In [1]:
```
#The below script can help us to identify the top 100 people's
    #wealth distribution
    #industry representation
    #geographical representation
    #age distribution.
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
dataframe=pd.read_csv('TopRichestInWorld.csv')
```

In [127…
```
#Top 5 richest people
dataframe.head(7)
```

Out[127…

	Name	NetWorth	Age	Country/Territory	Source	Industry
0	Elon Musk	$219,000,000,000	50	United States	Tesla, SpaceX	Automotive
1	Jeff Bezos	$171,000,000,000	58	United States	Amazon	Technology
2	Bernard Arnault & family	$158,000,000,000	73	France	LVMH	Fashion & Retail
3	Bill Gates	$129,000,000,000	66	United States	Microsoft	Technology
4	Warren Buffett	$118,000,000,000	91	United States	Berkshire Hathaway	Finance & Investments
5	Larry Page	$111,000,000,000	49	United States	Google	Technology
6	Sergey Brin	$107,000,000,000	48	United States	Google	Technology

In [154…
```
#The bottom 5 richest people
dataframe.tail()
```

Out[154…

	Name	NetWorth	Age	Country/Territory	Source	Indus
96	Vladimir Potanin	1.730000e+10	61	Russia	metals	Meta Mi
97	Harold Hamm & family	1.720000e+10	76	United States	oil & gas	Ene
98	Sun Piaoyang	1.710000e+10	63	China	pharmaceuticals	Health
99	Luo Liguo & family	1.700000e+10	66	China	chemicals	Manufactu
100	Peter Woo	1.700000e+10	75	Hong Kong	real estate	Real Es

In [129…

```
#Sorting people from youngest to oldest
print(dataframe.sort_values(by='Age', ascending=True))
```

```
                     Name          NetWorth  Age Country/Territory  \
59        Sam Bankman-Fried   $24,000,000,000   30     United States
14          Mark Zuckerberg   $67,300,000,000   37     United States
24             Zhang Yiming   $50,000,000,000   38             China
65         Guillaume Pousaz   $23,000,000,000   40       Switzerland
85     Yang Huiyan & family   $18,700,000,000   40             China
..                      ...               ...  ...               ...
64           Leonard Lauder   $23,100,000,000   89     United States
75  Rupert Murdoch & family   $20,800,000,000   91     United States
4            Warren Buffett  $118,000,000,000   91     United States
36              Li Ka-shing   $34,800,000,000   93         Hong Kong
38             Lee Shau Kee   $32,600,000,000   94         Hong Kong

                    Source              Industry
59  cryptocurrency exchange  Finance & Investments
14                 Facebook            Technology
24                  TikTok  Media & Entertainment
65                 fintech  Finance & Investments
85             real estate           Real Estate
..                     ...                   ...
64             Estee Lauder       Fashion & Retail
75   newspapers, TV network  Media & Entertainment
4        Berkshire Hathaway  Finance & Investments
36             diversified           Diversified
38             real estate           Real Estate

[101 rows x 6 columns]
```

In [130…

```
#details of people standing at position 50 to 55 using iloc
print(dataframe.iloc[50:55,:])
print(dataframe.loc[:4, ['Name', 'Age']])
```

```
                            Name         NetWorth  Age Country/Territory  \
50            Dietrich Mateschitz  $27,400,000,000   77            Austria
51  Leonardo Del Vecchio & family  $27,300,000,000   86              Italy
52                     Ken Griffin  $27,200,000,000   53     United States
53         Tadashi Yanai & family  $26,100,000,000   73              Japan
54                William Lei Ding  $25,200,000,000   50              China

            Source               Industry
50        Red Bull        Food & Beverage
51      eyeglasses       Fashion & Retail
52     hedge funds  Finance & Investments
53  fashion retail       Fashion & Retail
54    online games             Technology
                       Name  Age
0                 Elon Musk   50
1                Jeff Bezos   58
2  Bernard Arnault & family   73
3                Bill Gates   66
4             Warren Buffett   91
```

In [131…
```
#To check if the data is cleaned. this checks missing values
dataframe.isnull().sum()
```

Out[131…
```
Name                0
NetWorth            0
Age                 0
Country/Territory   0
Source              0
Industry            0
dtype: int64
```

In [132…
```
#Summary of statistical data
dataframe.describe()
```

Out[132…

	Age
count	101.000000
mean	67.118812
std	13.892651
min	30.000000
25%	57.000000
50%	67.000000
75%	77.000000
max	94.000000

