DATA ANALYSE

1. Load the file

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

2. Print first 5 rows of data

```
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

Next steps: Generate code with df | View recommended plots | New interactive sheet

3. Print last 5 rows of data

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

```
df.isnull().sum()
```

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

dtype: int64

5. Get some info about the data

```
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
```

6. Get some description about the data

```
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

```
df.shape
```

```
(2600, 8)
```

```
age = df['age'].value_counts()
age
```

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

76 rows × 1 columns

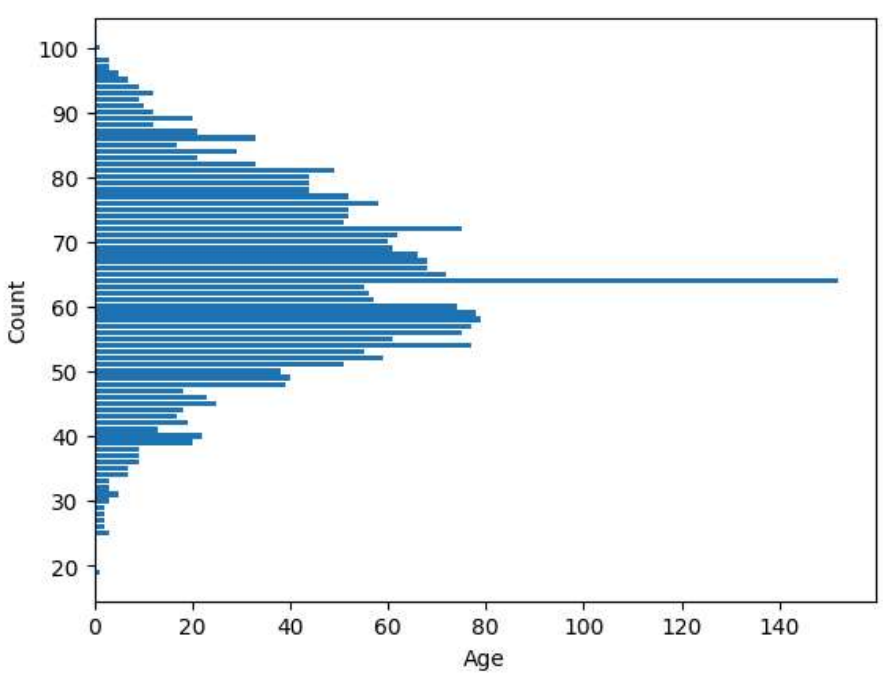
dtype: int64

DATA VISUALIZATION

---

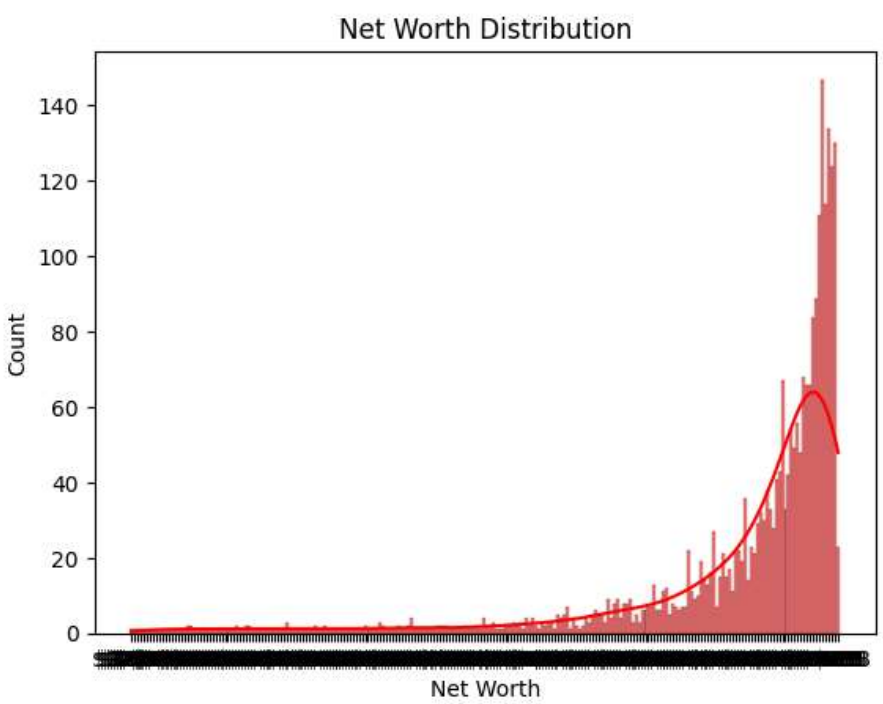
1. Show the Age distribution among the data using bar plot

```
plt.figure(figsize=(10,5))
plt.title('Age Distribution of Billionaires')
plt.barh(age.index,age)
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()
```



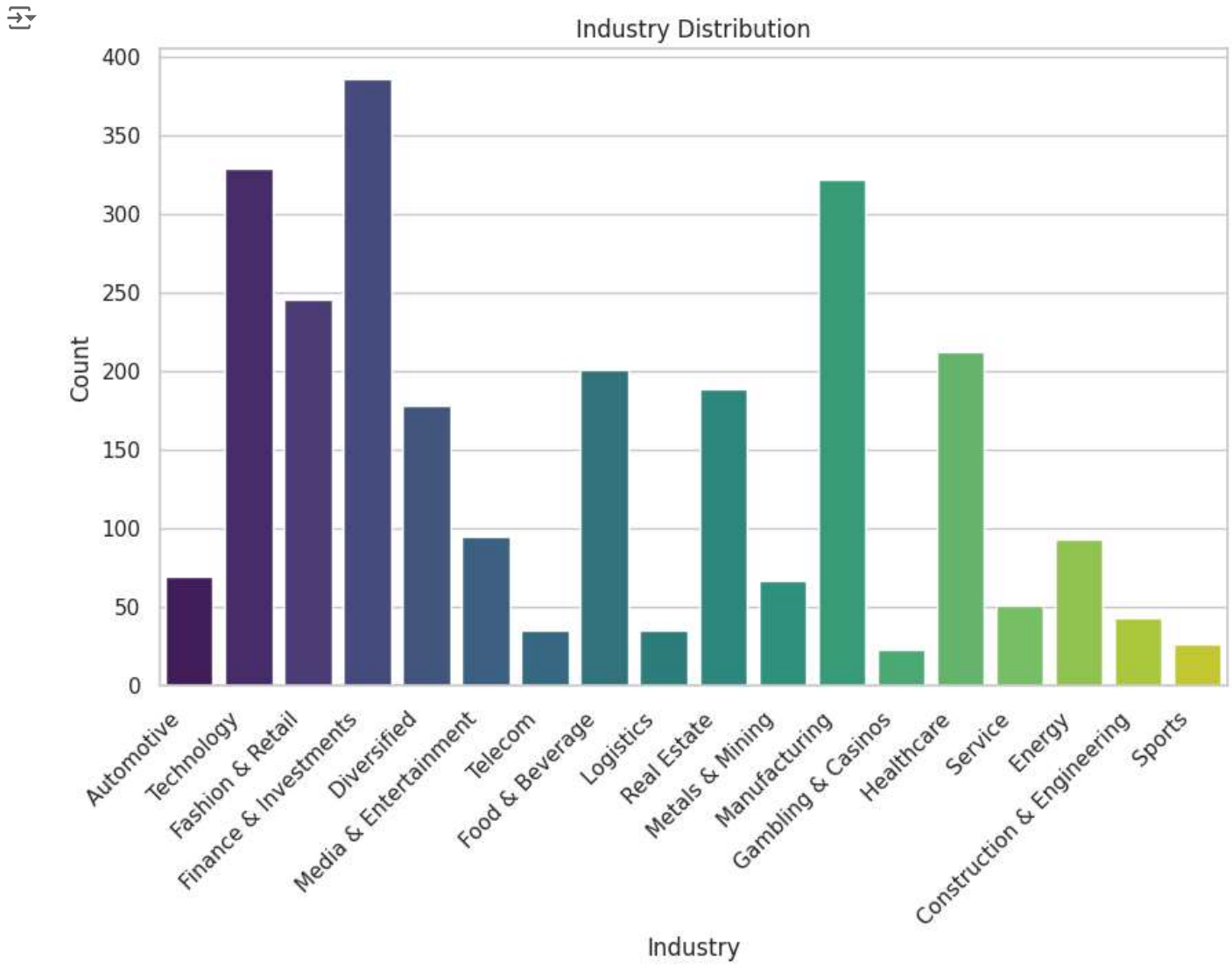
2. Show the Net Worth Vs Frequency using bar plot

```
sns.histplot(data=df, x='networth', bins=40, kde=True, color='red')
plt.title('Net Worth Distribution')
plt.xlabel('Net Worth')
plt.ylabel('Count')
plt.show()
```



