

Prueba - Simulación

Estudiante: Eduardo Zhizhpon

Fecha: 04/05/2022

```
In [2]: import pandas as pd
import matplotlib.pyplot as plt
import zipfile
import shutil
from os import path
import os
```

```
In [68]: df = pd.read_csv('dataset-sri-resumen.csv')
df.columns
```

```
Out[68]: Index(['Ranking (1) Tamaño 2021', 'Ranking (1) Tamaño 2020',
'Variación Ranking Tamaño 2020 - 2021', 'Nombre Grupo Económico',
'Reporta sus BENEFICIARIOS EFECTIVOS \n(*)',
'No. de Integrantes domiciliados en PARAÍSO FISCAL (**) ',
'No. de Integrantes y Empresas Offshore relacionadas con el Grupo
Económico publicados en PANAMÁ PAPERS (***)',
'Personas \nNaturales (a) ', 'Personas \nExtranjeras (b) ',
'Sociedades \nNacionales (c)', 'Sociedades \nExtranjeras (d)',
'Contribuyentes \n Especiales',
'Grandes \nContribuyentes e \nImportancia \nRecaudatoria',
'Instituciones \nFinancieras (2)', 'Medios de \nComunicación (3)',
'Total Integrantes e=a+b+c+d', 'Ranking Contribución I.R. 2020',
'Ranking Contribución I.R. 2019',
'Variación Ranking Contribución I.R. \n2019-2020',
'Impuesto a la Renta Causado (4) 2020', 'Total Ingresos 2020',
'Contribución I.R. 2020', 'Impuesto a la Renta Causado (4) 201
9',
'Total Ingresos 2019', 'Contribución I.R. 2019', 'Variación Ingr
esos',
'Variación Impuesto a la Renta Causado',
'Variación Contribución I.R. (2020-2019)', 'Impuesto Causado ISD
2020',
'Impuesto Causado ISD 2019', 'Impuesto Causado IAE 2020 (5)',
'Impuesto Causado IAE 2019 (5)', 'Total Activo \n(6) 2020',
'Total Activo \n(6) 2019', 'Total Patrimonio (7) 2020',
'Total Patrimonio \n(7) 2019', '\nTotal Recaudado (8) 2020',
'\nTotal Recaudado (8) 2019'],
dtype='object')
```

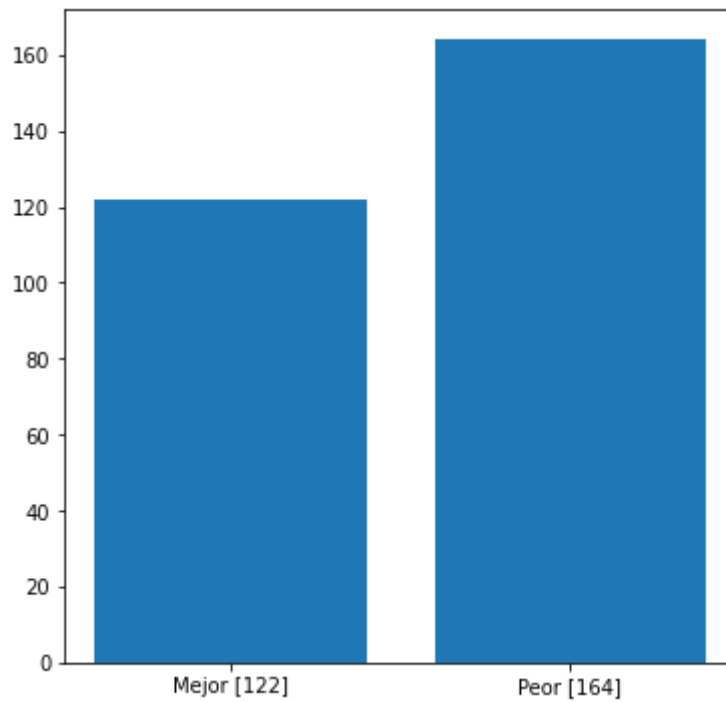
- Generar gráfico de barras de solo las empresas que han mejorado y/o empeorado su ranking con respecto al año anterior.

```
In [67]: df = pd.read_csv('dataset-sri-resumen.csv')

mejor = df[df['Variación Ranking Tamaño 2020 - 2021'] < 0.0].shape[0]
peor = df[df['Variación Ranking Tamaño 2020 - 2021'] > 0.0].shape[0]
y = [mejor, peor]
x = ['Mejor [' + str(mejor) + ']', 'Peor [' + str(peor) + ']']

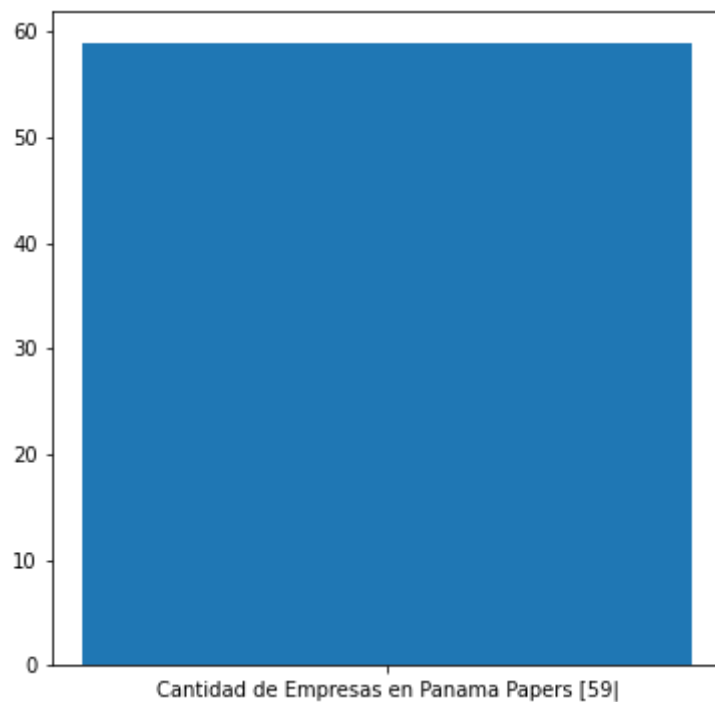
fig = plt.figure(figsize=(6,6))
ax = fig.add_subplot(1, 1, 1)
```

```
ax.bar(x,y)
plt.show()
```



