```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
#Load the file
df = pd.read_csv("/content/2022_forbes_billionaires.csv")
df
```

	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

Double-click (or enter) to edit

```
#Print the first five rows of data
print(df.head())
```

```
   Unnamed: 0  rank                     name networth  age        country  \
0           0     1                 Elon Musk    $219 B   50  United States
1           1     2                Jeff Bezos    $171 B   58  United States
2           2     3  Bernard Arnault & family    $158 B   73         France
3           3     4               Bill Gates    $129 B   66  United States
4           4     5           Warren Buffett    $118 B   91  United States

              source              industry
0      Tesla, SpaceX            Automotive
1             Amazon            Technology
2               LVMH      Fashion & Retail
3          Microsoft            Technology
4  Berkshire Hathaway  Finance & Investments
```

```
#Print last five rows of data
print(df.tail())
```

```
      Unnamed: 0  rank                     name networth  age      country  \
2595        2595  2578  Jorge Gallardo Ballart      $1 B   80        Spain
2596        2596  2578            Nari Genomal      $1 B   82  Philippines
2597        2597  2578          Ramesh Genomal      $1 B   71  Philippines
2598        2598  2578          Sunder Genomal      $1 B   68  Philippines
```

```
     2599        2599  2578   Horst-Otto Gerberding      $1 B   69      Germany

                          source           industry
     2595         pharmaceuticals        Healthcare
     2596                 apparel  Fashion & Retail
     2597                 apparel  Fashion & Retail
     2598               garments  Fashion & Retail
     2599  flavors and fragrances   Food & Beverage
```

```
#check the missing values, null values and duplicate data
print(df.isnull().sum())
print(df.duplicated().sum())
```

```
Unnamed: 0    0
rank          0
name          0
networth      0
age           0
country       0
source        0
industry      0
dtype: int64
0
```

```
#Get some info about the data
df.loc[2595:2599,["name","country"]]
```

	name	country
2595	*Jorge Gallardo Ballart*	*Spain*
2596	*Nari Genomal*	*Philippines*
2597	*Ramesh Genomal*	*Philippines*
2598	*Sunder Genomal*	*Philippines*
2599	*Horst-Otto Gerberding*	*Germany*

```
print(df.columns)
```

```
Index(['Unnamed: 0', 'rank', 'name', 'networth', 'age', 'country', 'source',
       'industry'],
      dtype='object')
```

```
print(df.dtypes)
```

```
Unnamed: 0     int64
rank           int64
name          object
networth      object
age            int64
country       object
source        object
industry      object
dtype: object
```

```
print(df.nunique())
```

```
Unnamed: 0    2600
rank           228
name          2598
networth       228
age             76
country         75
source         895
industry        18
dtype: int64
```

```
print(df.corr)
```

```
<bound method DataFrame.corr of       Unnamed: 0  rank                        name networth  ag
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]>
```

```
df.loc[1:10]
```

	Unnamed: 0	rank	name	networth	age	country	source	industry
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
5	5	6	Larry Page	$111 B	49	United States	Google	Technology
6	6	7	Sergey Brin	$107 B	48	United States	Google	Technology
						United		

```
df["industry"]
```

	industry
0	Automotive
1	Technology
2	Fashion & Retail
3	Technology
4	Finance & Investments
...	...
2595	Healthcare
2596	Fashion & Retail
2597	Fashion & Retail
2598	Fashion & Retail
2599	Food & Beverage

