Proyecto Forbes 2021 - 2022

Manipulación y transformación de datos con Pandas y Numpy

Autores

- Victor Manuel Arenas Lopez
- Marct Zullieth Isaza Yoja

In [ ]:
```
import os
import numpy as np
import pandas as pd
```

In [ ]:
```
os.chdir('C:/Users/Victor Manuel Arenas/Projects/IA/Inteligencia_artificial')
os.listdir()
```

Out[ ]:
```
['.git',
 '.gitignore',
 '.venv',
 '.vscode',
 '1. Introduccion.ipynb',
 '2. Numpy.ipynb',
 '3. PandasForbes.ipynb',
 'caracteristicas de vinos.csv',
 'Explicacion del modelo kmeans.ipynb',
 'Forbes 2000 2021.csv',
 'Forbes 2000 2022.csv',
 'forbes2023.csv',
 'Mall_Customers.csv',
 'Pandas_Forbes2022.ipynb',
 'Project kmeans customer.ipynb']
```

In [ ]:
```
forbes23 = pd.read_csv('forbes2023.csv',encoding= 'latin-1',sep=';')
```

In [ ]:
```
# Cambiar el nombre de las columnas del DF
forbes23.columns = ['Rank','Company','Country','Sales','Profits','Assets','Market_v
forbes23.columns
```

Out[ ]:
```
Index(['Rank', 'Company', 'Country', 'Sales', 'Profits', 'Assets',
       'Market_value'],
      dtype='object')
```

In [ ]:
```
# Quitar las ',' de todo el DF usando una expresion regular
forbes23['Profits'] = forbes23['Profits'].replace(',', '', regex=True)
```

In [ ]:
```
# Coersionando los datos de str a float de la columna Profits
forbes23.loc[:,'Profits'] = forbes23.loc[:,'Profits'].astype(float)
```

In [ ]:
```
# Validar si existe algun dato vacio en el DF
forbes23.isna().any()
```

Out[ ]:
```
Rank            False
Company         False
Country         False
Sales           False
Profits         False
Assets          False
Market_value    False
dtype: bool
```

In [ ]:
```
# Accediendo a las filas de la 2 a la 4 mostrando las columnas Company, Country y S
forbes23[['Company','Country','Sales']][2:4]
```

Out[ ]:

	Company	Country	Sales
2	ICBC	China	216770
3	China Construction Bank	China	203080

In [ ]:
```
# GroupBy Country
forbes23.groupby('Country').size()
```

Out[ ]:

Country	
Argentina	1
Australia	32
Austria	9
Belgium	6
Bermuda	6
Brazil	22
Canada	57
Cayman Islands	2
Chile	8
China	302
Colombia	4
Czech Republic	1
Denmark	10
Egypt	1
Finland	10
France	52
Germany	53
Greece	8
Hong Kong	44
Hungary	2
India	55
Indonesia	8
Ireland	22
Israel	11
Italy	28
Japan	192
Jordan	1
Kazakhstan	2
Kuwait	2
Luxembourg	6
Malaysia	8
Mexico	13
Morocco	2
Netherlands	25
Nigeria	2
Norway	9
Oman	1
Pakistan	1
Peru	1
Philippines	4
Poland	7
Portugal	4
Qatar	6
Saudi Arabia	17
Singapore	14
South Africa	12
South Korea	59
Spain	20
Sweden	24
Switzerland	44
Taiwan	45
Thailand	17
Turkey	9
United Arab Emirates	16
United Kingdom	67

```
United States           610
Uruguay                   1
Vietnam                   5
dtype: int64
```

Analisis Exploratorio

In [ ]:
```
# Primeras 5 empresas del ranking forbes 2020 2023
forbes23.head(5)['Company']
```

Out[ ]:
```
0                              JPMorgan Chase
1    Saudi Arabian Oil Company (Saudi Aramco)
2                                       ICBC
3                       China Construction Bank
4                    Agricultural Bank of China
Name: Company, dtype: object
```

In [ ]:
```
# Valorar el 1% de las empresas con mas ganancias en el mundo
p99p = np.percentile(forbes23['Profits'],99)
print(p99p)
```

```
24840.59999999999
```

In [ ]:
```
# Filtrar las compañias con ganancias mayores a p99p
forbes23[forbes23['Profits'] > p99p][['Company','Country','Profits']]
```

Out[ ]:

	Company	Country	Profits
0	JPMorgan Chase	United States	41800.0
1	Saudi Arabian Oil Company (Saudi Aramco)	Saudi Arabia	156360.0
2	ICBC	China	52470.0
3	China Construction Bank	China	48250.0
4	Agricultural Bank of China	China	37920.0
5	Bank of America	United States	28620.0
6	Alphabet	United States	58590.0
7	ExxonMobil	United States	61690.0
8	Microsoft	United States	69020.0
9	Apple	United States	94320.0
10	Shell	United Kingdom	43510.0
11	Bank of China	China	33230.0
13	Samsung Electronics	South Korea	34490.0
17	Chevron	United States	35780.0
27	BP	United Kingdom	25890.0
34	Tencent Holdings	China	27260.0
38	Pfizer	United States	29040.0
43	Taiwan Semiconductor	Taiwan	33570.0
51	Equinor	Norway	28980.0
57	Petrobras	Brazil	36470.0

In [ ]:
```
# Filtrar las empresas Colombianas que tuvieron ganancias mayoes $1.000
forbes23[(forbes23['Country']== 'Colombia') & (forbes23['Profits']> 1000)]
```

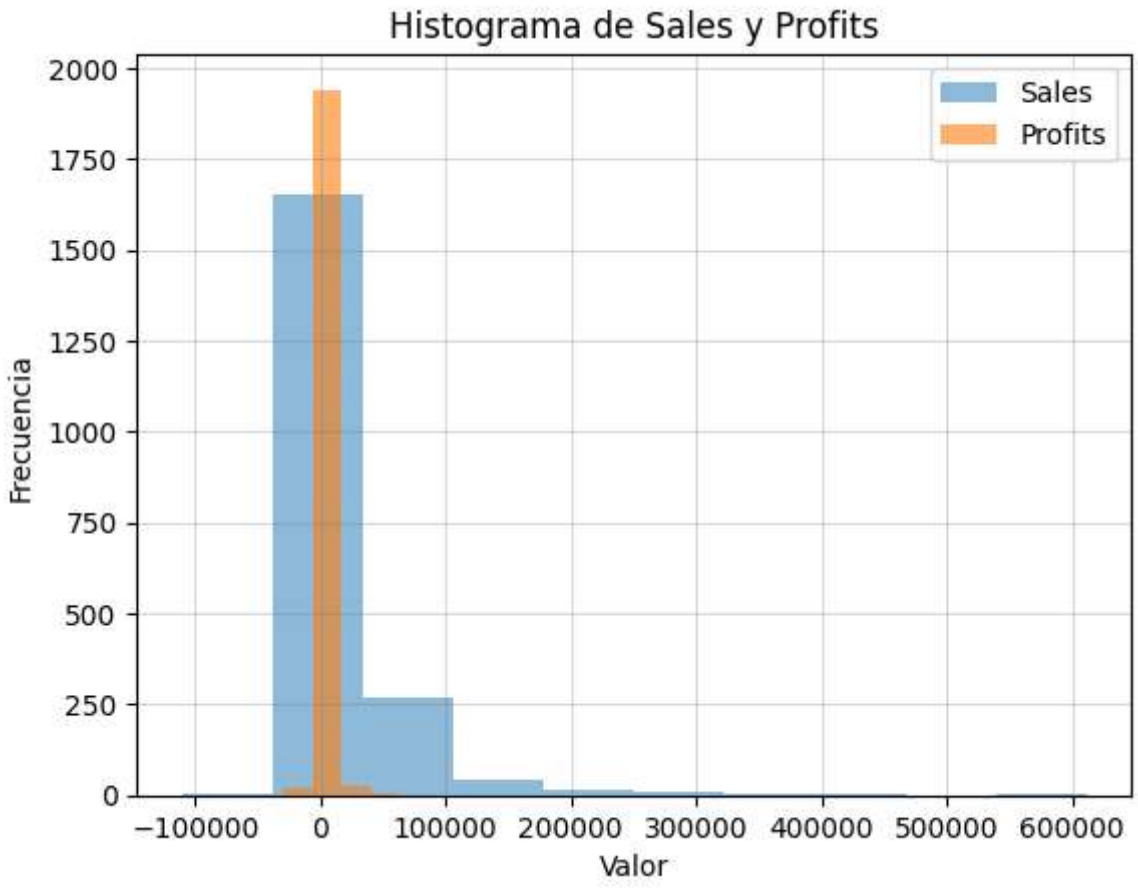
Out[ ]:

	Rank	Company	Country	Sales	Profits	Assets	Market_value
310	311	Ecopetrol	Colombia	33670	7850.0	62450	18760
896	897	Bancolombia	Colombia	7850	1590.0	72760	7380
1404	1405	Grupo Aval	Colombia	6580	5828.0	60960	2920
1977	1977	Grupo Bolivar	Colombia	5880	2492.0	41940	1110

In [ ]:
```
import matplotlib.pyplot as plt
```

In [ ]:
```
# Comparacion de las ventas vs ganancias de las compañias
# Crear un histograma para las ventas de todas las compañias
```

```
plt.hist(forbes23['Sales'], alpha=0.5, label='Sales')
# Crear un histograma para las ganancias de todas las compañias
plt.hist(forbes23['Profits'], alpha=0.6, label='Profits')
# Configurar las propiedades de leyenda del gráfico
plt.legend()
plt.xlabel('Valor')
plt.ylabel('Frecuencia')
plt.title('Histograma de Sales y Profits')
plt.grid(color='gray', linestyle='solid', alpha = 0.3)
plt.show()
```



In [ ]:
```
# Paises definidos para Sudamerica
suda = ['Argentina','Bolivia','Brazil','Chile','Colombia','Ecuador','Paraguay','Per
# Crear un DF solamente con las compañias que pertenezcan a  los paises definidos p
f_suda = forbes23[forbes23['Country'].isin(suda)]
f_suda.sample(20)
```

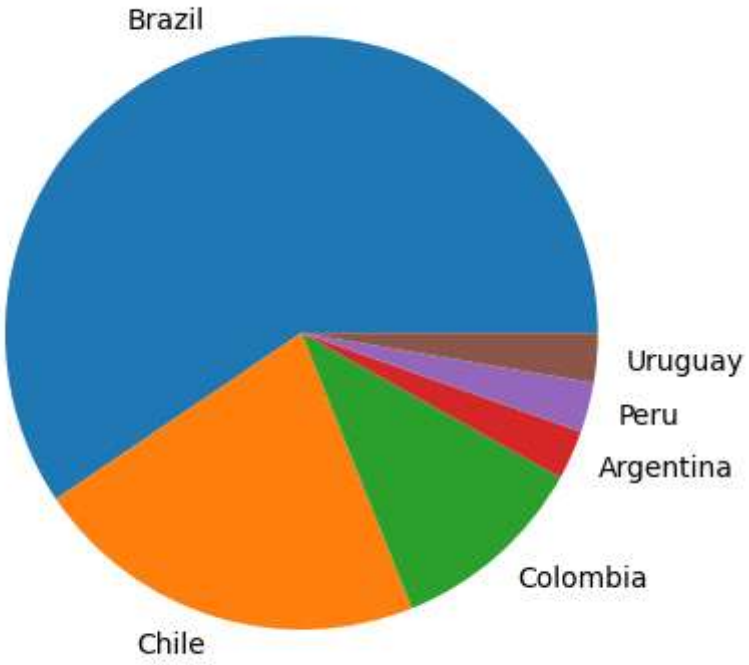
Out[ ]:

	Rank	Company	Country	Sales	Profits	Assets	Market_value
1435	1436	XP	Brazil	2650	6938.0	36370	8360
562	563	JBS	Brazil	72580	2990.0	39420	8160
1229	1230	BCI-Banco Credito	Chile	4470	8933.0	85310	5090
749	750	YPF	Argentina	18630	2210.0	25910	9050
773	774	Suzano	Brazil	9960	3550.0	26700	10950
1429	1430	Latam Airlines	Chile	9430	1480.0	13210	4520
1573	1574	B3	Brazil	1760	8183.0	9210	13840
1977	1977	Grupo Bolivar	Colombia	5880	2492.0	41940	1110
1222	1223	WEG	Brazil	5970	8865.0	5700	33130
927	928	Eletrobr□s	Brazil	6790	705.0	51180	17280
57	58	Petrobras	Brazil	124170	36470.0	184990	63030
949	950	Credicorp	Peru	5880	1210.0	62080	11240
1397	1398	Metalurgica Gerdau	Brazil	15700	751.0	15140	2390
805	806	SQM	Chile	10790	3940.0	10820	19320
371	372	Banco Btg Pactual	Brazil	15490	1520.0	85350	50770
1678	1679	Ultrapar Participacoes	Brazil	27670	3124.0	6690	3610
1014	1015	Marfrig Global Foods	Brazil	25290	8045.0	25780	894
1602	1602	Vibra Energia	Brazil	35130	2976.0	7920	3090
172	173	Banco Bradesco	Brazil	56260	3490.0	357450	31600
1404	1405	Grupo Aval	Colombia	6580	5828.0	60960	2920

In [ ]:

```
# Crear el DF para guardar las frecuencias e indices del DF de compañias sudamenric
table_suda=f_suda['Country'].value_counts()

# Grafico pie de las compañias sudamericanas
plt.pie(table_suda.values,labels=table_suda.index)
plt.show() #Imprime el gráfico
```


Brazil
Uruguay
Peru
Argentina
Colombia
Chile