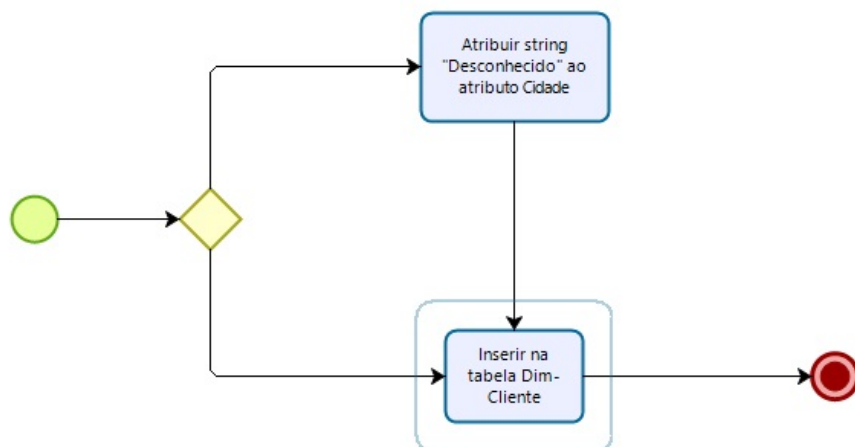


Figura 5: Limpeza de nulos

De seguida, sobre a fonte de dados em Excel, foram concatenados o primeiro nome com o último nome do Cliente, de forma a constituírem apenas um atributo no *data warehouse*, de acordo com a figura 6.

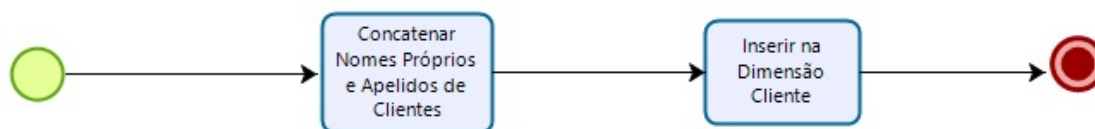


Figura 6: Processo de transformação para a Dim Cliente

A seguinte, foi elaborada sobre as duas fontes de dados, na qual as datas de aluguer dos veículos originam quatro novos atributos, referentes à data de aluguer, ao mês de aluguer, trimestre de aluguer e ano de aluguer, todas extraídas a partir da fonte de dados. Estas transformações foram feitas de acordo com as funções *date*, *monthname*, *quarter* e *year*, respetivamente, em SQL. A figura 7 ilustra este processo.



Figura 7: Processo de Transformação para a Dim Calendário

6.2.3 Carregamento

A última fase do processo ETL é o carregamento (*Loading*). Assim, efetuou-se o carregamento dos dados devidamente transformados para as respetivas tabelas destino, tabela de factos e dimensões, do *data warehouse*.

De acordo com a figura 8, a Preparação do Povoamento é o primeiro passo a ser elaborado. Este processo engloba os processos de extração dos dados, a limpeza dos dados nulos das fontes e a sua substituição pela string 'Desconhecido' e '999999999', de acordo com o exposto na figura 5, a substituição dos identificadores de cidade e país pelas respetivas designações, Veículo e Alugueres, de acordo com o exposto da figura 3, a concatenação dos primeiro e último nome do Cliente, de forma a constituírem apenas um atributo no *data warehouse*, de acordo com a figura 6 e o processo de transformação de datas, que origina quatro novas datas, de acordo com a figura 7.

Após o processo de Preparação de Povoamento acima explicado e, de acordo com a figura 8, podem ser efetuados os povoamentos das dimensões Dim-Cliente, Dim-Veiculo e Dim-Calendário, em paralelo.

Após o povoamento das dimensões três dimensões do *data warehouse*, é necessário povoar a tabela de factos aluguer, denominada de TF-Aluguer. Este povoamento, no final deste processo é então, direto.

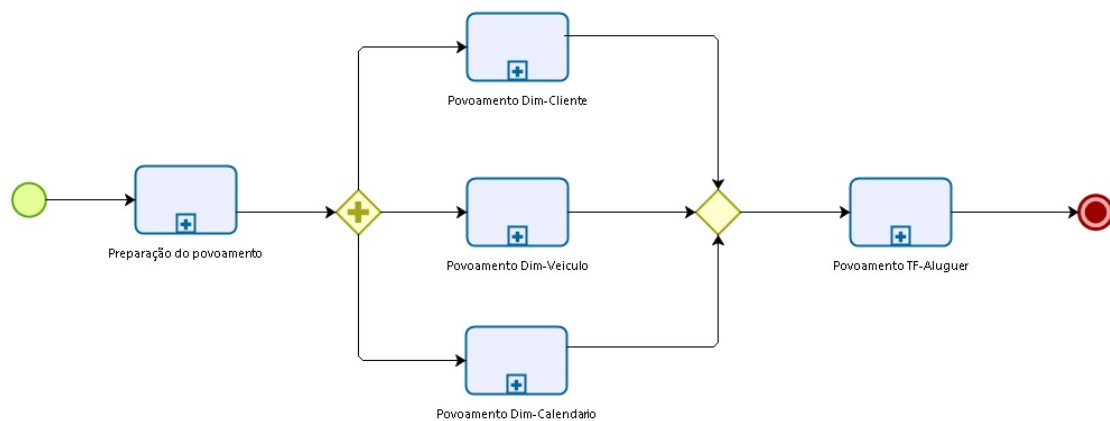


Figura 8: Processo de Povoamento do Data Warehouse

7. Implementação do Sistema de Data Warehousing

O presente capítulo destina-se a explicar todos os processos necessários elaborados a fim de implementar o *Data Warehouse* idealizado.

7.1 Implementação Física do *Data Warehouse*

De acordo com o esquema dimensional, definido anteriormente, procedeu-se ao desenvolvimento no MySQL Workbench do esquema lógico apresentado na figura abaixo. Através deste, o programa gerou a script necessária para a criação do *data warehouse* pretendido, com o auxílio da funcionalidade *forward engineer*. Assim, implementou-se o sistema pretendido, contendo a tabela de factos Aluguer, e as dimensões Calendário, Veículo e Cliente, tal como exposto na figura 9.