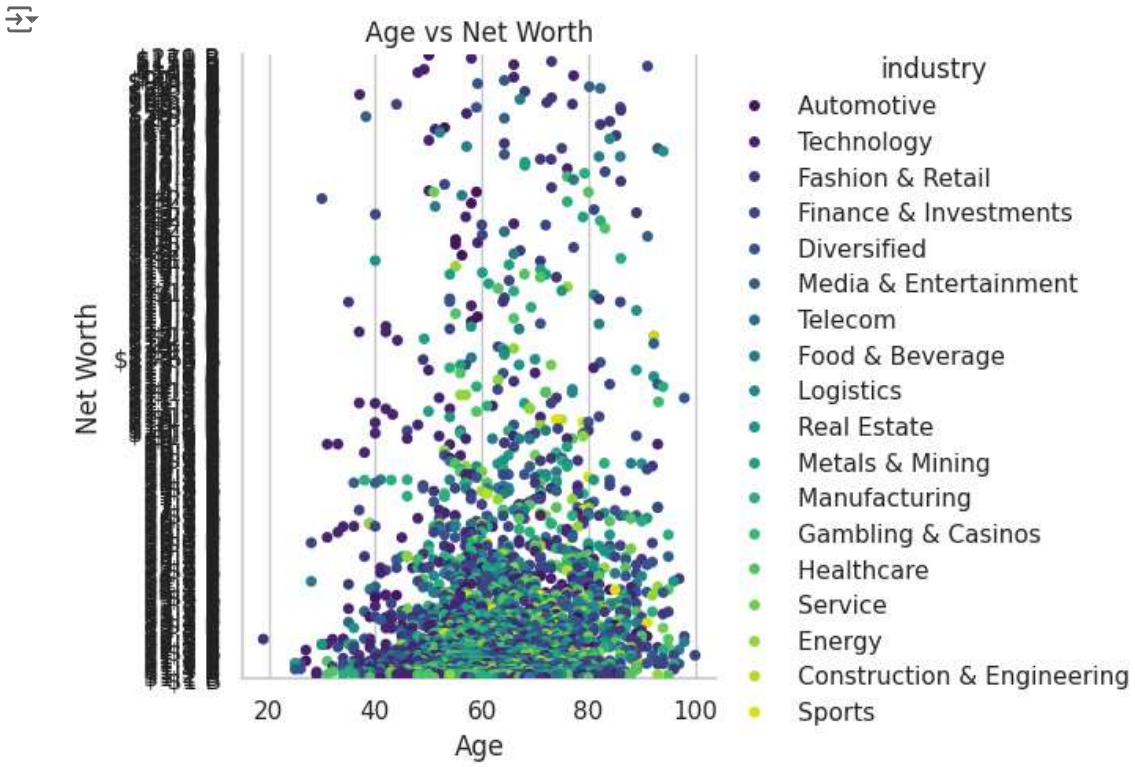
3. Show Industry Vs Frequency using bar plot

```
plt.figure(figsize=(10, 6))
sns.set(style='whitegrid')
sns.countplot(data=df, x='industry', palette='viridis')
plt.title('Industry Distribution')
plt.xlabel('Industry')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')

plt.show()
```



4. Show how does Net worth Change with age and industry using cat plot

```
sns.catplot(data=df, x='age', y='networth', hue='industry', palette='viridis')
plt.title('Age vs Net Worth')
plt.xlabel('Age')
plt.ylabel('Net Worth')
plt.show()
```



