

Relatório de Projeto – VisionOnEletrification

Análise Preditiva de Vendas de Veículos Eletrificados



VisionOnEletrification
SOFTWARE DEVELOPMENT

Composição do Grupo

Grupo Nº	Nome do Projeto
G02	VisionOnEletrification



João Pedro Brito Ramos – média: 12,50 valores

Sou aluno do terceiro ano do curso de Informática de Gestão no IADE, tenho 20 anos e escolhi este curso por ser uma área de interesse pessoal, além disso, venho de um curso profissional relacionado com hardware e com este curso pretendo complementar a área de software.

Hard Skills: SQL, Python, Power BI.



Pedro Domingues da Cunha – média: 15,28 valores

Sou aluno do terceiro ano do curso de Informática de Gestão no IADE, tenho 20 anos e escolhi este curso por ser uma área de interesse pessoal, além disso, do meu ponto de vista, é um curso bastante promissor para os anos que se sucedem. Áreas de interesse: Data Analytics, Big Data e Business Intelligence.

Resumo

Após algumas pesquisas, seguidas de uma análise, a equipa de projeto chegou à conclusão de que a sustentabilidade é uma preocupação crescente devido à escassez cada vez maior de recursos naturais a cada ano que passa.

Por este motivo, as fabricantes automóveis veem-se obrigadas a lançar produtos menos poluentes e também os consumidores adquirem cada vez mais veículos com motorizações eletrificadas. Este fenómeno faz com que haja um grande número de dados existentes que pode proporcionar informações uteis para empresas de diversos setores.

Por estes motivos decidimos criar a VisionOnEletrification, um conceito único em termos da conceção e visualização de indicadores personalizáveis relacionados com veículos, focados na vertente da eletrificação.

O propósito deste documento é reunir todos os elementos que foram necessários para o desenvolvimento e implementação deste projeto no âmbito da UC de Data Mining and Web Analytics onde o grande objetivo e desafio era o desenvolvimento de um algoritmo preditivo com recurso ao *Python* e a bibliotecas de *Data Science*.

Palavras-Chave: Eletrificação, Veículos, Sustentabilidade, *Data Science*

Índice

COMPOSIÇÃO DO GRUPO	2
RESUMO	2
1 INTRODUÇÃO AO ÂMBITO DO PROJETO.....	4
1.1 OBJETIVOS DO PROJETO	4
1.2 PLANEAMENTO DAS FASES DO PROJETO	5
2 ENQUADRAMENTO NA ÁREA DE <i>DATA SCIENCE</i>	5
2.1 DATA SCIENCE E NUMPY	6
2.2 DATA SCIENCE E PANDAS.....	6
2.3 DATA SCIENCE E MATPLOTLIB	6
3 METODOLOGIA	6
4 PROJETO DE SOFTWARE	7
4.1 FONTE DE DADOS.....	7
4.2 BIBLIOTECAS UTILIZADAS	7
4.3 ETAPAS DA EXECUÇÃO DO SOFTWARE	7
4.3.1 IMPORTAR E FILTRAR OS DADOS.....	7
4.3.2 VISUALIZAÇÃO DOS DADOS.....	9
4.3.3 ALGORITMO PREDITIVO	10
5 DESENVOLVIMENTO DE WEBSITE	11
6 SECÇÃO COM RESULTADOS	12
7 CONCLUSÕES E TRABALHO FUTURO	13
7.1 COMENTÁRIOS FINAIS	13
7.2 LIMITAÇÕES E DIFICULDADES	13
REFERÊNCIAS BIBLIOGRÁFICAS	14

1 Introdução ao âmbito do projeto

Nos últimos anos, um dos temas mais abordados em termos sociopolíticos é justamente a sustentabilidade – A capacidade de satisfazer as nossas necessidades no presente sem comprometer a capacidade de as gerações futuras satisfazerem as suas próprias necessidades e é hoje, um tema-chave para a competitividade das empresas, cada vez mais importante para as suas estratégias de curto, médio e longo prazo.

A rápida delapidação dos recursos naturais, os impactos que a atividade humana tem nos ecossistemas e na biosfera, os níveis elevados de desigualdades sociais e de pobreza que enfrentamos, e a falta de ética de gestão corporativa são inimigos de um modelo de desenvolvimento sustentável, capaz de assegurar às gerações futuras as oportunidades e qualidade de vida de que as atuais gerações (ainda) gozam.

Desta maneira, ao longo dos anos, diversas empresas dos mais variados setores viram-se obrigadas a reformular as suas políticas empresariais de modo a proporcionar serviços e produtos com o menor impacto ambiental possível, visto que também cada vez mais os clientes dão importância a este fator na hora da compra, por sentirem que estão a contribuir para o bem de toda a sociedade.

Um exemplo disto são as empresas do setor automóvel, que devido às normas e leis que vão sendo metidas em vigor sistematicamente, veem-se obrigadas a lançar produtos com menos emissões de gases poluentes e com o menor nível de ruído possível, desta forma, também os consumidores adquirem cada vez mais veículos com motorizações eletrificadas, quer sejam elétricos ou híbridos, devido às inúmeras vantagens que estes proporcionam relativamente aos veículos tradicionais movidos a combustíveis fósseis.

Além disto tudo, o tema da sustentabilidade está ainda fortemente presente porque ao contrário do que se pensava até há uns anos, o petróleo não é infinito e como tal é necessário racionalizar o existente para não comprometer as gerações futuras.

Este fenómeno faz com que haja um grande número de dados existentes que podem proporcionar informações úteis para empresas de diversos setores, nomeadamente, o energético, a mobilidade/estacionamento e o petrolífero - que pretendam efetuar o planeamento interno das suas atividades no médio e longo prazo, de maneira que o investimento feito tenha o maior retorno possível e seja o mais eficaz de modo a chegar ao maior número de clientes de modo a assegurar a sobrevivência da empresa.

Por todos estes motivos decidimos criar a VisionOnEletrification, um conceito único em termos da conceção e visualização de indicadores personalizáveis relacionados com veículos, focados na vertente da eletrificação, para ajudar no apoio à decisão das empresas a decidir o seu futuro, atuando assim como intermediários.

1.1 Objetivos do Projeto

- Utilização de bibliotecas de *Data Science* em *Python*;
- Analisar e filtrar os dados de forma a restarem apenas os dados úteis para a ótica e âmbito do projeto;
- Transformar os dados em informação através de elementos Visuais como Gráficos e Filtros para utilização e leitura intuitiva por parte do utilizador final;
- Desenvolver Algoritmo Preditivo relativo ao crescimento das vendas dos veículos eletrificados;
- Estruturar os elementos referidos num *Dashboard* intuitivo e de fácil leitura.

1.2 Planeamento das Fases do Projeto

O seguinte cronograma tem como objetivo mostrar as fases que foram necessárias para a implementação do projeto e para dar a noção do progresso que o projeto teve ao longo do semestre.

É também importante deixar claro que todas as tarefas do mesmo foram repartidas de igual forma por os elementos do grupo e que não houve quaisquer problemas relativamente à divisão das tarefas, ambos estiveram de igual forma empenhados para o sucesso do mesmo.

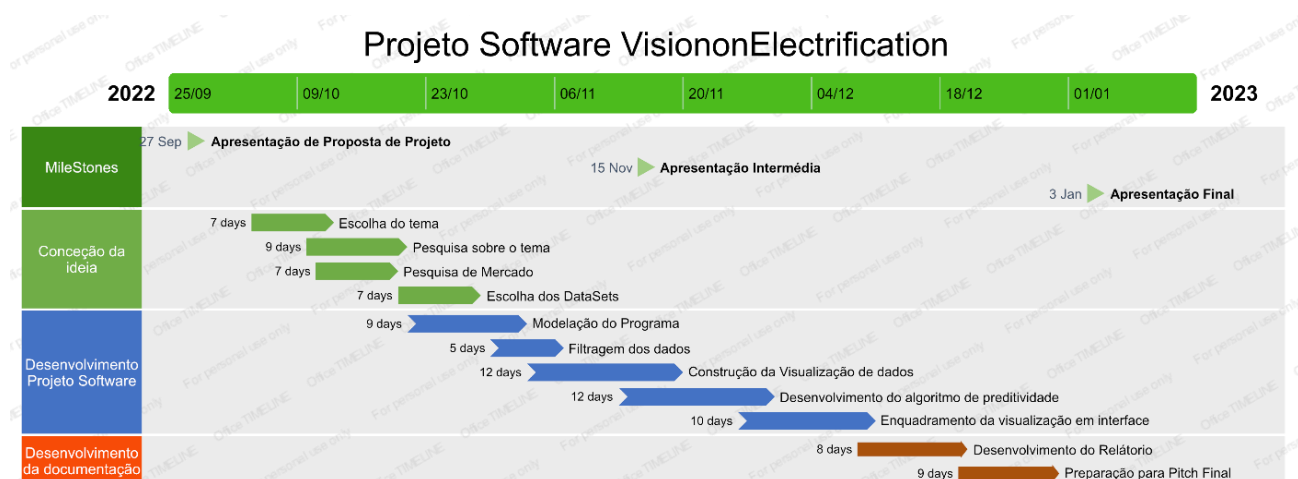


Figura 1 - Cronograma de planificação do projeto

2 Enquadramento na Área de Data Science

A *Data Science* é uma área interdisciplinar que resulta da combinação de diferentes técnicas e teorias provenientes de diversas áreas, nomeadamente: Matemática/Estatística, Ciências da Computação/Tecnologias de Informação e Conhecimento do Negócio.

