

Projeto Data Science – Análise Preditiva

Futuras Vendas de Veículos Eletrificados



VisionOnElectrification

SOFTWARE DEVELOPMENT

A Nossa Equipa e as Suas Competências



Pedro Cunha

20 anos - Analista de Dados

Bastante empenhado, persistente e dedicado nas tarefas que lhe são propostas.

Gosta de trabalhar em equipa e é bastante flexível e compreensivo nas tomadas de decisão.

Possui nível C1 de Inglês e interesse na áreas de marketing, automóvel e de data analytics.

SQL



Python



Power BI



Photoshop



João Ramos

20 anos - Analista de Dados

Líder criativo muito organizado, com talento para apresentar ideias e soluções únicas. Eficiente a planear e gerir tarefas o que torna um ativo valioso em trabalho de equipa.

Possui também nível 4 em manutenção de equipamentos informáticos e certificação reparador Apple.

SQL



Python



Power BI

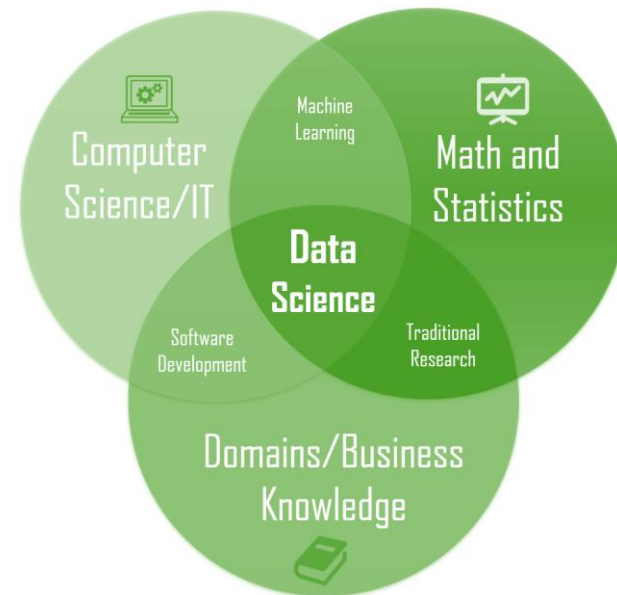


UML



Conceito de Data Science – O que é?

- É o **estudo e análise de um grande volume de informações** sobre uma determinada área para obter informações uteis para gerar ideias determinantes no **sucesso dos negócios**.
- É a **combinação** de diferentes técnicas e teorias provenientes de **diversas áreas**.
- Além disto, utiliza ainda **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.
- **Transformação Digital** em constante evolução



Introdução – **Background** e Motivações

Sustentabilidade

Preocupação crescente
por escassez de
recursos naturais

Mobilidade Sustentável

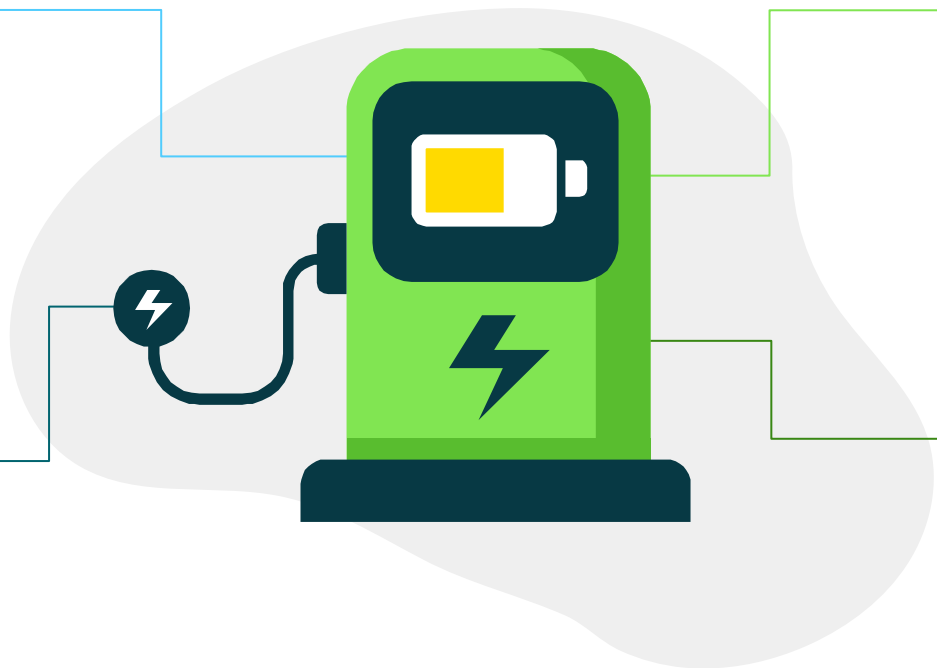
Influenciada pelas leis
criadas referentes às
emissões de gases devido as
preocupações ambientais