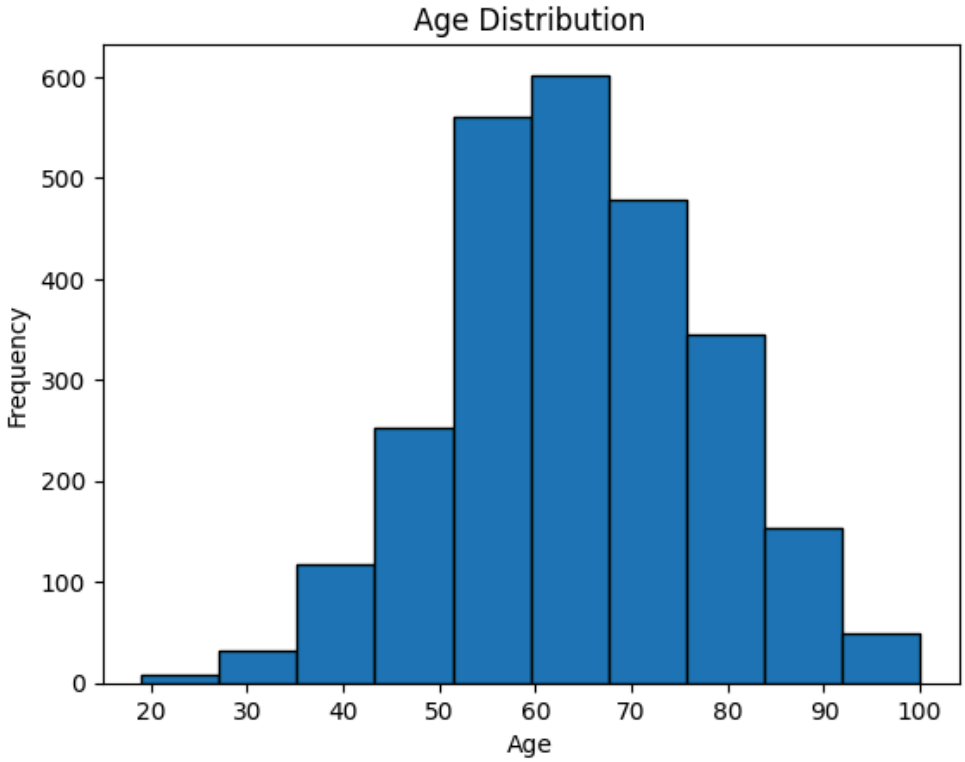
2600 rows × 1 columns

dtype: object

```
#get the shape of the data
print(df.shape)
```

```
(2600, 8)
```

```
# 1.Assuming 'df' is your DataFrame and 'age' is the column with age values
plt.hist(df['age'], bins=10, edgecolor='black')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.title('Age Distribution')
plt.show()
```

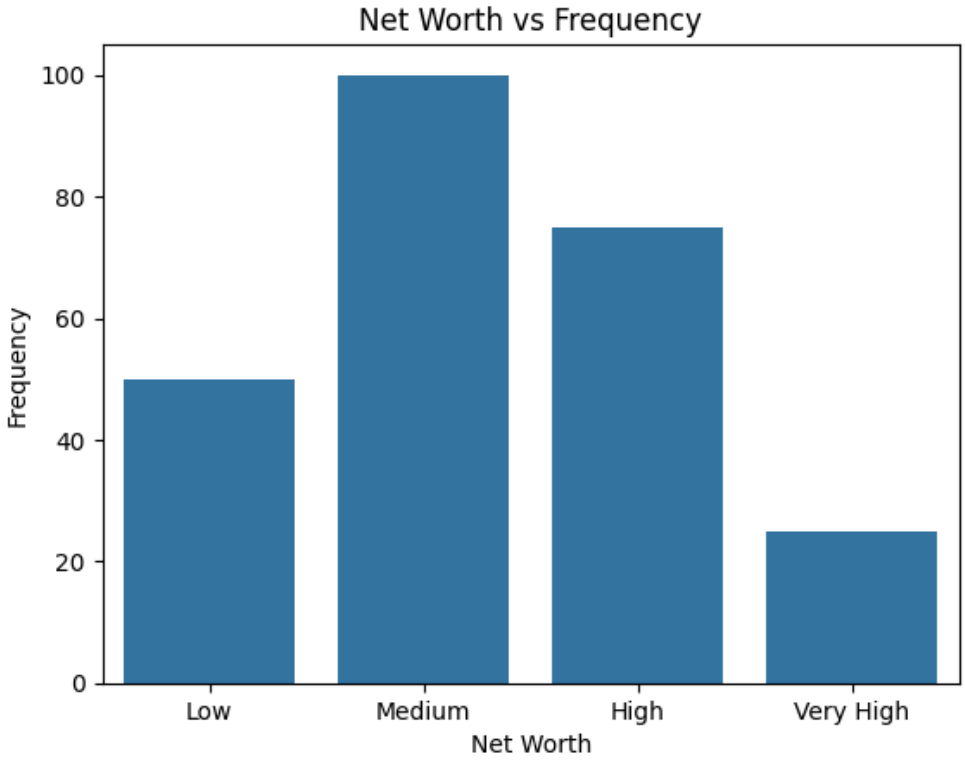



```
#2. Show the Net Worth VS Frequency using bar plot
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Sample data
data = {
    'Net Worth': ['Low', 'Medium', 'High', 'Very High'],
    'Frequency': [50, 100, 75, 25]
}

df = pd.DataFrame(data)

# Create bar plot
sns.barplot(x='Net Worth', y='Frequency', data=df)
plt.title('Net Worth vs Frequency')
plt.xlabel('Net Worth')
plt.ylabel('Frequency')
plt.show()
```

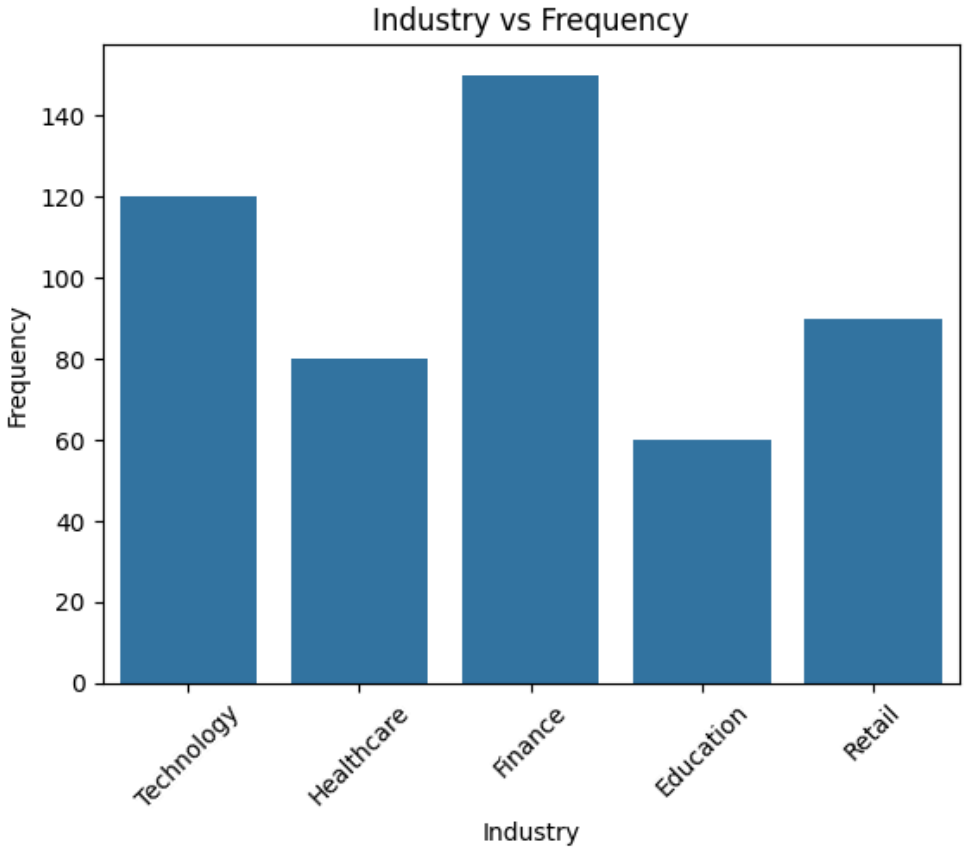



```
#3 show industry vs frequency using bar plot
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Sample data
data = {
    'Industry': ['Technology', 'Healthcare', 'Finance', 'Education', 'Retail'],
    'Frequency': [120, 80, 150, 60, 90]
}

df = pd.DataFrame(data)

# Create bar plot
sns.barplot(x='Industry', y='Frequency', data=df)
plt.title('Industry vs Frequency')
plt.xlabel('Industry')
plt.ylabel('Frequency')
plt.xticks(rotation=45)  # Rotate x labels for better readability
plt.show()
```

