

Enterprise Data Science and Analytics  
October 29th, 2019

## Score Car

David Oliveira<sup>1</sup>, Henrique Pereira<sup>2</sup> & Manuel Oom<sup>3</sup>

<sup>1</sup>M20181430: M20181430@novaims.unl.pt

<sup>2</sup> M20181395: M20181395@novaims.unl.pt

<sup>3</sup> M20181431: M20181431@novaims.unl.pt

### I. Introdução

Existe uma grande variedade de oferta por parte das marcas de automóveis existentes no mercado e a escolha de uma viatura pode ser uma tarefa mais difícil do que poderíamos antecipar. Desde a imensidão de especificações técnicas espalhadas pelos *websites* até aos diferentes preços apresentados para viaturas que, à primeira vista, poderão parecer iguais, a decisão de alguém que pretende selecionar uma viatura torna-se cada vez mais difícil.

O *Score Car* tem como objetivo a disponibilização de toda a informação necessária para ajudar na seleção de uma viatura num conjunto alargado de modelos e versões apresentadas no *website* do Entrepasto Auto. Esta solução pretende simplificar a tarefa de pesquisar a informação relativa a diferentes viaturas, agregando toda essa informação (marcas, modelos, versões, imagens das viaturas, preços, especificações técnicas, entre outros) numa única aplicação, sem necessidade de estar a fazer múltiplas pesquisas para comparar as viaturas. Para complementar essa informação, a nossa solução apresenta ainda preços relativos às fontes de energia e aos combustíveis tipicamente utilizados (Gasóleo, Gasolina, GPL e Eletricidade), de forma a apresentar o custo mensal associada a cada viatura tendo em conta o consumo associada à mesma.

Esta solução além de ajudar os clientes que querem escolher uma viatura, seja por pretenderem efetuar uma compra ou por simplesmente terem a oportunidade de ter uma viatura através das suas empresas, poderá também ajudar os colaboradores do Entrepasto Auto (ou de outros distribuidores automóveis, caso a solução seja alargada a mais viaturas) a terem a informação relativa às viaturas de forma agregada e resumida, permitindo responder de forma rápida a eventuais questões de potenciais clientes com informação real e precisa.

Este relatório apresenta a solução Score Car, desde a extração dos dados do *website* do Entrepasto Auto e a sua transformação até ao modelo final de dados, apresentando detalhadamente cada um dos *reports* da solução e as suas funcionalidades.

### II. Dados

Nesta secção será feita uma breve descrição dos dados utilizados, bem como da abordagem seguida. A abordagem foi composta pelas seguintes fases:

- Extração de Dados;
- Tratamento de dados;
- Modelo;
- Elaboração dos *reports*.

Para efetuar o estudo foram utilizadas 3 fontes de dados (Entrepasto Auto, Mais Gasolina e EuroStat), tendo sido utilizado um *script* Python para efetuar a extração de dados e o *software* Power BI para tratamento de dados e elaboração de *reports*. Para auxiliar o desenvolvimento do estudo pelos vários membros da equipa foi utilizada a plataforma de controlo de versões colaborativa Github (figura 1).



Figura 1 - Ferramentas utilizadas para desenvolvimento da aplicação

## a. Extração de Dados

Para o desenvolvimento da aplicação e do estudo associado à mesma, foram utilizados dados com origem em diferentes fontes:

- [www.entrepостоauto.pt/](http://www.entrepостоauto.pt/) – Dados e especificação de várias viaturas;

Para os dados retirados do *website* do Entrepосто Auto, foi desenvolvido um *script* em Python (*package scrapy*) que colocou os dados de todas as viaturas do *website* num ficheiro JSON. Os dados retirados do site incluem: preço, fotografias, combustível, consumo (do combustível fóssil gasolina ou gasóleo), aceleração 0-100 km/h, altura, comprimento, entre outros. As marcas dividem-se em modelos e os modelos dividem-se em versões.

