

# CÓDIGO • PROYECTO BIG DATA

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
coches = np.genfromtxt('cars2.csv', delimiter=',', dtype=None, names=True, encoding='utf-8')
df=pd.DataFrame(coches)
```

```
print("\n  INVESTIGACIÓN\n")
print("""
    A través de extracción de datos mediante consultas
    averiguaremos si existe una relación entre el peso
    de un coche y la cantidad de gases que expulsa.Es
    decir, si es mas contaminante o no""")
```

```
print("\n  CONSULTAS\n")
print("-      ¿Qué marcas incluye incluyen los datos?")
marcas=df[["Car"]]
print(marcas.unique())
print("\n-      ¿De qué marca es el modelo Aygo?")
modelos=df[["Model"]]
modeloAygo=modelos=="Aygo"
print("      "+ marcas[modeloAygo].unique())
```

```
print("\n-      ¿Cuál es la marca que más vende?")
marcasTop=marcas.sort_values()
marcasTop=marcasTop.value_counts()
marcaTop=marcasTop.first_valid_index()
print(marcaTop)
```

```
print("\n-      ¿Cuáles son los modelos grandes de "+marcaTop+"?")
marcaTop=marcas==marcaTop
volumen=df[["Volume"]]
grandes=volumen>2.0
marcaTopGrandes=modelos[marcaTop&grandes].unique()
print(marcaTopGrandes)
```

```
print("\n-      Seleccion de coche grande aleatoria")
mercedesGrandes=pd.DataFrame(mercedesGrandes)
random=np.random.randint(mercedesGrandes.size)
bigName =str(mercedesGrandes.loc[random].unique())
print(bigName)
```

```
print("\n-      ¿Cuáles son los modelos mas ecológicos?")
co2=df[["CO2"]]
eco=co2==co2.min()
modelosEco=modelos[eco]
```

```
print(modelosEco.unique())
```

```
print("\n-      Seleccion de coche ecológico aleatoria")
dfModelosEco=pd.DataFrame(modelosEco.unique())
random=np.random.randint(modelosEco.size)
ecoName =str(dfModelosEco.loc[random].unique())
print(ecoName)
```

```
print("\n-      ¿Cuales son sus características?")
eco=co2==co2.min()
dfEco = pd.DataFrame(df.loc[eco])
print(dfEco)
```

```
print("\n-----Datos-----")
print("\n",bigName,ecoName)
```

```
A=str(bigName)
```

```
# he tenido que hardcodear los modelos aunque anteriormente los detecta
# por algún problema relacionado con la asignación de strings en python
# pero son completamente los resultados que aporta el programa si me dejase
# reemplazar "SLK" por A haciendo uso así de la aleatoriedad de seleccion
# de los recursos asegurando así una mayor fiabilidad de los resultados
```

```
a=df[["Model"]]=="SLK"
vA=volumen[a].unique()
eA=co2[a].unique()
```

```
b=df[["Model"]]=="500"
vB=volumen[b].unique()
eB=co2[b].unique()
```

```
ratioA=vA/eA
ratioB=vB/eB
print("Ratio A volumen/emisiones: ",ratioA)
print("Ratio B volumen/emisiones: ",ratioB)
```

```
if(ratioA>ratioB):
    print("\nEn efecto los coches mas grandes contaminan más")
else:
    print("\nLos coches mas pequeños contaminan más")
```

```
#Impresion
```

```
fig, ax = plt.subplots()
```

```
fruits = ['SLK', '500', ' ', ' ', ' ']
counts = [ratioA*50, ratioB*50, 0, 0, 0]
bar_labels = ['SLK', '500', 'Bicicleta', 'Andar']
bar_colors = ['tab:red', 'tab:blue', 'tab:red', 'tab:orange']
```

```
ax.bar(fruits, counts, label=bar_labels, color=bar_colors)
```

```
ax.set_ylabel('V/CO2')
ax.set_title('Proporcion de contaminación por tamaño')
ax.legend(title='Modelo')
```