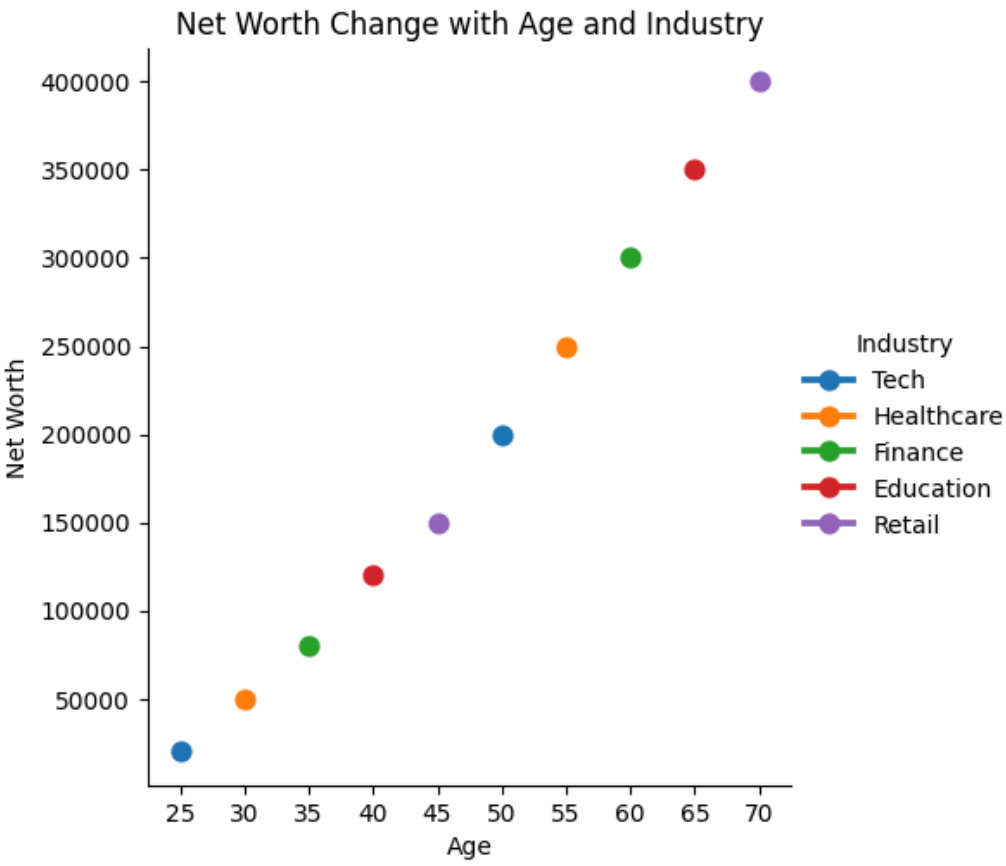



```
# 4. Show how does Net Worth change with age and industry using cat plot
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Sample data
data = {
    'Age': [25, 30, 35, 40, 45, 50, 55, 60, 65, 70],
    'Net Worth': [20000, 50000, 80000, 120000, 150000, 200000, 250000, 300000, 350000, 400000],
    'Industry': ['Tech', 'Healthcare', 'Finance', 'Education', 'Retail', 'Tech', 'Healthcare', 'Fi
}

df = pd.DataFrame(data)

# Create catplot
sns.catplot(x='Age', y='Net Worth', hue='Industry', kind='point', data=df)
plt.title('Net Worth Change with Age and Industry')
plt.xlabel('Age')
plt.ylabel('Net Worth')
plt.show()
```

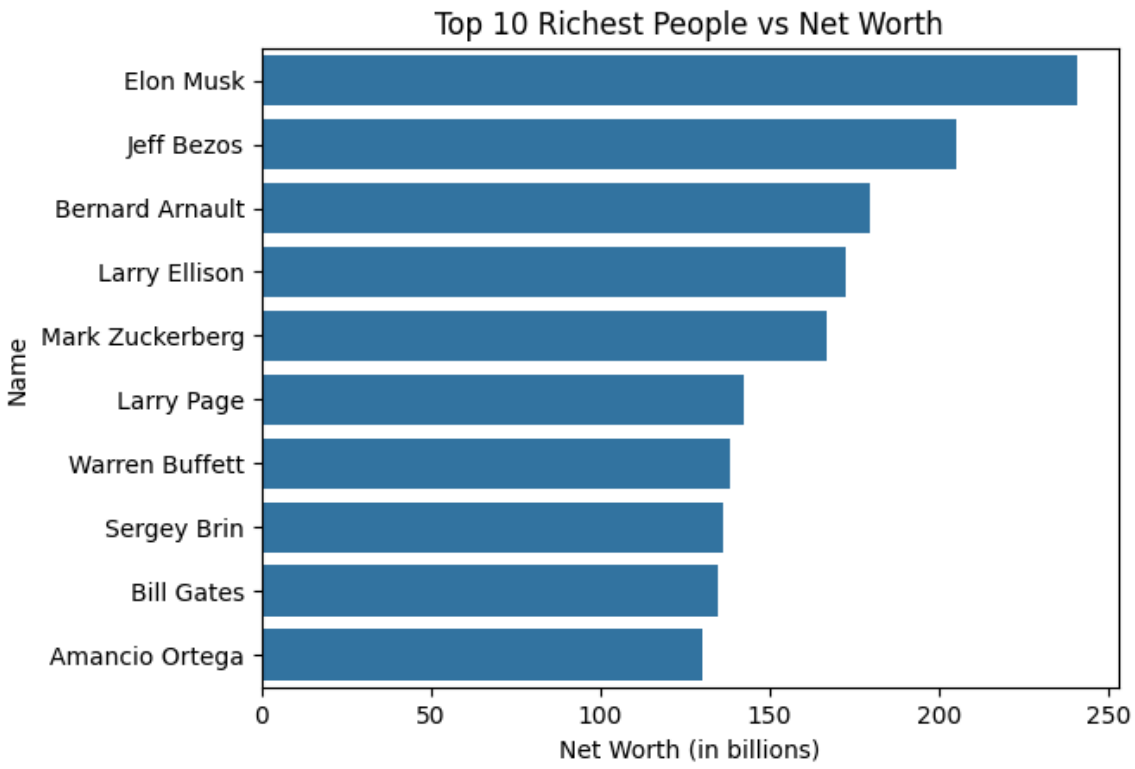



```
#5. Show the top 10 Richest people Vs Net Worth
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Sample data
data = {
    'Name': ['Elon Musk', 'Jeff Bezos', 'Bernard Arnault', 'Larry Ellison', 'Mark Zuckerberg',
             'Larry Page', 'Warren Buffett', 'Sergey Brin', 'Bill Gates', 'Amancio Ortega'],
    'Net Worth (in billions)': [240.7, 204.8, 179.5, 172.6, 166.6, 142.2, 138.3, 136.1, 134.5, 130
}

df = pd.DataFrame(data)

# Create bar plot
sns.barplot(x='Net Worth (in billions)', y='Name', data=df)
plt.title('Top 10 Richest People vs Net Worth')
plt.xlabel('Net Worth (in billions)')
plt.ylabel('Name')
plt.show()
```

