



project 3

File Edit View Insert Runtime Tools Help Last saved at 5:13 PM

Comment

Share



+ Code + Text

RAM Disk

Gemini

↑ ↓ ↻ ⌨ ⚙ 📄 🗑 ⋮

```
# 1. Loading the file
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
Forbes_df=pd.read_csv("/content/2022_forbes_billionaires.csv")
print(Forbes_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	food & beverages	Food & Beverages

Connected to Python 3 Google Compute Engine backend



Type here to search



05:20 PM
24-09-2024

project 3 - Colab

colab.research.google.com/drive/1ZArVXD3E34oMFcDgSzz7MtxN8jyhXkpp

+ Code + Text

RAM Disk Gemini

2s

1. Loading the file

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

Forbes_df=pd.read_csv("/content/2022_forbes_billionaires.csv")

print(Forbes_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]

Connected to Python 3 Google Compute Engine backend

05:20 PM 24-09-2024

project 3 - Colab

colab.research.google.com/drive/1ZArVXD3E34oMFcDgSzz7MtxN8jyhXkpp

+ Code + Text

2599 Germany flavors and fragrances Food & Beverage

[2600 rows x 8 columns]

[] # 2. Printing the first 5 rows of data:
Forbes_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

[] # 3. Printing last 5 rows of data
Forbes_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

[] # 4. Checking for missing values
Forbes_df.isnull().sum()

Connected to Python 3 Google Compute Engine backend

05:20 PM 24-09-2024

Type here to search

```
# 4. Checking for missing values
Forbes_df.isnull().sum()
```

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

dtype: int64

+ Code + Text

```
[ ] # 4. Check for null values
Forbes_df.isna()
```

	Unnamed: 0	rank	name	networth	age	country	source	industry
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False

Connected to Python 3 Google Compute Engine backend

```
[ ] source 0
industry 0
dtype: int64
```

```
# 4. Check for null values
Forbes_df.isna()
```

Unnamed: 0	rank	name	networkh	age	country	source	industry
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
...
2595	False	False	False	False	False	False	False
2596	False	False	False	False	False	False	False
2597	False	False	False	False	False	False	False
2598	False	False	False	False	False	False	False
2599	False	False	False	False	False	False	False

2600 rows x 8 columns

```
[ ] # 4.checking for duplicated values
Forbes_df.isnull().any().any()
```

project 3 - Colab

colab.research.google.com/drive/1ZArVXD3E34oMFcDgSzz7MtxN8jyhXkpp

+ Code + Text

RAM Disk Gemini

[] # 4.checking for duplicated values
Forbes_df.isnull().any().any()

False

4.checking for duplicated values
Forbes_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

[] # 5. Getting some info about the data

Connected to Python 3 Google Compute Engine backend

Type here to search

05:21 PM
24-09-2024


```
[ ] # 5. Getting some info about the data
Forbes_df.info()
```

```
# 6. Get some description about the data
Forbes_df.describe()
```

✓ Connected to Python 3 Google Compute Engine backend

project 3 - Colab

colab.research.google.com/drive/1ZArVXD3E34oMFcDgSzz7MtxN8jyhXkpp

+ Code + Text

RAM Disk

Gemini

[] 6 source 2600 non-null object
7 industry 2600 non-null object
dtypes: int64(3), object(5)
memory usage: 162.6+ KB

[] # 6. Get some description about the data
Forbes_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

7. Get the shape of the data
shape=Forbes_df.shape
print(shape)

(2600, 8)

[] pd.DataFrame(Forbes_df['rank'].value_counts())

count

project 3 - Colab

colab.research.google.com/drive/1ZArVXD3E34oMFcDgSzz7MtxN8jyhXkpp

+ Code + Text

RAM Disk

Gemini

```
[ ] shape=Forbes_df.shape
print(shape)

(2600, 8)

[ ] pd.DataFrame(Forbes_df['rank'].value_counts())
```

rank	count
1929	147
2190	134
2448	130
2324	124
2076	114
...	...
89	1
90	1
96	1
97	1
1	1

228 rows x 1 columns

```
[ ] pd.DataFrame(Forbes_df['age'].value_counts())
```

age	count
-----	-------

Connected to Python 3 Google Compute Engine backend

Type here to search

05:21 PM 24-09-2024

```
pd.DataFrame(Forbes_df['age'].value_counts())
```

count	
age	
64	152
58	79
59	78
54	77
57	77
...	...
27	2
26	2
29	2
19	1
100	1

76 rows x 1 columns

```
[ ] pd.DataFrame(Forbes_df['country'].value_counts())
```

count	
country	
United States	719
China	515

project 3 - Colab

colab.research.google.com/drive/1ZArVXD3E34oMFcDgSzz7MtxN8jyhXkpp

+ Code + Text

RAM Disk

Gemini

[] 100 1

76 rows × 1 columns

[] pd.DataFrame(Forbes_df['country'].value_counts())

count

country

United States	719
China	515
India	161
Germany	130
Russia	81
...	...
Venezuela	1
Portugal	1
Algeria	1
Eswatini (Swaziland)	1
Estonia	1

75 rows × 1 columns

[] #VISUALIZATION

#1.Age distribution among the data using bar plot

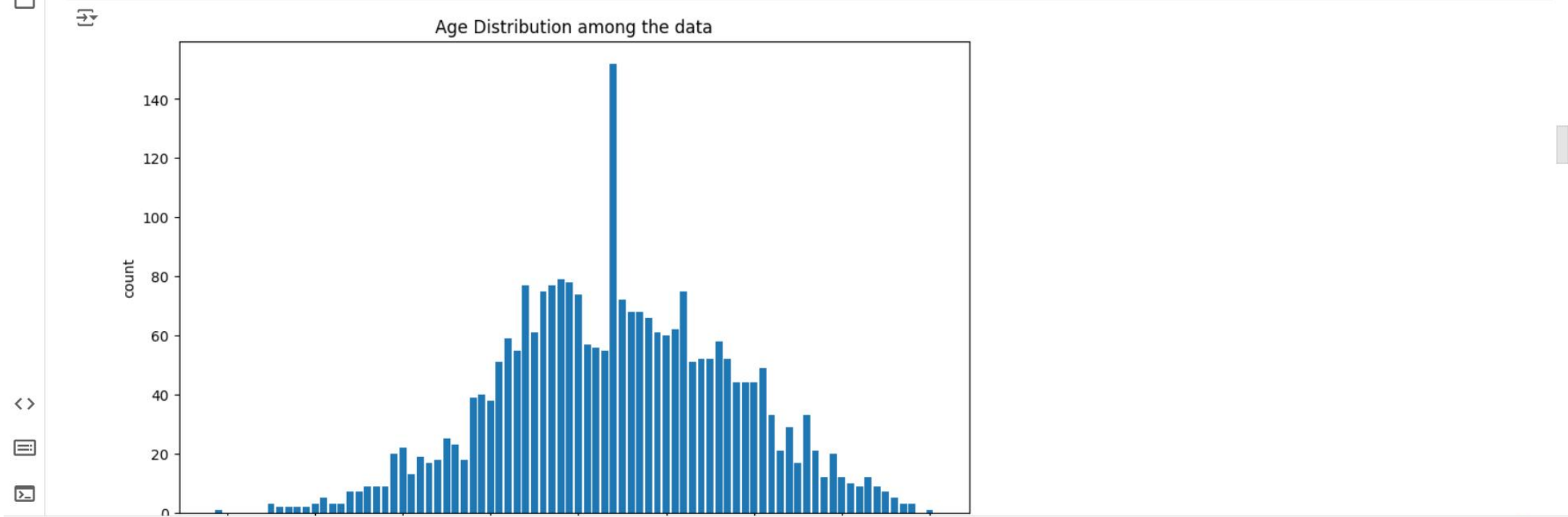
age=Forbes_df['age'].value_counts(ascending=True)

plt.figure(figsize=(10,6))

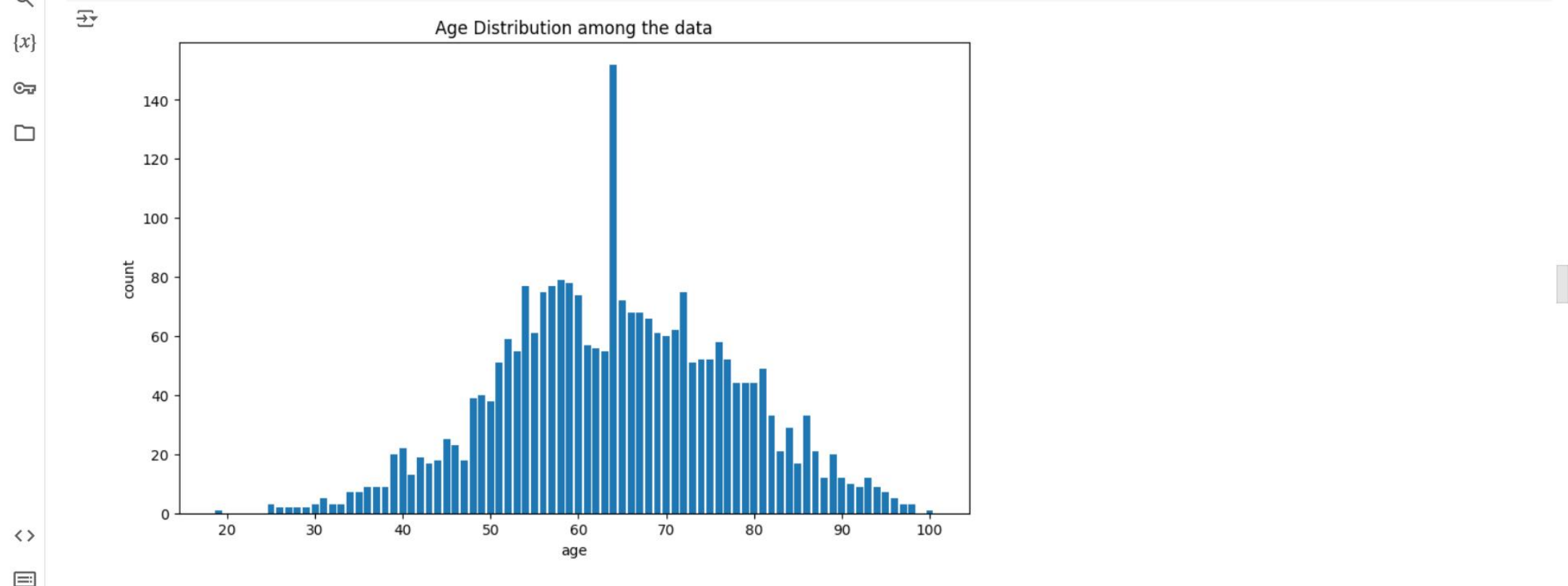
Connected to Python 3 Google Compute Engine backend

05:21 PM 24-09-2024

```
#1.Age distribution among the data using bar plot
age=Forbes_df['age'].value_counts(ascending=True)
plt.figure(figsize=(10,6))
plt.bar(age.index,age)
plt.xlabel("age")
plt.ylabel("count")
plt.title("Age Distribution among the data")
plt.show()
```



```
plt.title('Age Distribution among the data')  
plt.show()
```



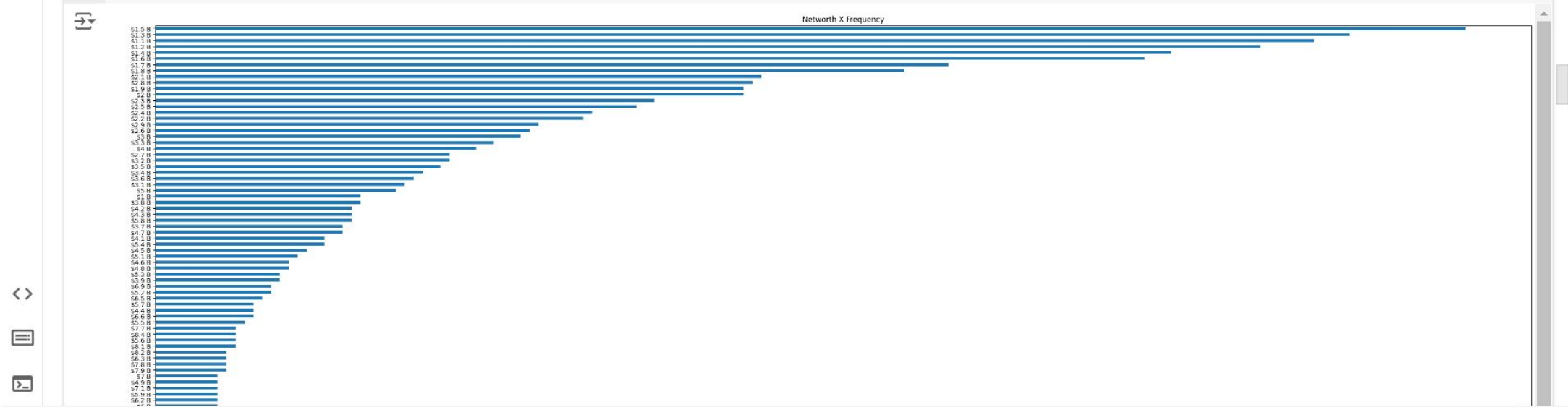
```
[ ] # 2. Net Worth Vs Frequency using bar plot  
from matplotlib.pyplot import figure
```

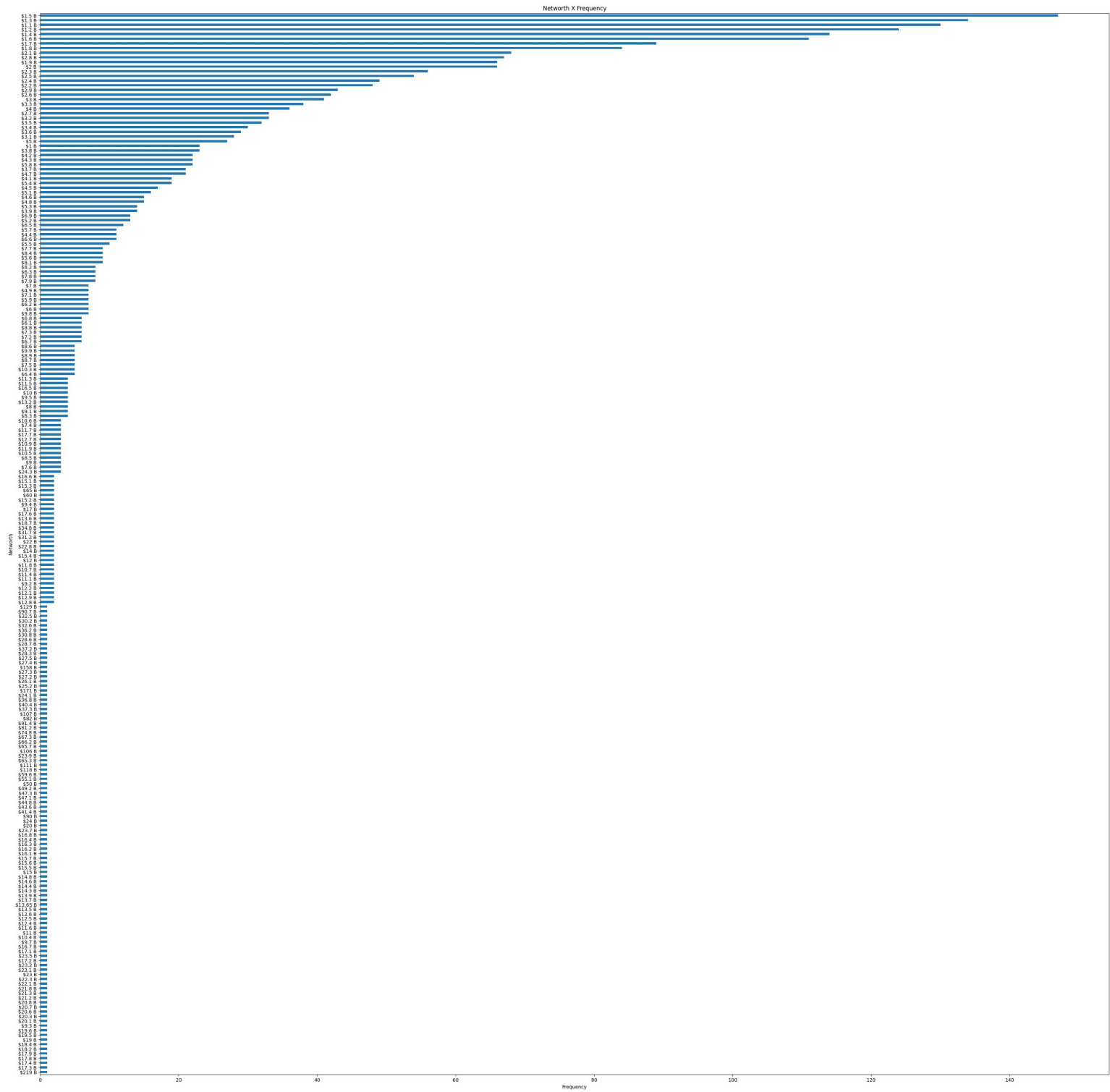
Connected to Python 3 Google Compute Engine backend