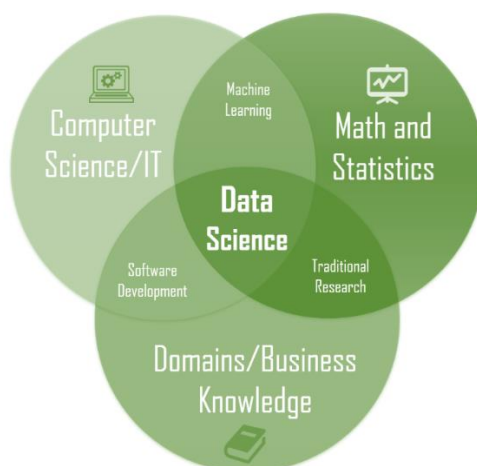


Figura 2 - Diagrama alusivo ao conceito de Data Science

Tem como objetivo o estudo e análise de grandes volumes de dados não-estruturados (*Big Data*), a fim de disponibilizar de maneira gráfica e compreensível ao olho humano grandes volumes de informações, de maneira a facilitar e otimizar as tomadas de decisão sobre problemas do mundo real, que por muitas vezes ajudam empresas a tomar decisões mais acertadas e benéficas, reduzir custos, reduzir riscos e a aumentar seus lucros, que no fundo é o que as empresas mais desejam.

Além disto, recorre ainda a técnicas de Inteligência Artificial e *Machine Learning* para extrair dados úteis e prever padrões e comportamentos futuros.

As atividades profissionais nesta área dividem-se em: *Data Analyst*, *Data Engineer* e *Data Scientist*. Com a transformação digital a evoluir a um ritmo cada vez mais acelerado, a procura por estes profissionais tem crescido no mercado, isto porque as empresas de todos os setores querem ser mais assertivas nas suas decisões de modo a tomar o melhor rumo possível.

Todos os dias, um grande volume de dados é criado. Nem todos esses dados são processados, muito menos analisados, e menos ainda, estão estruturados. Vivemos na Era da Informação. A Humanidade passou pela Revolução Agrícola, Industrial e vivemos a Revolução Digital.

Estima-se que 90% dos dados do mundo tenham sido criados nos últimos dois anos. Os utilizadores do Facebook, por exemplo carregam cerca de 10 milhões de fotos por hora. Todos estes dados podem ser muito interessantes para as empresas. Com isso criam-se diversos de novos paradigmas, como os de cyber-segurança, privacidade e segurança de dados, que acarretam na necessidade de adaptarmos o mundo que vivemos à essa nova realidade.

2.1 Data Science e NumPy

NumPy é uma biblioteca de *Python* que se popularizou por ser capaz de processar e analisar *arrays* e matrizes multidimensionais facilmente.

2.2 Data Science e Pandas

Pandas é também uma biblioteca de *Python* que permite analisar estruturas e operações de dados para manipular tabelas numéricas e séries temporais. Ou seja, permite processar bases de dados estruturados (SQL) em tabelas, rapidamente, sem ter que selecionar "linhas", e com capacidade de processamento muito superior ao número de 1.048.576 linhas máximo suportado pelo Excel para descobrir informações chave de um conjunto de dados.

2.3 Data Science e Matplotlib

Utilizando esta biblioteca em conjunto com *NumPy* e *Pandas*, além de analisar dados, é possível criar gráficos personalizados para visualização de dados e transformar todos os resultados em *story-telling*.

3 Metodologia

O grande e principal objetivo deste projeto passava por elaborar um algoritmo de análise preditiva, no nosso caso, do número de vendas de veículos eletrificados para os próximos anos na zona da União Europeia para poder chegar à conclusão de qual o rumo que ia tomar, ou seja, se efetivamente as vendas irão continuar a aumentar, ou se por outro lado, irão diminuir.

Para alcançar estes resultados, a metodologia passou, inicialmente, pela pesquisa e posterior recolha dos dados inerentes aos veículos que foram vendidos na Europa em diferentes anos, nomeadamente 2018, 2019 e 2020.

Com esta primeira etapa concluída, procedemos à importação dos mesmos para o *Python*, onde filtrámos as informações recolhidas de modo que restassem apenas os dados que nos seriam úteis na ótica do âmbito do projeto.

Com isto feito, restava apenas proceder à criação de gráficos relevantes para a visualização da informação e também criar o nosso algoritmo preditivo e juntar todos estes elementos num *website*, devido à complexidade acrescida relativamente às outras componentes do projeto, o desenvolvimento do algoritmo foi o elemento que nos levou mais tempo, mas mesmo assim conseguimos atingir esse objetivo e demonstrar os resultados.

4 Projeto de software

4.1 Fonte de Dados

A nossa fonte dos nossos dados é a European Environment Agency (EEA), uma agência da União Europeia cuja missão é fornecer informações sólidas e independentes sobre o ambiente. A EEA visa apoiar o desenvolvimento sustentável, ajudando a alcançar uma melhoria significativa e mensurável no meio ambiente europeu, através do fornecimento de informações oportunas, relevantes e confiáveis direcionadas para os agentes de formulação de políticas e o público.

Para tal, a mesma dispõe de diversos gráficos sobre os mais variados temas ambientais, estes por sua vez, possuem por trás um dataset que pode ser consultado e transferido por qualquer pessoa. Um dos *datasets* disponíveis é denominado de [CO2 emissions from new passenger cars](#), e foi este que utilizamos para realizar o nosso projeto, o mesmo possui os dados dos veículos ligeiros de passageiros que foram vendidos nos países pertencentes à União Europeia desde 2010, alguns dos dados que são disponibilizados são por exemplo: marca, modelo, combustível, emissões de CO2 e país de registo do veículo.

Para a execução do projeto considerámos apenas os dados dos anos de 2018, 2019 e 2020 e não os usámos na sua totalidade devido ao grande número de dados, isto torna o ficheiro bastante pesado e lento de ler e tornava-se inviável trabalhar com a totalidade do ficheiro no *Python*.

4.2 Bibliotecas Utilizadas

As bibliotecas utilizadas no projeto foram as seguintes:

- **Pandas:** Utilizado para importar os datasets e fazer filtragem dos dados dos mesmos
- **Matplotlib, Plotly e Seaborn:** Utilizados para efetuar a vertente de visualização dos dados em gráficos
- **Streamlit:** utilizado para criar o web server para exibir o *dashboard*, de modo a permitir a visualização dos gráficos.

4.3 Etapas da Execução do Software

4.3.1 Importar e Filtrar os Dados

No processo de importar e filtrar os dados foi utilizada a biblioteca Pandas como já foi referido anteriormente. Na figura abaixo, que contém um pedaço de código, podemos observar as linhas que correspondem à importação dos dados, assim como o *merge* dos vários *datasets* dos diferentes anos que utilizamos.

```
#Tratamento de dados e importar dados
@st.cache
def get_data_from_csv():
    cars_2018=pd.read_csv('dados/2018.csv',low_memory=False,nrows=100000) #
    carregar datasets
    cars_2019=pd.read_csv('dados/2019.csv',low_memory=False,nrows=100000)
    cars_2020=pd.read_csv('dados/2020.csv',low_memory=False,nrows=100000)
    cars=pd.concat([cars_2018,cars_2019,cars_2020]) # Juntar os datasets
```

Figura 3 - Importar Dados

Na figura que se segue, podemos ver a filtragem de dados efetuada no *dataset*. Iniciamos com o primeiro comando que serve para eliminarmos os valores a *null* das colunas indicadas.

De seguida, na segunda instrução retirámos as colunas desnecessárias e sem semântica para o projeto e após isso renomeamos algumas das colunas que sobraram e corrigimos os valores que possuíam a sintaxe errada ou com outro tipo de erro de modo a melhorar a qualidade dos dados.

```

cars.dropna(subset=['ID', 'Country', 'Mk', 'Cn', 'm (kg)', 'Ft'], inplace=True)
# Eliminar os nulls nas colunas discriminadas
# Eliminar colunas indesejadas
cars.drop(['VFN', 'Mp', 'Mh', 'Man', 'MMS', 'Tan', 'T', 'Va', 'Ve', 'Ct', 'Cr', 'Mt',
'Ewltp (g/km)', 'W (mm)', 'At1 (mm)', 'At2 (mm)', 'Fm', 'z (Wh/km)', 'IT', 'Ernedc
(g/km)', 'Erwltp (g/km)', 'De', 'Vf', 'Status'], axis=1)
# Renomear as colunas
cars.rename(columns = {'m (kg)': 'massa', 'Enedc (g/km)': 'emissoes', 'ec
(cm3)': 'cilindrada', 'ep (KW)': 'Kw'}, inplace = True)

# Corrigir palavras mal escritas
cars=cars.replace({'Ft': {'Petrol': 'PETROL', 'petrol': 'PETROL', 'Diesel': 'DIE
SEL', 'diesel': 'DIESEL', 'Electric': 'ELECTRIC', 'electric': 'ELECTRIC', 'Petrol/Ele
ctric': 'PETROL/ELECTRIC', 'petrol/electric': 'PETROL/ELECTRIC', 'ng': 'NG', 'lpg': '
LPG', 'hydrogen'
: 'HYDROGEN', 'diesel/electric': 'DIESEL/ELECTRIC'}})
cars=cars.replace({'Mk': {'?KODA': 'SKODA', 'MERCEDES-AMG': 'Mercedes-
Benz', 'Kia': 'KIA', 'BMW I': 'BMW', 'FORD-CNG-
TECHNIK': 'FORD', 'SsangYong': 'SSANGYONG', 'VOLKSWAGEN, VW': 'VOLKSWAGEN', 'Ã...Ã
KODA': 'SKODA', 'Volkswagen': 'VOLKSWAGEN', 'CATERHAM CARS
LTD': 'CATERHAM', 'Toyota': 'TOYOTA', 'BMW i': 'BMW', 'Ã...Ã
KODA': 'SKODA', 'Ford': 'FORD', 'FORD W GMBH': 'FORD', 'FORD-CNG
TECHN': 'FORD', 'JAGUAR LAND ROV': 'JAGUAR', 'MITSUBISHI J': 'MITSUBISHI', 'FORD
D': 'FORD', 'FORD USA': 'FORD', 'MERCEDES AMG': 'MERCEDES-
BENZ', 'ŠKODA': 'SKODA', 'Ssangyong': 'SSANGYONG', 'VOLKSWAGEN
AG': 'VOLKSWAGEN', 'VOLKSWAGEN V W': 'VOLKSWAGEN', 'Jeep': 'JEEP', 'MERCEDES
BENZ': 'MERCEDES-BENZ', 'FORD-CNG TECHNIK': 'FORD', 'VOLKSWAGEN
VW': 'VOLKSWAGEN'}})

return cars
cars=get_data_from_csv()

