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

```
dataset.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

```
dataset=pd.read_csv("/content/2022_forbes_billionaires.csv")
dataset.head()
```

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

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

dtype: int64

```
dataset.isnull()
```

	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
...	...	...	...	...	...	...	...	...
2595	False	False	False	False	False	False	False	False
2596	False	False	False	False	False	False	False	False
2597	False	False	False	False	False	False	False	False
2598	False	False	False	False	False	False	False	False
2599	False	False	False	False	False	False	False	False

2600 rows × 8 columns

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

0

```
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

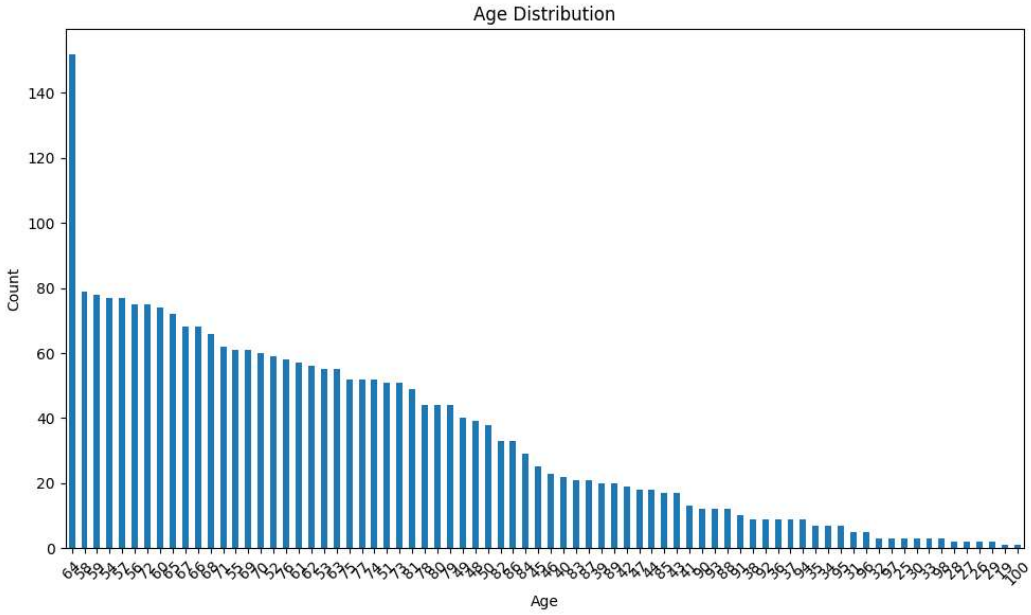
```
shape=dataset.shape
print(shape)
```

```
(2600, 8)
```

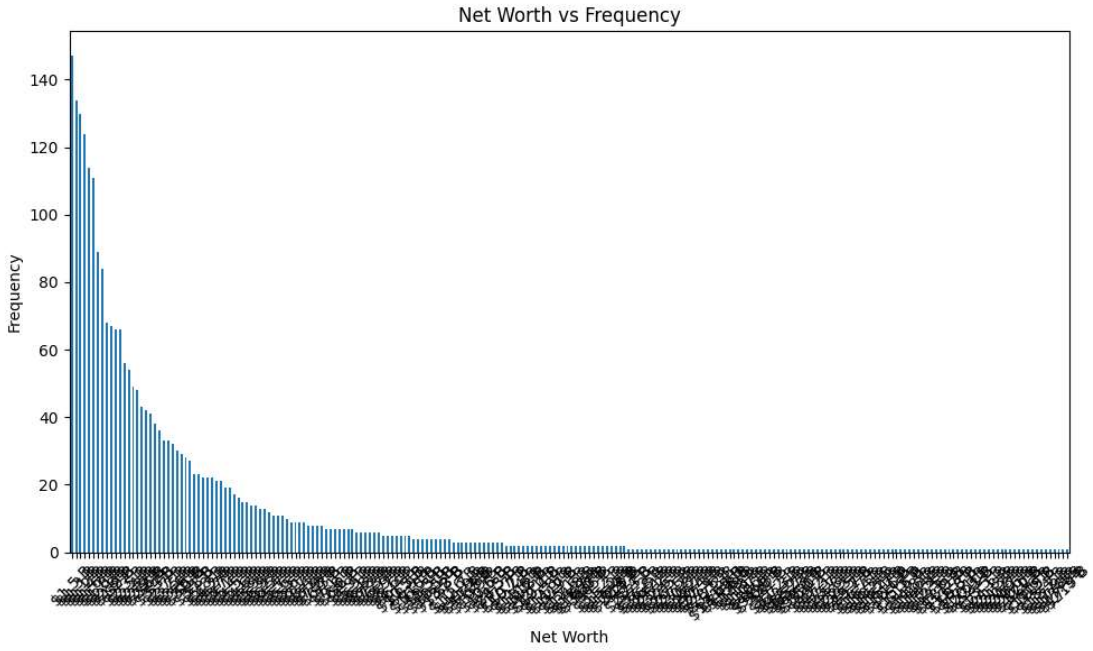
```
dataset.info()
```

```
<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
```

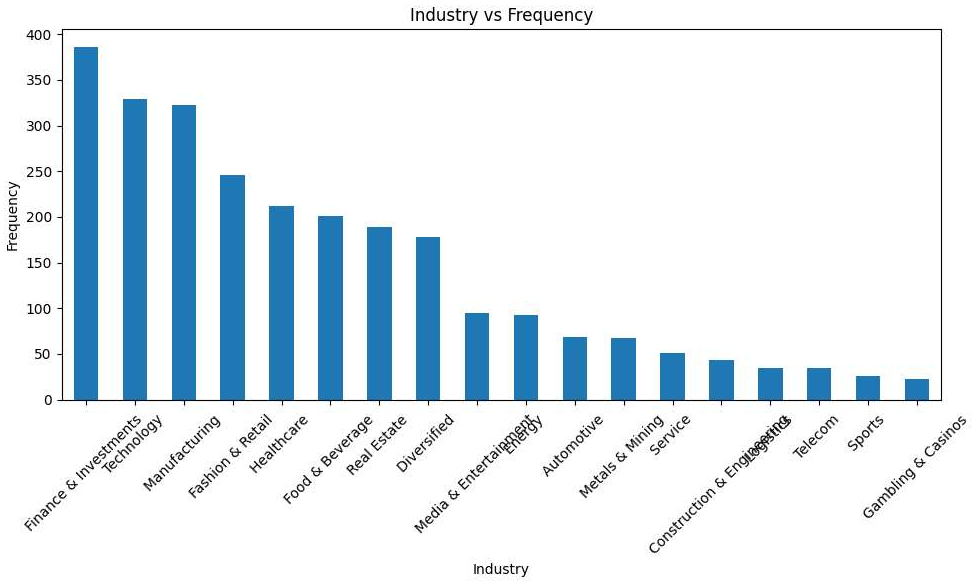
```
# calculate age distribution
age_distribution = dataset['age'].value_counts()
plt.figure(figsize=(10,6))
age_distribution.plot(kind='bar')
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



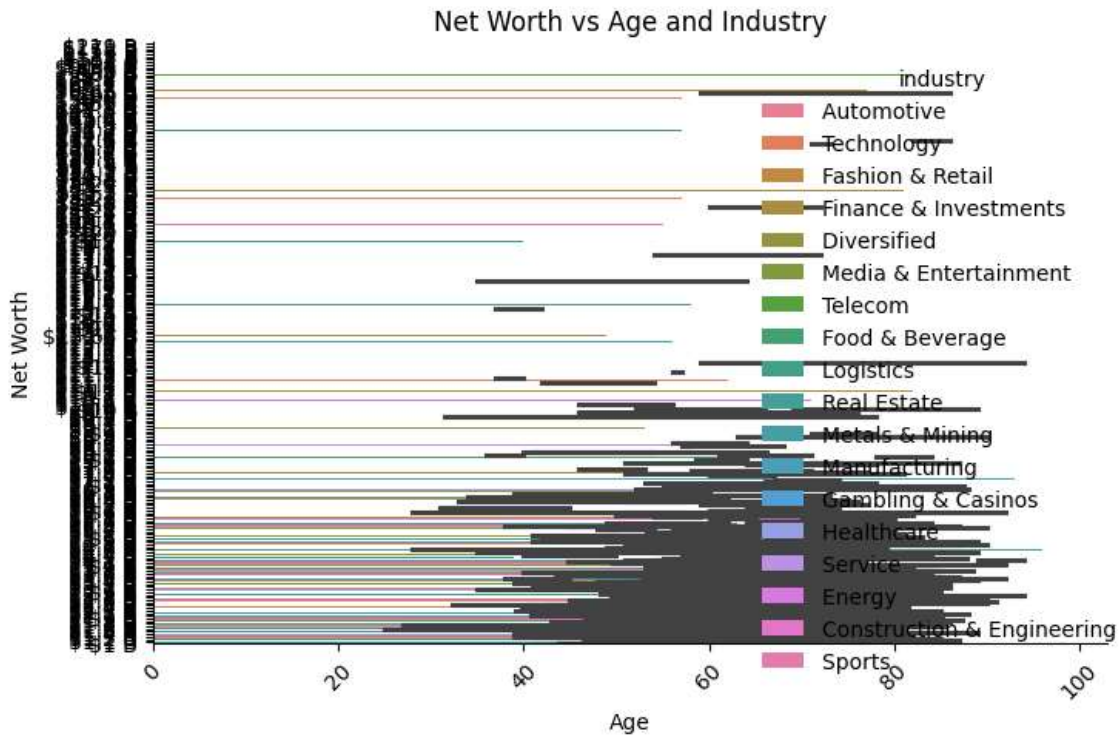

```
show the net worth vs frequency using bar plot

net_worth_frequency = dataset['networth'].value_counts()
plt.figure(figsize=(10,6))

net_worth_frequency.plot(kind='bar')
plt.title("Net Worth vs Frequency")
plt.xlabel('Net Worth')
plt.ylabel('Frequency')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```




```
show Industry vs frequency using bar plot

industry_frequency = dataset['industry'].value_counts()
plt.figure(figsize=(10,6))

industry_frequency.plot(kind='bar')
plt.title("Industry vs Frequency")
plt.xlabel('Industry')
plt.ylabel('Frequency')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
print(dataset.columns)
```

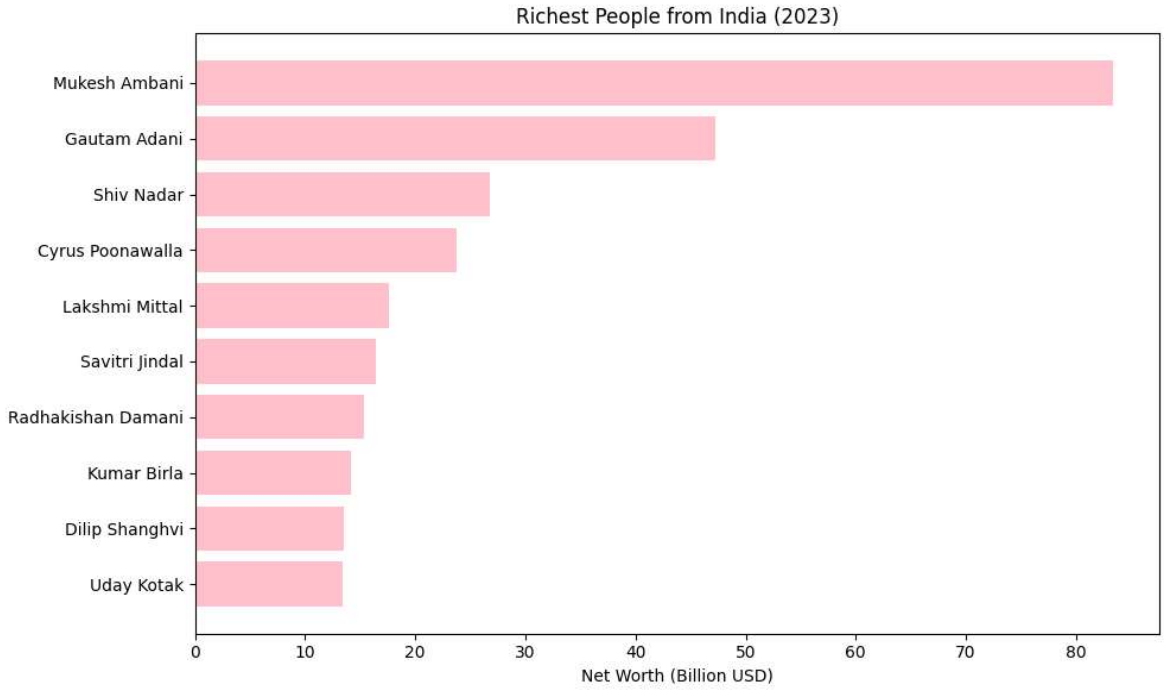


```
show how does net worth change with age and industry using cat plot

plt.figure(figsize=(12,8))
sns.catplot(x='age',y='networth',hue='industry',data=dataset,kind='bar')
plt.title("Net Worth vs Age and Industry")
plt.xlabel('Age')
plt.ylabel('Net Worth')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
print(dataset.columns)
```

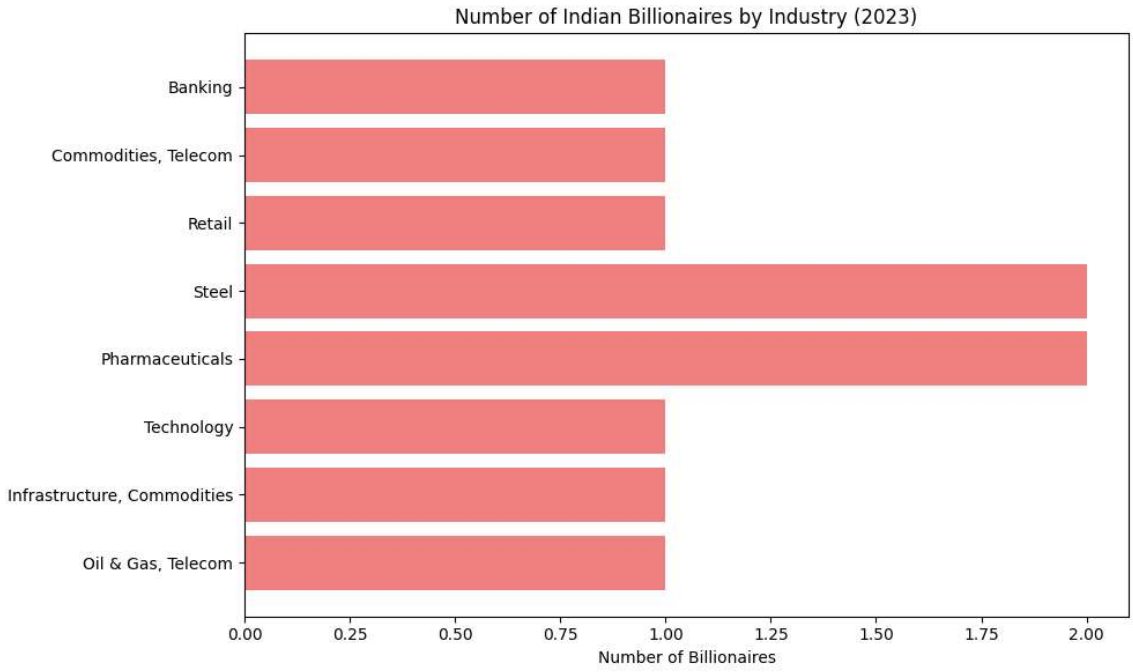



```
import matplotlib.pyplot as plt

# Data: Names of the richest people and their net worth (in billion USD)

wealth = [83.4, 47.2, 26.8, 23.7, 17.6, 16.4, 15.3, 14.2, 13.5, 13.4]


plt.figure(figsize=(10,6))
plt.barh(names, wealth, color='pink')
plt.xlabel('Net Worth (Billion USD)')
plt.title('Richest People from India (2023)')
plt.gca().invert_yaxis()
```

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




```
import matplotlib.pyplot as plt
from collections import Counter



industry_count = Counter(industries)


industry_names = list(industry_count.keys())
industry_values = list(industry_count.values())


plt.figure(figsize=(10,6))
plt.barh(industry_names, industry_values, color='lightcoral')
plt.xlabel('Number of Billionaires')
plt.title('Number of Indian Billionaires by Industry (2023)')
plt.tight_layout()

plt.show()
```




```
import pandas as pd

df = pd.read_csv('/content/2022_forbes_billionaires.csv')

young_billionaires=df[df['age']<=50]
min_age= young_billionaires['age'].min()
min_age_billionaire = young_billionaires.loc[young_billionaires['age'] == min_age]

print("name:", min_age_billionaire['name'].values[0])
print("industry:", min_age_billionaire['industry'].values[0])
print("age:", min_age)
```

```
name: Kevin David Lehmann
industry: Fashion & Retail
age: 19
```

```
import matplotlib.pyplot as plt



wealth = [83.4, 47.2, 26.8, 23.7, 17.6, 16.4, 15.3, 14.2, 13.5, 13.4]
```

```
plt.figure(figsize=(10,6))
plt.bar(names, wealth, color='skyblue')
plt.xticks(rotation=45, ha='right')
plt.ylabel('Net Worth (Billion USD)')
plt.title('Top 10 Richest People in India (2023) vs Net Worth')


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