Inicialmente, foram realizadas tentativas para que o *webscrapping* fosse realizado através do Power BI, no entanto, como o *website* tem um carregamento de dados por *javascript*, não foi possível obter nenhum tipo de dados através desta ferramenta.

- [www.maisgasolina.com/](http://www.maisgasolina.com/) - Dados relativos aos preços dos combustíveis;

Os preços dos combustíveis gasolina 95 simples, 95 “Plus”, 98 simples, 98 “Plus”, gasóleo simples e gasóleo “Plus” foram obtidos da mesma forma que no ponto anterior devido a constrangimentos técnicos idênticos.

- [ec.europa.eu/eurostat/](http://ec.europa.eu/eurostat/) – Dados actuais e históricos de preço da eletricidade praticados na união europeia;

Estes dados foram obtidos através de uma API dedicada da Eurostat. Esta API foi utilizada como *query* dinâmica pela ferramenta de obtenção de dados web do Power BI. Foram retirados apenas os dados do preço do kWh com e sem impostos em Portugal.

## b. Transformação de Dados

Os dados foram obtidos na origem sob diversas formas, nomeadamente formatos JSON e tabulares. Estas origens foram tratadas de forma a normalizar a informação, de modo a que fosse possível construir um modelo de dados com factos e respetivas dimensões.

Outras transformações efectuadas incluem: substituição de valores erróneos, eliminação de linhas com valores nulos, transposição/dinamização de colunas, alteração de tipo de dados, renomeação de títulos de coluna, entre outras alterações. Todas as transformações efectuadas tiveram como objectivo a eficiente e correcta visualização da informação nos seus factos e dimensões anexas, bem como a fácil leitura das mesmas.

Ainda ao nível do Power BI, a componente *Query Editor* foi amplamente utilizada (inclusivamente através da linguagem M) para efectuar estas transformações, sendo que também foi necessário efectuar posteriores transformações através de DAX, com o objetivo de obter medidas, variáveis, parâmetros e colunas adicionais.

## c. Modelo

Na Figura 2, é apresentando o modelo de dados final, após a extracção e transformação dos dados.