```
# 2. Net Worth Vs Frequency using bar plot

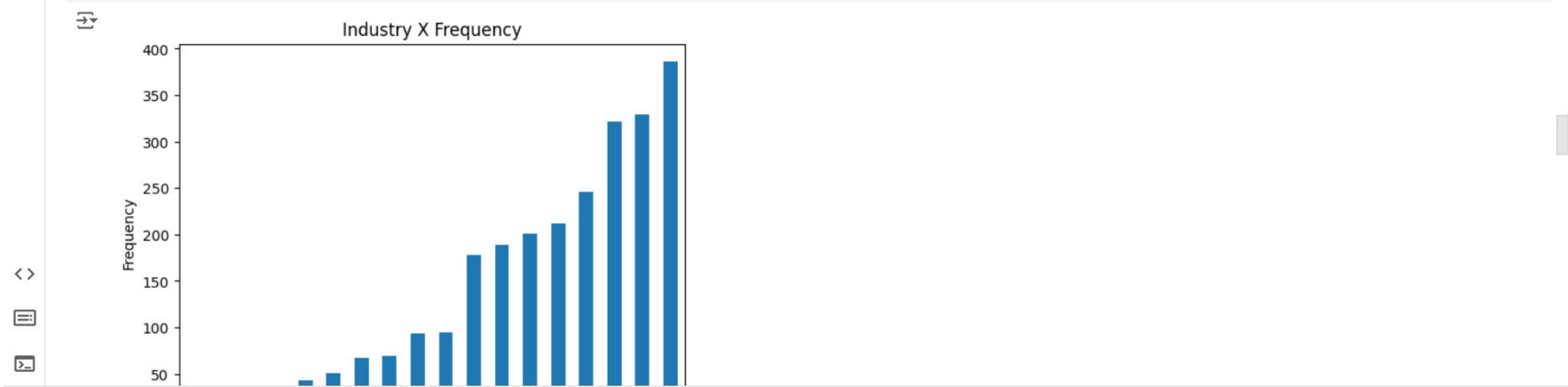
from matplotlib.pyplot import figure
figure(num=None, figsize=(40, 40))
Forbes_df["networth"].value_counts(ascending=True).plot.barh()
plt.title("Networth X Frequency")
plt.ylabel("Networth")
plt.xlabel("Frequency")
plt.show()
```

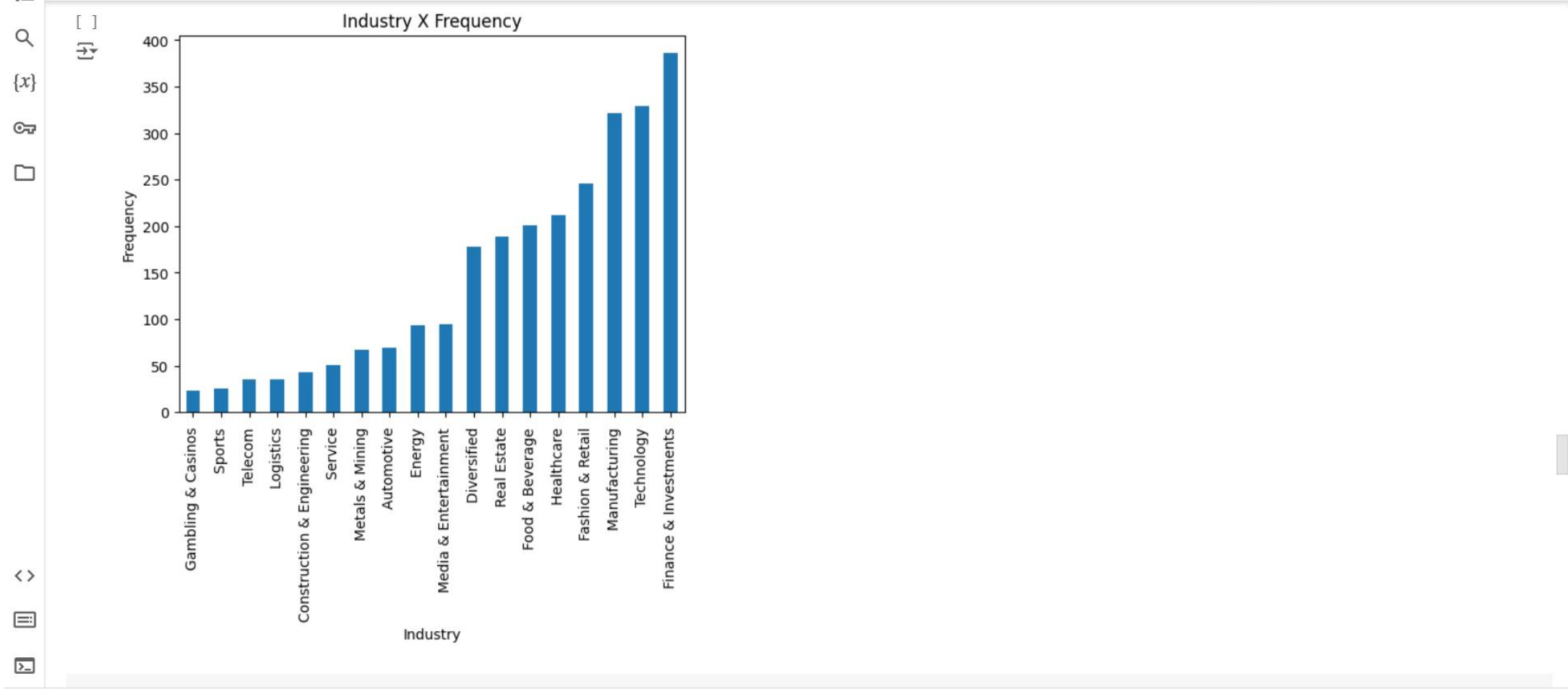






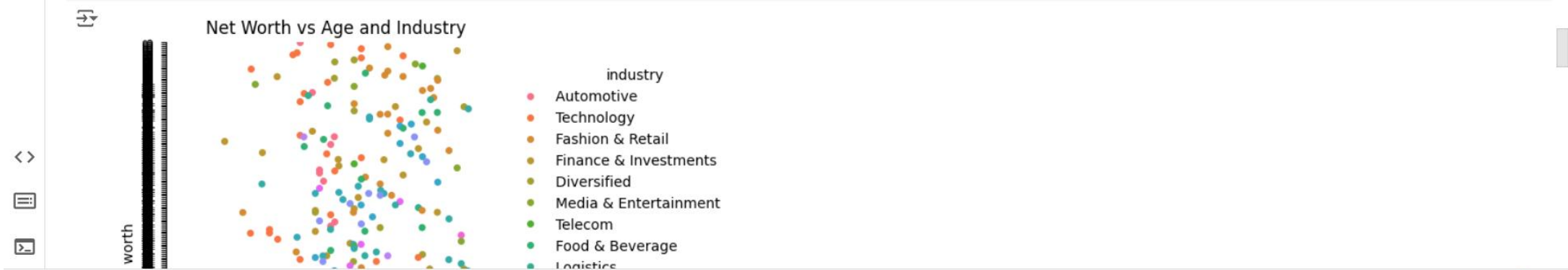
```
[ ] # 3. Showing Industry Vs Frequency using bar plot
from matplotlib.pyplot import figure
figure(num=None)
Forbes_df["industry"].value_counts(ascending=True).plot.bar()
plt.title("Industry X Frequency")
plt.xlabel("Industry")
plt.ylabel("Frequency")
plt.show()
```