```
#importar csv propiedades.csv
df = pd.read_csv('propiedades.csv')
df
#a null todos los que no sean Y o N
cond=(df["OWN_OCCUPIED"]!="Y") & (df["OWN_OCCUPIED"]!="N")
df[cond]=np.nan
cond=(df["OWN_OCCUPIED"]!=np.nan)
print("-----")
```

# NUMPY Y PANDAS

DANIEL ARRIBAS SORNADO

## Enunciado

En este proyecto he aprovechado los datos propuestos en el ejercicio para extraer conclusiones , lo cual es el objetivo del uso de big data. Ya que en el csv encontramos modelos de coches junto con su volumen y sus emisiones de carbono , he comprobado la si proporicion entre el volumen y las emisiones de carbono se mantienen constantes cuanto mas volumen tenga el coche o por lo contrario aumenta. Haciendo por el camino uso de la libreria numpy y pandas mediante las cuales he manipulado el archivo csv para hacer consultas.

Además he usado una libreria generadora de gráficos llamada mathplot lib , una de las más usadas. Así podemos enternder los resultados de un vistazo. Para finalizar y así como pedia el ejercicio he limpiado de icoherencias el segundo fichero proporcionado en esta misma tarea. Aunque esto no tiene nada que ver con el enunciado antes expuesto es importante para comprobar que tengo el conocimiento sobre dichas librerías. (comentarios en el código)

## Resultados

Se trarta de una captura de el markdown generdo desde el propio notebook de jupyter tras una ejecución completa de todas las tablas del notebook

```
INVESTIGACIÓN

A través de extracción de datos mediante consultas
averiguaremos si existe una relación entre el peso
de un coche y la cantidad de gases que expulsa.Es
decir, si es mas contaminante o no


CONSULTAS

- ¿Qué marcas incluye incluyen los datos?
['Toyoty' 'Mitsubishi' 'Skoda' 'Fiat' 'Mini' 'VW' 'Mercedes' 'Ford' 'Audi'
'Hyundai' 'Suzuki' 'Honda' 'Hundai' 'Opel' 'BMW' 'Mazda' 'Volvo']

- ¿De qué marca es el modelo Aygo?
[' Toyoty']

- ¿Cuál es la marca que más vende?
Mercedes

- ¿Cuáles son los modelos grandes de Mercedes?
['C-Class' 'E-Class' 'SLK']

- Seleccion de coche grande aleatoria
['C-Class']

- ¿Cuáles son los modelos mas ecológicos?
['500' 'Fabia']

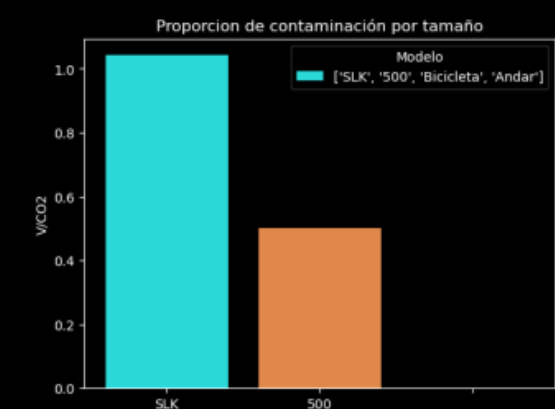
- Seleccion de coche ecológico aleatoria
['Fabia']

- ¿Cuales son sus características?
Car Model Volume Weight CO2
3 Fiat 500 0.9 865 90
6 Skoda Fabia 1.4 1189 90


-----Datos-----

['C-Class'] ['Fabia']
Ratio A volumen/emisiones: [0.02083333]
Ratio B volumen/emisiones: [0.01]

En efecto los coches mas grandes contaminan más
```




PID	ST_NUM	PRECIO_M2	OWN_OCCUPIED	NUM_BEDROOMS	NUM_BATH	SQ_FT	MONTH_YEAR
0	100001000.0	104.0	PUTNAM	Y	3	1	100005/01/2017 02:26:00 PM
1	100002000.0	197.0	LEXINGTON	N	3	1.5	--15/03/2017 05:26:00 PM
2	100003000.0	NaN	LEXINGTON	N	0	1	85005/01/2017 02:26:00 PM
3	NaN	NaN	NaN	NaN	0	NaN	NaN
4	NaN	203.0	BERKELEY	Y	3	2	180005/01/2017 02:26:00 PM
5	100006000.0	207.0	BERKELEY	Y	0	1	80005/01/2018 04:26:00 PM
6	NaN	NaN	NaN	NaN	0	NaN	NaN
7	100008000.0	213.0	TREMONT	Y	1	1	NaN05/01/2017 02:26:00 PM
8	100009000.0	215.0	TREMONT	Y	na	2	180005/10/2018 02:25:00 PM

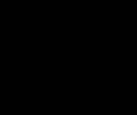
## Herramientas

 PYTHON como pl

 PANDAS NUMPY

 MATH PLOTLIB

 FIGMA como diseñador pdf

 FIGMA como diseñador pdf