```

Figura 4 - Filtragem de Dados

Ambas as figuras fazem parte de uma função que guarda os *datasets* em *cache*, para que desta maneira, só necessitem de serem carregados uma única vez ao longo da execução do código.

4.3.2 Visualização dos Dados

Para esta componente, depois de uma análise que efetuamos, seleccionámos alguns gráficos que achámos que seriam relevantes e importantes de exibir para o contexto do nosso projeto, isto é, caso estivéssemos a apresentar o nosso produto para alguma empresa que mostrasse interesse no nosso produto.

Nesta secção podemos observar a utilização das bibliotecas *Matplotlib*, *Plotly*, *Seaborn* e *Sklearn*. A última enunciada, foi responsável por colocar os dados numa regressão linear e os passos de desenvolvimento da mesma serão explicados mais à frente no relatório.

```
vendasTotais= int(cars["r"].sum())
with esquerda:
    st.subheader(f"Vendas totais {vendasTotais}") # Indicador de vendas
    totais/Registos apresentados
    fr=px.scatter(cars_selection1,x='Mk',y='Country', title='Vendas por Marca
em Países') # grafico de vendas por marca em paises
    st.write(fr)# mostrar o grafico acima
    plt.scatter(preditct['ELECTRIC'],preditct['COMBUSTAO'])# grafico de pontos
    plt.plot(preditct['ELECTRIC'],y_predict,color='red') # linha de tendencia
    plt.title(" Algoritmo preditividade de vendas de carros eletricos")
    st.pyplot(plt)# mostrar o grafico acima

with centro:
    fig=px.scatter(cars_selection2,
x="Country",y='emissoes',title='Distribuição de Emissoes(G/Km) por pais')
    st.write(fig)
    #Determina a correlação entre variaveis
    st.text('Correlação entre vendas de veiculos combustão e eletricos')
    cor=sns.heatmap(preditct.corr(),cmap="YlGnBu",annot=True,linewidth=1)
    st.write(cor.get_figure())

with direita:
    fog=px.box(cars_selection2, x="Country",y='massa',title='Distribuição de
Massa(Kg) veiculos por pais')
    st.write(fog)
    circ = px.pie(cars_selection3, values='r', names='Ft', title='Total de
vendas por combustivel e ano')
    st.write(circ)
```

Figura 5 - Visualização dos Dados

4.3.3 Algoritmo Preditivo

Para o desenvolvimento do algoritmo preditivo procedemos ao isolamento da coluna de combustível e fizemos a contagem tanto dos veículos que utilizam combustíveis fósseis, nomeadamente, gasolina e gasóleo, como dos veículos eletrificados, onde estão incluídos os elétricos e os híbridos.

