Billionaires Analysis using Python

Sreeparna Ray ¶

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 I want to find more about how we can find relationships among billionaires around the world. Here I will walk you through the task of billionaires analysis with Python.

In [1]:
```
# The dataset that I am using to analyze the data about billionaires around the
# and is downloaded from Kaggle. The dataset contains information about global

# Names
# Net Worth
# Country
# Source
# Rank
# Age
# Industry
```

Start with the task of billionaires analysis usin python

In [2]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

%matplotlib inline
```

Data Collection

In [3]:
```
data = pd.read_csv(r".\Billionaries Analysis.csv")
data.head()
```

Out[3]:

	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

In [4]:
```
print(data.head())
```

```
                     Name NetWorth        Country         Source  Rank  \
0               Jeff Bezos   $177 B  United States         Amazon     1
1                Elon Musk   $151 B  United States  Tesla, SpaceX     2
2  Bernard Arnault & family   $150 B         France           LVMH     3
3               Bill Gates   $124 B  United States      Microsoft     4
4          Mark Zuckerberg    $97 B  United States       Facebook     5

    Age          Industry
0  57.0        Technology
1  49.0        Automotive
2  72.0  Fashion & Retail
3  65.0        Technology
4  36.0        Technology
```

In [5]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2755 entries, 0 to 2754
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   Name      2755 non-null   object 
 1   NetWorth  2755 non-null   object 
 2   Country   2755 non-null   object 
 3   Source    2755 non-null   object 
 4   Rank      2755 non-null   int64  
 5   Age       2676 non-null   float64
 6   Industry  2755 non-null   object 
dtypes: float64(1), int64(1), object(5)
memory usage: 150.8+ KB
```

In [6]:
```
print(data.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2755 entries, 0 to 2754
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   Name      2755 non-null   object 
 1   NetWorth  2755 non-null   object 
 2   Country   2755 non-null   object 
 3   Source    2755 non-null   object 
 4   Rank      2755 non-null   int64  
 5   Age       2676 non-null   float64
 6   Industry  2755 non-null   object 
dtypes: float64(1), int64(1), object(5)
memory usage: 150.8+ KB
None
```

In [7]:
```
data.describe()
```

Out[7]:

	Rank	Age
count	2755.000000	2676.000000
mean	1345.663521	63.113602
std	772.669811	13.445153
min	1.000000	18.000000
25%	680.000000	54.000000
50%	1362.000000	63.000000
75%	2035.000000	73.000000
max	2674.000000	99.000000

In [8]:
```
data.columns
```

Out[8]:
```
Index(['Name', 'NetWorth', 'Country', 'Source', 'Rank', 'Age', 'Industry'], d
type='object')
```