Fabricantes

Investir em soluções
eletrificadas para
reduzir emissões e ruído

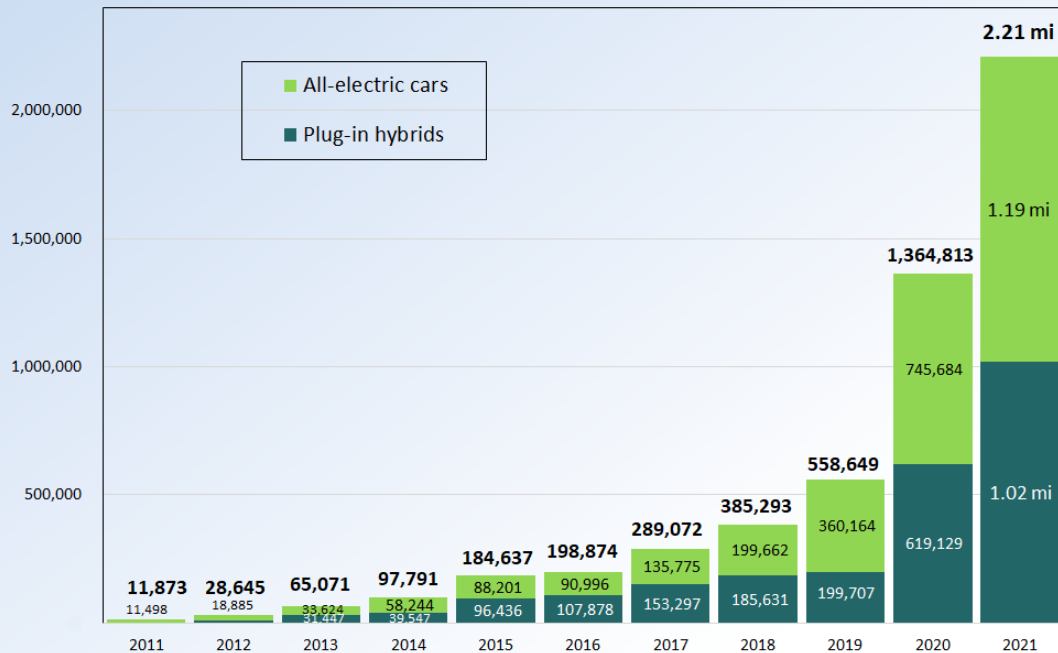
Pessoas

Adquirem mais veículos
deste tipo fazendo
aumentar as vendas



Introdução – Dados Existentes

Annual registrations of plug-in electric passenger cars in Europe*
(2011-2021)



*Note: Combined registration figures EU + EFTA + UK. EFTA countries included are Iceland, Norway, and Switzerland

Fonte: [Wikipedia](#)

Metodologia Utilizada – Como?

- O objetivo do projeto era **efetuar uma análise preditiva** do número de vendas de veículos eletrificados para os próximos anos na 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 **recolha dos dados** inerentes aos veículos que foram vendidos na Europa em diferentes anos (2018, 2019 e 2020) - EEA.
- Com a primeira etapa concluída, **filtrámos as informações recolhidas (pandas)** de modo a que restassem apenas os dados que nos seriam úteis na ótica do âmbito do projeto.
- Para a realização do algoritmo recorreremos ao **Python** com auxílio a algumas **bibliotecas**, nomeadamente Pandas e sklearn.

CO₂ emissions from new passenger cars

CO₂ emissions from new passenger cars registered in EU27, Iceland (from 2018) and Norway (from 2019) - Regulation (EU) 2019/631.

Data table explorer Dashboard

Search term

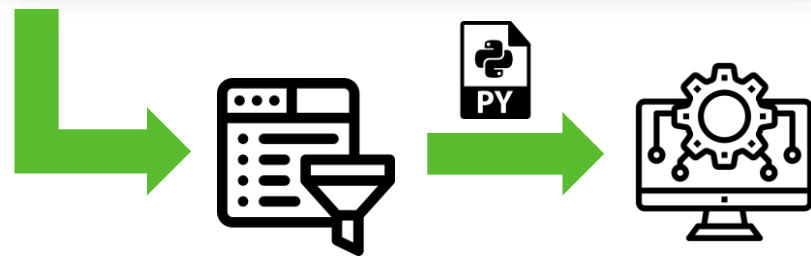
Current filters applied: 2 Reset filters

Status: Final Registration year: 2020

ID	Country	Vehicle family identification n...	Pool	Manufacturer name (ISO 4217)	Manufacturer name (OEI)	Manufacturer name (OEI)...	Type approval number	Type	Variant	Version	Make	Commercial name	Category of the...	Category of the...	Registration year
W03479	GB	IP- W03479_2018_00004 W031	FORD- VOLVO	FORD- W03479	FORD- W03479	FORD	49P2001046P10142113	34H	MSB115	SCORPIONA350	FORD	Fiesta ST-Line Edition Turbodiesel	M1	M1	2018
W03480	GB	IP-B W03480_2018-0025	FORD- VOLVO	FORD- W03480	FORD- W03480	FORD	49P2001046P10010208	34H	M1J1115	SAFARIW03480	FORD	ECOSPORT ST-Line	M1	M1	2018
W03481	GB	IP-B W03481_2018-0025	FORD- VOLVO	FORD- W03481	FORD- W03481	FORD	49P2001046P10010208	34H	M1J1115	SAFARIW03481	FORD	ECOSPORT ST-Line	M1	M1	2018
W03482	GB	IP- C034_2018_00001 W031	FORD- VOLVO	FORD- W03482	FORD- W03482	FORD	41P2001046P10111106	DEH	870A1N6	SAFARIW03482	FORD	FOCUS ST-Line	M1	M1	2018
W03483	GB	IP-11 W03483_2017-0028	FORD- VOLVO	FORD- W03483	FORD- W03483	FORD	41P2001046P10111106	DEH	870A1N6	SAFARIW03483	FORD	FOCUS ST-Line	M1	M1	2018
W03484	GB	IP- W03484_2018_00004 W031	FORD- VOLVO	FORD- W03484	FORD- W03484	FORD	49P2001046P10142113	34H	870C1K	SCORPIONA350	FORD	Fiesta Titanium X Turbodiesel	M1	M1	2018
W03485	GB	IP-11 W03485_2017-0028	FORD- VOLVO	FORD- W03485	FORD- W03485	FORD	41P2001046P10111106	DEH	870A1N6	SAFARIW03485	FORD	FOCUS Titanium X	M1	M1	2018
W03486	GB	IP-5 W03486_2018-0003	FORD- VOLVO	FORD- W03486	FORD- W03486	FORD	41P2001116P1000P108	BA7	11C11Y5	SLAZARW03486	FORD	MONDEO Titanium Edition Turbodiesel	M1	M1	2018

Country Pool Manufacturer name Commercial name

Activate Windows
Go to Settings to activate Windows
Mass in running order (kg)