De seguida colocámos estes dados num novo *data frame* que posteriormente foi utilizado para inserir os dados na regressão linear e desta forma chegamos ao nosso algoritmo preditivo.

```

E2018=cars[(cars.Ft=='ELECTRIC') | (cars.Ft=='PETROL/ELECTRIC') |
(cars.Ft=='DIESEL/ELECTRIC')& (cars.year==2018)]
E2019=cars[(cars.Ft=='ELECTRIC') | (cars.Ft=='PETROL/ELECTRIC') |
(cars.Ft=='DIESEL/ELECTRIC') & (cars.year==2019)]
E2020=cars[(cars.Ft=='ELECTRIC') | (cars.Ft=='PETROL/ELECTRIC') |
(cars.Ft=='DIESEL/ELECTRIC') & (cars.year==2020)]

#Pesquisa por veiculos combustao por ano e guardar variaveis
P2018=cars[(cars.Ft=='PETROL')| (cars.Ft=='DIESEL') & (cars.year==2018)]
P2019=cars[(cars.Ft=='PETROL')| (cars.Ft=='DIESEL') & (cars.year==2019)]
P2020=cars[(cars.Ft=='PETROL')| (cars.Ft=='DIESEL') & (cars.year==2020)]

#fazer a contagem de cada valor de forma a obter numeros totais por ano
df=[[E2018['Ft'].count(),P2018['Ft'].count()], [E2019['Ft'].count(),P2019['Ft']
.count()], [E2020['Ft'].count(),P2020['Ft'].count()]]

#para ilustrar melhor de forma tabelar
ind=[['2018','2019','2020']]
preditct=pd.DataFrame(df,index=ind,columns=['ELECTRIC','COMBUSTAO'])

#GUARDAR CADA VALOR EM UMA VARIÁVEL
n1=preditct.drop('ELECTRIC',axis='columns')
n=preditct.drop('COMBUSTAO',axis='columns')
# Fim Construção de um subdataset para isolar os dados para analise preditiva

#Colocar as variaveis na regressao linear
lr=linear_model.LinearRegression()
lr.fit(n,n1)

#Definir varivel a prever
y_predict=lr.predict(n)
  
```

Figura 6 - Algoritmo

5 Desenvolvimento de website

Para o desenvolvimento do website recorreremos à biblioteca *Streamlit* que simula um web server, poupando imenso trabalho, agilizando ainda mais o desenvolvimento do projeto.

Nesta instrução simples, uma das vantagens desta biblioteca, definimos o título e o *icon* da *webpage*, assim como a largura da mesma. Para além disto, já dentro da página HTML em si, criámos um título e uma *sidebar* para podermos aplicar filtros aos dados existentes.

```
#Config pagina
st.set_page_config(page_title="Eletrico
dash",page_icon=":bar_chart:",layout="wide")

#Dar um titlo à pagina
st.title("DashBoard Veiculos na Europa ")
#SIDE BAR
st.sidebar.header("Insira os parametros")
```

Construímos então as opções (filtros) da *sidebar* e definimos também quais deveriam ser os dados atualizados em cada gráfico quando se aplicasse os filtros.

```
Country=st.sidebar.multiselect(
    "Selecione os Países",
    #Definir valores a estar nas opções
    options=cars["Country"].unique(),
    default=cars["Country"].unique()
)

year=st.sidebar.multiselect(
    "Selecione os Anos",
    options=cars["year"].unique(),
    default=cars["year"].unique()
)

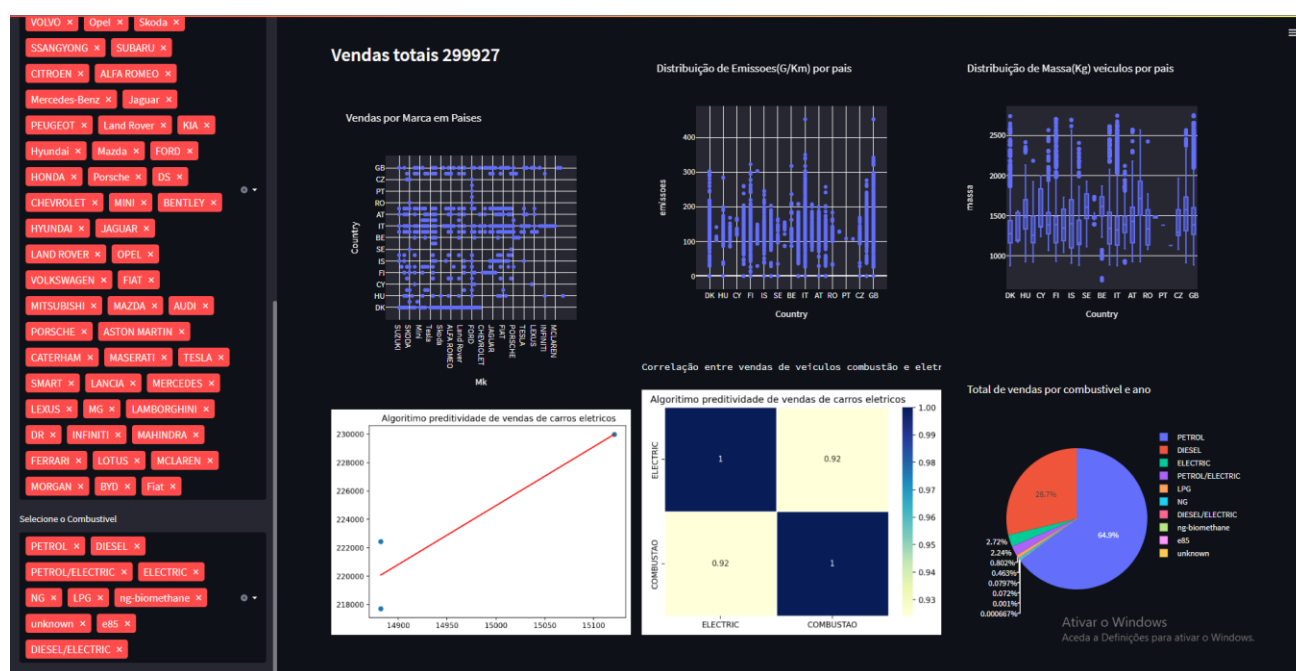
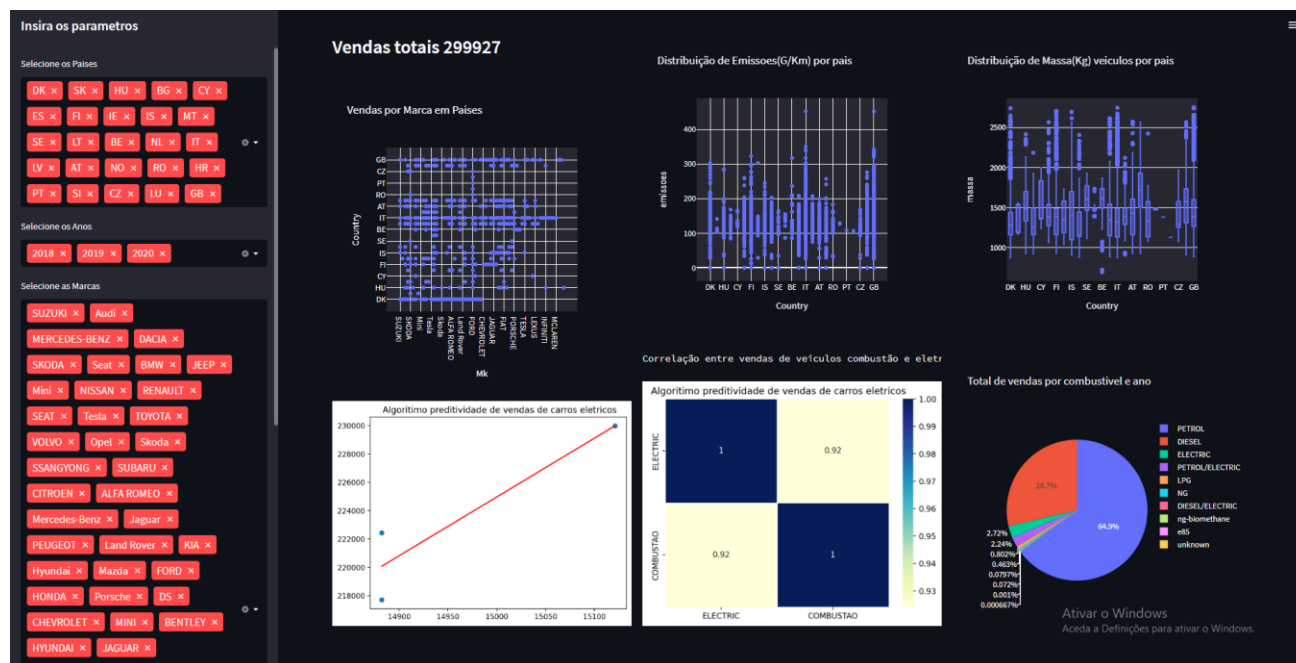
Mk=st.sidebar.multiselect(
    "Selecione as Marcas",
    options=cars["Mk"].unique(),
    default=cars["Mk"].unique()
)

Ft=st.sidebar.multiselect(
    "Selecione o Combustivel",
    options=cars["Ft"].unique(),
    default=cars["Ft"].unique()
)

#criar seleções para os graficos
cars_selection1=cars.query(
    "Country == @Country and Mk == @Mk "
)
cars_selection2=cars.query(
    "Country == @Country"
)
cars_selection3=cars.query(
    "year == @year"
)
```

6 Secção com resultados

Nesta secção encontram-se os resultados que obtemos com base no código explicado na secção anterior, em que no lado esquerdo podemos observar a *sidebar* com os filtros que criamos e na restante página podemos observar os diversos gráficos que elaboramos, bem como o algoritmo preditivo.



