U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Modelo relacional, Linguagem SQL e Visualização de dados

Estudo qualitativo da unidade curricular:

Bases de Dados

Curso de Ciência da Informação

Docente: Gabriel de Sousa Torcato David

VENDA DE CARROS

Grupo S

Simão Machado: up201704685

Abril de 2020

FEUP

Sumário

FEUP

1. Nota introdutória

O presente relatório foi realizado no âmbito da unidade curricular de Bases de Dados do terceiro ano do segundo semestre da licenciatura em Ciência da Informação.

De forma implícita, tem como objetivo secundário reforçar as capacidades dos instruendos em analisar, pensar e escolher a melhor estratégia de armazenamento dos dados da organização a adotar de modo a resolver um dado problema da melhor forma possível. De forma concreta, tem como objetivo principal a experimentação das matérias expostas na disciplina, em particular o modelo relacional, a linguagem *SQL* e, neste caso em específico, técnicas de visualização de dados.

Para o caso do modelo relacional, será usado o *SGBD Oracle* como forma de implementação do modelo e utilizar-se-á a *IDE SQL developer*. Em relação à linguagem *SQL*, usar-se-á a extensão do *PL/SQL* da *Oracle*. Relativamente à visualização, será utilizado uma ferramenta que se encontra no pódio das ferramentas mais bem classificadas pelo mercado: *Microsoft Power BI*. Note-se que, aquando da tradução do modelo relacional, construir-se-á, primeiramente, o diagrama de classes respetivo e, uma vez feito o modelo relacional e estiver construído as tabelas no Oracle, irá ser feito a extração e a transformação dos dados de um documento *excel* e o posterior carregamento numa base de dados *Oracle* – processo *ETL* – com a linguagem *Python*.

Assim sendo, para se aplicar tais competências e conhecimentos, escolheu-se, como recomendação da empresa *Data Science Academy*, o conjunto de dados "Venda de carros" retirado do sítio *Web* de governo de dados abertos do Brasil.

Todo este trabalho foi uma mais valia, visto que servirá como um relatório de referência para o meu futuro como gestor de informação.

Portanto, o relatório encontra-se divido do seguinte estado: descrição do problema, modelo de classes, modelo relacional, processo *ETL*, elaboração de perguntas e visualização de dados. Na última parte, será também feita uma conclusão geral sobre o que se usufruiu na realização deste trabalho.

2. Descrição do problema

Nesta parte irá ser feita uma descrição sucinta do problema, nomeadamente, origem, processos organizacionais envolvidos, contexto e objetivo.

O conjunto de dados teve como fonte de dados o sítio *Web* de governo de dados abertos do Brasil (Governo, 2015) a recomendo da empresa *Data Science academy* (Academy, 2016). Inicialmente, o *dataset* selecionado tinha o formato *CSV* que depois foi convertido em *xlsx* (Figura 1 - Conjunto de dados).

Posto isto, o dicionário dos dados é o seguinte:

Nome da coluna	Descrição	Tipo de dado
Data	Data da efetuação da venda	Data (DD-MM-AAAA)
Fabricante	Vendedor do automóvel	Conjunto de caracteres
Estado	Estado onde reside o cliente	Conjunto de caracteres
Valor_venda	Valor total da venda de um automóvel em reais	Inteiro
Valor_custo	Valor total do custo de um automóvel em reais	Inteiro
Desconto	Desconto na venda de um automóvel em reais	Inteiro
Custo_entrega	Custo da entrega de um automóvel em reais	Inteiro
Custo_de_mao_de_obra	Custo da produção de um automóvel em reais	Inteiro
Cliente	Cliente que compra o automóvel	Conjunto de caracteres
Modelo	Modelo do automóvel	Conjunto de caracteres
Cor	Cor do automóvel	Conjunto de caracteres

Analisando o dicionário e tendo sempre em atenção a perpsetiva da visão de processos e de comportamento da modelação organizacional de *Eriksson & Penker* (Vernadat, 2001) – perspetiva que está na base da visão de negócio –, repara-se que o processo representado implicitamente acima é o da venda de carros.

Assim, o contexto do problema será o de avaliar a *performance* do vendedor *Aston Martin* comparativamente aos demais vendedores.

Para esse caso em particular, o objetivo será o de fazer um estudo sobre o total ganho do fabricante em questão numa determinada altura e entregar o relatório de resultados ao departamento de *Marketing* da empresa de forma a poderem elaborar

uma estratéga mais forte – caso seja preciso – ou manter a estratégia – caso estejam num bom caminho[1].

3. Modelo de classes

Depois da descrição do problema, elaborou-se o modelo de classes. Para ter um bem representativo, tentou-se sempre coadjuvar o diagrama com o corrente processo. O meio da sua representação foi a utilização do modelo da perspetiva de estrutura *UML* de *Eriksson & Penker* (Vernadat, 2001). O plano assumido nesta secção foi o seguinte:

1. Indicar as classes [associação] e atributos
2. Estabelecer associações
   2.1. Especiais
   2.2. Normais n-árias

Relativamente ao primeiro ponto, identificou-se as classes dos fabricantes, clientes, automóveis e a classe-associação vendas: para os fabricantes, associou-se o atributo do nome do fabricante; para os clientes, associou-se o atributo do nome dos clientes e o atributo do estado; para os automóveis, associou-se os atributos da cor, modelo, valor do custo, custo da entrega, custo da mão de obra e o desconto e; para as vendas, associou-se os atributos do valor da venda e a data. Os nomes das classes passaram todas para o singular.