In [133…
```
dataframe.nlargest(1,'Age')
```

Out[133…

	Name	NetWorth	Age	Country/Territory	Source	Industry
38	Lee Shau Kee	$32,600,000,000	94	Hong Kong	real estate	Real Estate

In [134…
```
dataframe.nsmallest(1,'Age')
```

Out[134…

	Name	NetWorth	Age	Country/Territory	Source	Indust
59	Sam Bankman-Fried	$24,000,000,000	30	United States	cryptocurrency exchange	Finance Investmer

In [135…
```
 #information of dataset
dataframe.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 101 entries, 0 to 100
Data columns (total 6 columns):
 #   Column             Non-Null Count  Dtype 
---  ------             --------------  ----- 
 0   Name               101 non-null    object
 1   NetWorth           101 non-null    object
 2   Age                101 non-null    int64 
 3   Country/Territory  101 non-null    object
 4   Source             101 non-null    object
 5   Industry           101 non-null    object
dtypes: int64(1), object(5)
memory usage: 4.9+ KB
```

In [136…
```
#get a randomly selected row, column, or both from a dataset
dataframe.sample()
```

Out[136…

	Name	NetWorth	Age	Country/Territory	Source	Industry
0	Elon Musk	$219,000,000,000	50	United States	Tesla, SpaceX	Automotive

In [137…
```
#To view columns of the
dataframe.columns
```

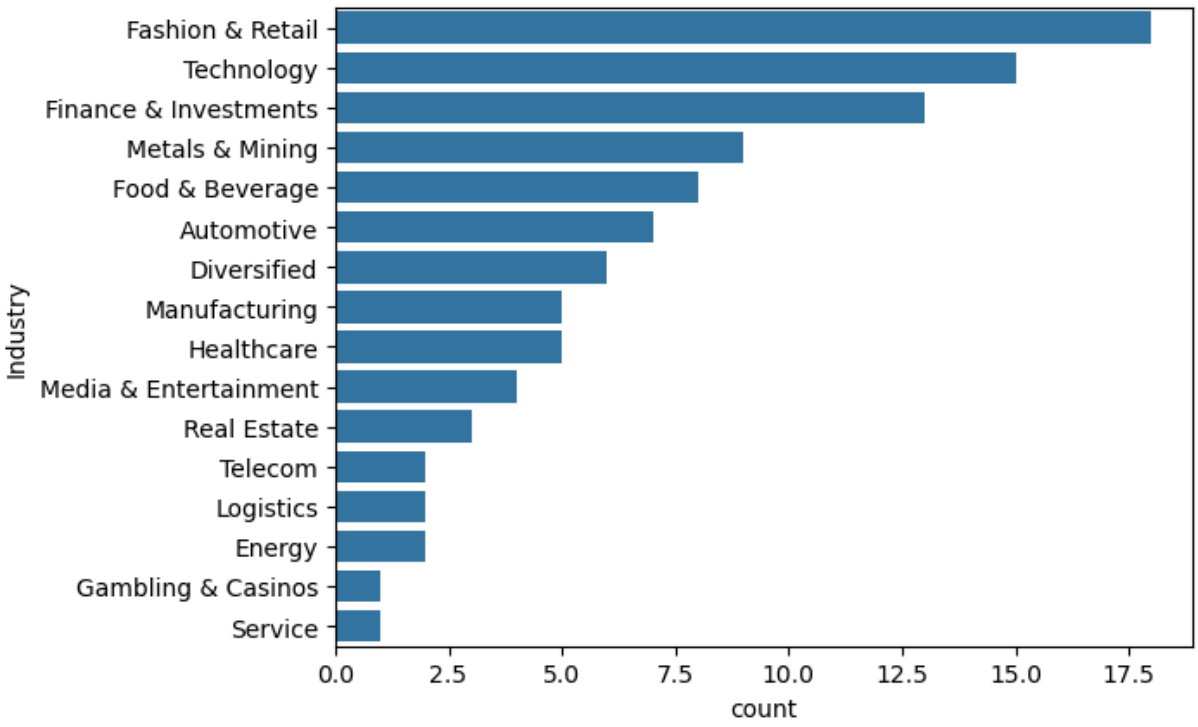
Out[137…
```
Index(['Name', 'NetWorth', 'Age', 'Country/Territory', 'Source', 'Industr
y'], dtype='object')
```

In [138…
```
#The value_counts() function in pandas is used to count the occurrences of e
print(dataframe.value_counts())
```

```
Name                    NetWorth          Age  Country/Territory  Source
Industry
Abigail Johnson         $21,200,000,000   60   United States      money manage
ment             Finance & Investments     1
Alain Wertheimer        $31,200,000,000   73   France             Chanel
Fashion & Retail          1
Alice Walton            $65,300,000,000   72   United States      Walmart
Fashion & Retail          1
Amancio Ortega          $59,600,000,000   86   Spain              Zara
Fashion & Retail          1
Andrew Forrest          $17,800,000,000   60   Australia          mining
Metals & Mining           1

..
Warren Buffett          $118,000,000,000  91   United States      Berkshire Ha
thaway           Finance & Investments     1
William Lei Ding        $25,200,000,000   50   China              online games
Technology                1
Yang Huiyan & family    $18,700,000,000   40   China              real estate
Real Estate               1
Zhang Yiming            $50,000,000,000   38   China              TikTok
Media & Entertainment     1
Zhong Shanshan          $65,700,000,000   67   China              beverages, p
harmaceuticals  Food & Beverage           1
Name: count, Length: 101, dtype: int64
```

In [139…

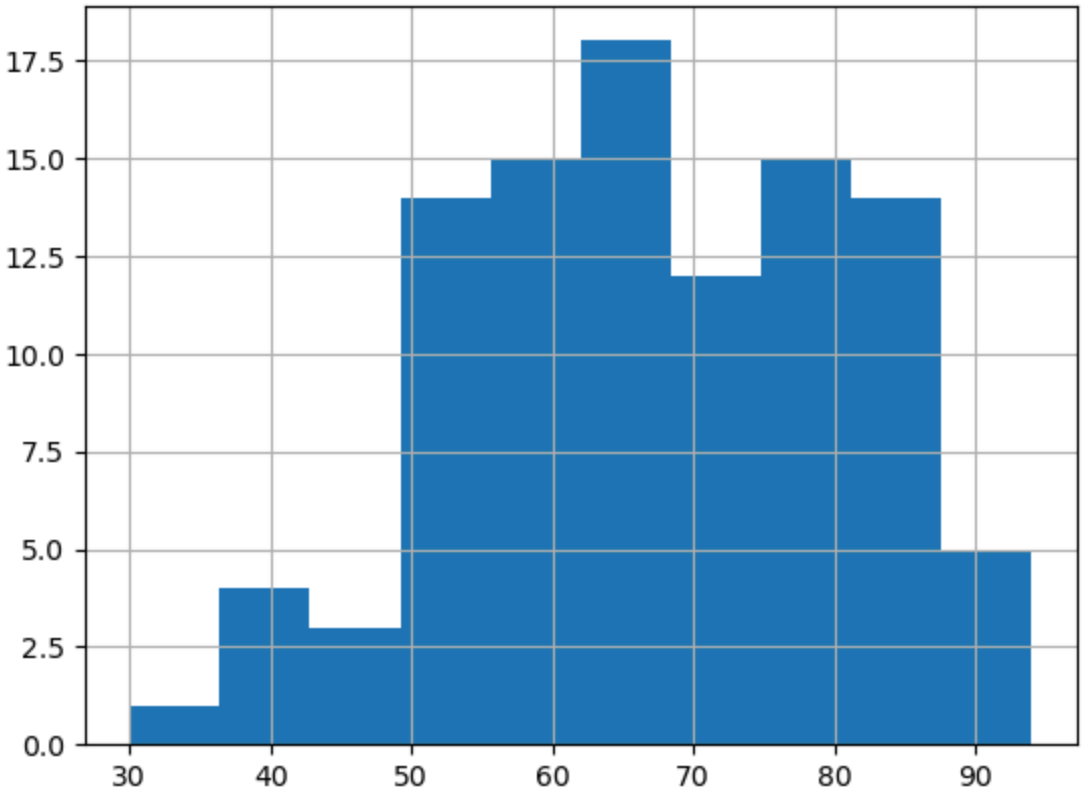
```
#The below bargraph represents which industry has

sns.countplot(dataframe["Industry"] , order=dataframe["Industry"].value_cour
plt.show()
```



In [140…

```
dataframe['Age'].hist()
```

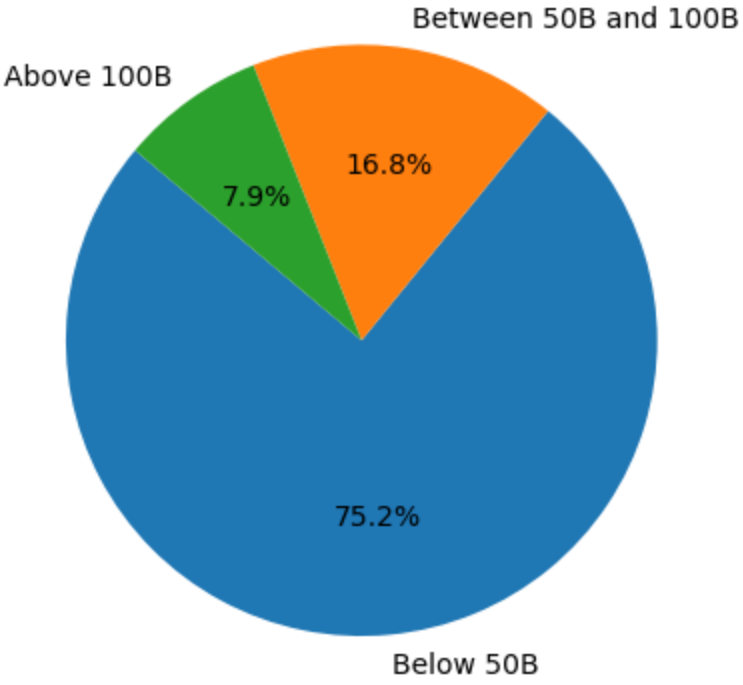
Out[140… <Axes: >



In [141…
```
#for pie chart
# Preprocess the data
#Setting regex=False disables regular expression syntax and treats the patte
dataframe['NetWorth'] = dataframe['NetWorth'].astype(str)
dataframe['NetWorth'] = dataframe['NetWorth'].str.replace('$', '', regex=Fal
dataframe['NetWorth'] = dataframe['NetWorth'].str.replace(',', '', regex=Fal
filtered_df1 = dataframe[(dataframe['NetWorth'] >= 0) & (dataframe['NetWorth
net_worth1 = filtered_df1['NetWorth'].shape[0]
print(net_worth1)

filtered_df2 = dataframe[(dataframe['NetWorth'] >= 50000000000) & (datafram
net_worth2 = filtered_df2['NetWorth'].shape[0]
print(net_worth2)

filtered_df3 = dataframe[(dataframe['NetWorth'] >= 100000000000)]
net_worth3 = filtered_df3['NetWorth'].shape[0]
print(net_worth3)

labels = ['Below 50B', 'Between 50B and 100B', 'Above 100B']
values = [net_worth1,net_worth2,net_worth3]

plt.pie(values, labels=labels, autopct='%1.1f%%', startangle=140)
```

```
76
17
8
```

Out[141…
```
([<matplotlib.patches.Wedge at 0x7df7ddf9b800>,
  <matplotlib.patches.Wedge at 0x7df7ddf9be90>,
  <matplotlib.patches.Wedge at 0x7df7ddf0b140>],
 [Text(0.10438966273027035, -1.095035523768549, 'Below 50B'),
  Text(0.1685098281895781, 1.0870163006153675, 'Between 50B and 100B'),
  Text(-0.6425587533082865, 0.8928147896103092, 'Above 100B')],
 [Text(0.056939816034692906, -0.597292103873754, '75.2%'),
  Text(0.09191445173976986, 0.5929179821538367, '16.8%'),
  Text(-0.35048659271361077, 0.48698988524198683, '7.9%')])
```



In [142…
```
dataframe.fillna(value={'Column1': 'DefaultValue', 'Column2': 0}, inplace=Tr
print(dataframe.head())
```

```
                      Name      NetWorth  Age Country/Territory  \
0                 Elon Musk  2.190000e+11   50     United States
1               Jeff Bezos   1.710000e+11   58     United States
2  Bernard Arnault & family  1.580000e+11   73            France
3               Bill Gates   1.290000e+11   66     United States
4           Warren Buffett   1.180000e+11   91     United States

               Source              Industry
0       Tesla, SpaceX             Automotive
1              Amazon             Technology
2                LVMH       Fashion & Retail
3           Microsoft             Technology
4  Berkshire Hathaway  Finance & Investments
```

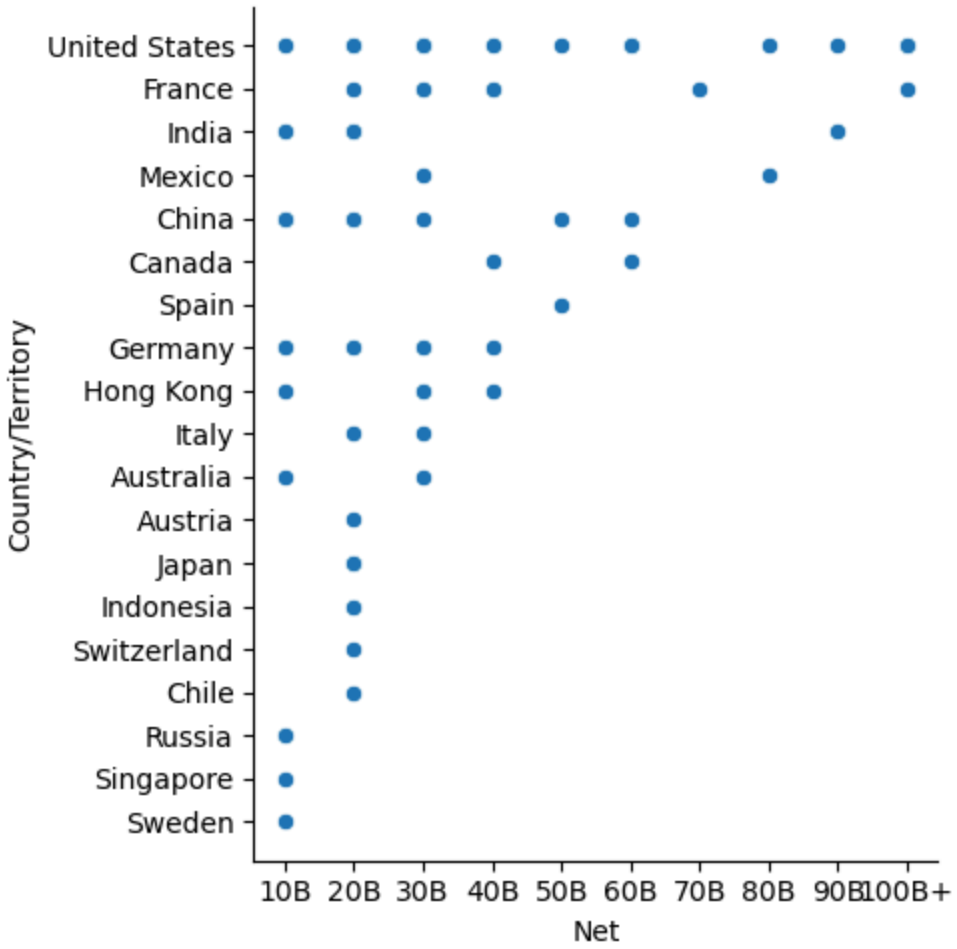
In [143…
```
# Count the number of billionaires per country/territory
country_distribution = dataframe['Country/Territory'].value_counts()
country_distribution
```

Out[143…

```
Country/Territory
United States    37
China            18
France            7
India             7
Germany           7
Hong Kong         4
Japan             3
Italy             2
Canada            2
Australia         2
Mexico            2
Russia            2
Indonesia         2
Spain             1
Austria           1
Switzerland       1
Chile             1
Singapore         1
Sweden            1
Name: count, dtype: int64
```

In [145…

```
# relationg graph of networth vs country
bins = [0, 10e9, 20e9, 30e9, 40e9, 50e9, 60e9, 70e9, 80e9, 90e9, 100e9, df['
labels = ['0B', '10B', '20B', '30B', '40B', '50B', '60B', '70B', '80B', '90B
df['Net'] = pd.cut(df['NetWorth'], bins=bins, labels=labels, right=False)
sns.relplot(x="Net", y="Country/Territory", data=df)
plt.figure(figsize=(300, 8))
plt.show()
```




```
<Figure size 30000x800 with 0 Axes>
```

In [152…

```python
# barh graph

df_grouped = df['Industry'].value_counts().reset_index()
df_grouped.columns = ['Industry', 'Count']

plt.figure(figsize=(10, 6))
bars = plt.bar(df_grouped['Industry'], df_grouped['Count'], color='skyblue')

plt.title('Count of Companies by Industry')
plt.xlabel('Industry')
plt.ylabel('Count')

for index, value in enumerate(df_grouped['Count']):
    plt.text(index, value, str(value), ha='center', va='bottom')

plt.xticks(rotation=45, ha='right')

plt.tight_layout()
plt.show()
```