5. Show the top 10 richest people vs net worth.

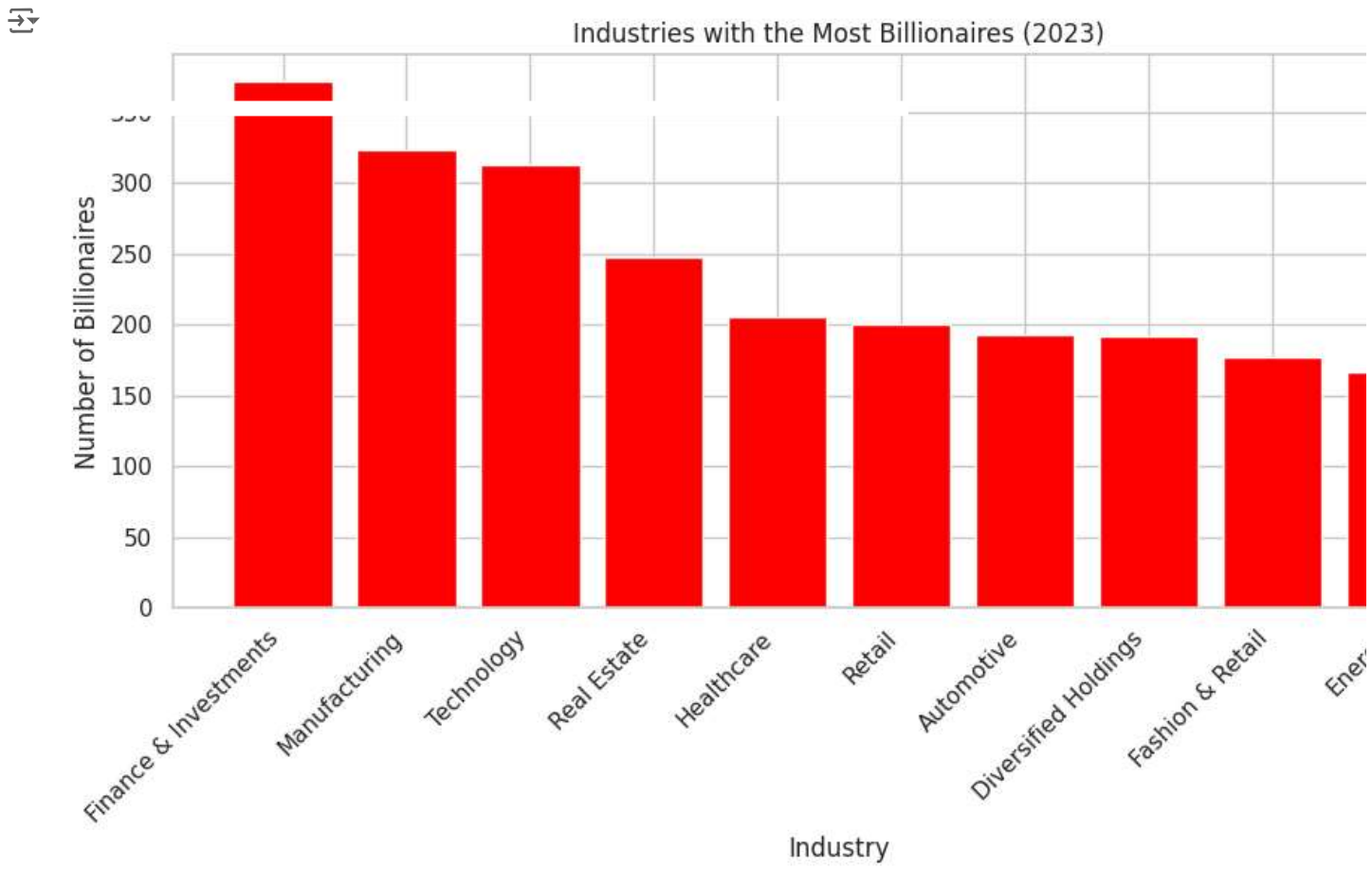
```
names = [
    "Elon Musk", "Jeff Bezos", "Larry Ellison", "Mark Zuckerberg",
    "Bernard Arnault", "Warren Buffett", "Bill Gates", "Larry Page",
    "Amancio Ortega", "Sergey Brin"
]
net_worths = [
    251.5, 204.6, 203.7, 185.4, 172.4, 143.4, 138.5, 132.5, 128.0, 126.9
]

print("Top 10 Richest People:")
for rank, (person, net_worth) in enumerate(zip(names, net_worths), start=1):
    print(f"{rank}. {person}: ${net_worth:.1f} billion")
```

```
Top 10 Richest People:
1. Elon Musk: $251.5 billion
2. Jeff Bezos: $204.6 billion
3. Larry Ellison: $203.7 billion
4. Mark Zuckerberg: $185.4 billion
5. Bernard Arnault: $172.4 billion
6. Warren Buffett: $143.4 billion
7. Bill Gates: $138.5 billion
8. Larry Page: $132.5 billion
9. Amancio Ortega: $128.0 billion
10. Sergey Brin: $126.9 billion
```