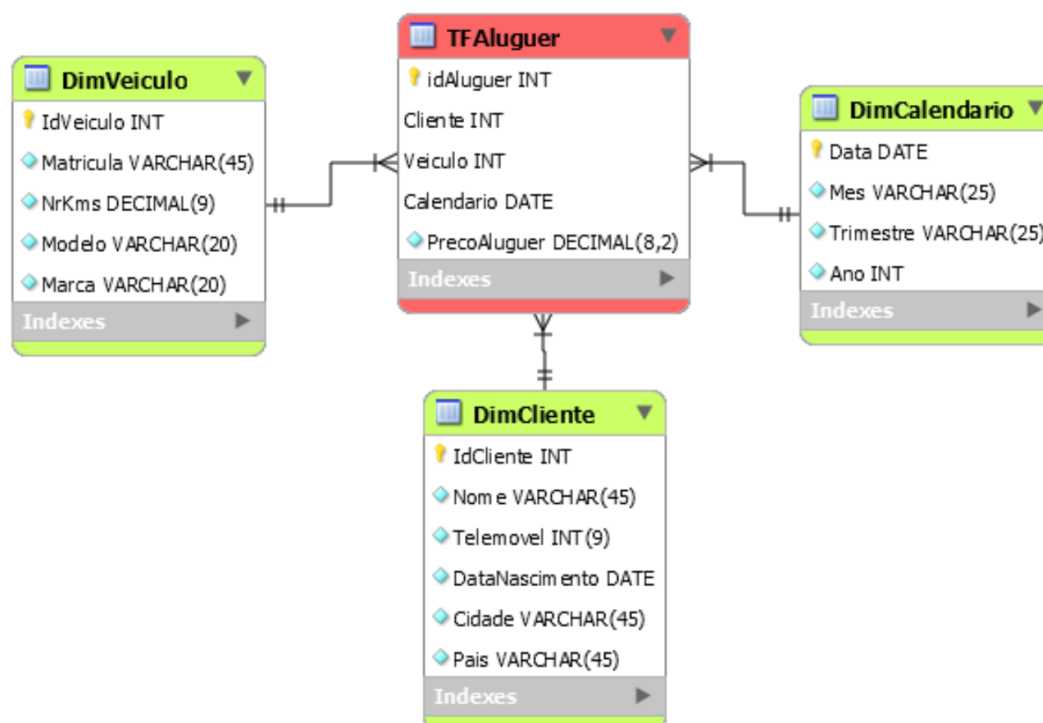


Figura 9: Esquema do *Data Warehouse*

7.2 Implementação Física do Sistema de Povoamento

Após a modelação do *data warehouse*, ed a validação por parte do cliente, inicia-se a implementação do sistema, com o auxílio da ferramenta *Pentaho Kenttle*. Esta permite desenvolver todo o processo ETL de forma simples e razoavelmente rápida.

7.3 Análise da Execução do Sistema de Povoamento

Com a finalidade de executar o sistema de povoamento foi utilizada a ferramenta *Spoon*. Foram considerados aspetos básicos que permitem o povoamento do sistema de forma correta. É esperado que antes do carregamento dos dados não haja qualquer registo inserido.

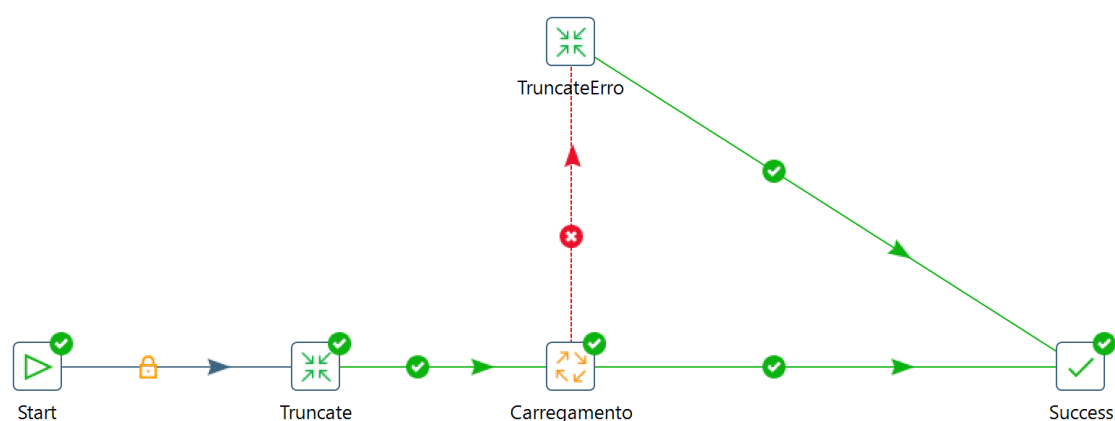


Figura 10 -Esquema Carregamento Inicial

De acordo com o exposto na figura 10, observa-se que para que o processo seja concluído, isto é, para que o *sucess* seja obtido, é necessário um transformador responsável por limpar todos os registos (*truncate*), seguido de um *job* encarregado de efetuar o carregamento dos dados.

Relativamente à limpeza dos registos, foi utilizado um transformador que executa o carregamento das dimensões em paralelo, enquanto que o carregamento dos dados é efetuado em sequência.

O carregamento dos dados, como já referido, foi carregado em paralelo, permitindo assim, o carregamento de todas as dimensões seguindo-se do carregamento

da tabela de factos. Evidencia-se que, apenas será possível efetuar o carregamento da tabela de factos após terem sido carregadas as três dimensões do caso do estudo.

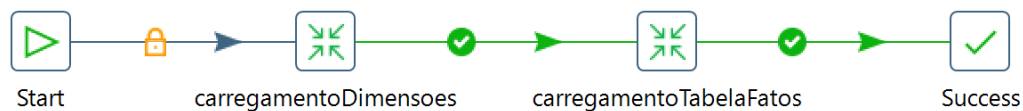


Figura 11: Carregamento em Sequência

De seguida, serão apresentados os esquemas elaborados no *Spoon* que permitiram elaborar o povoamento do *data warehouse*, conciliando os dados da base de dados, correspondente à fonte um, com os dados provenientes do ficheiro excel, referentes à segunda fonte.

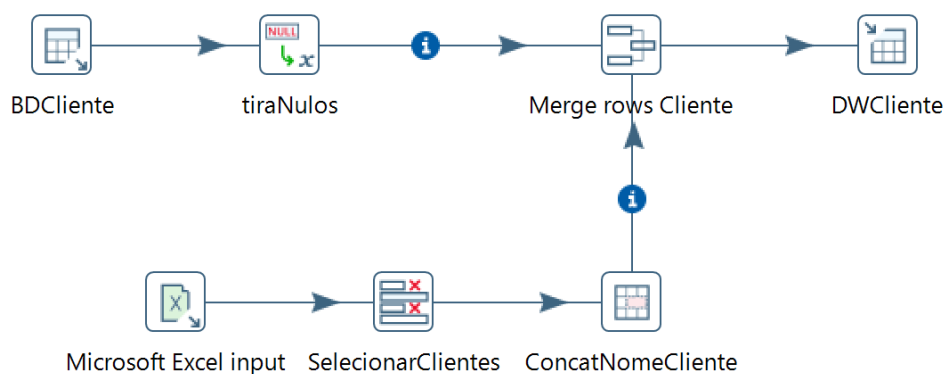


Figura 12: Carregamento dados Clientes

Em seguimento, a figura 12 ilustra a forma como foi carregada a dimensão cliente no *data warehouse*.

Sucintamente, os dados foram extraídos de uma base de dados relacional e de um ficheiro *excel*. Relativamente à extração dos dados oriundos da base de dados, foi aplicado um *step*, responsável pela transformação dos registos. Esta transformação substitui os campos do cliente cujo número de telemóvel é *null* por “999999999” e, da mesma forma, os campos da tabela cidade *null*, pela *string* “Desconhecido”.

Relativamente ao ficheiro Excel, proveniente da empresa emViagem, os dados fornecidos consistiam em folhas no interior do ficheiro Excel, isto é, continham uma folha para cada tipo de dados. Neste caso, acedendo à folha de cálculo “DadosCliente”, foi

possível observar o conteúdo do mesmo. O *step* “*Select Values*”, serviu para selecionar os dados do cliente importantes para a análise do *data mart*. Uma vez que os dados dos clientes apresentavam uma coluna para o primeiro e para o último nome e, os dados provenientes da base de dados, não o apresentavam, foi necessário adicionar o *step* “*ConcatNomeCliente*”, com o objetivo de concatenar o primeiro e o último nome. Este *step* possibilitou a conciliação dos dados no *data warehouse*. Desta forma, através do *step* “*Merge rows (diff)*”, foi possível conciliar as duas fontes e povoar o *data warehouse*.

Relativamente ao carregamento dos dados dos veículos para o *data warehouse*, de acordo com o exposto na figura 13, observa-se que, tal como sucedido na dimensão cliente, também nesta foram considerados como *inputs* um ficheiro excel, correspondente a uma folha de cálculo que contém os dados que caracterizam os veículos da empresa, e a base de dados relacional correspondente à *OnRoad*. Posto isto, foi executada a junção direta, através do *step* “*Merge rows (diff)*”, entre a base de dados com o conteúdo selecionado do ficheiro excel dos dados dos veículos.

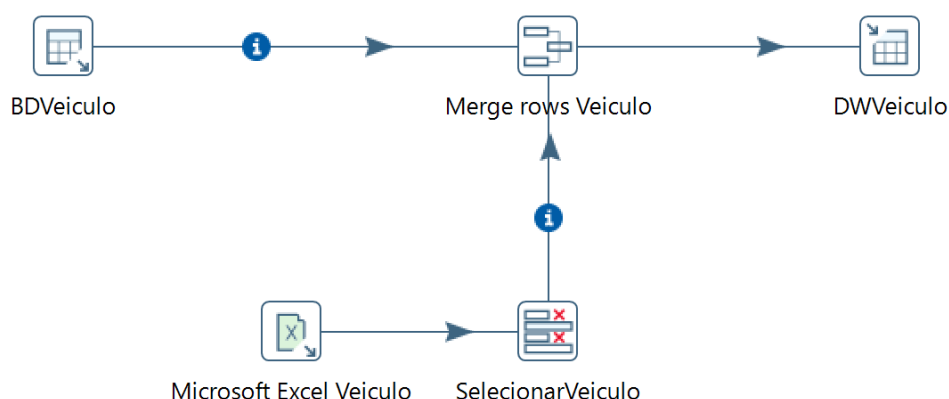


Figura 13 - Carregamento dos dados Veiculo para data warehouse

Em seguimento do trabalho desenvolvido, foi necessário povoar a dimensão calendário. Novamente foram utilizadas as duas fontes. Visto que a ferramenta *Spoon*, por defeito, admite o tempo, em horas, minutos e segundos, juntamente com as datas, considerou-se necessário remover essa componente de tempo antes de serem selecionados os campos que viriam a compor a dimensão Calendário, uma vez que esta contém apenas datas. Para este fim, foi criado no excel, um calendário que compreende nove anos, desde o ano 2015 até ao ano de 2024. De igual modo, foi necessário selecionar as colunas relevantes do ficheiro excel, tais como a data, o trimestre, o mês e o ano, remover as horas automaticamente colocadas pela ferramenta e, novamente, efetuar uma seleção das mesmas. Posto isto, na fase de conciliação, os dados obtidos do ficheiro excel são fundidos com os dados provenientes da base de dados, a fim de serem transportados para o data warehouse.

Este processo descrito acima está explícito na figura 14.

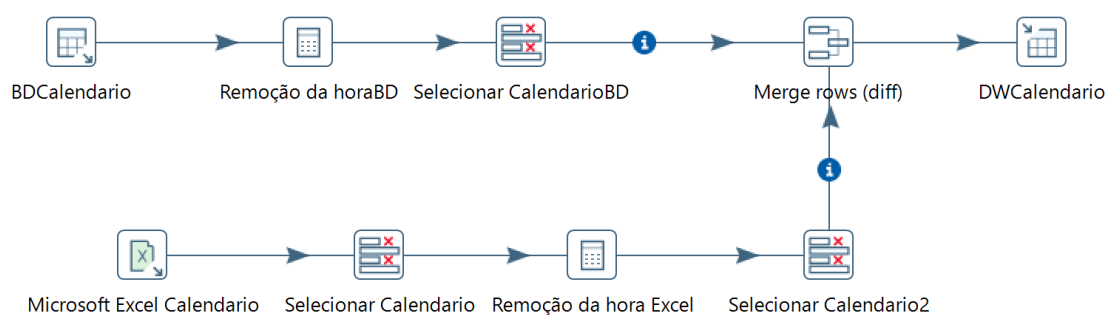


Figura 14 - Carregamento de dados para a tabela DWCalendario

A tabela de factos é a tabela dominante de um esquema de modelação do tipo estrela. É, também, o foco de análise do *data warehouse* em questão. Após o povoamento das dimensões, segue-se o carregamento da tabela de factos. O raciocínio para este carregamento foi semelhante aos utilizados anteriormente. De realçar que, para o carregamento da tabela de factos, apenas foram seleccionadas as colunas data, cliente, veiculo e preço do aluguer, pois são os factos e a medida de desempenho da tabela de factos, respetivamente.

A figura 15 ilustra o carregamento de dados para a tabela de factos em questão.

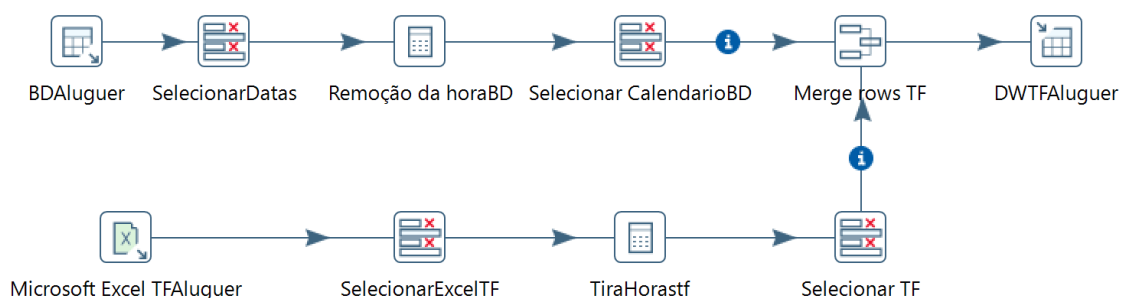


Figura 15 - Carregamento de dados para a tabela TFAluguer

8. *Business Intelligence*

O grupo de desenvolvimento, após a implementação e primeiro povoamento do *Data Warehouse*, utilizou a ferramenta do *Power BI* de forma a visualizar os dados e extrair informação pertinente dos mesmos.

Deste modo, com a finalidade em responder às questões já definidas cujo *data warehouse* tem de dar resposta, apresentam-se as mesmas, juntamente com o respetivo gráfico e explicação.

1. Quais os clientes que fazem mais alugueres?

Perceber os clientes que efetuam mais alugueres na empresa propicia um tratamento mais personalizado aos mesmos se assim a empresa o desejar. De qualquer das formas, esta informação

De acordo com o exposto na figura 16, a cliente que alugou mais veículos denomina-se por Rita Pereira, com 4 veículos.

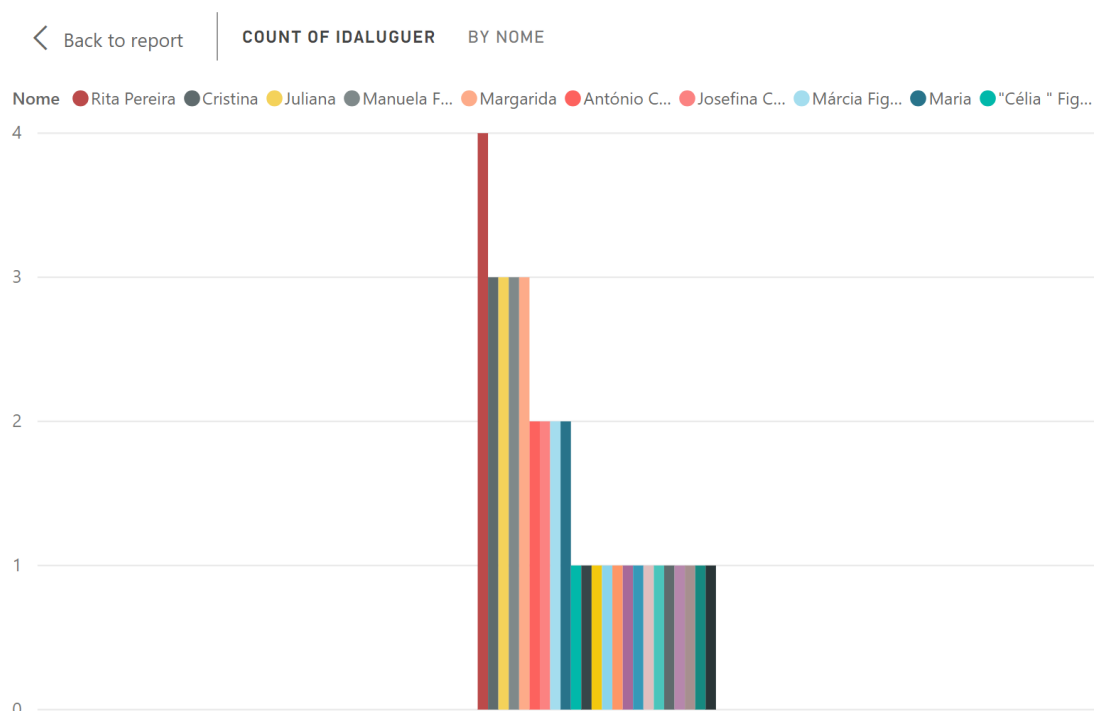


Figura 16: Clientes que mais alugaram veículos

2. Quais os clientes que investiram mais verbas na empresa?

Compreender quais os clientes que investem mais verbas e o valor das mesmas na empresa, possibilita à mesma aplicar promoções em alugueres futuros, tal como pretendido.

Assim sendo, os clientes que investiram mais verbas na empresa foram o cliente Else Costa, com 1250€, o cliente Ricardo Ribeiro com 1120€, a cliente Rita Pereira com 1000€ e a cliente Margarida Marques com 850€. Os clientes já mencionados estão por ordem decrescente, de acordo com as verbas investidas na empresa, como evidenciado na figura 17.

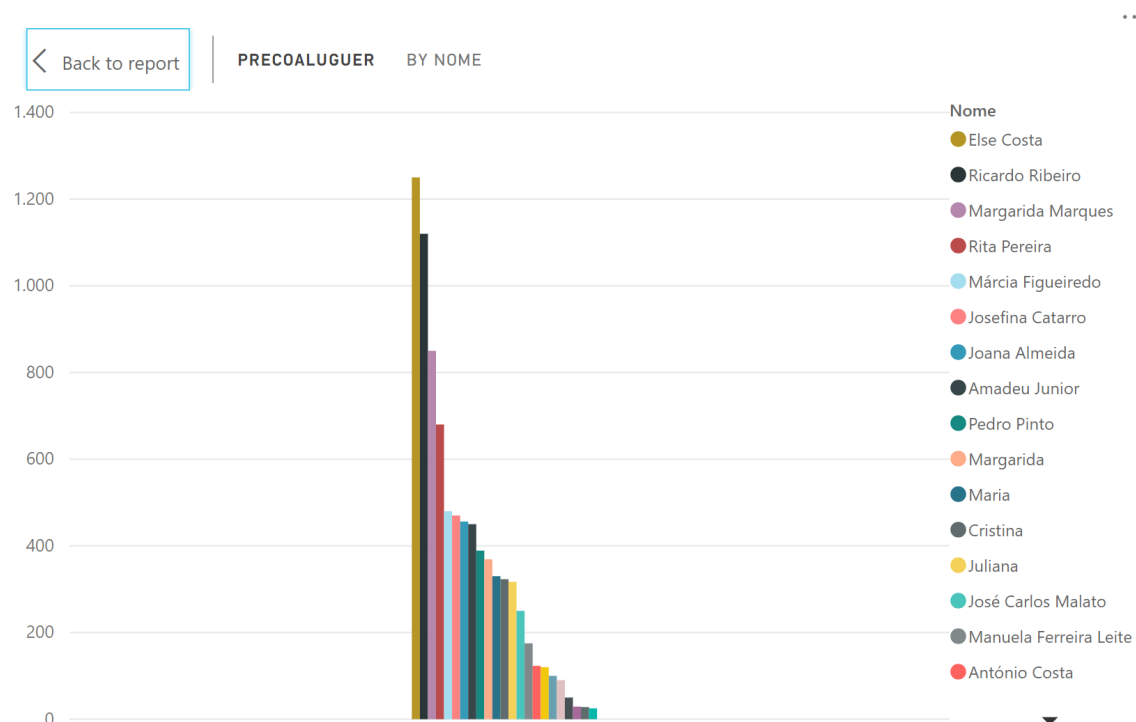


Figura 17: Clientes que investiram mais verbas na empresa

Como se verifica na figura 18, quando os dados são agregados pelo valor médio do valor pago pelos mesmos, verificou-se que o Else Costa, com 1250€, o cliente Ricardo Ribeiro com 1120€, a cliente Rita Pereira com 1000€ e a cliente Margarida Marques com 850€. Os clientes acima mencionados correspondem aos mesmos da figura 17.

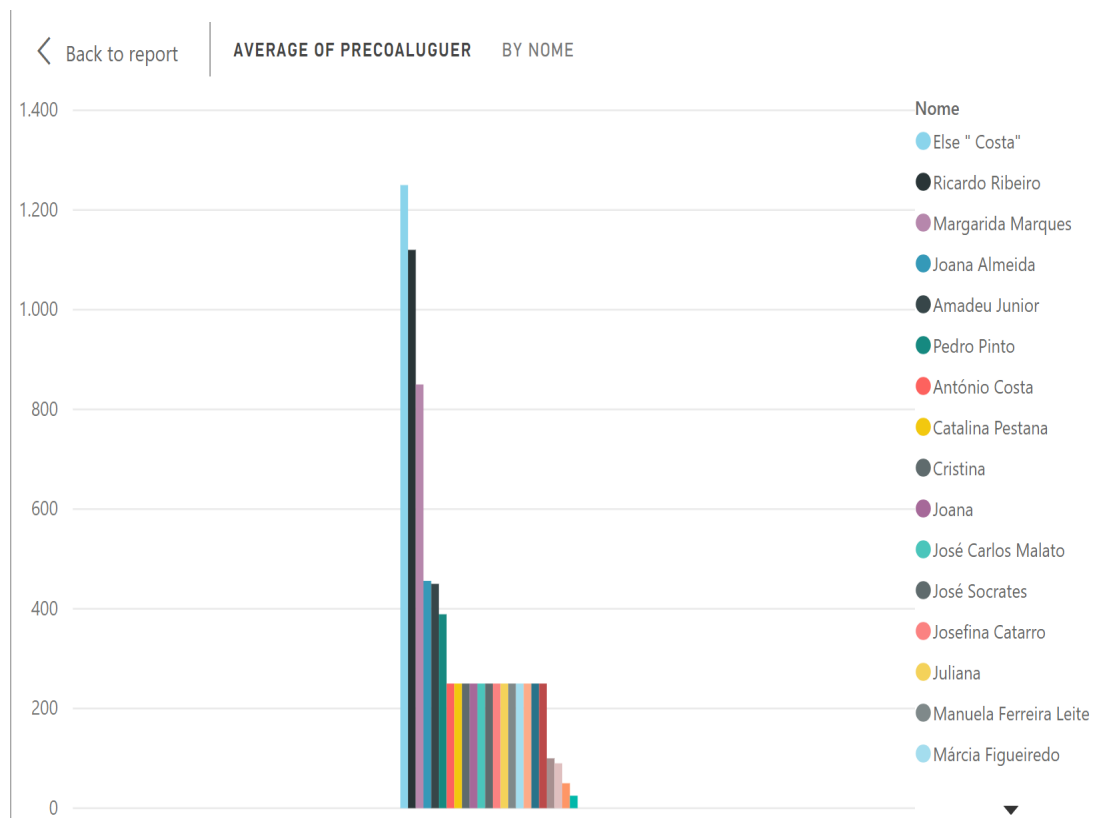


Figura 18 – Clientes que, em média, mais verbas investiram na empresa

3. Quais as marcas de veículos mais alugados?

Como já mencionado, as marcas de carros mais alugadas pelos clientes traduzem a sua preferência, permitindo à empresa gerir de forma mais apoiada, a sua gama de veículos. Posto isto, cumpre-se mais um dos objetivos pretendidos pela OnRoad ao grupo de desenvolvimento do *data warehouse*.

De acordo com a figura 19, as marcas de carros preferidas pelos clientes são a mercedes, com 6 alugueres, seguida pelas marcas Fiat, Hyundai e Renault, com três alugueres cada. Por outro lado, a marca menos alugada é a Kia com apenas dois alugueres.