- Generar gráfico de las empresas que tienen panamá papers.

```
In [96]: df = pd.read_csv('dataset-sri-resumen.csv')
df.columns
cantidad_empresas = df[df['No. de Integrantes y Empresas Offshore relacionadas'] > 0]
cantidad_empresas = len(cantidad_empresas['Nombre Grupo Económico'].unique())
fig = plt.figure(figsize=(6,6))
ax = fig.add_subplot(1, 1, 1)
ax.bar('Cantidad de Empresas en Panama Papers [' + str(cantidad_empresas) +
plt.show()
```



- Generar un pie de las empresas que más aportan a la economía del país.

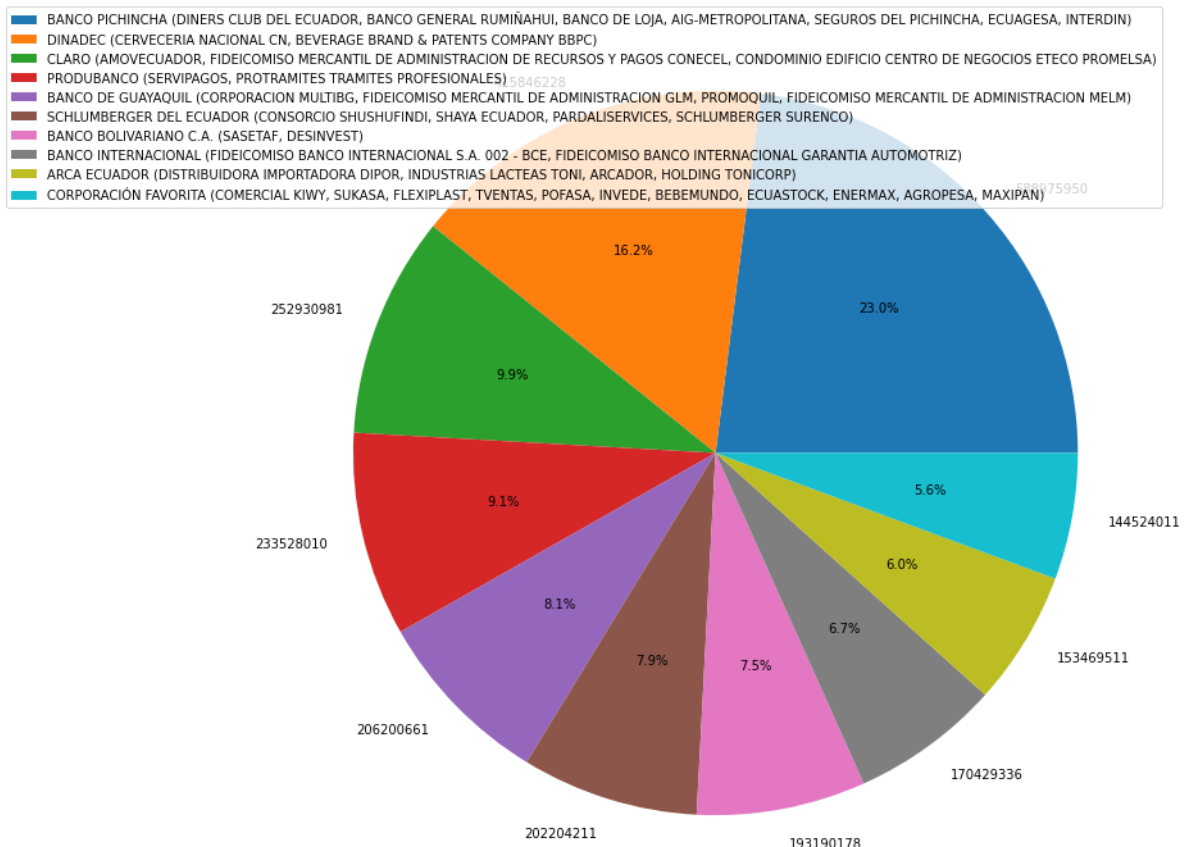
In [119...

```
column = '\nTotal Recaudado (8) 2020'

df = pd.read_csv('dataset-sri-resumen.csv')
df = df.dropna(subset=[column])
df[column] = df[column].str.replace(',', '')
df[column] = df[column].astype(int)
df_aporte = df.sort_values(column, ascending=False)[1:11]
aporte_values = df_aporte[column].values

empresas = df_aporte['Nombre Grupo Económico'].values

fig = plt.figure(figsize=(15,13))
ax = fig.add_subplot(1, 1, 1)
ax.pie(aporte_values, labels = aporte_values, autopct='%1.1f%%')
ax.legend(empresas)
plt.show()
```



