Car Scorcard

David Oliveira1, Henrique Pereira2 & Manuel Oom3

1M20181430: M20181430@novaims.unl.pt

2 M20181395: M20181395@novaims.unl.pt

3 M20181431: M20181431@novaims.unl.pt

**Abstract/Resumo:** XXX

**Palavras-chave**: Car Scorecard; comparação de veículos; comparação de consumos; comparação de preços; comparação de poluição

# Introdução

XXXXX – David Oliveira

O futebol tem uma presença relevante na população portuguesa. Tendo em conta o foco dado ao futebol no nosso dia-a-dia: em notícias espalhadas pelas diversas plataformas digitais; nos programas desportivos que ocupam uma grande parte da programação dos canais de televisão portuguesas; nas conversas entre amigos e colegas de trabalho, pode-se afirmar que se trata do desporto rei em Portugal. Este foco não é uma regra para toda a população, no entanto, todos acabam por ser afetados indiretamente pelo mediatismo deste desporto.

Em março de 2018, o Jornal de Notícias dava conta de que “*o futebol profissional contribui com 456 milhões para o PIB*” na época de 2016/2017, representando cerca de 0,25% do PIB nacional. Já em 2013, António Mexia (Presidente do Conselho de Administração da EDP) dizia que *“era bom para Portugal que o Benfica fosse campeão. Acho que isso tinha um efeito positivo no PIB*”.

O PIB é um indicador macroeconômico que demonstra a criação de riqueza de um país. Este indicador representa o valor de bens e serviços produzidos no ano. Vários setores da sociedade e diversos fatores externos contribuem direta ou indiretamente para o PIB, razão pela qual poderá não ser possível quantificar a influência isolada do Futebol. Assim, para além do futebol, foram considerados outros factores como a inflação, os rendimentos e taxas de desemprego da população portuguesa.

O objectivo deste projecto é estudar a influência que a equipa de futebol vencedora do campeonato nacional tem no Produto Interno Bruto (PIB) de Portugal desse ano. Esta influência, a existir, poderá estar relacionada com uma maior propensão da população adepta a aumentar o consumo devido ao entusiasmo das vitórias dos seus clubes.

# 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.



Figure 1 - substituir tabelas utilizadas no power BI

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

**Pagina 1 – 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.).

**Pagina 2 -**

XXXXX

**Pagina 3 – Comparação de veículos**

XXXX-

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.

# Referências

1 <https://www.jn.pt/desporto/interior/futebol-profissional-contribui-com-456-milhoes-para-o-pib-9210445.html>, notícia publicada a 23 março 2018.

2 <https://www.publico.pt/2018/03/23/desporto/noticia/futebol-portugues-contribuiu-com-456-milhoes-de-euros-para-o-pib-1807859>, notícia publicada a 23 março 2018.

3 <https://tribunaexpresso.pt/futebol-nacional/2019-04-03-Contributo-do-futebol-portugues-para-o-PIB-baixa-13-para-396-milhoes-na-ultima-epoca>, notícia publicada a 3 abril 2019.

4 <https://www.publico.pt/2004/06/20/jornal/a-excessiva-importancia-social-e-politica-que-damos-ao-futebol-e-ditatorial-e-asfixiante-189923>, artigo publicada a 20 junho 2004.

5 Kutner, M.H., Nachtsheim, C.J., Neter, J., Li,W. (2004) Applied Linear Statistical Models, 5th edition, McGraw-Hill.

6 McKinney, Wes. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. "O'Reilly Media, Inc.", 2012.