Falando agora do segundo ponto, relativamente às associações especiais, não se identificou nenhuma associação do tipo generalização/especialização, agregação, decomposição nem de dependência, pelo menos nenhuma relevante o suficiente para este problema em específico.

Ainda no mesmo ponto, mas agora em relação às associações normais, identificou-se uma associação ternária entre fabricantes, clientes e automóveis e uma classe-associação vendas: a associação fabricante-automovel é de um para muitos; a associação cliente-automovel é de um para muitos e; a associação fabricante-cliente é de um para um[2].

Assim sendo, e verificando a Figura 2 - Diagrama de classes UML, a cardinalidade pode-se ler da seguinte maneira: um objeto da classe fabricante pode vender vários objetos da classe automóvel ou um objeto da classe automóvel pode ser vendido por um objeto da classe fabricante e; a um objeto da classe cliente pode ser vendido vários objetos da classe automóvel ou um objeto da classe automóvel pode ser vendido a um objeto da classe cliente. Como em todas as associações está sempre presente a ação da venda, cria-se uma classe-associação com os seus próprios atributos.

1 De salientar o facto de que se podia ter escolhido outro problema e que, se fosse o caso, o objetivo tenderia a mudar.
2 De notar que existe várias formas de se desenhar um diagrama.

FEUP

4. Modelo relacional

Posteriormente ao diagrama, fez-se a respetiva tradução para um esquema relacional. Para ter um bem representativo, tentou-se sempre coadjuvar o esquema com o corrente diagrama.

O plano assumido nesta secção foi o seguinte:

1. Traduzir classes para relações
2. Traduzir associações
   2.1. Especiais para relações
   2.2. Normais n-árias para relações

No que diz respeito ao primeiro tópico, para cada uma das classes, acrescentou-se um atributo referente ao identificador de objeto – chave artificial primária[3] – e atribuiu-se a cada atributo um domínio. Os nomes das classes – agora denominadas como relações – passaram todas para o plural.

Em relação ao segundo tópico, não se traduziu nenhuma associação especial para relações e fez-se a tradução da classe-associação – nascida a partir da associação ternária – para uma relação. Nesta, acrescentou-se três atributos que eram os referentes ao identificador de objeto das outras relações – chave estrangeira – e que funcionaram como identificador de objeto da mesma – chave artificial conjunta primária. Da mesma forma que numa classe normal, atribuiu-se a cada atributo um domínio.

Com isto, e verificando a Figura 3 - Modelo relacional, o esquema pode-se ler da maneira seguinte: um objeto da classe fabricante que vende a um objeto da classe cliente, poderá vender vários objetos da classe automóvel ao cliente, sendo que cada um deles tem um valor de venda e uma data que são valores únicos.

3 Não se escolheu o nome do fabricante nem o nome do cliente como chave natural primária para as duas diferentes relações pois seguiu-se um específico processo de tradução. Se se tivesse escolhido aqueles atributos como chaves, o resultado seria o mesmo pois os valores de ambos devem ser únicos .

FEUP

5. ETL

Aqui será utilizado o processo *ETL* em bases de dados.

Após se ter construído o modelo relacional, é preciso, agora, extraír os dados da fonte, modificá-los e, seguidamente, armazená-los na base de dados (Figura 4 - Processo *ETL*). Porém, antes de começar este processo, faça-se, em primeiro lugar, a construção do esquema relacional na *Oracle* (Figura 6 - Criação de tabelas).

Utilizou-se a linguagem *SQL* como opção para a criação das tabelas – mais concretamente, a linguagem de definição de dados (LDD). Por conseguinte, tendo em conta o que se fez no modelo relacional, passou-se o mesmo para a base de dados, não descurando as restrições que foram implementadas.

De seguida, começou-se com o processo *ETL*.

O *ETL* pode ser descrito, por outras palavras, como a extração de dados de diversos sistemas, transformação desses dados conforme regras de negócios e, por fim, o carregamento dos dados geralmente para um *Data Mart* e/ou *Data Warehouse*. No entanto, nada impede que também seja para enviar os dados para um determinado sistema da organização (Ferreira, Miranda, Abelha, & Machado, 2010).

Por isso, é constituído por três fases cruciais:

1. Extração
2. Transformação
3. Carregamento

A ferramenta utilizada para este processo foi a linguagem *Python*, através do *Jupyter Notebook* – uma espécie de *IDE* no navegador *Web*.