Parte Prática – Bibliotecas

```
import pandas as pd
import streamlit as st
import matplotlib.pyplot as plt
import plotly.express as px
import seaborn as sns
from sklearn import linear_model
```

Parte Prática – Filtragem de Dados

```
#Tratamento de 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

    cars.dropna(subset=['ID','Country','Mk','Cn','m (kg)','Ft'],inplace=True) # Eliminar os nulls nas colunas discriminadas
    #Eliminar columnas 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 columnas
    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':'DIESEL','diesel':'DIESEL','Electric':'ELECTRIC','electric':'ELECTRIC','Petrol/Electric':'PETROL/ELECTRIC','petrol/electric':'PETROL/ELECTRIC',
    : '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','Ä...Å KODA':'SKODA','Volkswagen':'VO

    return cars
cars=get_data_from_csv()
```


Parte Prática – Construção da Sidebar e Seleções

```
#SIDE BAR
st.sidebar.header("Insira os parametros")

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"
)

#Dar um titulo à pagina
st.title("DashBoard Veiculos na Europa ")
st.markdown("###")

#Calculo de vendas totais
vendasTotais= int(cars["r"].sum())

#Definir colunas de forma a organizar o espaço onde os graficos ficam
esquerda,centro,direita=st.columns(3)
```

Parte Prática – Construção do Algoritmo de Preditividade

```
#Construção de um subdataset para isolar os dados para análise preditiva
#Pesquisa por veículos eletrificados por ano e guardar variáveis
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 veículos combustao por ano e guardar variáveis
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 análise preditiva

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

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

Parte Prática – Disposição dos Gráficos Estruturados

```
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 países
    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 Emissões(G/Km) por país')
    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 país')
    st.write(fog)
    circ = px.pie(cars_selection3, values='r', names='Ft', title='Total de vendas por combustivel e ano')
    st.write(circ)
```

Sprint Retrospective – Visão Geral



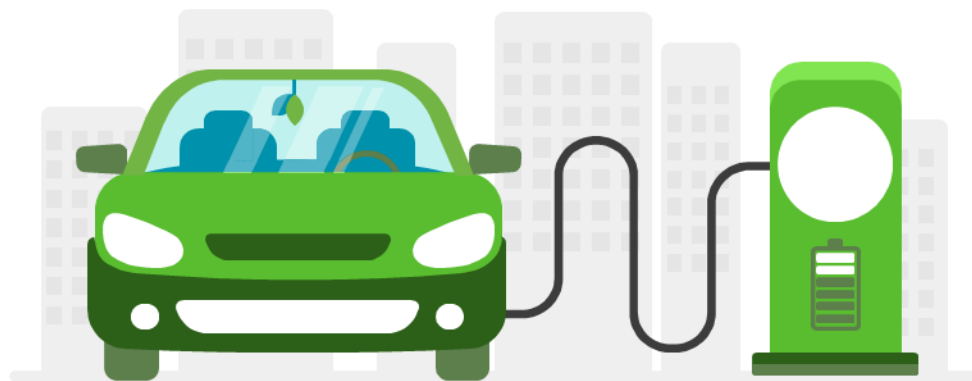
● O que correu bem:

- Conseguimos alcançar as metas estabelecidas
- Divisão equitativa das tarefas
- Aplicabilidade de bibliotecas de Data Science
- Aprofundar conhecimentos na área de Data Science



● Trabalho Futuro:

- Qualidade dos dados analisados
- Aperfeiçoar e ajustes do algoritmo de preditividade
- Adicionar mais variáveis ao algoritmo
- Estruturação do Dashboard



Obrigado pela Atenção!

Grupo 2 – Data Mining and Web Analytics

João Ramos (20200255)
Pedro Cunha (20200908)

03/01/2023



VisionOnElectrification
SOFTWARE DEVELOPMENT



Faculdade de Design,
Tecnologia e Comunicação
Universidade Europeia

Referencias 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].

DEMONSTRAÇÃO EM PYTHON