7 Conclusões e Trabalho Futuro

Para concluir e com base nos resultados obtidos podemos então dizer que as vendas deste tipo de veículos eletrificados tem tendência a aumentar visto que a cada ano que passa o número tem vindo a aumentar substancialmente, e apesar de em 2019 ter havido uma descida, no ano a seguir houve um aumento explosivo. Este fenómeno pode ser explicado pelo efeito da pandemia, isto porque uma vez que as pessoas estavam maioritariamente em casa não sentiam tanta necessidade de utilizar os seus veículos e desta forma também não se venderam tantos veículos como num ano considerado como normal.

É também importante salientar que o projeto de software que nos foi proposto desenvolver foi concluído, na nossa opinião, com êxito e atingiu todos os objetivos e metas estabelecidas. Desde o início até à conclusão do mesmo, a equipa de desenvolvimento sempre trabalhou de forma focada, eficiente e dedicada de modo a garantir que a qualidade do software fosse a melhor possível e também que fosse entregue dentro do prazo e de acordo com os requisitos estabelecidos.

Contudo, existe espaço para melhorias, as quais se encontram especialmente focadas na qualidade do algoritmo Preditivo, o qual gostaríamos de aperfeiçoar relativamente à sua precisão e leitura. No que toca ao nível de visualização dos dados gostaríamos de oferecer uma melhor *user experience* de leitura para os utilizadores, de modo a ser ainda mais intuitivo e *clean*. Finalmente, gostaríamos ainda de poder oferecer uma granularidade maior, isto é, possuir mais quantidade de dados relacionados com os veículos, como por exemplo a data de registo que seria bastante útil para dividir o ano em trimestres.

7.1 Comentários Finais

Ao concluirmos este projeto, gostaríamos de expressar a nossa gratidão a todos os envolvidos na sua realização, especialmente aos outros grupos e ao professor orientador do projeto, que demonstrou sempre disponibilidade para nos ajudar em qualquer questão e nos aconselhou sempre o melhor caminho a seguir.

É importante também destacar que este projeto permitiu-nos aprofundar os conhecimentos já existentes de programação, mais concretamente em *Python* – permitiu estar em contacto e trabalhar com bibliotecas de *data science*, aprofundar conhecimentos de *data analytics* e ainda ajudar a perceber como funciona um algoritmo preditivo.

7.2 Limitações e Dificuldades

Durante o desenvolvimento deste projeto, encontrámos, porém, algumas limitações ao nível do desempenho do hardware que nos impediram de alcançar o potencial máximo do software, ou seja, como já dito anteriormente, o peso dos nossos *datasets* era bastante grande o que tornava o programa bastante lento na parte da leitura dos dados, desta maneira tivemos de limitar o numero de dados para 100,000 linhas para cada ano, o que pode influenciar bastante os resultados visto que é apenas uma amostra dos *datasets* que extraímos.

Além disto, tivemos dificuldade em encontrar e agrupar mais dados úteis para os conjuntos de dados que estávamos a utilizar, ou seja, em encontrar outros *datasets* que pudéssemos integrar e relacionar com aqueles que já possuímos.

Apesar destas limitações e dificuldades, fizemos sempre tudo para as conseguir ultrapassar e superar com as ferramentas que tínhamos ao nosso dispor, nomeadamente o facto de reduzir o número de dados por cada ano e assim poder englobar mais diversidade de dados importantes, é importante referir que esta condição é facilmente alterável no código deixando assim o projeto preparado para a futura escalabilidade.

Referências Bibliográficas

1. C. Murphy, "Investing in electric vehicles and Green Transportation," *Investopedia*, 27-Aug-2022. [Online]. Available: <https://www.investopedia.com/investing-in-electric-vehicles-and-green-transportation-5220604>. [Accessed: 25-Sep-2022].
2. J. Poliscanova , "Electric car sales sky-rocket in Europe," *Transport & Environment*, 27-Jul-2021. [Online]. Available: <https://www.transportenvironment.org/discover/electric-car-sales-sky-rocket-europe/>. [Accessed: 25-Sep-2022].
3. Iea, "Trends in electric light-duty vehicles – global EV outlook 2022 – analysis," IEA. [Online]. Available: <https://www.iea.org/reports/global-ev-outlook-2022/trends-in-electric-light-duty-vehicles>. [Accessed: 25-Sep-2022].
4. "Carros Elétricos: 7 razões para dar o passo," *Observador*, 24-Jan-2018. [Online]. Available: <https://observador.pt/2018/01/24/carros-eletricos-7-razoes-para-dar-o-passo/>. [Accessed: 25-Sep-2022].
5. "Monitoring of CO2 emissions from Passenger Cars – Regulation (EU) 2019/631," European Environment Agency, 21-Sep-2022. [Online]. Available: <https://www.eea.europa.eu/data-and-maps/data/co2-cars-emission-20>. [Accessed: 25-Sep-2022].