No que diz respeito à extração, as tarefas desenvolvidas foram as seguintes:

1. Importar pacote do *Oracle* e do *Pandas*
2. Importar o conjunto de dados “Venda de carros”
3. Mostrar o conjunto de dados para verificação de eventuais erros

No que toca à transformação, as tarefas desenvolvidas foram as seguintes:

1. Mini pré-processamento
   1.1. Dividir o *dataset* em várias tabelas, cada uma com os seus atributos e os seus próprios registos únicos
      1.1.1. Alterar dados peculiares
      1.1.2. Criar uma nova coluna da data
      1.1.3. Substituir coluna velha da data pela nova
   1.2. Preparar os dados para serem inseridos

No carregamento, as tarefas desenvolvidas foram as seguintes:

1. Conectar à base de dados

2. Criar um cursor para executar comandos *SQL*
3. Criar os vários comandos para inserir registos nas várias tabelas (LMD)
4. Funções para inserir registos
5. Inserir registos automaticamente

Com isto, o processo *ETL* foi dado como concluído com sucesso.

FEUP

6. Elaboração de perguntas

Como forma de se explorar o conjunto de dados na base de dados, utilizou-se a linguagem *SQL* – mais concretamente, a linguagem de interrogação de dados (LID). Não obstante, a mesma foi utilizada também para se avaliar a complexidade das *queries* e a competência em se fazer consultas[4]. Para cada uma das consultas será apresentado o resultado.

As perguntas elaboradas à base de dados foram as seguintes (Figura 5 - Perguntas):

- Somente para os casos em que o total da soma do valor ganho de cada fabricante é superior à média do total da soma do valor ganho de todos os fabricantes, qual o fabricante que tem o maior valor? Indique o nome e o máximo.
- Para cada ano, para cada estado, quantas vendas foram feitas e qual o valor total ganho somente para o fabricante que tem o maior valor? Indique o ano, estado, numero de vendas e respetivo valor total.
- De todos os clientes, quais foram aqueles que compraram, pelo menos, um carro a todos os fabricantes e voltaram a comprar a, pelo menos, um fabricante? Indique quais os clientes e valor total gasto
- Crie uma nova coluna com a soma do custo de entrega e mão de obra em termos qualitativos. Calcule a frequência para cada categoria.
- Sabendo que os modelos mais atuais dos dois fabricantes com mais modelos têm como nome "DB9" e "Wraith", ordene a lista dos carros mais atuais dos dois fabricantes que têm mais modelos da cor que tenha como início da palavra a letra “V”.
- Quais foram os top 3 fabricantes que mais arrecadaram por ano?

Relativamente à primeira pergunta, de uma forma geral, utilizou-se uma agregação da agregação por linhas com filtro. O resultado foi o seguinte:

	NOME_FABRICANTE	VALOR
1	Aston Martin	3897325

Relativamente à segunda pergunta, de uma forma geral, utilizou-se uma subpergunta variável e uma vista. O resultado foi o seguinte:

4 Aqui o foco foi maior para a construção de comandos com um elevado grau de dificuldade, visto que, quanto mais complexa a *query* se tornava, menor e mais específico se tornava o resultado, não se podendo, assim, extrair o máximo de conhecimento do conjunto de dados de uma forma mais resumida. Para tal, será utilizada uma outra abordagem mais eficaz na demonstração do resultado que responda ao problema que foi definido.

	ANO	ESTADO	NUMERO_DE_VENDAS	VALOR
1	2012	São Paulo	8	932000
2	2012	Minas Gerais	2	220000
3	2013	São Paulo	12	1641000
4	2013	Minas Gerais	9	445670
5	2013	Espírito Santo	6	268990
6	2014	Rio de Janeiro	12	1264190
7	2014	São Paulo	7	799250
8	2014	Minas Gerais	2	199000
9	2015	Rio de Janeiro	26	2457970
10	2015	São Paulo	11	1561250
11	2015	Minas Gerais	8	622540
12	2015	Espírito Santo	7	274180

Relativamente à terceira pergunta, de uma forma geral, utilizou-se a quantificação universal e operadores de conjunto. O resultado foi o seguinte:

	NOME_CLIENTE	VALOR
1	Wheels are us	3121500
2	Honest John	2692600
3	Sporty Types Corp	1987500
4	Embassy Motors	1245450

Relativamente à quarta pergunta, de uma forma geral, utilizou-se o *CASE*. O resultado foi o seguinte:

	CATEGORIA	FREQUENCIA
1	1. Muito baixo	55
2	2. Baixo	256
3	3. Normal	106
4	4. Elevado	33
5	5. Muito elevado	7

Relativamente à quinta pergunta, de uma forma geral, utilizou-se uma consulta hierárquica. O resultado foi o seguinte:

	MODELO	COR
1	DB4	Verde
2	DB7	Vermelho
3	DB9	Verde
4	DB9	Vermelho
5	Rapide	Vermelho
6	Vanquish	Verde
7	Vanquish	Vermelho
8	Vantage	Verde
9	Vantage	Vermelho
10	Zagato	Verde

	MODELO	COR
1	Camargue	Verde
2	Camargue	Vermelho
3	Phantom	Verde
4	Prata Ghost	Verde
5	Prata Ghost	Vermelho
6	Prata Seraph	Vermelho
7	Prata Shadow	Verde
8	Prata Shadow	Vermelho

Relativamente à sexta pergunta, de uma forma geral, utilizou-se uma consulta analítica. O resultado foi o seguinte:

	NOME_FABRICANTE	ANO	SOMA	POSICAO
1	Aston Martin	2012	1152000	1
2	Bentley	2012	919500	2
3	Jaguar	2012	647500	3
4	Aston Martin	2013	2355660	1
5	Rolls Royce	2013	1947300	2
6	Jaguar	2013	1380000	3
7	Aston Martin	2014	2262440	1
8	Bentley	2014	1811500	2
9	Jaguar	2014	1352000	3
10	Aston Martin	2015	4915940	1
11	Rolls Royce	2015	3982600	2
12	Jaguar	2015	2939500	3

7. Visualização de dados

Nesta etapa, apresentar-se-á a solução ao problema "avaliar a performance do vendedor *Aston Martin* comparativamente aos demais vendedores". Para isso, recorreu-se à ferramenta do *Microsoft Power BI*. De igual modo, utilizar-se-á a ferramenta também para explicar, de uma forma geral, o conjunto de dados. Por isso, entender-se-á primeiro o conjunto de dados e depois será dada uma resposta ao problema.

Depois de se ter analisado o conjunto de dados, reparou-se que o mesmo pode ser visto por várias dimensões, nomeadamente: por fabricante; por cliente e; por automóvel. Embora não esteja explícito no modelo relacional, também se pode ver a dimensão do tempo[5]. Assim sendo, foi com esta perspetiva multidimensional que se construiu um *dashboard* que pudesse responder ao problema. Normalmente, se existe tabelas de dimensão, haverá, por consequência, uma tabela de factos. Neste caso, a tabela é a das vendas.

De acordo com a Figura 7 - Dashboard do conjunto de dados" Venda de carros", o conjunto de dados pode ser explicado através das seguintes perguntas principais:

1. Estados dos clientes
2. Valor de vendas dos carros por fabricante e cor
3. Valor de vendas dos carros por ano e fabricante
4. Valor dos custos dos carros por fabricante
5. Valor dos custos dos carros por ano e fabricante
6. Top 3 clientes com mais compras
7. Top 3 modelos com mais vendas

Existe ainda a data das vendas e o valor total ganho por fabricante.

Se se filtrar os resultados pelo vendedor Aston Martin, vemos uma resposta para cada pergunta (Figura 8 - Resposta ao problema 1).

Sobre a primeira pergunta, repara-se que o vendedor Aston Martin tem vendas com clientes em todos os estados, exceto no Paraná e na Bahia.

Sobre a segunda pergunta, repara-se que o vendedor Aston Martin tem um valor de vendas de carros de mais de dez milhões e vende mais carros de cor azul, prata e vermelho.

Sobre a terceira pergunta, repara-se que o vendedor Aston Martin manteve o seu valor de vendas ao longo do tempo com um aumento positivo, exceto entre os anos de 2013-2014.

5 Note-se que se podia ter criado uma tabela só para o tempo. Porém, seguindo um processo rigoroso de tradução do diagrama de classes para o modelo relacional, optou-se por não se criar tal tabela.

Sobre a quarta pergunta, repara-se que o vendedor Aston Martin tem um valor de custos de carros de 7 milhões, aproximadamente.

Sobre a quinta pergunta, repara-se que o vendedor Aston Martin, da mesma forma que na terceira pergunta, manteve o seu valor de custos ao longo do tempo com um aumento positivo. Porém, existe uma ligeira diferença com a terceira pergunta pois o aumento positivo também se manteve entre os anos 2013-2014, ao contrário do valor das vendas.

Sobre a sexta pergunta, dos top 3 clientes com mais compras, cinco carros foram comprados pelo Wheels are Us, dez pelo Bright Orange e nove pelo Aldo Moors.

Sobre a sétima pergunta, dos top 3 modelos com mais vendas, um deles pertence ao vendedor Aston Martin, isto é, o DB9.

Em relação ao valor total ganho do vendedor Aston Martin ao longo do tempo, repara-se que o valor foi de, aproximadamente, quatro milhões, o maior valor ganho entre todos os fabricantes6.

Posto isto, como a performance do vendedor Aston Martin comparativamente aos demais vendedores é muito superior, logo o departamento de Marketing da empresa apenas precisa de manter a excelente estratégia que tem vindo a adotar.

[6] De reforçar a ideia de que existe mais informação a extrair deste dashboard. Apenas se extraiu a informação necessária para responder a um problema de negócio.

FEUP

8. Nota conclusiva

Em síntese, desde a descrição do problema– com a explicação da origem, processos organizacionais envolvidos, contexto e objetivo que funcionaram como a luz guiadora no alcance da tão desejada resposta ao problema de negócio – da elaboração do modelo de classes– com a indicação das classes e atributos e o estabelecimento das associações, quer especiais, quer normais – da tradução para um modelo relacional – com a tradução de classes para relações, bem como a tradução das associaões, quer especiais, quer normais – do processo *ETL* – com a construção do esquema relacional e a extração, transformação e carregamento dos dados – da elaboração de perguntas – de elevado grau de complexidade – até à visualização de dados – com a solução do problema – cumpriu-se o objetivo de experimentar as matérias expostas na disciplina, em particular o modelo relacional, a linguagem *SQL* e, neste caso em específico, técnicas de visualização de dados.

Apesar de alguns problemas terem surgido ao longo do desenvolvimento do trabalho, com esforço, empenho, tempo e perseverança, conseguiu-se realizar o relatório e ultrapassar todos os obstáculos que se depararam no caminho. De facto, a experiência foi benéfica, uma vez que permitiu adquirir um maior conhecimento acerca da realidade nas organizações a diversos níveis. Em qualquer instituição, seja ela privada ou pública, a abundância de problemas não respondidos está em crescimento. Por este motivo, cabe ao gestor de informação o tratamento e recuperação da informação de modo a aumentar a eficiência de uma organização. Todas as organizações enfrentam diversos problemas quando se trata da competitividade no mercado. De forma a que as estas se consigam manter competitivas no mercado, torna-se necessário a utilização criteriosa de técnicas de modelação que permitam responder às necessidades de cada de forma correta.

Após a realização deste trabalho, sentiu-se que pôde olhar para a realidade de uma instituição de forma diferente, assim como as necessidades da mesma, admitindo que esta capacidade de observação será essencial na preparação de futuros profissionais na área da gestão da informação. Assim, aprendeu-se com os erros e foi possível chegar a conclusões.

FEUP

9. Referências

Academy, D. S. (2016). Data Science Academy. Retrieved March 29, 2020, from https://www.datascienceacademy.com.br/pages/home

Ferreira, J., Miranda, M., Abelha, A., & Machado, J. (2010). O Processo ETL em Sistemas Data Warehouse. *INForum 2010 - II Simpósio de Informática*, 757–765.

Governo. (2015). Portal Brasileiro de Dados Abertos. Retrieved March 29, 2020, from http://www.dados.gov.br/

Vernadat, F. (2001). Enterprise modelling. *Production Planning and Control*, *12*(2 SPEC.), 107–109. https://doi.org/10.1080/09537280150501202

10. Anexos

A	B	C	D	E	F	G	H	I	J	K
Data	Fabricante	Estado	Valor_venda	Valor_custo	Desconto	Custo_entrega	Custo_mao_de_obra	Cliente	Modelo	Cor
04/10/2012	Rolls Royce	São Paulo	95000	50000	500	750	750	Aldo Motors	Camargue	Vermelho
01/01/2012	Aston Martin	São Paulo	120000	75000	0	1500	550	Honest John	DBS	Azul
02/02/2012	Rolls Royce	São Paulo	88000	75000	750	1000	550	Bright Orange	Prata Ghost	Verde
03/03/2012	Rolls Royce	São Paulo	89000	88000	0	1000	550	Honest John	Prata Ghost	Azul
04/04/2012	Rolls Royce	São Paulo	92000	62000	0	1500	550	Wheels'R'Us	Camargue	Prata
04/05/2012	Rolls Royce	São Paulo	102500	125000	0	1000	550	Cut'n'Shut	Camargue	Verde

