Billionaires Analysis with Python

The number of billionaires in a country says a lot about the business environment, startup success rate, and many other economic features of a Country. So if you want to learn more about how we can find relationships among billionaires around the world, this article is for you. In this article, I will walk you through the task of billionaires analysis with Python.

dataset

The dataset that I am using to analyze the data about billionaires around the world,is Contained :

- Names
- Net Worth
- Country
- Source
- Rank
- Age
- Industry

let's start .....

importing the necessary Python libraries and the dataset:

In [16]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [17]:
```
data = pd.read_csv("Billionaire.csv")
data.head()
```

Out[17]:

	Name	NetWorth	Country	Source	Rank	Age	Industry
0	Jeff Bezos	$177 B	United States	Amazon	1	57.0	Technology
1	Elon Musk	$151 B	United States	Tesla, SpaceX	2	49.0	Automotive
2	Bernard Arnault & family	$150 B	France	LVMH	3	72.0	Fashion & Retail
3	Bill Gates	$124 B	United States	Microsoft	4	65.0	Technology
4	Mark Zuckerberg	$97 B	United States	Facebook	5	36.0	Technology

Before we go ahead, let's see whether or not this dataset contains missing values:

In [18]:
```
data.isnull().sum()
```

Out[18]:
```
Name          0
NetWorth      0
Country       0
Source        0
Rank          0
Age          79
Industry      0
dtype: int64
```

So this dataset has 79 missing values in the Age column, let's remove these rows:

In [19]:
```
data = data.dropna()
data.isnull().sum()
```

Out[19]:
```
Name         0
NetWorth     0
Country      0
Source       0
Rank         0
Age          0
Industry     0
dtype: int64
```

Note that !

The NetWorth column in this dataset has a "dollar sign" at the beginning of Billionaires' Net worth and B at the end. So we need to remove these signs and convert the NetWorth column to float:

In [20]:
```
data["NetWorth"] = data["NetWorth"].str.strip("$")
data["NetWorth"] = data["NetWorth"].str.strip("B")
data["NetWorth"] = data["NetWorth"].astype(float)
```

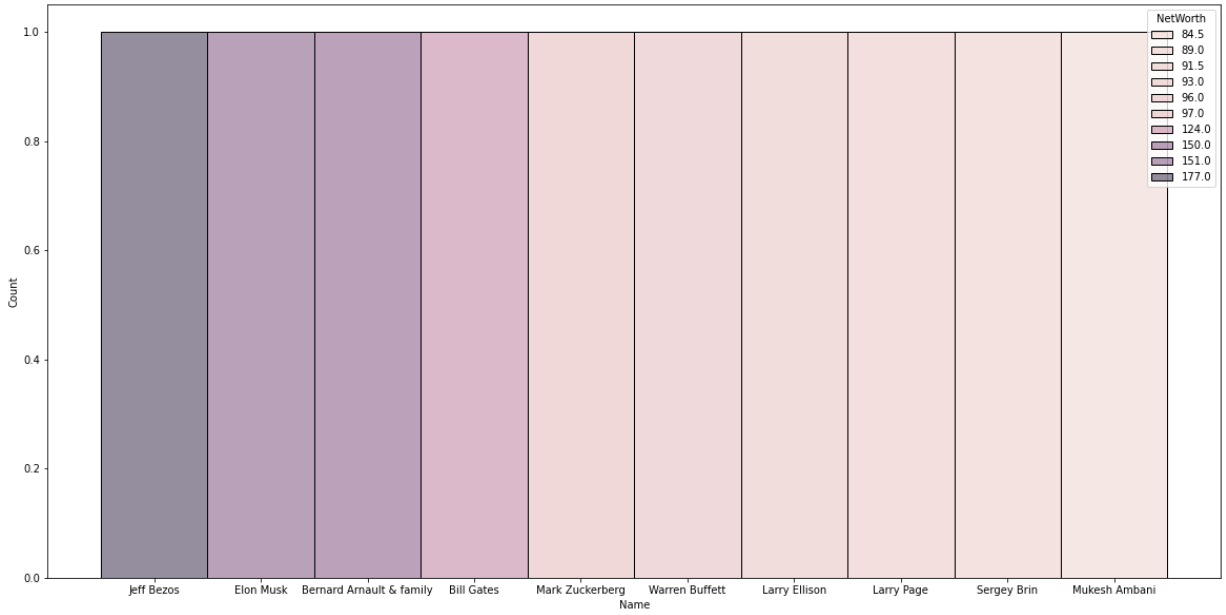
In [21]:
```
data.head()
```

Out[21]:

	Name	NetWorth	Country	Source	Rank	Age	Industry
0	Jeff Bezos	177.0	United States	Amazon	1	57.0	Technology
1	Elon Musk	151.0	United States	Tesla, SpaceX	2	49.0	Automotive
2	Bernard Arnault & family	150.0	France	LVMH	3	72.0	Fashion & Retail
3	Bill Gates	124.0	United States	Microsoft	4	65.0	Technology
4	Mark Zuckerberg	97.0	United States	Facebook	5	36.0	Technology

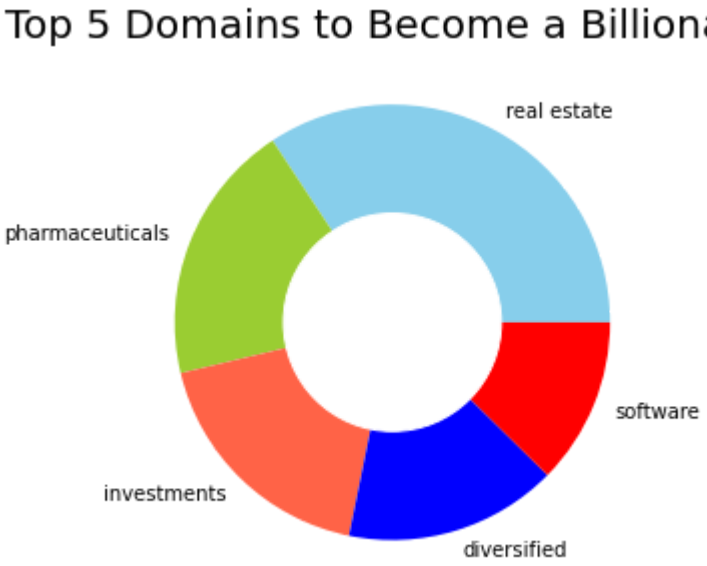
Now let's have a look at the top 10 billionaires according to their NetWorth:

In [22]:
```
df = data.sort_values(by = ["NetWorth"], ascending=False).head(10)
plt.figure(figsize=(20, 10))
sns.histplot(x="Name", hue="NetWorth", data=df)
plt.show()
```



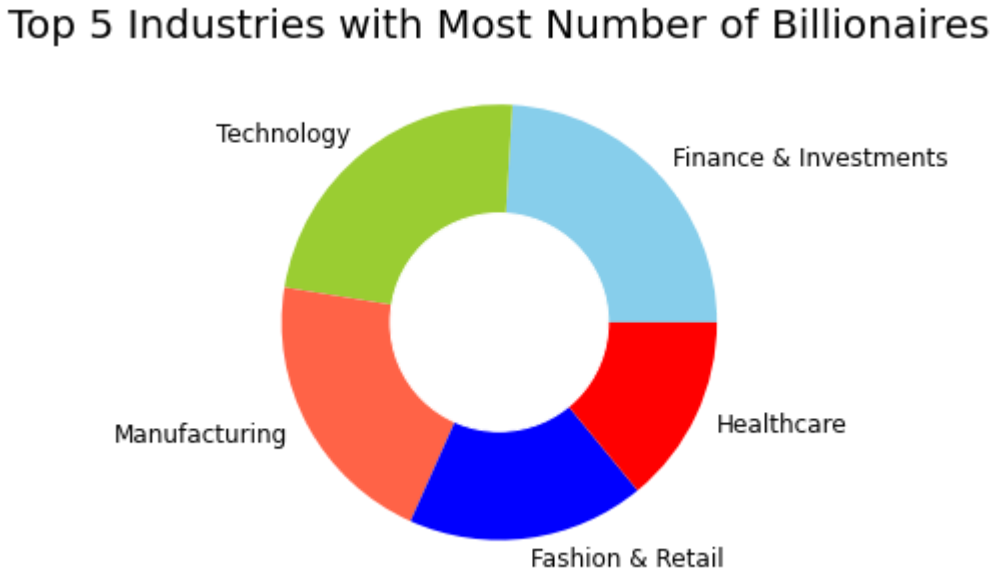
Now let's have a look at the top 5 domains with the most number of billionaires:

In [23]:
```
a = data["Source"].value_counts().head()
index = a.index
sources = a.values
custom_colors = ["skyblue", "yellowgreen", 'tomato', "blue", "red"]
plt.figure(figsize=(5, 5))
plt.pie(sources, labels=index, colors=custom_colors)
central_circle = plt.Circle((0, 0), 0.5, color='white')
fig = plt.gcf()
fig.gca().add_artist(central_circle)
plt.rc('font', size=12)
plt.title("Top 5 Domains to Become a Billionaire", fontsize=20)
plt.show()
```



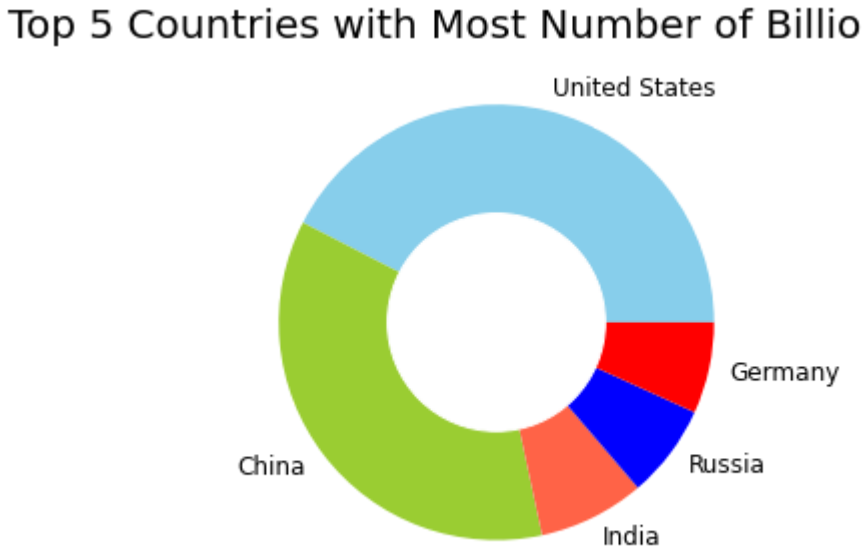
Now let's have a look at the top 5 industries with the most number of billionaires:

In [24]:
```
a = data["Industry"].value_counts().head()
index = a.index
industries = a.values
custom_colors = ["skyblue", "yellowgreen", 'tomato', "blue", "red"]
plt.figure(figsize=(5, 5))
plt.pie(industries, labels=index, colors=custom_colors)
central_circle = plt.Circle((0, 0), 0.5, color='white')
fig = plt.gcf()
fig.gca().add_artist(central_circle)
plt.rc('font', size=12)
plt.title("Top 5 Industries with Most Number of Billionaires", fontsize=20)
plt.show()
```



Now let's have a look at the top 5 countries with the most number of billionaires:

In [25]:
```
a = data["Country"].value_counts().head()
index = a.index
Countries = a.values
custom_colors = ["skyblue", "yellowgreen", 'tomato', "blue", "red"]
plt.figure(figsize=(5, 5))
plt.pie(Countries, labels=index, colors=custom_colors)
central_circle = plt.Circle((0, 0), 0.5, color='white')
fig = plt.gcf()
fig.gca().add_artist(central_circle)
plt.rc('font', size=12)
plt.title("Top 5 Countries with Most Number of Billionaires", fontsize=20)
plt.show()
```



In [ ]: