


```
#Load the file
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
richest_df=pd.read_csv('/content/archive.zip')
print(richest_df)
```




	Unnamed: 0	rank	name	networth	age	\
0	0	1	Elon Musk	\$219 B	50	
1	1	2	Jeff Bezos	\$171 B	58	
2	2	3	Bernard Arnault & family	\$158 B	73	
3	3	4	Bill Gates	\$129 B	66	
4	4	5	Warren Buffett	\$118 B	91	
...	
2595	2595	2578	Jorge Gallardo Ballart	\$1 B	80	
2596	2596	2578	Nari Genomal	\$1 B	82	
2597	2597	2578	Ramesh Genomal	\$1 B	71	
2598	2598	2578	Sunder Genomal	\$1 B	68	
2599	2599	2578	Horst-Otto Gerberding	\$1 B	69	

	country	source	industry
0	United States	Tesla, SpaceX	Automotive
1	United States	Amazon	Technology
2	France	LVMH	Fashion & Retail
3	United States	Microsoft	Technology
4	United States	Berkshire Hathaway	Finance & Investments
...
2595	Spain	pharmaceuticals	Healthcare
2596	Philippines	apparel	Fashion & Retail
2597	Philippines	apparel	Fashion & Retail
2598	Philippines	garments	Fashion & Retail
2599	Germany	flavors and fragrances	Food & Beverage


[2600 rows x 8 columns]

```
#First five rows of data
richest_df.head()
```



	Unnamed: 0	rank	name	networth	age	country	source	industry
0	0	1	Elon Musk	\$219 B	50	United States	Tesla, SpaceX	Automotive
1	1	2	Jeff Bezos	\$171 B	58	United States	Amazon	Technology
2	2	3	Bernard Arnault & family	\$158 B	73	France	LVMH	Fashion & Retail
3	3	4	Bill Gates	\$129 B	66	United States	Microsoft	Technology
4	4	5	Warren Buffett	\$118 B	91	United States	Berkshire Hathaway	Finance & Investments

```
#Last five rows of data
richest_df.tail()
```



	Unnamed: 0	rank	name	networth	age	country	source	industry
2595	2595	2578	Jorge Gallardo Ballart	\$1 B	80	Spain	pharmaceuticals	Healthcare
2596	2596	2578	Nari Genomal	\$1 B	82	Philippines	apparel	Fashion & Retail
2597	2597	2578	Ramesh Genomal	\$1 B	71	Philippines	apparel	Fashion & Retail

2598	2598	2578	Sunder Genomal	\$1 B	68	Philippines	garments	Fashion & Retail
2599	2599	2578	Horst-Otto Gerberding	\$1 B	69	Germany	flavors and fragrances	Food & Beverage

```
#check for missing values
richest_df.isnull().sum()
```

	0
Unnamed: 0	0
rank	0
name	0
networth	0
age	0
country	0
source	0
industry	0

dtype: int64

```
#check for null values
richest_df.isnull().any().any()
```

	False
---	-------

```
#check for duplicate values
richest_df.duplicated()
```

	0
0	False
1	False
2	False
3	False
4	False
...	...
2595	False
2596	False
2597	False
2598	False
2599	False

2600 rows × 1 columns

dtype: bool

```
#info about the data
richest_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2600 entries, 0 to 2599
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype  
---  -
 0   Unnamed: 0   2600 non-null   int64  
 1   rank         2600 non-null   int64  
 2   name         2600 non-null   object  
 3   networth     2600 non-null   object  
 4   age          2600 non-null   int64  
 5   country      2600 non-null   object  
 6   source       2600 non-null   object  
 7   industry     2600 non-null   object  
dtypes: int64(3), object(5)
memory usage: 162.6+ KB
```

```
#description about the data
richest_df.describe()
```

```

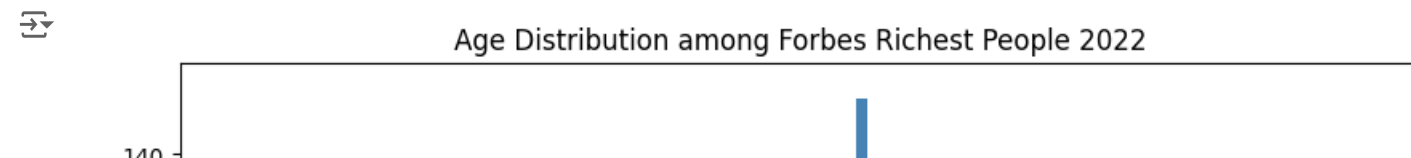
      Unnamed: 0      rank      age
count  2600.000000  2600.000000  2600.000000
mean    1299.500000   1269.570769    64.271923
std      750.699674    728.146364    13.220607
min         0.000000     1.000000    19.000000
25%     649.750000    637.000000    55.000000
50%    1299.500000   1292.000000    64.000000
75%    1949.250000   1929.000000    74.000000
max    2599.000000   2578.000000   100.000000
```

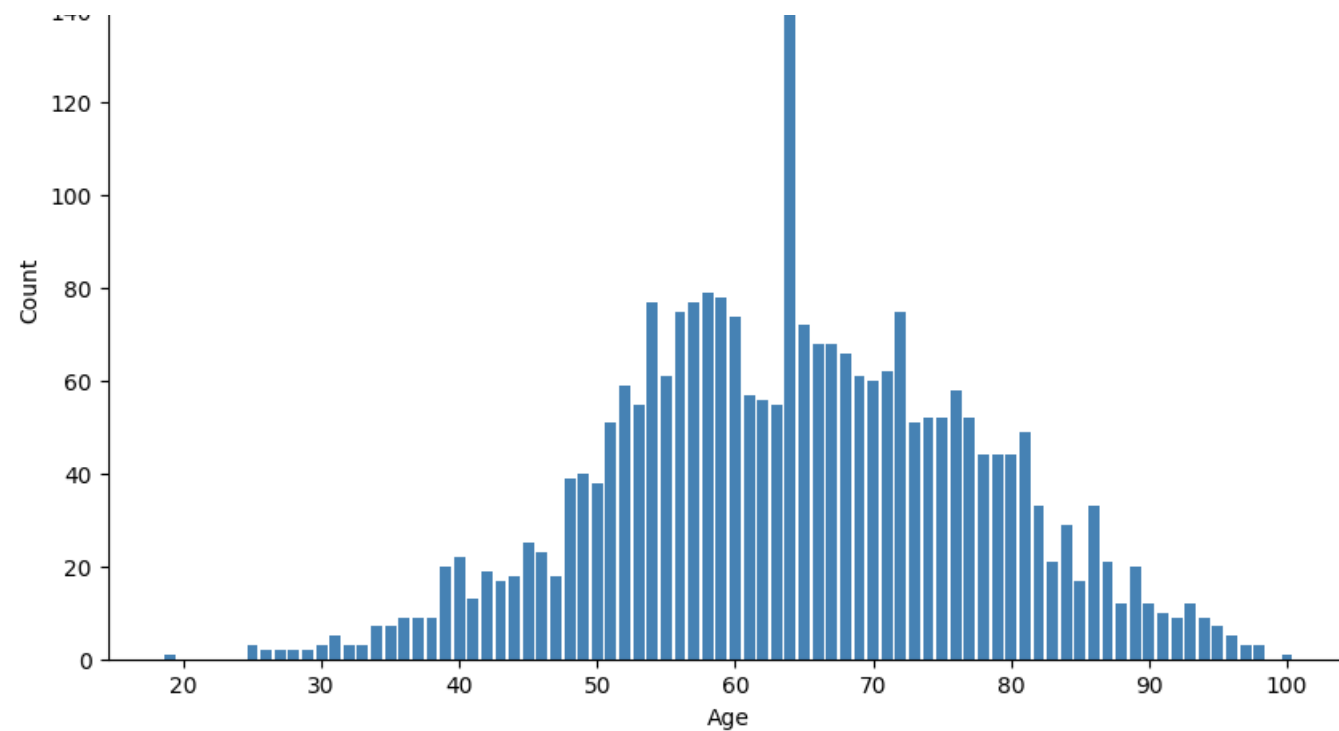
```
#shape of the data
richest_df.shape
```

```
(2600, 8)
```

```
#Age distribution among the data using bar plot
import matplotlib.pyplot as plt
import pandas as pd

age_counts = richest_df['age'].value_counts()
plt.figure(figsize=(10, 6))
plt.bar(age_counts.index, age_counts.values, color='steelblue')
plt.xlabel('Age')
plt.ylabel('Count')
plt.title('Age Distribution among Forbes Richest People 2022')
plt.show()
```





```
#Age distribution among the data using bar plot
import numpy as np
import matplotlib.pyplot as plt
plt.figure(figsize=(8,6),dpi=80)
sns.barplot(data=richest_df, x='age',y='industry',color='pink')
plt.title('Age distribution of the billionaires',fontsize=20)
plt.show()
```



```
-----
NameError                                Traceback (most recent call last)
<ipython-input-1-75ea4f1fac50> in <cell line: 5>()
      3 import matplotlib.pyplot as plt
      4 plt.figure(figsize=(8,6),dpi=80)
----> 5 sns.barplot(data=richest_df, x='age',y='industry',color='pink')
      6 plt.title('Age distribution of the billionaires',fontsize=20)
      7 plt.show()

NameError: name 'sns' is not defined

<Figure size 640x480 with 0 Axes>
```

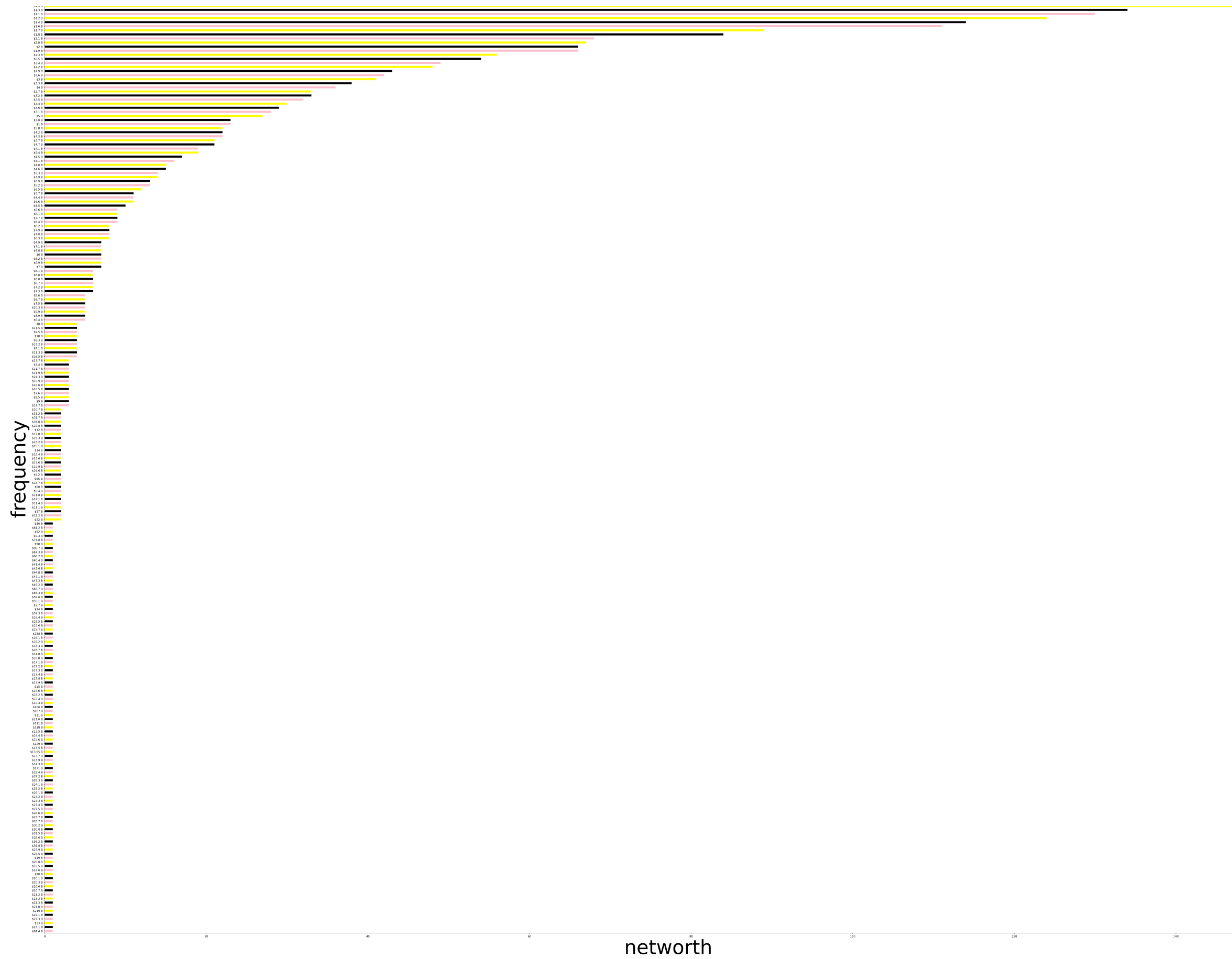
Double-click (or enter) to edit

```
#Networth Vs Frequency
from matplotlib.pyplot import figure
figure(num=None, figsize=(80,60), dpi=200, facecolor='w', edgecolor='k')
richest_df["networth"].value_counts(ascending=True).plot.barh(color=["pink","black","yellow"])
plt.title("Net Worth Vs Frequency",fontsize=70)
plt.xlabel("networth",fontsize=70)
plt.ylabel("frequency",fontsize=70)
```




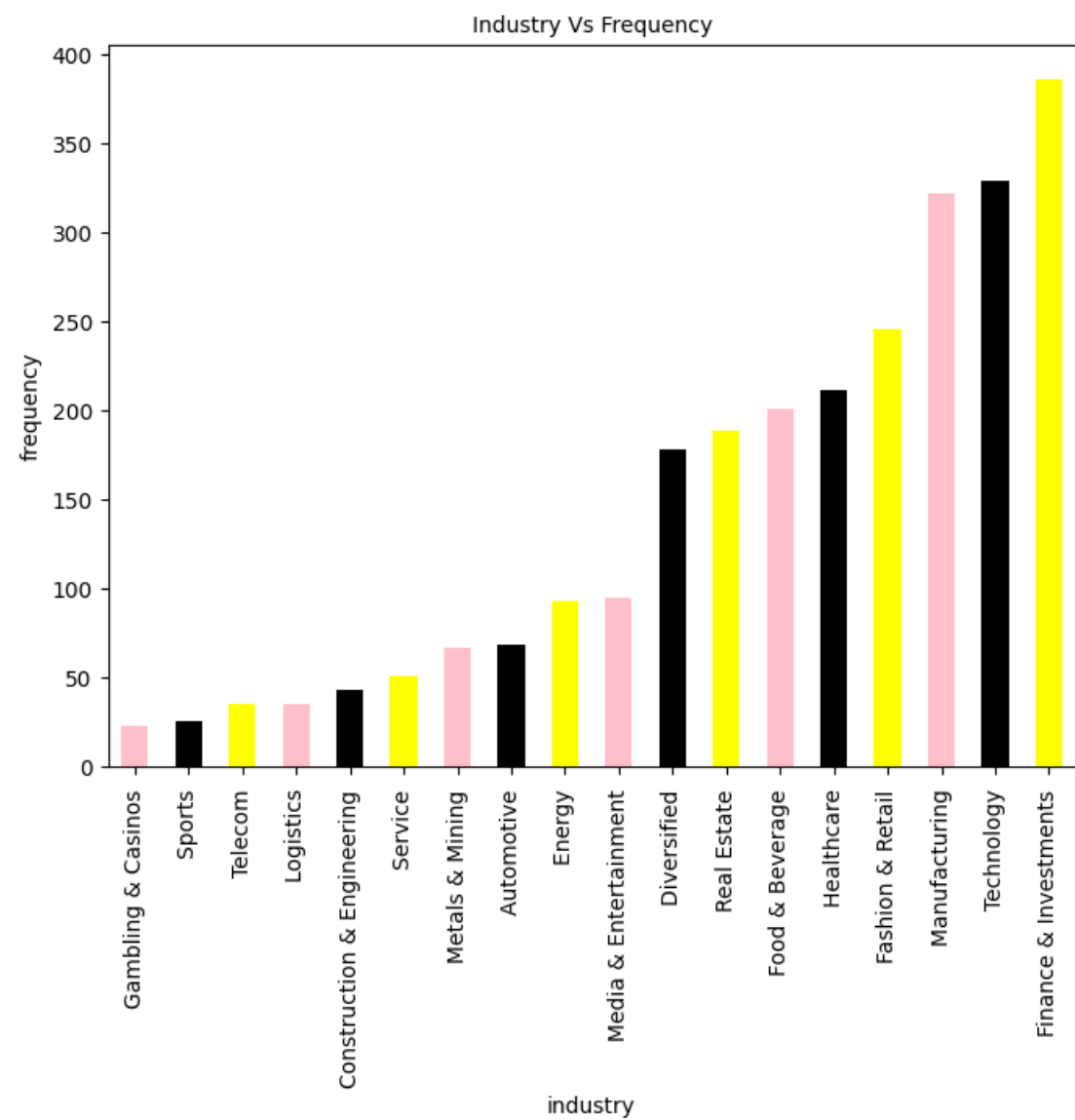
```
Text(0, 0.5, 'frequency')
```

Net Worth Vs Frequency




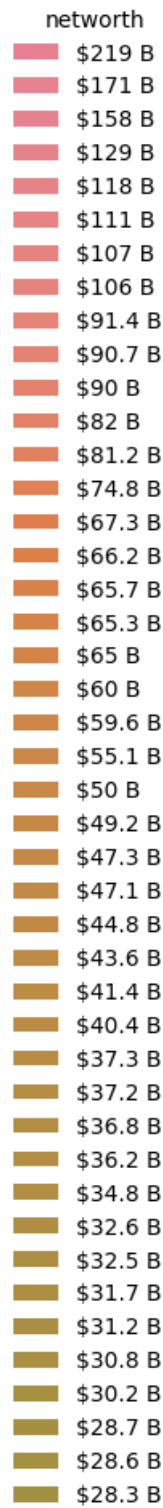
```
#Industry Vs frequency
from matplotlib.pyplot import figure
figure(num=None, figsize=(8,6), dpi=100, facecolor='w', edgecolor='k')
richest_df["industry"].value_counts(ascending=True).plot.bar(color=["pink","black","yellow"])
plt.title("Industry Vs Frequency",fontsize=10)
plt.xlabel("industry",fontsize=10)
plt.ylabel("frequency",fontsize=10)
```

 Text(0, 0.5, 'frequency')

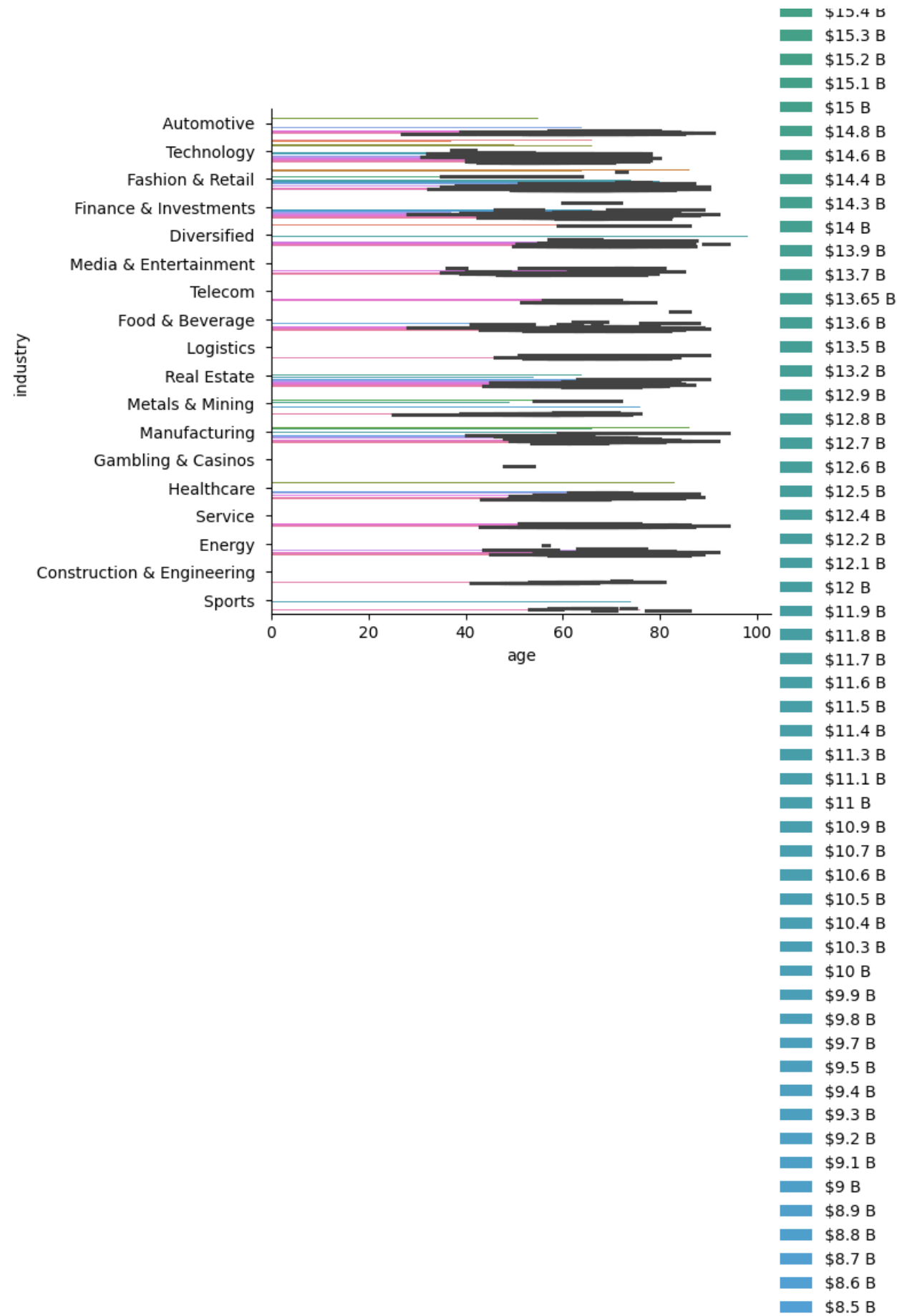


```
#Networth change with age and industry
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
sns.catplot(data=richest_df,x="age",y="industry",hue="networth",kind="bar")
```

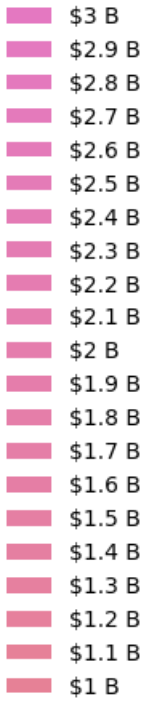
 /usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:123: UserWarning: Tight layout not applied. The bottom and top margins cannot be made large enough to accommodate all axes decorations.
self._figure.tight_layout(*args, **kwargs)
<seaborn.axisgrid.FacetGrid at 0x79c9ae530340>



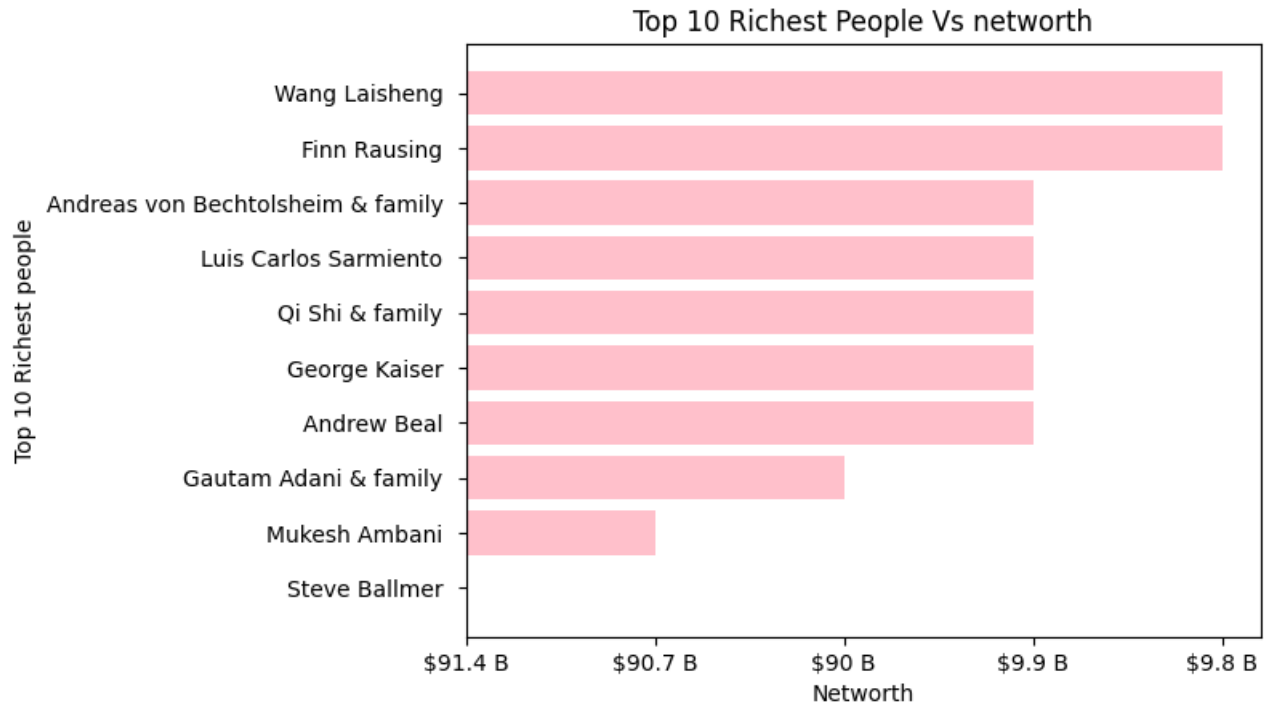
\$27.5 B
\$27.4 B
\$27.3 B
\$27.2 B
\$26.1 B
\$25.2 B
\$24.3 B
\$24.1 B
\$24 B
\$23.9 B
\$23.7 B
\$23.5 B
\$23.2 B
\$23.1 B
\$23 B
\$22.8 B
\$22.3 B
\$22.1 B
\$22 B
\$21.8 B
\$21.3 B
\$21.2 B
\$20.8 B
\$20.7 B
\$20.6 B
\$20.3 B
\$20.1 B
\$20 B
\$19.6 B
\$19.5 B
\$19 B
\$18.7 B
\$18.4 B
\$18.2 B
\$17.9 B
\$17.8 B
\$17.7 B
\$17.6 B
\$17.4 B
\$17.3 B
\$17.2 B
\$17.1 B
\$17 B
\$16.8 B
\$16.7 B
\$16.6 B
\$16.5 B
\$16.4 B
\$16.3 B
\$16.2 B
\$16.1 B
\$15.7 B
\$15.6 B
\$15.5 B
\$15.4 B



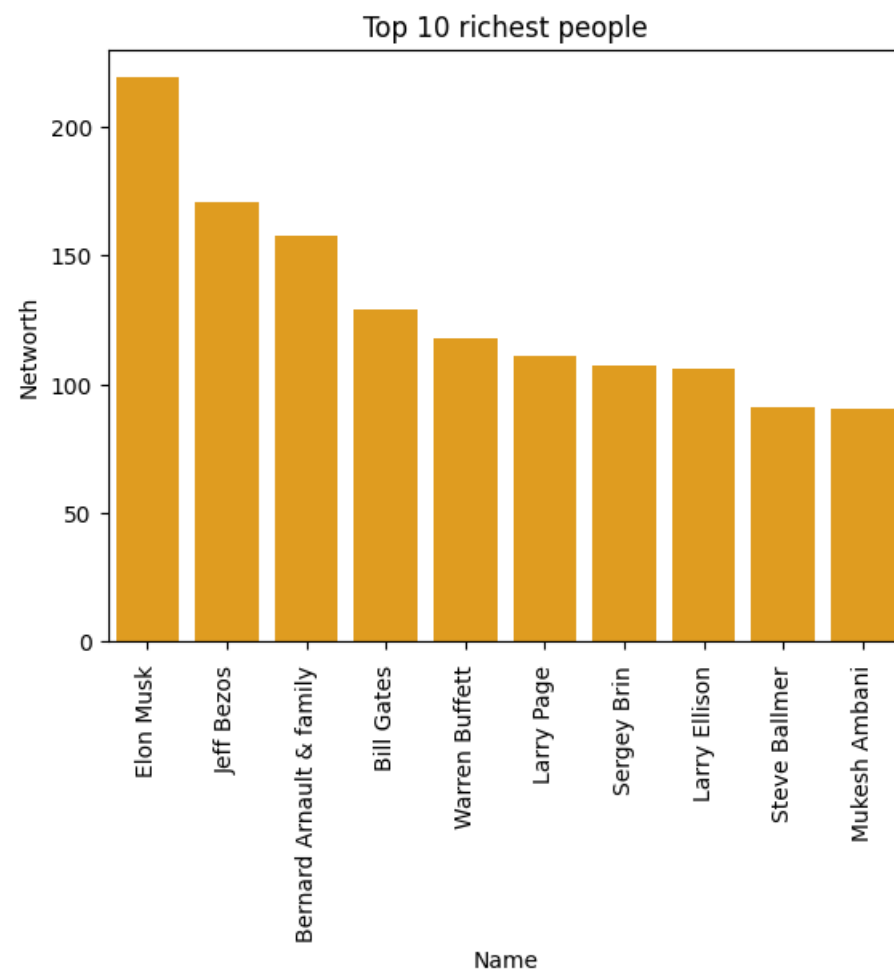




```
#Top 10 Richest people Vs networkh
import pandas as pd
import matplotlib.pyplot as plt
data = richest_df.sort_values(by='networth', ascending=False)
top_10 = data.head(10)
plt.barh(top_10['name'], top_10['networth'],color="pink")
plt.xlabel('Networth')
plt.ylabel('Top 10 Richest people')
plt.title('Top 10 Richest People Vs networkh')
plt.show()
```



```
#Top 10 richest people
rich=richest_df.copy()
rich['networth']=rich['networth'].str.replace('$','').str.replace('B','').astype(float)
rich=rich.nlargest(10,'networth')
sns.barplot(x='name',y='networth',data=rich,color='orange')
plt.title('Top 10 richest people')
plt.xlabel('Name')
plt.xticks(rotation=90)
plt.ylabel('Networth')
plt.show()
```



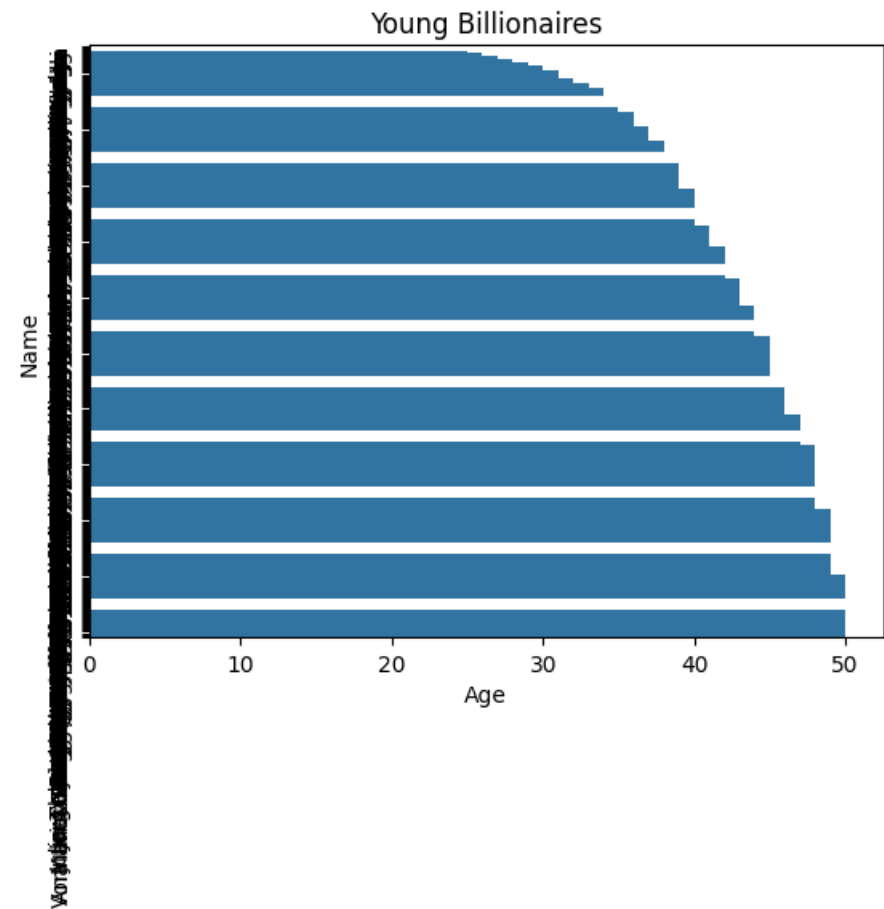
```
#Minimum age billionaires <=50
young_billionaires=richest_df[richest_df['age']<=50]
young_billionaires=young_billionaires[['name','age','industry']].sort_values(by='age')
print(young_billionaires)
sns.barplot(x='age',y='name',data=young_billionaires)
plt.title('Young Billionaires')
plt.xlabel('Age')
plt.ylabel('Name')
plt.yticks(rotation=90)
plt.show()
```



	name	age	industry
1311	Kevin David Lehmann	19	Fashion & Retail

2190	Alexandra Andresen	25	Diversified
1975	Pedro Franceschi	25	Finance & Investments
2062	Wang Zelong	25	Metals & Mining
2191	Katharina Andresen	26	Diversified
...
2395	Park Kwan-ho	50	Media & Entertainment
1572	Evan Williams	50	Technology
575	Daniel Ziff	50	Finance & Investments
1485	David Mindus	50	Real Estate
0	Elon Musk	50	Automotive

[359 rows x 3 columns]



```
#Industry Vs Billionaires
import pandas as pd
import matplotlib.pyplot as plt
industry_counts = richest_df['industry'].value_counts()
plt.figure(figsize=(10, 6))
industry_counts.head(10).plot(kind='bar')
plt.title('Top 10 Industries with the more Billionaires')
plt.xlabel('Industry')
plt.ylabel('Number of Billionaires')
plt.show()
```

