

✓ Forbest Richest People Analysis

1.Load the file.

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
dataset=pd.read_csv('/content/2022_forbes_billionaires.csv')
```

2.Print first 5 rows of data

```
dataset.head(5)
```



	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.Print last 5 row of data

```
dataset.tail(5)
```



	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

3.Check for missing values, null values and duplicate data.

```
dataset.isnull().sum()
```



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

dtype: int64

```
dataset.duplicated().sum()
```



0

```
dataset.drop_duplicates()
```



	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
...
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

2600 rows × 8 columns

5. Get some info about the data.

```
dataset.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

```

6. Get some description about the data.


```
dataset.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.

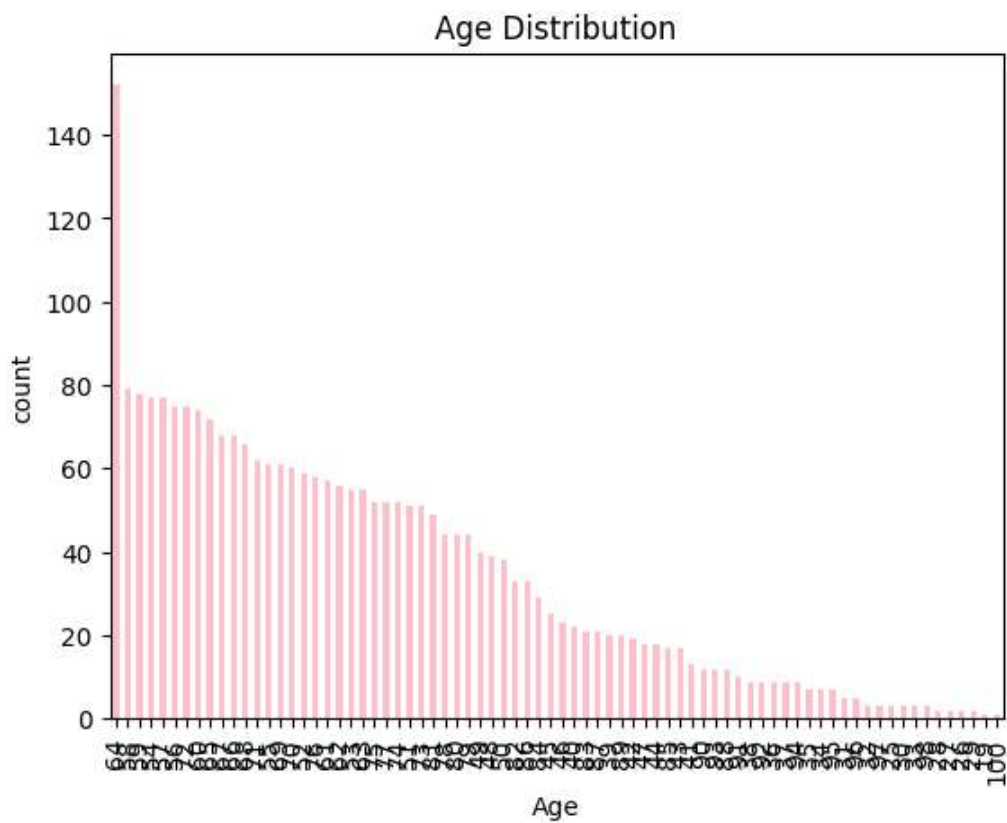
```
dataset.shape
```

 (2600, 8)

Visualization

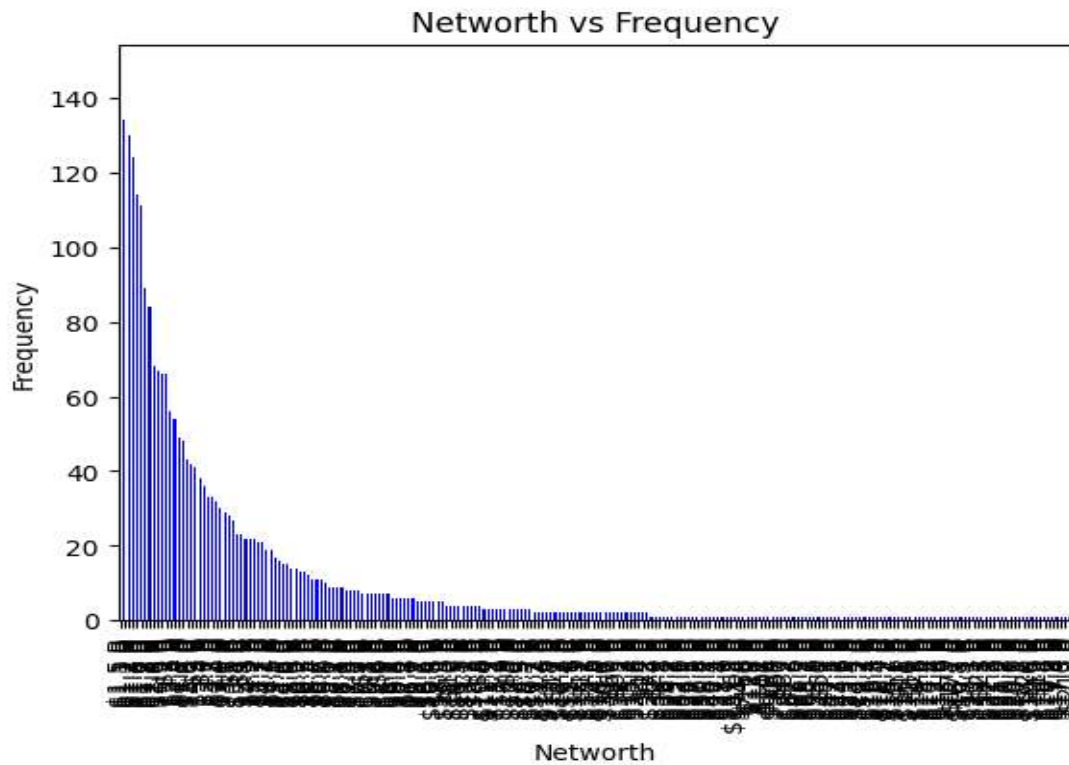
1.Age Distribution among the data using bar plot.

```
age=dataset["age"].value_counts()
plt.figure()
age.plot(kind='bar',color='pink')
plt.title('Age Distribution')
plt.xlabel("Age")
plt.ylabel("count")
plt.show()
```



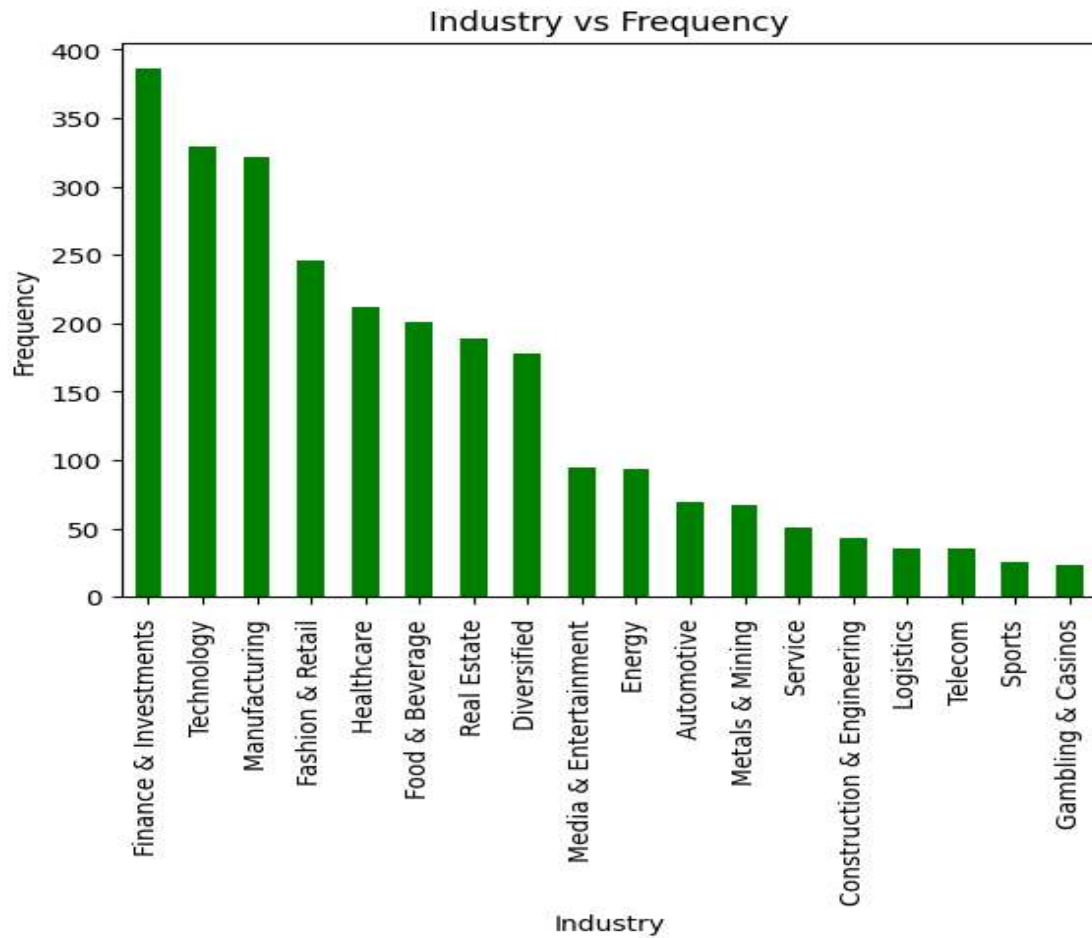
2.Net Worth vs Frequency.

```
net_worth=dataset['networth'].value_counts()
plt.figure()
net_worth.plot(kind='bar',color="blue")
plt.title('Networth vs Frequency')
plt.xlabel('Networth')
plt.ylabel('Frequency')
plt.show()
```



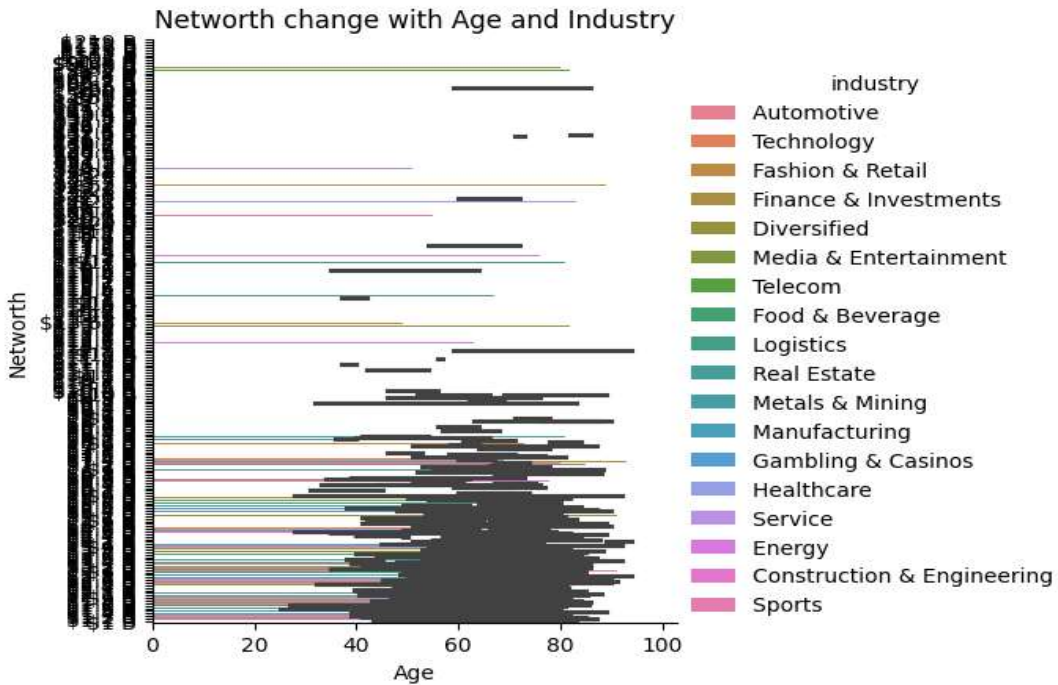
3. Industry vs Frequency.

```
industry=dataset['industry'].value counts()
plt.figure()
industry.plot(kind='bar',color="green")
plt.title('Industry vs Frequency')
plt.xlabel('Industry')
plt.ylabel('Frequency')
plt.show()
```



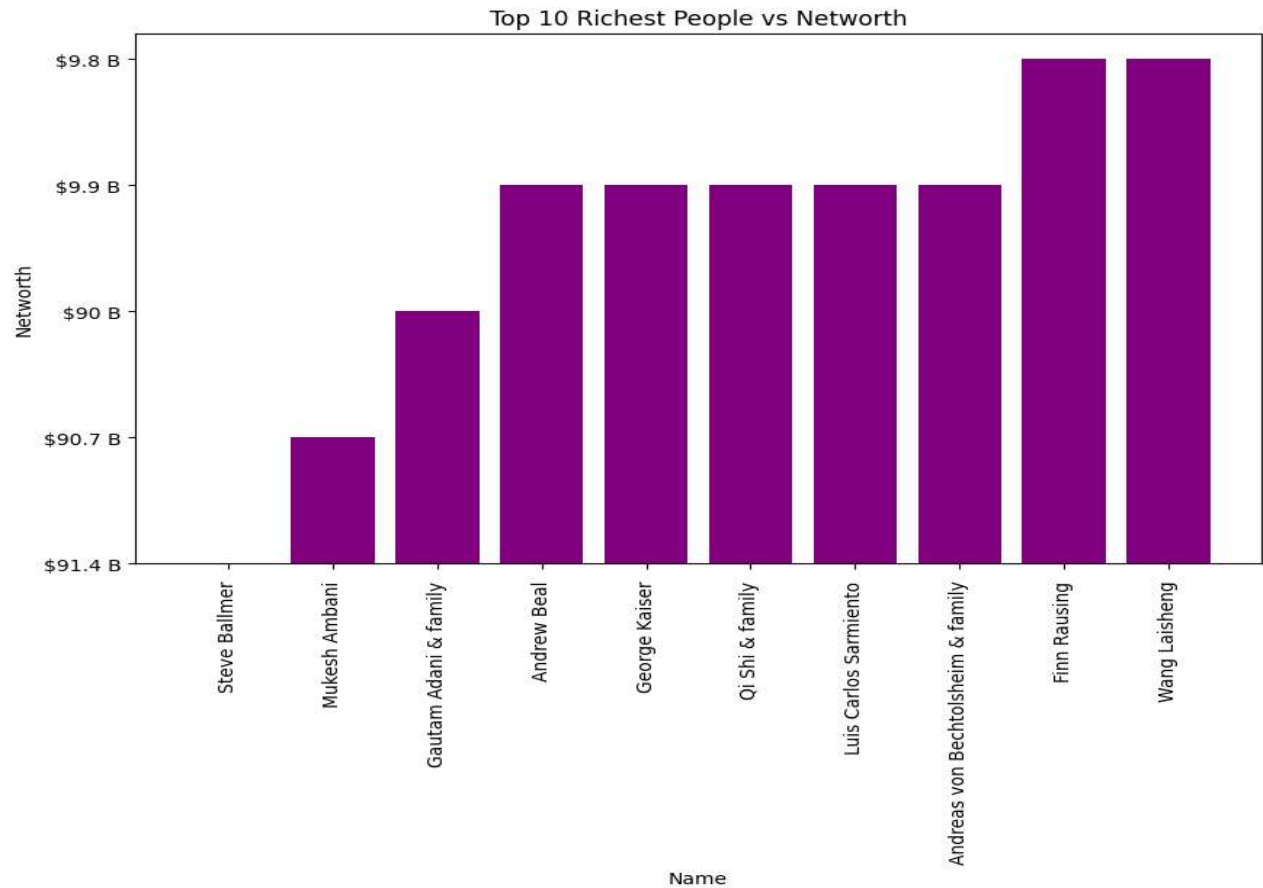
4.How does Net Worth Change with age and industry using cat plot.

```
plt.figure(figsize=(10,5))
sns.catplot(x='age',y='networth',hue='industry',data=dataset,kind='bar')
plt.xlabel('Age')
plt.ylabel('Networth')
plt.title('Networth change with Age and Industry')
plt.show()
```



5.Top 10 richest people vs network.

```
richest_people=dataset.sort_values(by='networth',ascending=False).head(10)
plt.figure(figsize=(10,6))
plt.bar(richest_people['name'],richest_people['networth'],color='purple')
plt.xticks(rotation=90)
plt.xlabel('Name')
plt.ylabel('Networth')
plt.title('Top 10 Richest People vs Network')
plt.show()
```