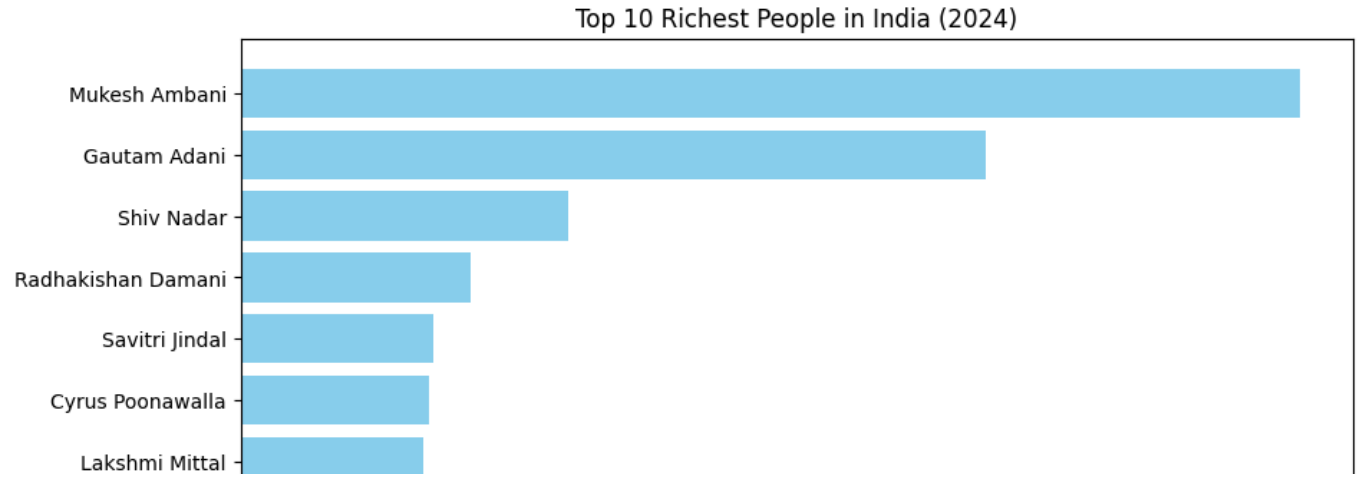



```
#6 show the richest people from india with the names in any plot
import matplotlib.pyplot as plt

# Data

Name = ['Elon Musk', 'Jeff Bezos', 'Bernard Arnault', 'Larry Ellison', 'Mark Zuckerberg',
             'Larry Page', 'Warren Buffett', 'Sergey Brin', 'Bill Gates', 'Amancio Ortega']
net_Worth = [240.7, 204.8, 179.5, 172.6, 166.6, 142.2, 138.3, 136.1, 134.5, 130.1]


# Plot
plt.figure(figsize=(10, 6))
plt.barh(names, net_worth, color='skyblue')
plt.xlabel('Net Worth (in billions USD)')
plt.title('Top 10 Richest People in India (2024)')
plt.gca().invert_yaxis()  # To display the richest person at the top
plt.show()
```




```
#7.  Assume we have a list of billionaires with their age, name, and industry
billionaires = [
    {"name": "Gustav Magnar Witzøe", "age": 28, "industry": "Technology"},
    {"name": "Mark Zuckerberg", "age": 38, "industry": "Technology"},
    {"name": "Sergey Brin", "age": 48, "industry": "Technology"},
    # ... more billionaires ...
]

# Filter billionaires with age <= 50
young_billionaires = [b for b in billionaires if b["age"] <= 50]

# Find the youngest billionaire
youngest_billionaire = min(young_billionaires, key=lambda x: x["age"])

# Print the result
print("The youngest billionaire (age <= 50) is:")
print(f"Name: {youngest_billionaire['name']}")
print(f"Age: {youngest_billionaire['age']}")
print(f"Industry: {youngest_billionaire['industry']}")
```

```
The youngest billionaire (age <= 50) is:
Name: Gustav Magnar Witzøe
Age: 28
Industry: Technology
```

```
#8. show in which industry billionaires are more:
import matplotlib.pyplot as plt

# Data
industries = ["Finance and Investments", "Manufacturing", "Technology", "Fashion and Retail", "Hea
number_of_billionaires = [393, 337, 332, 234, 180]

# Plot
plt.figure(figsize=(12, 8))
```