```
[ ] # 4.Showing how does Net worth Change with age and industry using cat plot
sns.catplot(data=Forbes_df, x="age", y="networth",hue="industry")
plt.title("Net Worth vs Age and Industry")
plt.xlabel("Age")
plt.ylabel("Net worth")
plt.yticks(rotation=90)
plt.show()
```



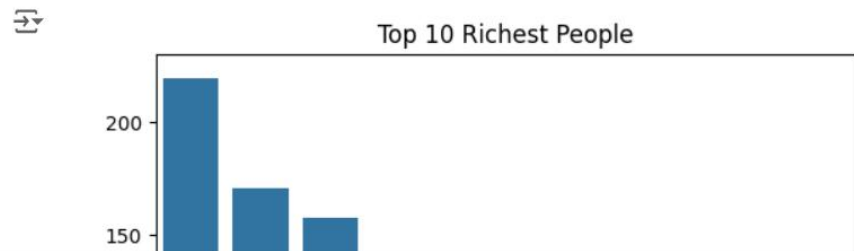
```
[ ] plt.title("Net Worth vs Age and Industry")
plt.xlabel("Age")
plt.ylabel("Net worth")
plt.yticks(rotation=90)
plt.show()
```



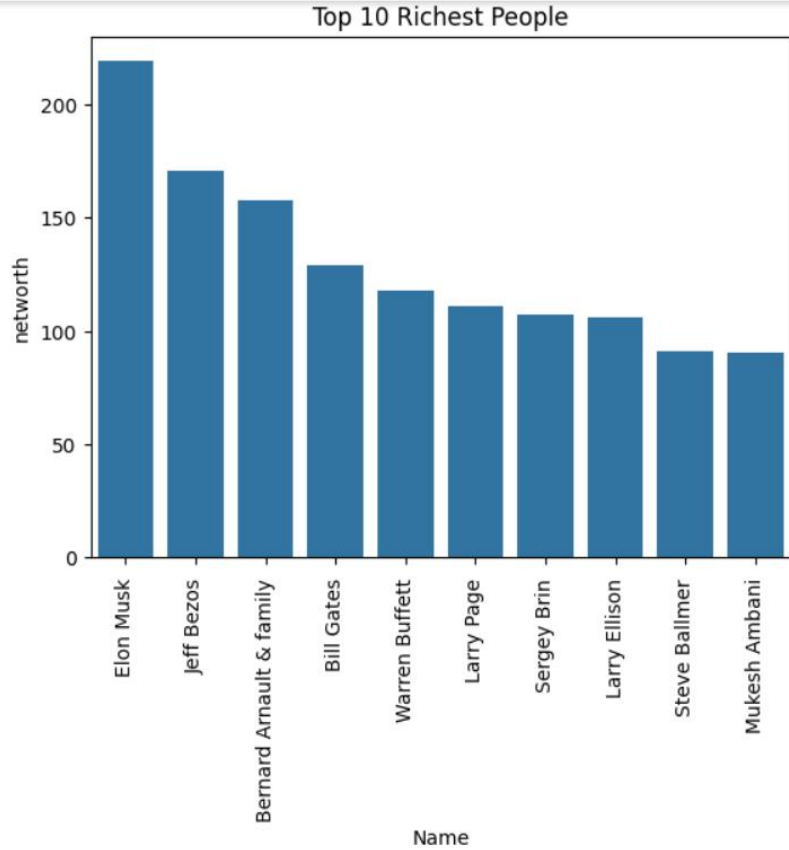
```
[ ] # 5. Showing the top 10 richest people vs net worth
```



```
[ ] # 5. Showing the top 10 richest people vs net worth
rich=Forbes_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)
plt.xticks(rotation=90)
plt.title('Top 10 Richest People')
plt.xlabel("Name")
plt.ylabel("networth")
plt.show()
```



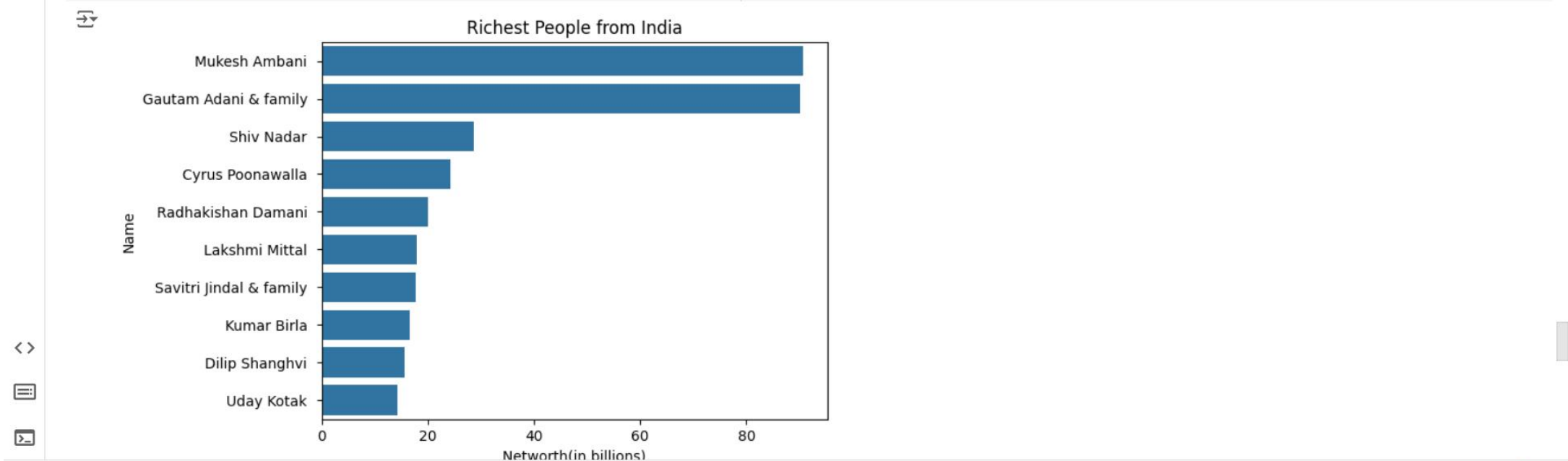
Connected to Python 3 Google Compute Engine backend



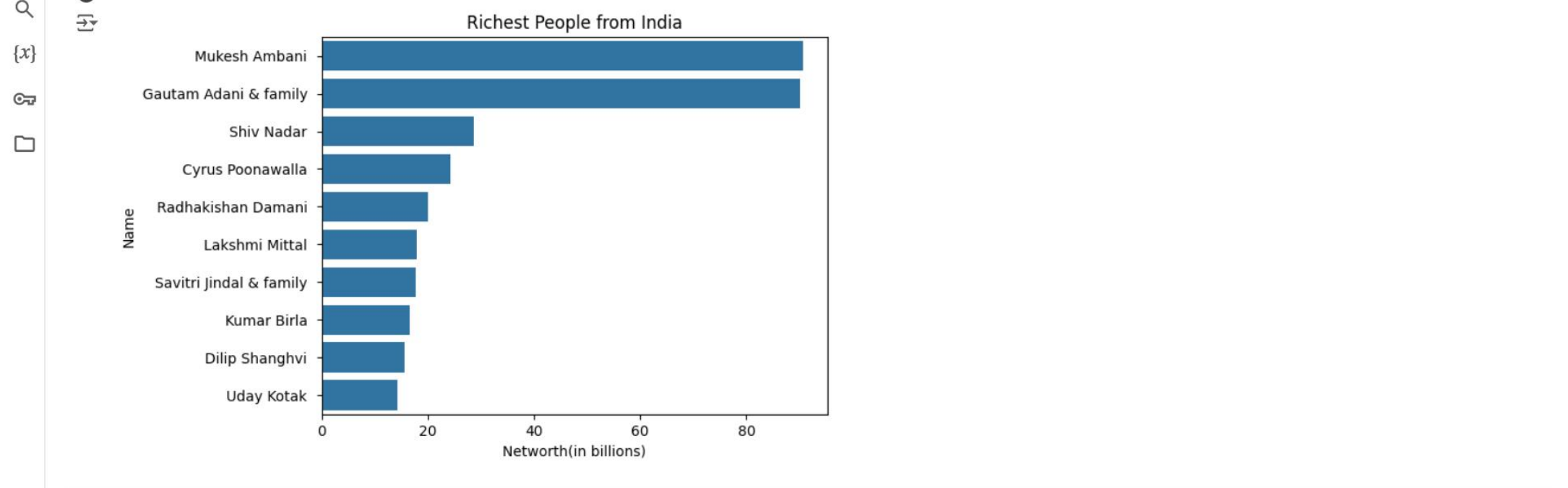
```
[ ] # 6.Showing the richest people from India
richest=Forbes_df[Forbes_df['country']=='India'].conv()
```

Connected to Python 3 Google Compute Engine backend

```
# 6.Showing the richest people from India
richest=Forbes_df[Forbes_df['country']=='India'].copy()
richest['networth']=richest['networth'].str.replace('\n','').str.replace('B','').str.replace('$','').astype(float)
richest=richest.nlargest(10,'networth')
sns.barplot(x='networth',y='name',data=richest)
plt.title("Richest People from India")
plt.xlabel("Networth(in billions)")
plt.ylabel("Name")
plt.show()
```



plt.show()



```
[ ] # 7.Showing the minimum age billionaire <=50 with name and industry.
young=Forbes_df[Forbes_df['age']<=50]
young=young[['name','age','industry']].sort_values(by='age')
print(young)
```

	name	age	industry
1311	Kevin David Lehmann	19	Fashion & Retail
2190	Alexandra Andresen	25	Diversified
1075	Padra Enneschi	25	Finance & Investments



```
[ ]  
# 7.Showing the minimum age billionaire <=50 with name and industry.  
young=Forbes_df[Forbes_df['age']<=50]  
young=young[['name','age','industry']].sort_values(by='age')  
print(young)
```

	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]

```
[ ] # 8.Showing in which industry billionaires are more  
industry_cnts=Forbes_df['industry'].value_counts()  
sns.barplot(x=industry_cnts.index,y=industry_cnts.values)  
plt.title("Number of Billionaires by Industry")  
plt.xlabel("Industry")  
plt.ylabel("Number of Billionaires")  
plt.xticks(rotation=90)  
plt.show()
```