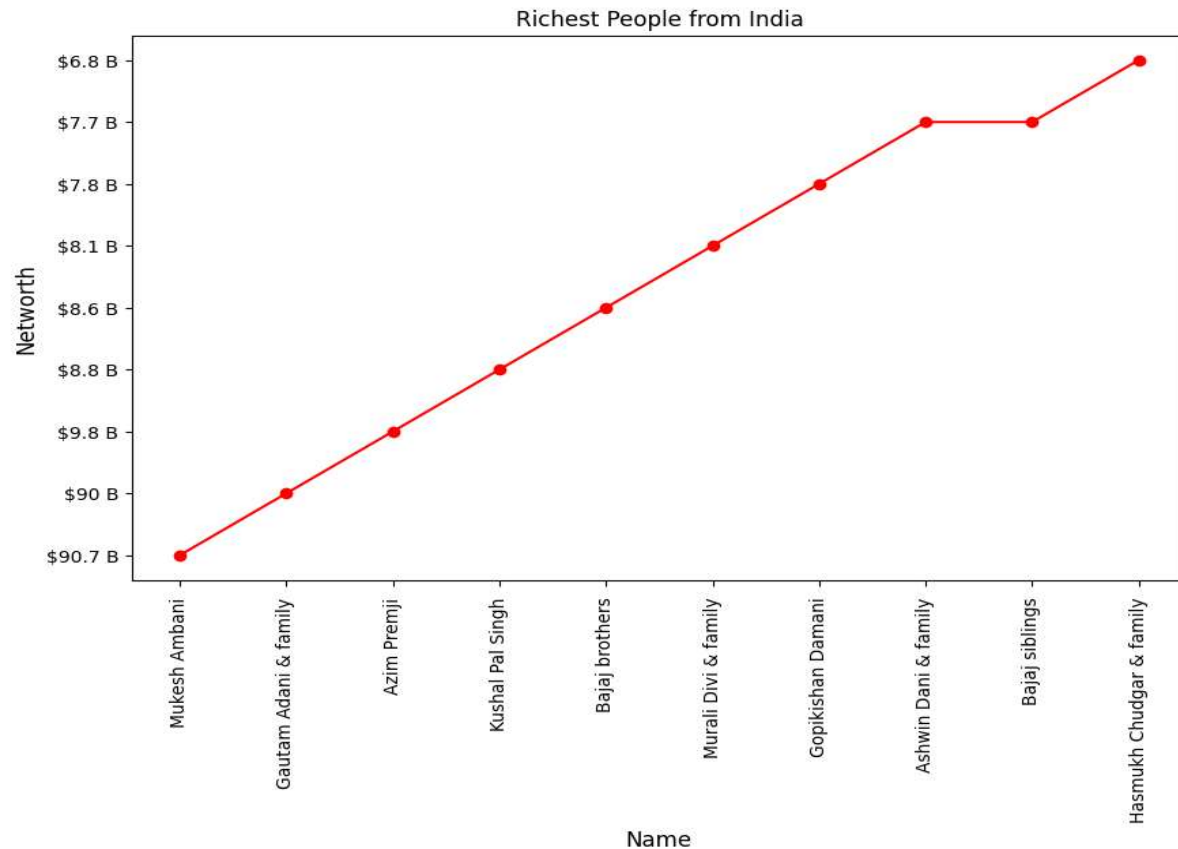



6. Richest People from India with names.

```

billionaires_india=dataset[dataset['country']=='India']
richest_in_india=billionaires_india.sort_values(by='networth',ascending=False)
richest_indian=richest_in_india.head(10)
richest_indian['Net worth']=richest_indian['networth'].apply(lambda x:f'${x}')
plt.figure(figsize=(10,6))
plt.plot(richest_indian['name'],richest_indian['networth'],marker='o',linestyle='-',color='b')
plt.xticks(rotation=90)
plt.title('Richest People from India')
plt.xlabel('Name',fontsize=12)
plt.ylabel('Networth',fontsize=12)
#plt.grid(True)
plt.show()

```



7. Minimum Age Billionaire <=50 with Name and Industry.

```
young_billionaires=dataset[dataset['age']<=50]
min_age=young_billionaires['age'].min()
min_age_young_billionaires=young_billionaires.loc[young_billionaires['age']==min_age]
print('Name=',min_age_young_billionaires['name'].values[0])
print('Industry=',min_age_young_billionaires['industry'].values[0])
print("Age=",min_age)
```

Name= Kevin David Lehmann
Industry= Fashion & Retail
Age= 19

8. Show in which industry billionaires are more.

```
industry=dataset['industry'].value_counts()
plt.figure(figsize=(10,6))
plt.pie(industry.values,labels=industry.index,autopct='%1.1f%%')
plt.title('Industry Distribution')
plt.show()
```

Industry Distribution