Figura 1 - Conjunto de dados

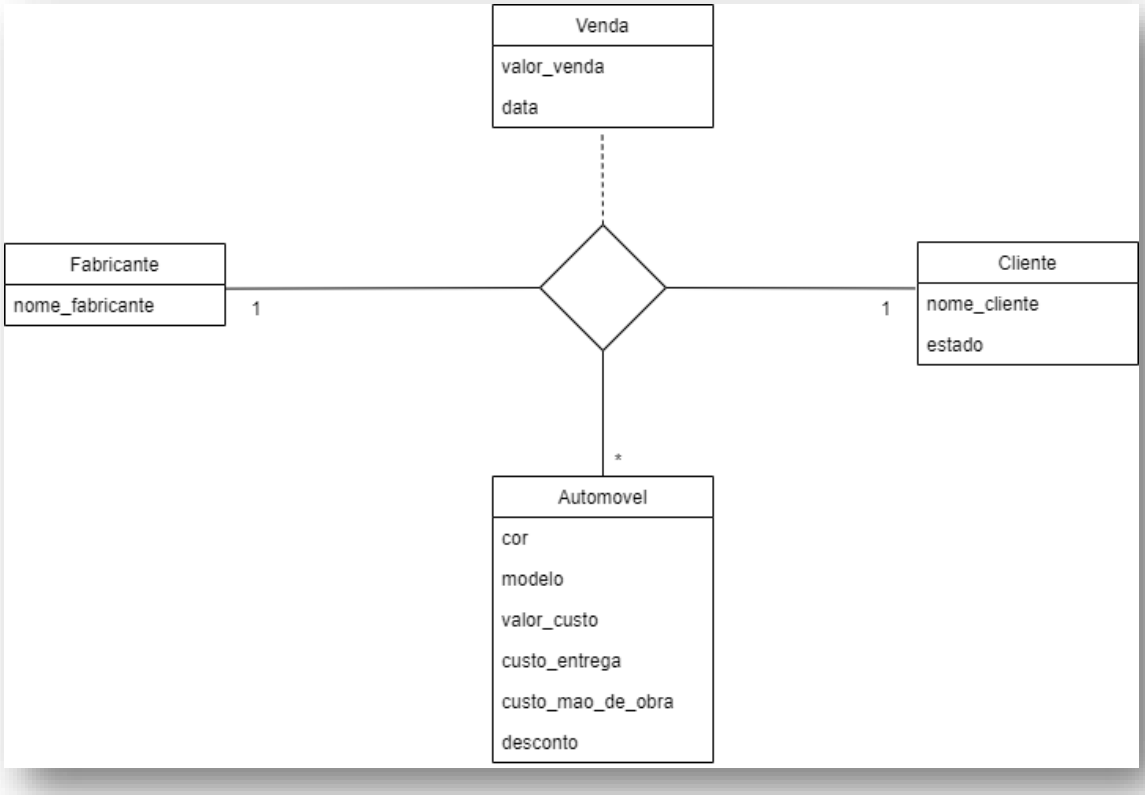

Figura 2 - Diagrama de classes UML

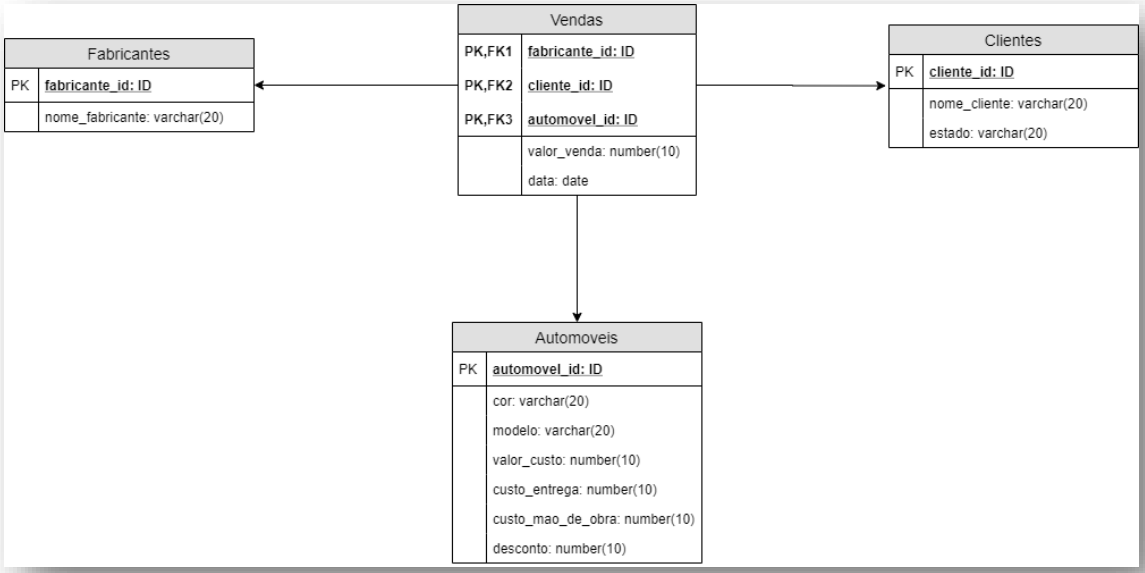

Figura 3 - Modelo relacional

Etl_jupyter_notebook

March 29, 2020

Extração, Transformação e Carregamento dos dados (ETL)

Extração no Excel

[1]:
```
# Importar pacote do Oracle e o Pandas
import cx_Oracle
import pandas as pd
```

[2]:
```
# Importar o dataset "Venda de carros"
dataset = pd.read_excel("Venda_de_carros.xlsx")
```

[5]:
```
# Mostrar o conjunto de dados
dataset.head()
```

[5]:

	Data	Fabricante	Estado	Valor_venda	Valor_custo	Desconto
0	2012-10-04	Rolls Royce	São Paulo	95000	50000	500.0
1	2012-01-01	Aston Martin	São Paulo	120000	75000	0.0
2	2012-02-02	Rolls Royce	São Paulo	88000	75000	750.0
3	2012-03-03	Rolls Royce	São Paulo	89000	88000	0.0
4	2012-04-04	Rolls Royce	São Paulo	92000	62000	0.0

	Custo_entrega	Custo_mao_de_obra	Nome_cliente	Modelo	Cor
0	750	750	Aldo Motors	Camargue	Vermelho
1	1500	550	Honest John	DBS	Azul
2	1000	550	Bright Orange	Prata Ghost	Verde
3	1000	550	Honest John	Prata Ghost	Azul
4	1500	550	Wheels'R'Us	Camargue	Prata

Transformação com Python

[ ]:
```
# Mini pré-processamento dos dados
## Dividir o dataset em várias tabelas, cada uma com os seus atributos e os seus␣
 ↪próprios registos únicos
tabela_fabricantes = pd.read_excel("Venda_de_carros.xlsx",␣
 ↪sheet_name="Tb_fabricantes")

### Alterar dados peculiares
tabela_clientes = pd.read_excel("Venda_de_carros.xlsx", sheet_name="Tb_clientes")
for x, y in enumerate(tabela_clientes["nome_cliente"]):
    if (y == "Wheels'R'Us"):
        tabela_clientes["nome_cliente"][x] = "Wheels are us"
```

1

Figura 4 - Processo ETL

LDD.sql

```
-- LDD
-- Criar tabelas
create table Fabricantes(
    fabricante_id    number(10),
    nome_fabricante    varchar2(20)    not null,
    constraint restricao_fabricante_id primary key (fabricante_id),
    constraint restricao_nome_fabricante unique (nome_fabricante)
);

create table Clientes(
   cliente_id    number(10),
   nome_cliente    varchar(30)    not null,
   estado    varchar2(20)    not null,
   constraint restricao_cliente_id primary key (cliente_id),
   constraint restricao_nome_cliente unique (nome_cliente)
);

create table Automoveis(
    automovel_id    number(10),
    cor    varchar(20)    not null,
    modelo    varchar(20)    not null,
    valor_custo    number(10)    not null,
    custo_entrega    number(10)    not null,
    custo_mao_de_obra    number(10) not null,
    desconto    number(10)    not null,
    constraint restricao_automovel_id primary key (automovel_id)
);

create table Vendas(
    fabricante_id    number(10)    references Fabricantes(fabricante_id),
    cliente_id    number(10)    references Clientes(cliente_id),
    automovel_id    number(10)    references Automoveis(automovel_id),
    valor_venda    number(10)    not null,
    data    date    not null,
    constraint restricao_vendas primary key (fabricante_id, cliente_id, automovel_id)
);
```

Figura 6 - Criação de tabelas

LID.sql

```
-- LID
-- Q1 - Somente para os casos em que o total da soma do valor ganho de cada fabricante
-- é superior à média do total da soma do valor ganho de todos os fabricantes,
-- qual o fabricante que tem o maior valor? Indique o nome e o máximo.
create view fabricante_maior_valor as (select nome_fabricante, soma as valor from (sele
                                                                                    fro
                                                                                    gro
                                                                                    hav


                                       where soma = (select max(soma) as maximo from (se
                                                                                    fro
                                                                                    gro
                                                                                    hav


select * from fabricante_maior_valor;

-- Q2 - Para cada ano, para cada estado, quantas vendas foram feitas e qual o valor tot
select extract(year from v1.data) as ano, c1.estado, count(*) as numero_de_vendas, sum(
from vendas v1 natural join clientes c1 natural join fabricantes f1
where extract(year from v1.data) in (select extract(year from v2.data) from vendas v2
                                     where extract(year from v2.data) = extract(year from

group by extract(year from v1.data), c1.estado
order by extract(year from v1.data), sum(v1.valor_venda) desc;
```

Figura 5 - Perguntas

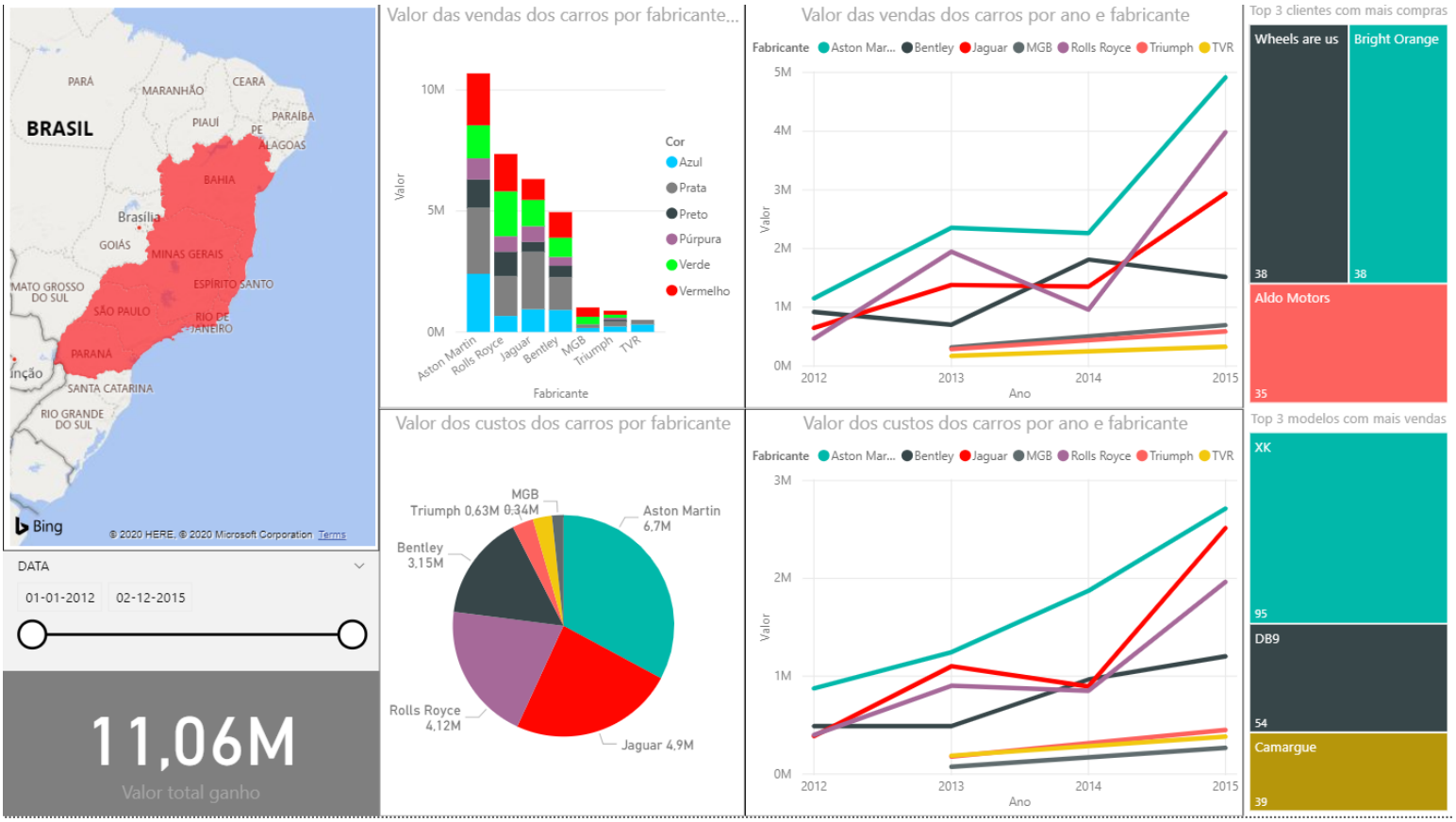

Figura 7 - Dashboard do conjunto de dados" Venda de carros"

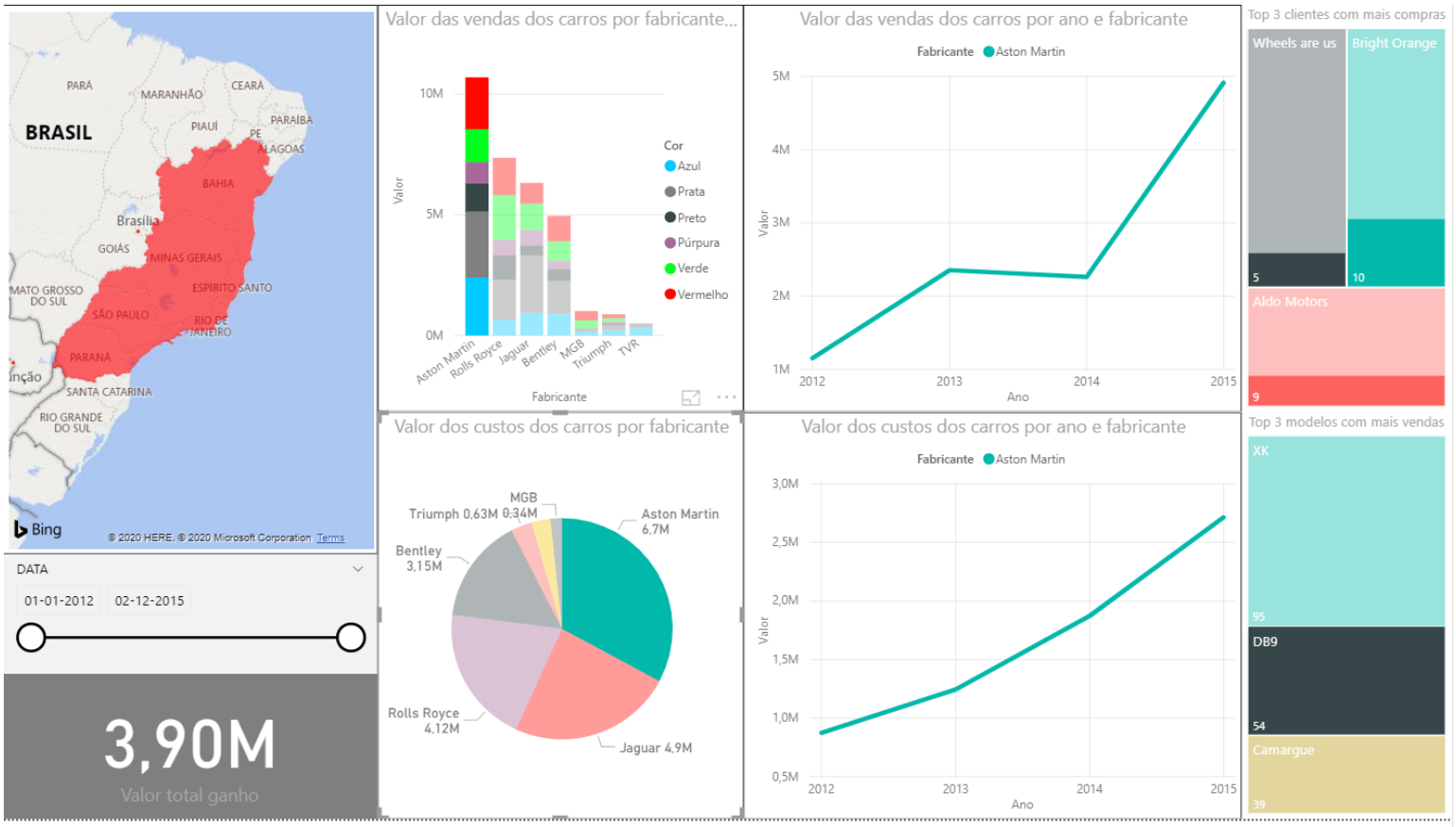

Figura 8 - Resposta ao problema 1