Check whether or not this dataset contains missing values or null values

In [9]:
```
print(data.isnull().sum())
```

```
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, Remove these rows

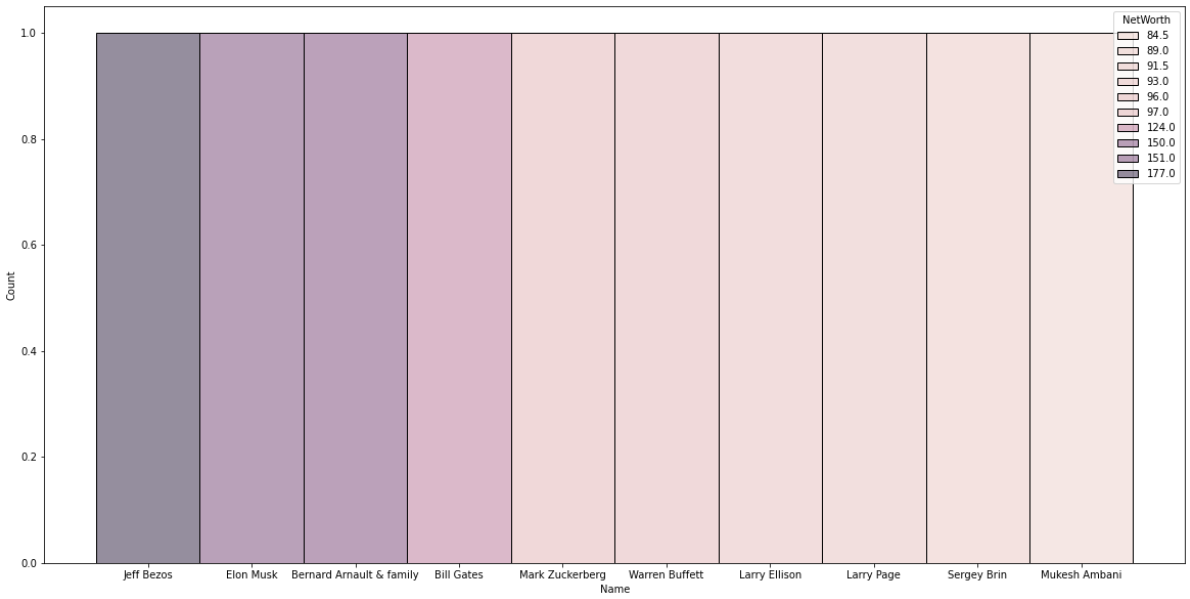
In [10]:
```
data = data.dropna()
```

The NetWorth column in this dataset has a $ 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 [11]:
```
data["NetWorth"] = data["NetWorth"].str.strip("$")
data["NetWorth"] = data["NetWorth"].str.strip("B")
data["NetWorth"] = data["NetWorth"].astype(float)
```

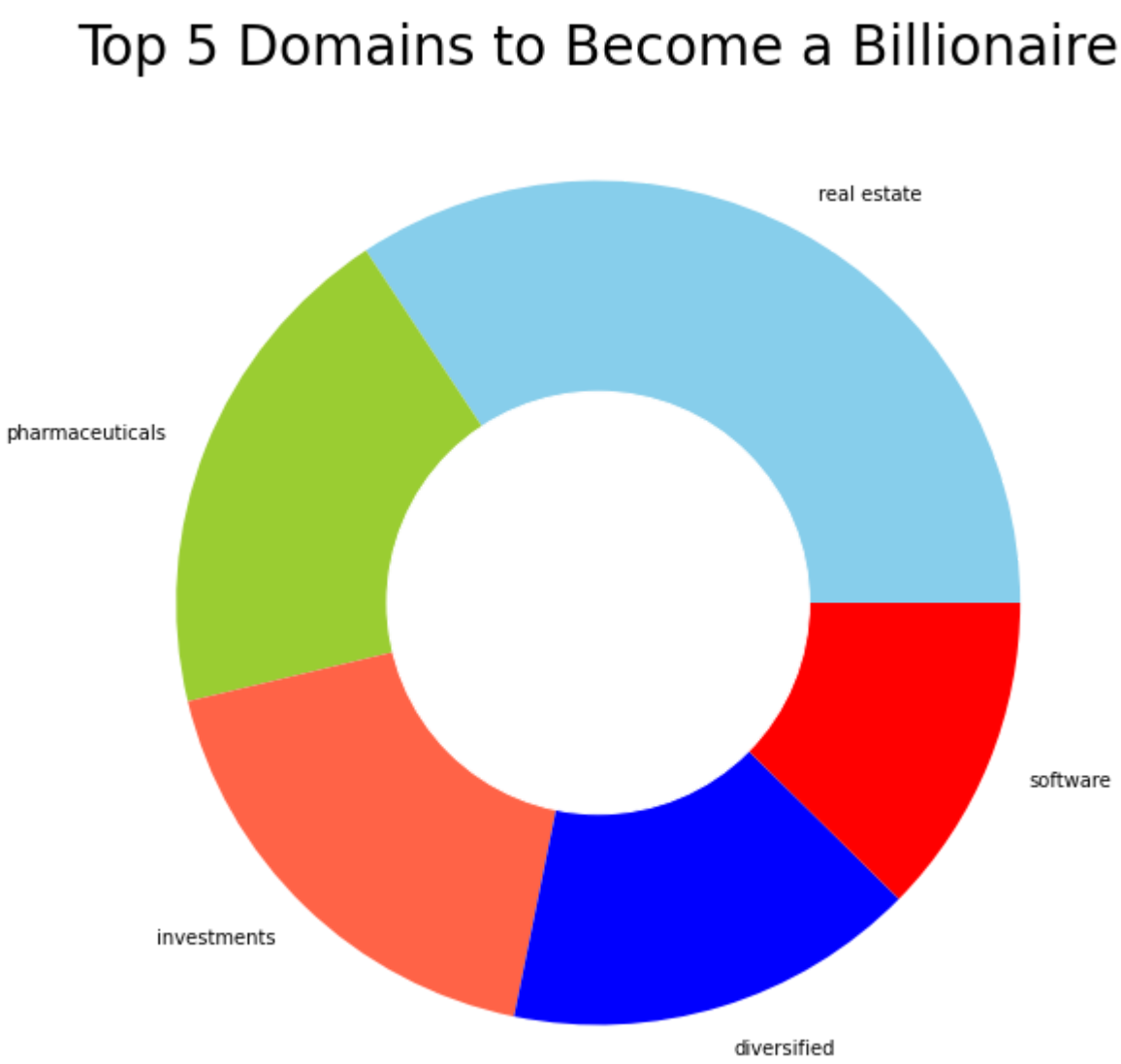
Check the top 10 billionaires according to their NetWorth:

In [12]:
```
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()
```



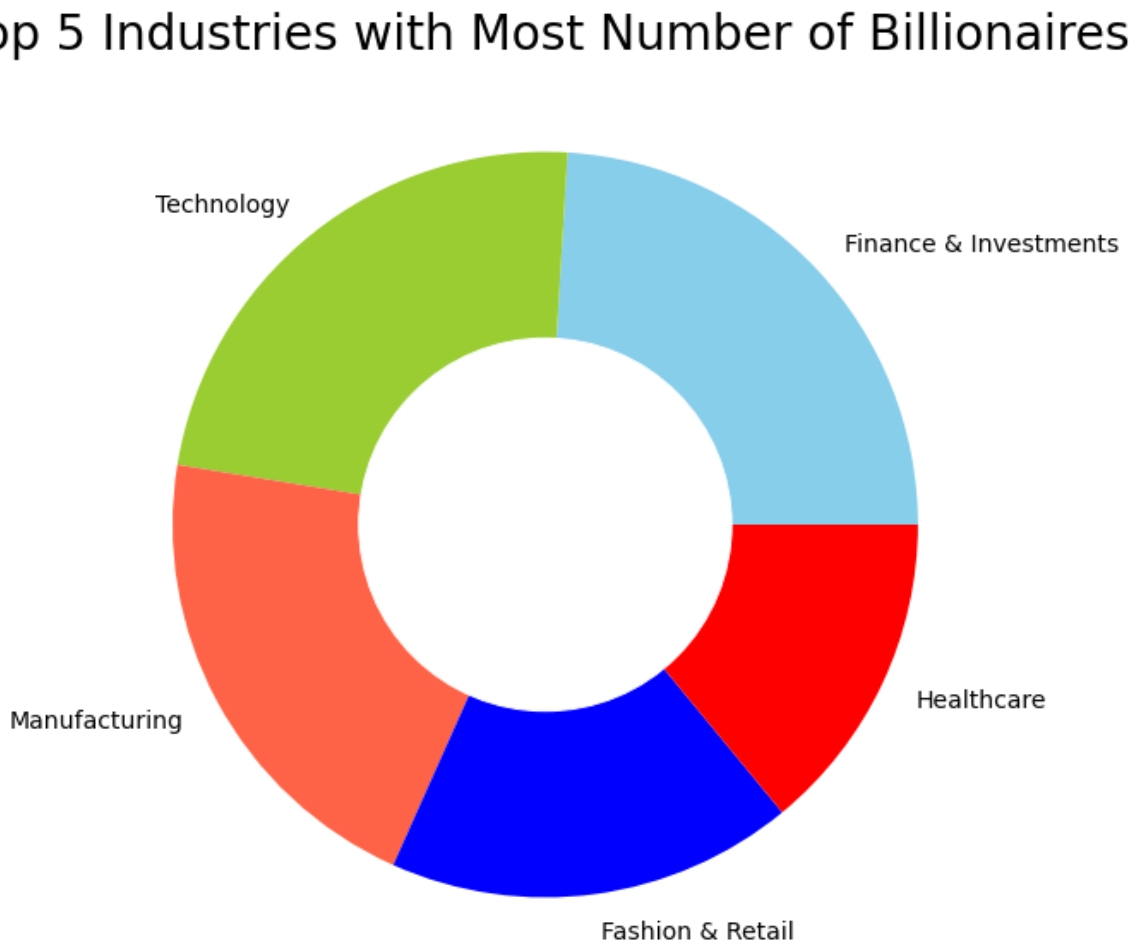
Now let's have a look at the top 5 domains with the most number of billionaires

In [13]:
```
a = data["Source"].value_counts().head()