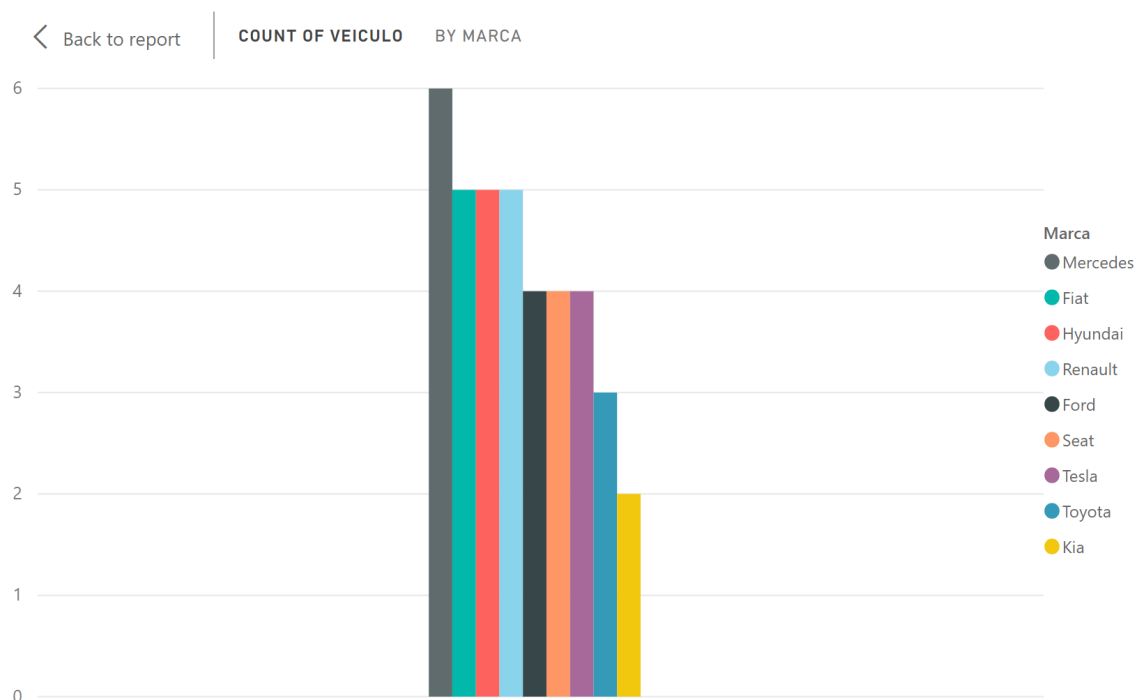


Figura 19: Marcas de carros mais alugadas

4. Quais os meses com mais procura?

De acordo com a figura 20, os meses com mais alugueres de veículos são dezembro de 2018 com sete alugueres e junho de 2018, com seis alugueres. A empresa pretende esta informação para aumentar o número de alugueres nos meses em que o número de alugueres é mais pequeno e, desta forma, aumentar o lucro da empresa. Segundo a figura 20, os meses com menos procura são janeiro e fevereiro, entre outros.

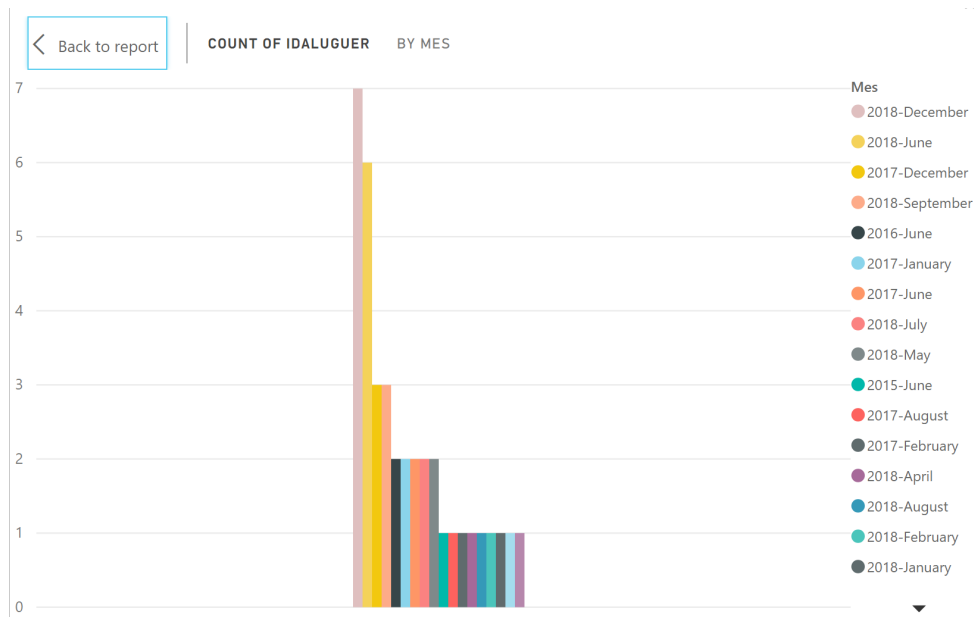


Figura 20: Meses com mais procura

5. Qual o modelo de carros mais alugados?

Como já mencionado, os modelos de carros mais alugados pelos clientes traduzem a sua preferência, permitindo à empresa gerir de forma mais apoiada, a sua gama de veículos. Posto isto, cumpre-se mais um dos objetivos pretendidos pela OnRoad ao grupo de desenvolvimento do *data warehouse*.