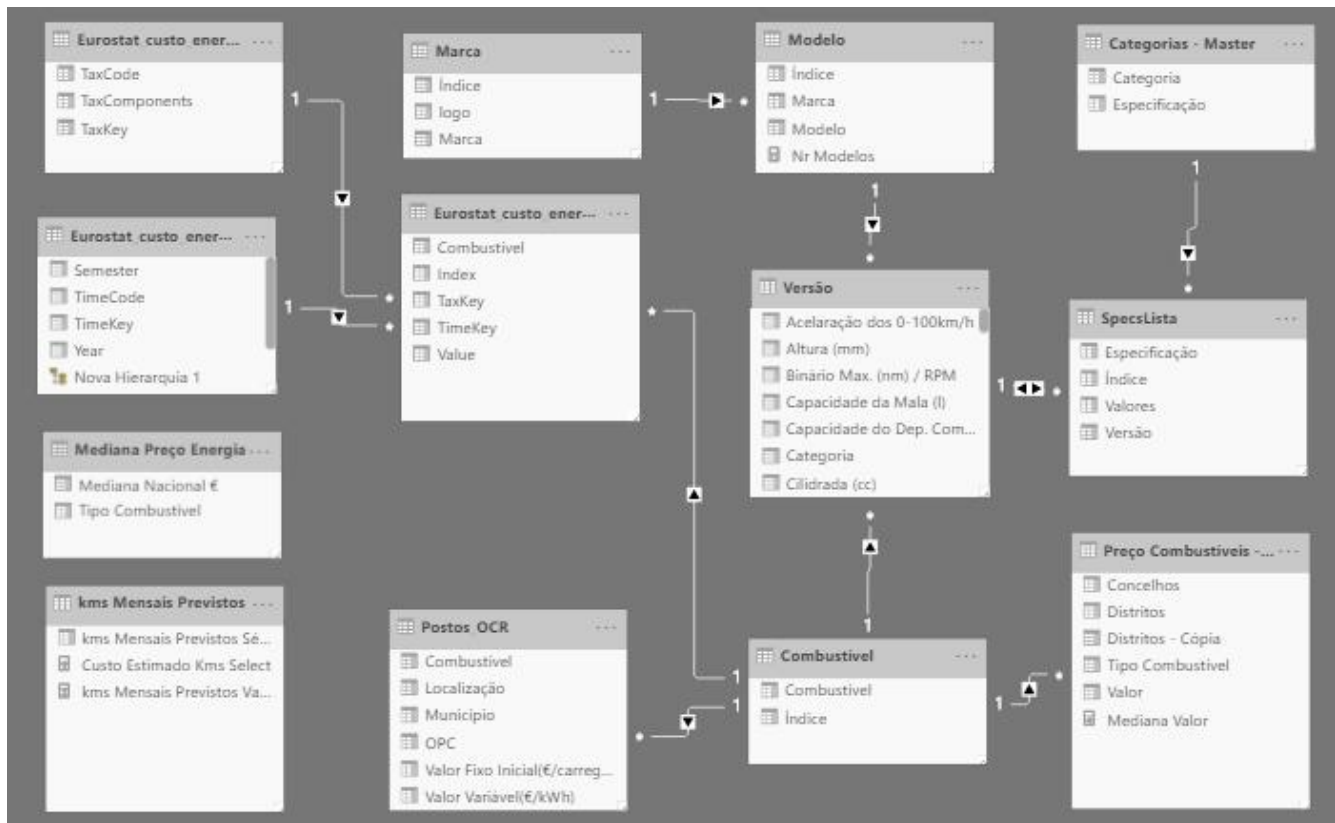


Figura 2 – Modelo de Dados

Os factos dos dados obtidos ficaram congregados nas tabelas “Versão” e “SpecsLista” de modo intercambiável. Isto é, estas tabelas diferem somente na dinamização das suas colunas, o que é útil para algumas visualizações apresentadas nos *reports*. Todas as restantes tabelas são dimensões das tabelas de factos ou tabelas acessórias, também essenciais.

### III. Reports

#### a. Secção “Resumo Geral”

Na Figura 3, é apresentando o *report* do Resumo Geral, onde são exibidas características agregadas do *dataset*: As marcas; o número de modelos por marca; os preços médios por marca; os preços das versões e o número de versões para cada modelo. É ainda possível filtrar as visualizações por marca e modelo ou definir um intervalo para o *budget*. Ao definir o intervalo para o *budget*, o mesmo será aplicado em todos os *reports*, através da sincronização de filtros.



Figura 3 - Report Resumo Geral

## b. Secção “Especificações Técnicas”

Após a apresentação do *report* Resumo Geral, que, tal como expressado anteriormente, permite ter uma visão geral das viaturas extraídas do *website* Entrepasto Auto, os *reports* referentes às Especificações Técnicas apresentam detalhes específicos de cada viatura. A solução *Score Car* apresenta as especificações técnicas de viaturas elétricas e não elétricas separadamente, dado que há detalhes que fazem sentido analisar apenas nas viaturas elétricas, enquanto outro tipo de detalhes apenas fazem sentido avaliar em viaturas não elétricas.

Estes *reports* permitem, tal como no *report* Resumo Geral, definir um intervalo de preço que permite filtrar as viaturas (as imagens das viaturas são apresentadas consoante o filtro aplicado), tal como é possível apresentar apenas as viaturas das marcas seleccionadas no elemento visual *Image Grid*.

Uma característica dos *reports* apresentados está relacionada com a tabela/matriz que apresenta as versões das viaturas e as suas características, sendo que essas características poderão ser filtradas dinamicamente através do *slicer* colocado à esquerda no *report*, possibilitando ao utilizador analisar as especificações que mais lhe interessem. Adicionalmente, a nossa solução apresenta as especificações técnicas agrupados por categoria, nomeadamente: Consumo, Dimensões, Mecânica e Performance.

Finalmente, o *report* das viaturas não elétricas apresenta o preço dos combustíveis (Diesel, Gasolina e GPL) por distrito de Portugal. O *report* das viaturas elétricas apresenta o preço do kWh (com e sem impostos), apresentando ainda a evolução temporal dos últimos dez anos.

Qualquer um destes *reports* dá também possibilidade de navegar pela solução através dos ícones no canto superior direito. Para aceder ao *report* referente às viaturas elétricas é necessário clicar no ícone com o formato de uma lâmpada, sendo que o *report* associado às viaturas elétricas permite também navegar para qualquer um dos outros *reports*.

**Nota:** Nestes *reports* foram aplicados filtros ao nível da página, para que a solução apresentasse as viaturas de acordo com o seu combustível no respetivo *report*.

## i. Viaturas Não Eléctricas



Figura 4 - Report de Especificações Técnicas - não eléctricos

## ii. Viaturas Eléctricas



Figura 5 - Report de Especificações Técnicas - Eléctricos



## c. Secção “Comparação de Viaturas”

No ultimo *report* podemos efetuar a comparação entre duas viaturas. Apesar de nos *reports* anteriores podermos comparar os veículos na tabela de especificações, neste *report*, o objectivo é apresentar uma comparação mais visual. Após filtrar pela versão (podendo esse filtro ser auxiliado pelo filtro de *budget*, marca e/ou modelo), o Score Car exibe algumas características dos dois veículos. Os filtros aplicados ao veículo do lado esquerdo não influenciam os veículos do lado direito apesar da fonte (tabela) dos dados ser a mesma. Estes filtros foram manipulados através das relações entre os elementos visuais, de forma a tornar a seleção dos veículos independente.

As características que podemos comparar são as que se consideraram mais importantes na escolha final de um veículo: Combustível, caixa de velocidades, preço, cilindrada, potência, velocidade máxima, consumo combinado, previsão de custo mensal e custo por cada 100 kms. No caso dos carros elétricos, optou-se por não calcular o custo por 100 km pelo facto de que a informação relativa à eficiência dos mesmos (kWh/100km) não ter sido possível obter para a totalidade das viaturas com este tipo de força motriz no *dataset* (não obstante, o modelo de dados estaria preparado para acomodar tal tipo de informação no futuro).

O retângulo na parte inferior do *report* permite ao utilizador seleccionar o número de quilómetros mensais e, dado o consumo do veículo e o preço do combustível utilizado, fazer uma previsão do custo mensal em termos de combustível.

Em rodapé, existe ainda informação das medianas dos preços dos combustíveis e do kWh.

Adicionalmente, existe ainda um botão de *reset* para limpar os filtros das versões, veículos e marcas, permitindo ao utilizador efetuar uma nova comparação.

À semelhança dos *reports* anteriores, este *report* mantém os ícones para facilitar a navegação dos utilizadores.