index = a.index
sources = a.values
custom_colors = ["skyblue", "yellowgreen", "tomato", "blue", "red"]
plt.figure(figsize=(10,10))
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=14)
plt.title("Top 5 Domains to Become a Billionaire", fontsize=28)
plt.show()
```



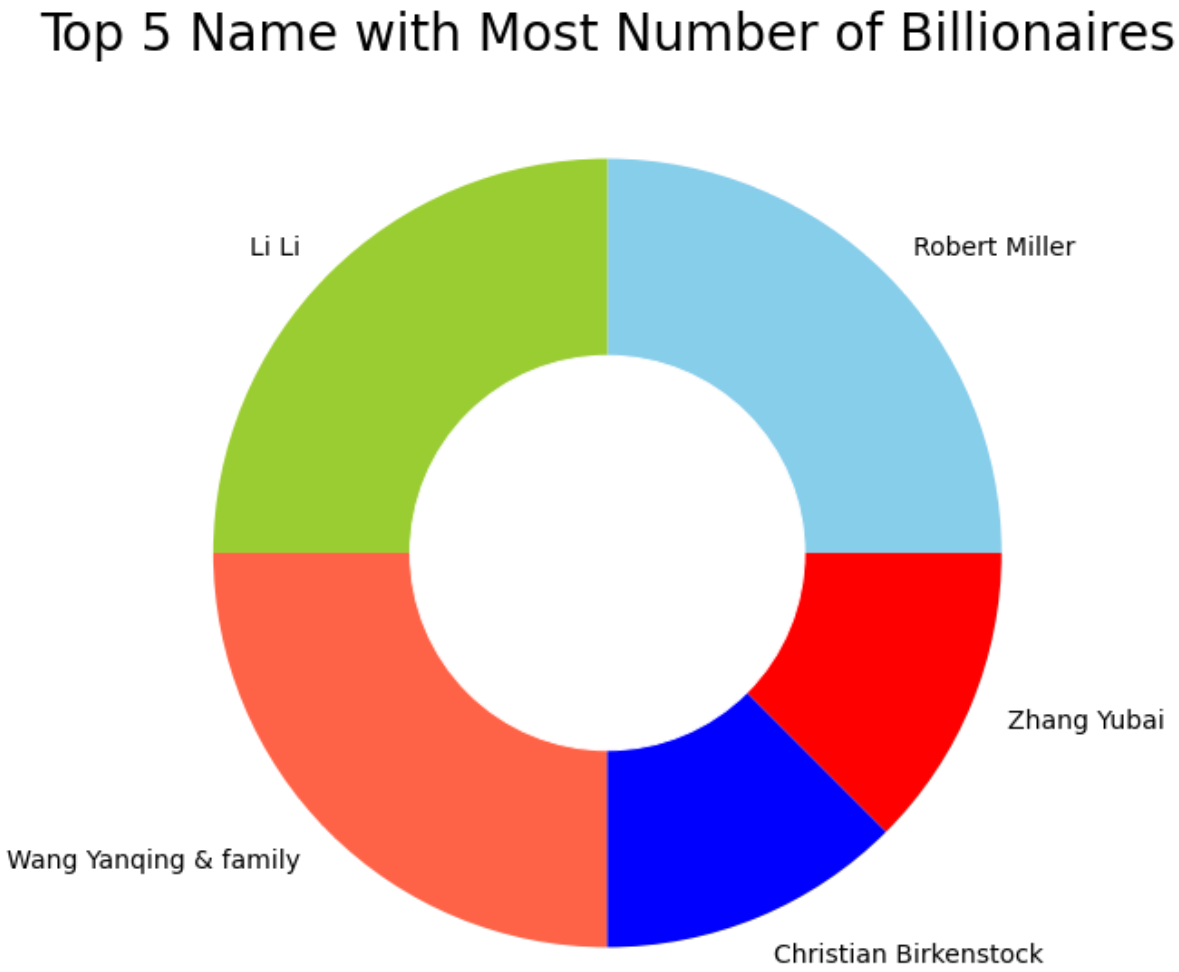
Now let's have a look at the top 5 industries with the most number of billionaires

In [14]:
```
a = data["Industry"].value_counts().head()
index = a.index
industtries = a.values
custom_colors = ["skyblue", "yellowgreen", "tomato", "blue", "red"]
plt.figure(figsize=(10,10))
plt.pie(industtries, 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=14)
plt.title("Top 5 Industries with Most Number of Billionaires", fontsize=28)
plt.show()
```



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

In [15]:
```
a = data["Name"].value_counts().head()
index = a.index
names = a.values
custom_colors = ["skyblue", "yellowgreen", 'tomato', "blue", "red"]
plt.figure(figsize=(10,10))
plt.pie(names, labels=index, colors=custom_colors)
central_circle = plt.Circle((0,0), 0.5, color="white")
fig = plt.gcf()
plt.gca().add_artist(central_circle)
plt.rc("font", size=14)
plt.title("Top 5 Name with Most Number of Billionaires", fontsize=28)
plt.show()
```



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

In [16]:
```
a = data["Country"]. value_counts().head()
index = a.index
countries = a.values
custom_colors = ["skyblue", "yellowgreen", 'tomato', "blue", "red"]
plt.figure(figsize=(10,10))
plt.pie(countries, labels=index, colors=custom_colors)
central_circle = plt.Circle((0,0), 0.5, color="white")
fig = plt.gcf()
plt.gca().add_artist(central_circle)
plt.rc('font', size=14)
plt.title("Top 5 Countries with Most Number of Billionaires", fontsize=28)
plt.show()
```