6. Show the richest people from India with the names in any plot

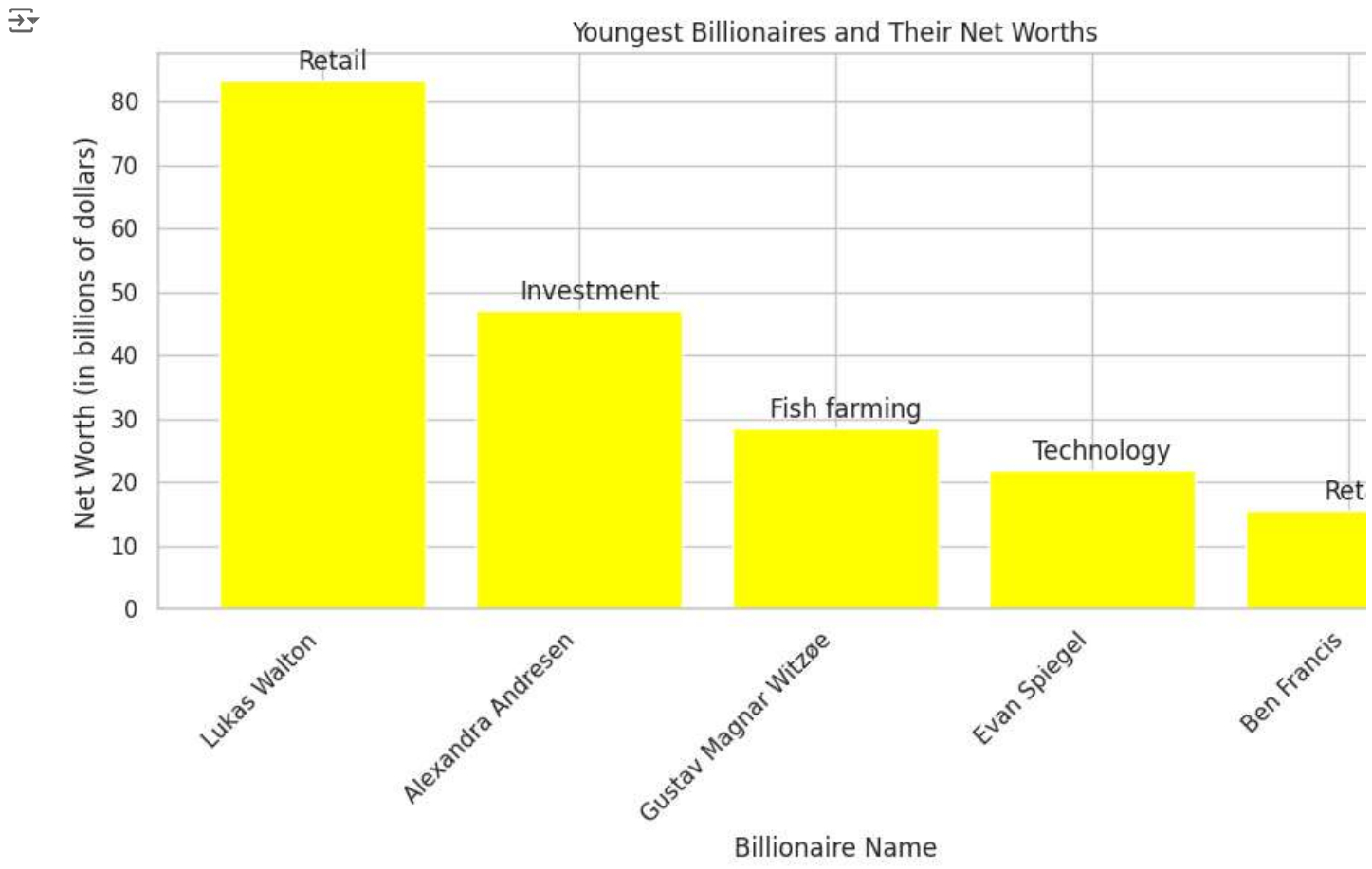
```
industries = ["Finance & Investments", "Manufacturing", "Technology", "Real Estate", "Healthcare", "Retail", "Automoti
num_billionaires = [372, 324, 313, 248, 206, 200, 193, 192, 177, 167]
plt.figure(figsize=(10, 6))
```

```
plt.bar(industries, num_billionaires, color='red')
plt.xlabel('Industry')
plt.ylabel('Number of Billionaires')
plt.title('Industries with the Most Billionaires (2023)')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```



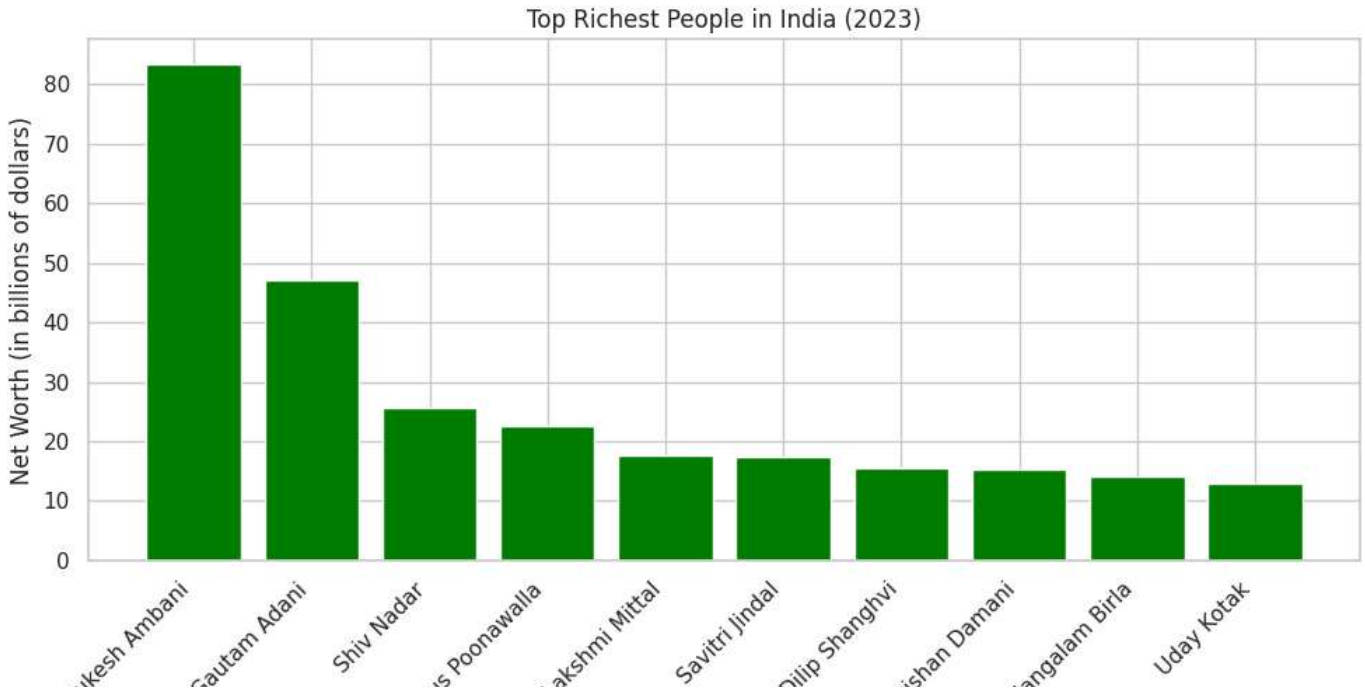
7. Show the minimum age billionaire <=50 with name and industry.

```
import matplotlib.pyplot as plt
names = ["Lukas Walton", "Alexandra Andresen", "Gustav Magnar Witzøe", "Evan Spiegel", "Ben Francis"]
ages = [37, 25, 28, 33, 31]
net_worths = [83.4, 47.2, 28.6, 22.0, 15.6]
industries = ["Retail", "Investment", "Fish farming", "Technology", "Retail"]
plt.figure(figsize=(10, 6))
plt.bar(names, net_worths, color='yellow')
plt.xlabel('Billionaire Name')
plt.ylabel('Net Worth (in billions of dollars)')
plt.title('Youngest Billionaires and Their Net Worths')
plt.xticks(rotation=45, ha='right')
for i, name in enumerate(names):
    plt.annotate(industries[i], xy=(name, net_worths[i]), xytext=(5, 5), textcoords="offset points", ha='center')
plt.tight_layout()
plt.show()
```



8. Show in which industry billionaires are more:-

```
import matplotlib.pyplot as plt
names = [
    "Mukesh Ambani", "Gautam Adani", "Shiv Nadar", "Cyrus Poonawalla",
    "Lakshmi Mittal", "Savitri Jindal", "Dilip Shanghvi", "Radhakishan Damani",
    "Kumar Mangalam Birla", "Uday Kotak"
]
net_worths = [
    83.4, 47.2, 25.6, 22.6, 17.7, 17.5, 15.6, 15.3, 14.2, 12.9
]

plt.figure(figsize=(10, 6))
plt.bar(names, net_worths, color='green')
plt.xlabel('Richest People in India')
plt.ylabel('Net Worth (in billions of dollars)')
plt.title('Top Richest People in India (2023)')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```


Top Richest People in India (2023)
Net Worth (in billions of dollars)
0
10
20
30
40
50
60
70
80
Shiv Nadar
Savitri Jindal
Uday Kotak