Score Car

David Oliveira1, Henrique Pereira2 & Manuel Oom3

1M20181430: M20181430@novaims.unl.pt

2 M20181395: M20181395@novaims.unl.pt

3 M20181431: M20181431@novaims.unl.pt

# Introdução

Actualmente, existe uma grande variedade de oferta por parte 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* e anúncios televisivos das marcas até aos diferentes preços apresentados para viaturas que, à primeira vista, poderão parecer iguais, embora apresentem certos detalhes distintos, a decisão de alguém que pretende seleccionar uma viatura torna-se cada vez mais difícil.

O *Score Car* tem como objectivo a disponibilização de toda a informação necessária para ajudar na selecção de uma viatura num conjunto alargado de modelos e versões apresentadas no *website* do Entreposto Auto. A nossa 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, tornando difícil a comparação entre 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 Electricidade), de forma a poder apresentar o custo mensal associada a cada viatura, tendo em conta o consumo associada à viatura.

Esta solução além de ajudar os clientes que querem escolher uma viatura, seja por pretenderem efectuar uma compra ou por simplesmente terem a oportunidade de ter uma viatura através das suas empresas, poderá também ajudar os colaboradores do Entreposto Auto (ou de outros distribuidores automóveis, caso a nossa 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 nossa solução, desde a extracção dos dados do *website* do Entreposto Auto e a sua transformação até ao modelo final de dados, apresentando, de seguida, detalhadamente cada um dos *dashboards* da solução *Score Car* e as suas funcionalidades.

# Dados

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

* Extração;
* Transformação de dados;
* Análise de dados; e
* Modelação.

## Extração e transformação de dados

Neste estudo foram utilizados dados com origem em diferentes fontes:

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

Para os dados retirados do site do entreposto foi efetuado um *script* em *python* *(packgage scrapy)* que colocou os dados de todos os carros 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, cumprimento, entre outros.

* [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 com o Power BI através da web. Como os valores foram obtidos através do Power BI os valores são dinâmicos e são atualizados assim que se abre o *dashboard.*

* ev-database.org/ - Dados relativos ao consumo dos veículos elétricos;

O site do entreposto, no que toca a consumos de carros elétricos, disponibiliza a informação de 0 l / 100km. O que corresponde à verdade se estivermos a considerar apenas os combustíveis fosseis. Para enriquecer o nosso data set, retiramos o consumo elétrico do site ev-database através de um outro script de python.

* <http://www.erse.pt> – Tabelas de preços da eletricidade praticados em Portugal em 2019

A informação dos preços da eletricidade são tabeladas pela ERSE. O método utilizado foi…

A informação retirada dos sites foi tratada nas *queries* do Power BI com o intuito de reclassificar os dados e renomeá-los para uma utilização mais ágil e eficaz. Algumas colunas não relevantes foram eliminadas.

## Análise e seleção de variáveis

Como referido anteriormente, os dados do entreposto não continha os consumos elctricos dos veiculos, pelo que, os consumos dos carros elétricos foram combinados com os dados do entreposto da seguinte forma:

De todas as imagens dos carros, trabalhámos apenas com uma. A seleção aimagem foi feita da seguinte forma:

Na Imagem 1 Podemos ver a informação retirada dos *websites* e as suas ligações.

As tabelas X e Y ligam-se através das colunas A e B…

As variáveis obtidas das várias fontes de informação foram agrupadas num único *dataframe*, chamado dataset, de forma a permitir uma análise conjunta de toda a informação recolhida. O dataset tem como *index* o ano, que corresponde ao ano de medição dos índices utilizados e ao ano de final de época das épocas desportivas (e.g. ano 2018 para a época 2017/2018). Considerou-se o ano de fim da época desportiva, por ser o ano em que se celebra a vitória do campeonato relativo à época.

O dataset inclui informação de XX veiculos

# Dashboards

## Resumo Geral

Analisando o *subdataset* com um período de 57 anos, verificamos que a nossa variável alvo tem correlação linear positiva com o rendimento disponível em percentagem do PIB (R=0,6), negativa com o total da população (R=-0,6). Adicionalmente, embora a variável consumo privado em percentagem do PIB não tenha uma correlação forte com a variável alvo (R=0,1), verificamos que esta correlação linear é notória se a análise for realizada por equipa vencedora. Neste caso, verificamos uma correlação positiva quando a equipa vencedora é o Sporting e negativa quando é o Benfica ou o Porto (vide *Jupyter Notebook* II.1.2.).

## Especificações Técnicas

1. Viaturas Não Eléctricas
2. Viaturas Eléctricas

## Comparação de Viaturas

Especificações Técnicas

tem o pior SSR, e analisando os gráficos dos resíduos (vide *Jupyter Notebook* III.3.) não se verifica o cumprimento dos requisitos LINE.

# Conclusões

Com estes dashboards podemos fazer estas comprações....

Podiamos acrescentar reviews dos utilizadores com ratings e assim....

Caso um vendedor, por eemplo o entreposto, quisesse utilizar os dashboards poderia ainda incluir informação de vendes por vendedor, solicitar reviews aos utlizadores para enriquecer a base de dados de reviews oua té reviews com ostest drives.