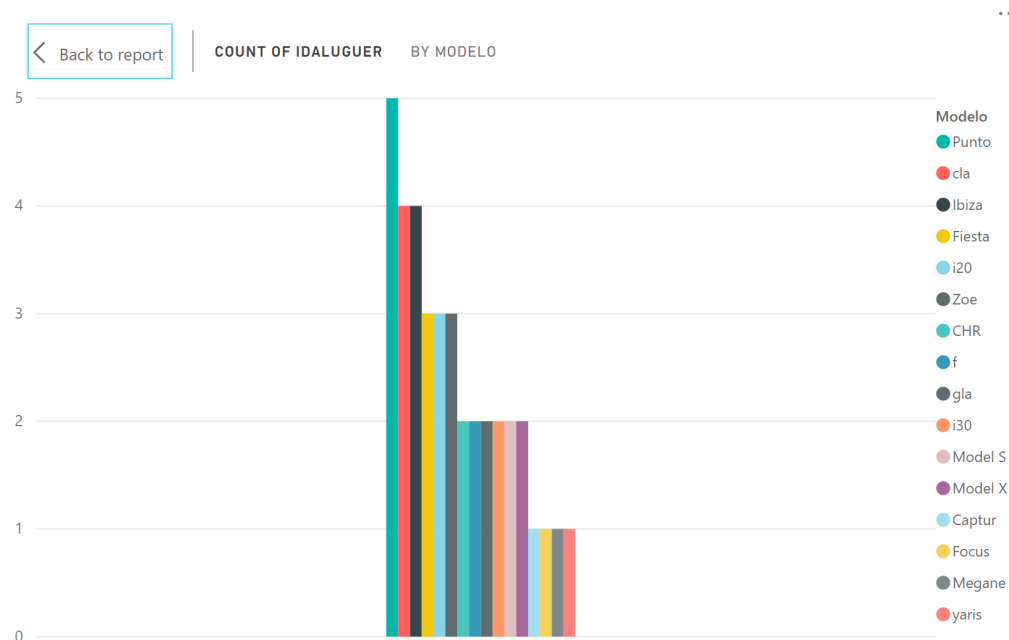


Figura 21: Modelo de carros mais alugados

De acordo com o a figura 21, o modelo de carros mais alugado pelos clientes é o Punto, com cinco alugueres, o cla e o Ibiza com quatro alugueres.

6. Qual o trimestre com maior frequência de alugueres?

A OnRoad pretende também obter a informação relativa ao número de alugueres agregada por trimestres de modo a verificar os trimestres em que o número de alugueres é mais baixo e, através de campanhas promocionais nestes meses, aumentar o lucro da empresa. Através da figura 22, é possível verificar que os trimestres com menor número de alugueres são o segundo trimestre de 2015, o terceiro trimestre de 2017 e o segundo trimestre de 2019.

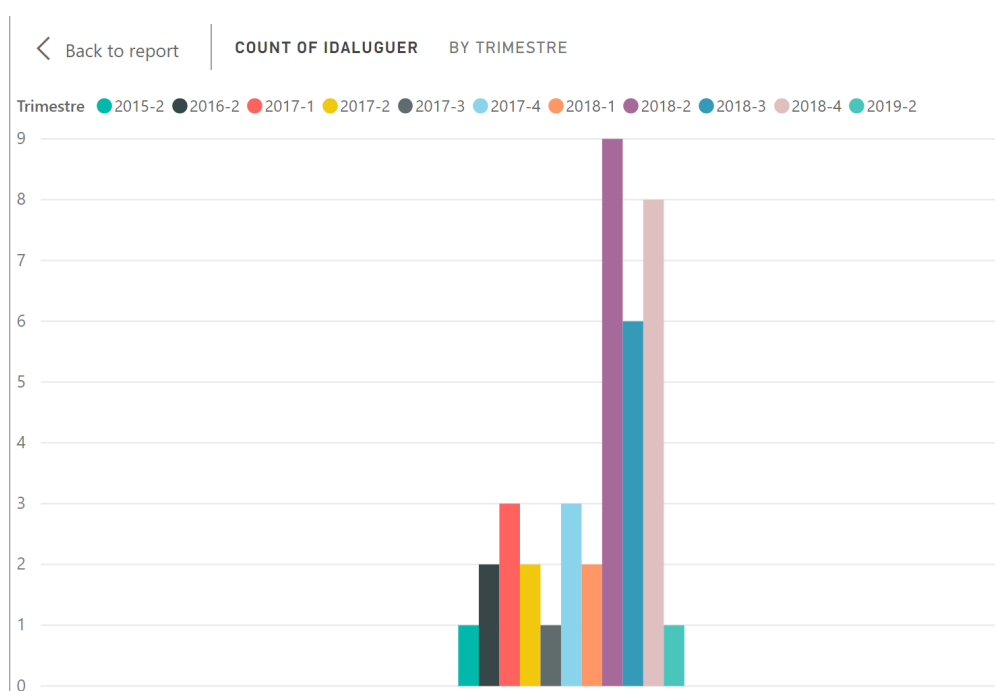


Figura 22: Trimestre de determinado ano com maior frequência de alugueres

7. Quais as cidades com maior frequência de alugueres?

Como as empresas pretendem aumentar o seu volume de negócios, e pretendem abrir mais uma loja de alugueres, verificou-se que seria importante verificar a que localidades pertenciam os seus clientes, com a finalidade de estar mais perto dos seus clientes. Conclui-se que 15,79% é de Guimarães, 13,16 é de Lisboa e 10,53% de Barcelos, portanto seria uma boa opção tentar abrir uma loja num local que fosse vantajoso para estas duas cidades.

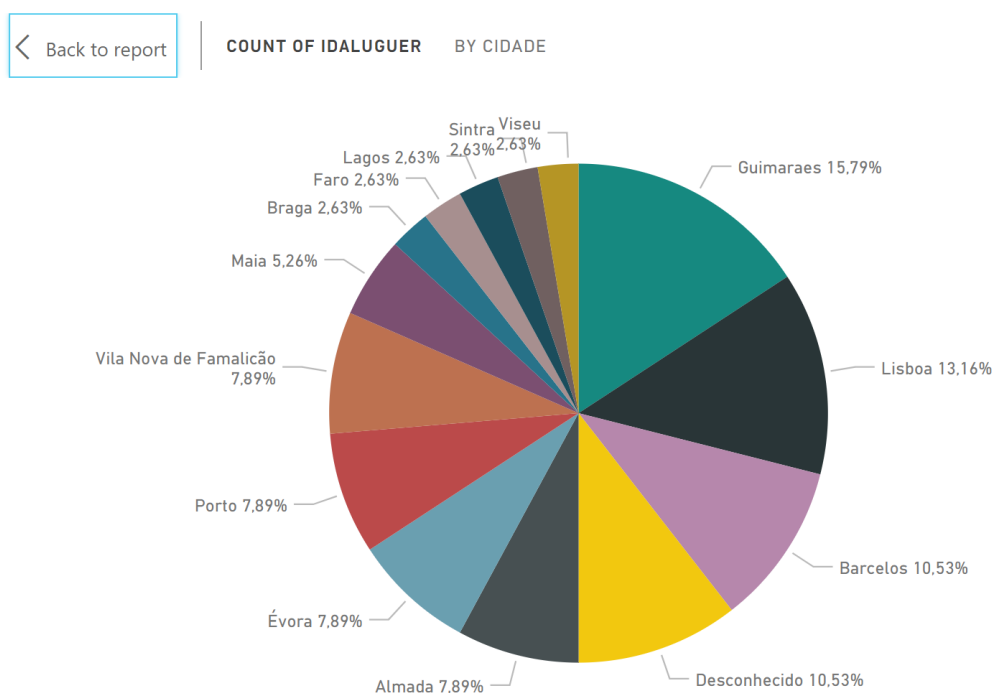


Figura 23 – Gráfico circular representativo das cidades com maior frequência de alugueres

9. Conclusões e Trabalho Futuro

A modelação concetual e a implementação física do *data warehouse* revelaram-se processos complexos, uma vez que os requisitos definidos englobavam alguma dificuldade. O processo de povoamento do *data warehouse* depende, significativamente, da qualidade dos dados existentes nas fontes. Fontes com valores nulos ou que contenham informação incoerente, tornam este processo de povoamento mais demorado, envolvendo mais etapas na preparação dos dados.

Verificou-se também que, o *data warehouse* implementado cumpria os requisitos previamente estabelecidos e respondia às questões pretendidas, permitindo assim analisar os dados existentes de acordo com as pretensões do cliente.

Para trabalhos futuros, sugere-se todo o processo que envolve o segundo povoamento, também conhecido como povoamento regular. O povoamento seria elaborado diariamente, pela ativação de um *trigger* que dispararia sempre que se verificasse um *insert*, *update* ou *delete* em qualquer uma das tabelas das fontes de dados, entre as 0h e as 8h da manhã.

Todo o processo de ETL envolveria a construção e implementação de tabelas de auditoria em áreas de retenção, que acolheriam todos os dados a serem inseridos o no *data warehouse*. Esta área destinar-se-ia a verificar se estes estariam de acordo com o pretendido. Para além da tabela mencionada, teriam de ser incorporadas tabelas de quarentena, cuja utilização reside em etiquetar todos os dados identificados como errados, tabelas de Jornal (LOG), nas quais seriam colocados todos os movimentos que seriam executados no processo ETL e, por fim tabelas de erros, com o objetivo de guardar a informação de todos os erros que foram feitos no processo de extração, transformação e carregamento de dados.

Seria também essencial criar vistas distintas para os diferentes utilizadores do sistema, nomeadamente o administrador e o cliente.

Referências

Golfarelli, M. and Rizzi, S. (2009). *Data Warehouse Design: Modern Principles and Methodologies*. 1st ed. New York: McGraw-Hill.

Lista de Siglas e Acrónimos

BD	Base de Dados
DW	Data Warehouse
OLTP	<i>On-Line Analytical Processing</i>
SDW	Sistema de Data Warehouse

Anexos

Script de criação de DW

I. Script de criação do data mart

-- MySQL Workbench Forward Engineering

```
SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0;
SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS,
FOREIGN_KEY_CHECKS=0;
SET @OLD_SQL_MODE=@@SQL_MODE,
SQL_MODE='TRADITIONAL,ALLOW_INVALID_DATES';
```

-- Schema dwonroad

```
DROP SCHEMA IF EXISTS `dwonroad` ;
```

-- Schema dwonroad

```
CREATE SCHEMA IF NOT EXISTS `dwonroad` DEFAULT CHARACTER SET utf8 ;
USE `dwonroad` ;
```

-- Table `dwonroad`.`DimVeiculo`

```
CREATE TABLE IF NOT EXISTS `dwonroad`.`DimVeiculo` (
  `IdVeiculo` INT NOT NULL AUTO_INCREMENT,
  `Matricula` VARCHAR(45) NOT NULL,
  `NrKms` DECIMAL(9) NOT NULL,
  `Modelo` VARCHAR(20) NOT NULL,
  `Marca` VARCHAR(20) NOT NULL,
  PRIMARY KEY (`IdVeiculo`))
ENGINE = InnoDB;
```

-- Table `dwonroad`.`DimCliente`

```
CREATE TABLE IF NOT EXISTS `dwonroad`.`DimCliente` (
  `IdCliente` INT NOT NULL AUTO_INCREMENT,
  `Nome` VARCHAR(45) NOT NULL,
  `Telemovel` INT(9) NOT NULL,
```

```

`DataNascimento` DATE NOT NULL,
`Cidade` VARCHAR(45) NOT NULL,
`Pais` VARCHAR(45) NOT NULL,
PRIMARY KEY (`IdCliente`))
ENGINE = InnoDB;

```

```

-----
-- Table `dwonroad`.`DimCalendario`
-----

```

```

CREATE TABLE IF NOT EXISTS `dwonroad`.`DimCalendario` (
  `Data` DATE NOT NULL,
  `Mes` VARCHAR(25) NOT NULL,
  `Trimestre` VARCHAR(25) NOT NULL,
  `Ano` INT NOT NULL,
  PRIMARY KEY (`Data`))
ENGINE = InnoDB;

```

```

-----
-- Table `dwonroad`.`TFAluguer`
-----

```

```

CREATE TABLE IF NOT EXISTS `dwonroad`.`TFAluguer` (
  `idAluguer` INT NOT NULL AUTO_INCREMENT,
  `DataAluguer` DATE NOT NULL,
  `Cliente` INT NOT NULL,
  `Veiculo` INT NOT NULL,
  `PrecoAluguer` DECIMAL(8,2) NOT NULL,
  PRIMARY KEY (`idAluguer`, `DataAluguer`, `Cliente`, `Veiculo`),
  INDEX `fk_TFAluguer_DimVeiculo_idx` (`Veiculo` ASC),
  INDEX `fk_TFAluguer_DimCliente1_idx` (`Cliente` ASC),
  CONSTRAINT `fk_TFAluguer_DimVeiculo`
    FOREIGN KEY (`Veiculo`)
      REFERENCES `dwonroad`.`DimVeiculo` (`IdVeiculo`)
      ON DELETE NO ACTION
      ON UPDATE NO ACTION,
  CONSTRAINT `fk_TFAluguer_DimCliente1`
    FOREIGN KEY (`Cliente`)
      REFERENCES `dwonroad`.`DimCliente` (`IdCliente`)
      ON DELETE NO ACTION
      ON UPDATE NO ACTION,
  CONSTRAINT `fk_TFAluguer_DimCalendario1`

```



```
FOREIGN KEY (`DataAluguer`)  
REFERENCES `dwonroad`.`DimCalendario` (`Data`)  
ON DELETE NO ACTION  
ON UPDATE NO ACTION)  
ENGINE = InnoDB;  
  
SET SQL_MODE=@OLD_SQL_MODE;  
SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS;  
SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS;  
SET @@global.time_zone = '+00:00';  
SET @@session.time_zone = '+00:00';
```

II. Script de criação do data mart