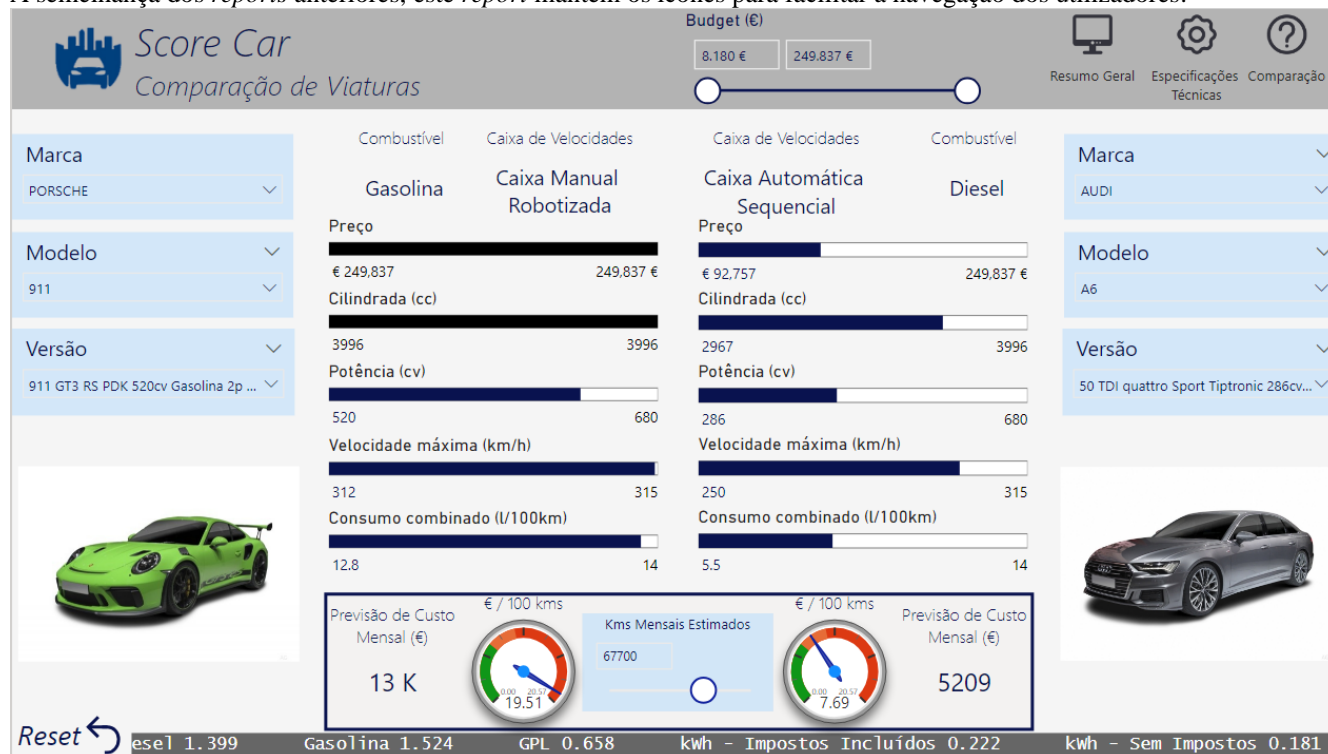


Figura 6 - Report de comparação de viaturas

## IV. Conclusões

Com esta solução, podemos retirar *insights* valiosos sobre os veículos e as suas características num só lugar, sem necessitar de abrir múltiplos separadores ou múltiplos *websites* nem abrir folhas de Excel para efetuar comparações de custos mensais dos veículos. Auxilia-se assim tanto o utilizador final como o vendedor na seleção da viatura a adquirir através de *reports* dinâmicos, actualizáveis e fáceis de utilizar.

A título de desenvolvimento futuro, os seguintes pontos poderão ser abordados:

- Como para esta solução não foi efetuada nenhuma parceria com nenhum vendedor automóvel, não são apresentados valores de vendas. No entanto, seria bastante interessante fazer tal parceria e utilizar técnicas de *advanced analytics* para sugerir conselhos aos clientes ou mostrar os veículos que pessoas nas mesmas condições e que procuram as mesmas especificações acabaram por adquirir;
- A par do anterior acréscimo, adicionar *datasets* da concorrência (externos) seria extremamente útil no caso de se pretender fazer um *benchmark* eficiente, e determinar a posição comercial actual;
- Seria também interessante classificar os carros através de *reviews* de utilizadores enriquecendo o *dataset* com informação de clientes e não apenas com informações fornecidas pelos fabricantes/revendedores. As *reviews* podiam ser efetuadas pelos clientes que efetuaram *test drives*, mas também pelos utilizadores que adquiriram as viaturas e conseguiriam dar *reviews* de quem já possui o veículo há mais tempo, enriquecendo ainda mais o *dataset*;
- Importante poderia ser também a inclusão de KPIs específicos de venda por segmentos, a fim de determinar quais são os veículos que têm melhor performance e quais necessitam de mais investimento a fim de serem melhorados.