- Obtener las empresas que mayor patrimonio tienen (10).

In [121...

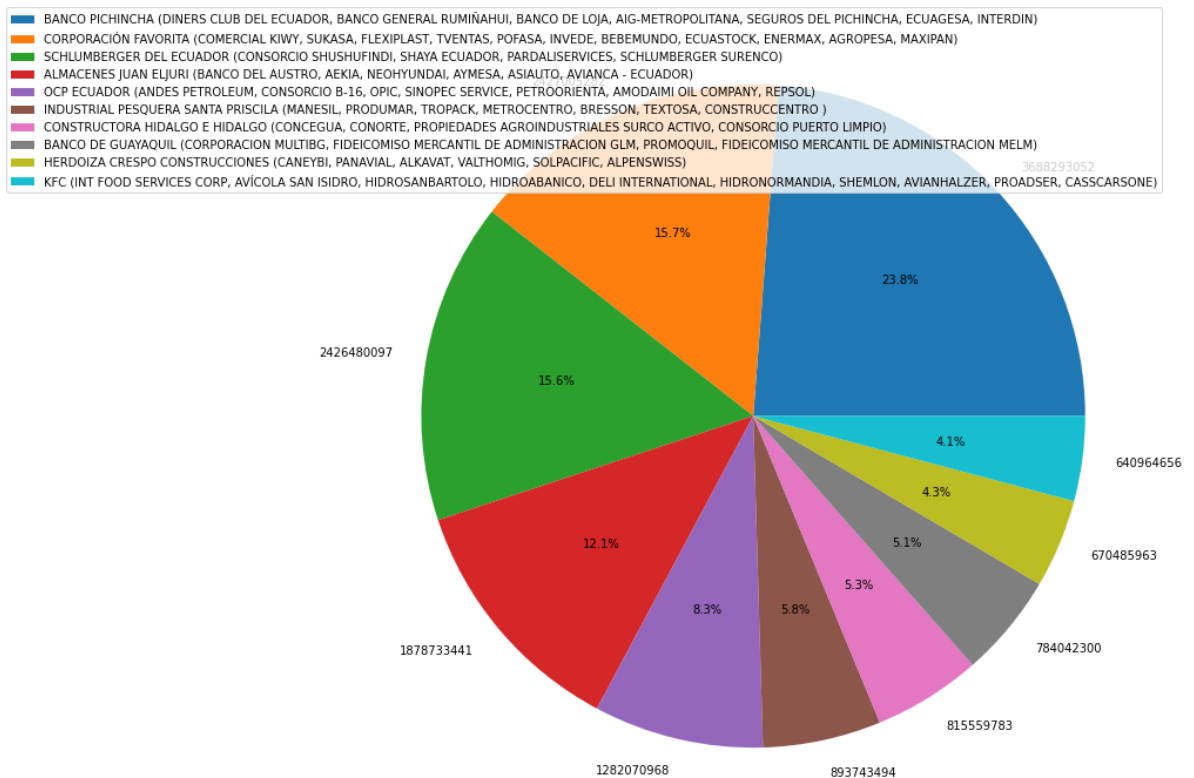
```
column = 'Total Patrimonio (7) 2020'

df = pd.read_csv('dataset-sri-resumen.csv')
df = df.dropna(subset=[column])
df[column] = df[column].str.replace(',', '')
df[column] = df[column].astype(int)
df_aporte = df.sort_values(column, ascending=False)[1:11]
aporte_values = df_aporte[column].values

empresas = df_aporte['Nombre Grupo Económico'].values

fig = plt.figure(figsize=(15,13))
ax = fig.add_subplot(1, 1, 1)
ax.pie(aporte_values, labels = aporte_values, autopct='%1.1f%%')
```

```
ax.legend(empresas)
plt.show()
```



Recomendaciones

La actividad más complicada para estas actividades fue filtrar la información. Ya que pasar de un archivo Excel a un formato que el módulo "pandas" pueda trabajar sin problemas es complicado. Para ello, se recomienda eliminar algunos atributos y filas con valores nulos para poder realizar las operaciones y validaciones necesarias.

Además, las gráficas realizadas proporcionan la información necesaria para su respectivo análisis posterior. Lo que permitirá obtener criterios, estrategias o predicciones a futuro de los aportes de las empresas al país.

In []: