Billionaires_Analysis_with_Python

In [30]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt


data = pd.read_csv("https://raw.githubusercontent.com/amankharwal/Website-data/master/Billionaire.csv")
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
```

Missing values:

In [32]:
```
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, let's remove these rows:

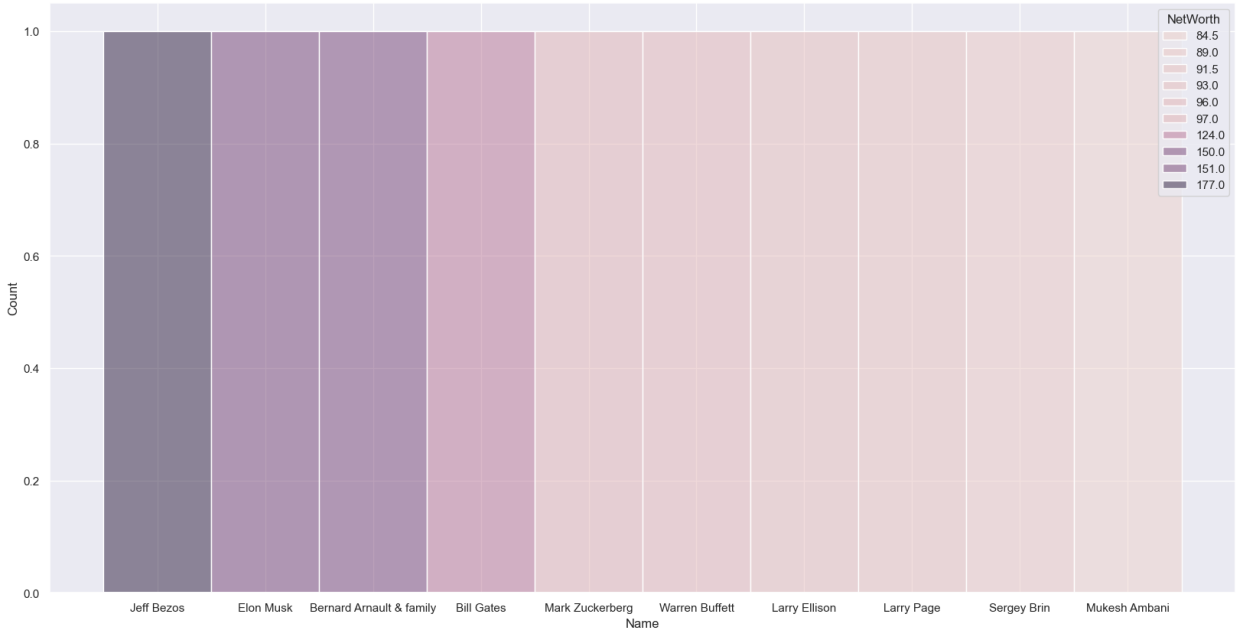
In [33]:
```
data = data.dropna()
```

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 [34]:
```
data["NetWorth"] = data["NetWorth"].str.strip("$")
data["NetWorth"] = data["NetWorth"].str.strip("B")
data["NetWorth"] = data["NetWorth"].astype(float)
```

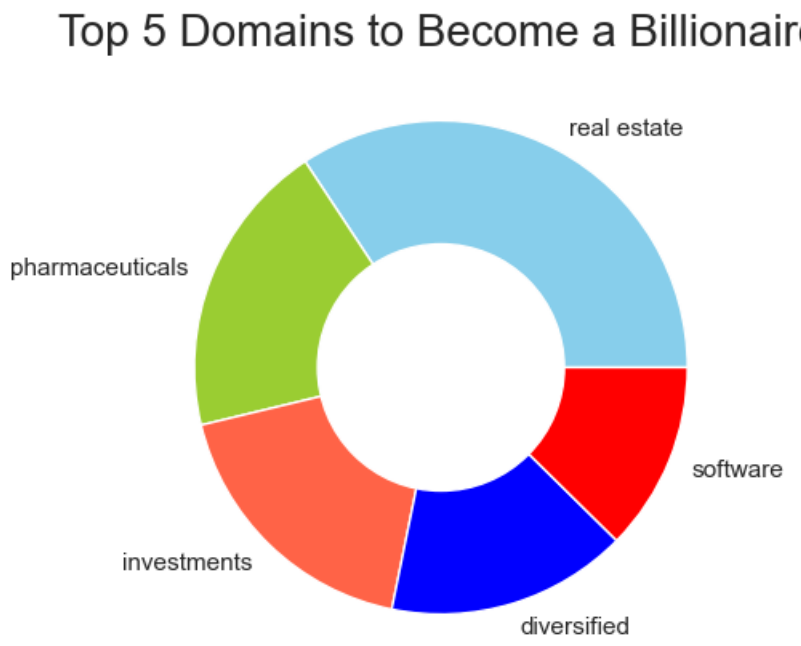
Now let's have a look at the top 10 billionaires according to their NetWorth:

In [35]:
```
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:

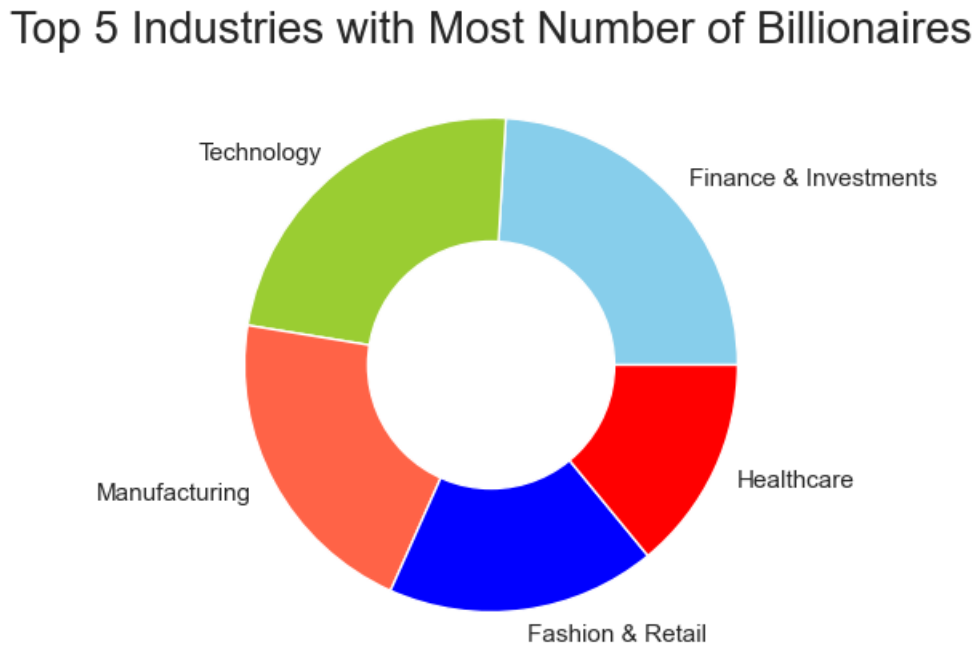
In [36]:
```
a = data["Source"].value_counts().head()
index = a.index
sources = a.values
custom_colors = ["skyblue", "yellowgreen", 'tomato', "blue", "red"]
plt.figure(figsize=(5, 5))
plt.pie(sources, labels=index, colors=custom_colors)
central_circle = plt.Circle((0, 0), 0.5, color='white')
fig = plt.gcf()
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:

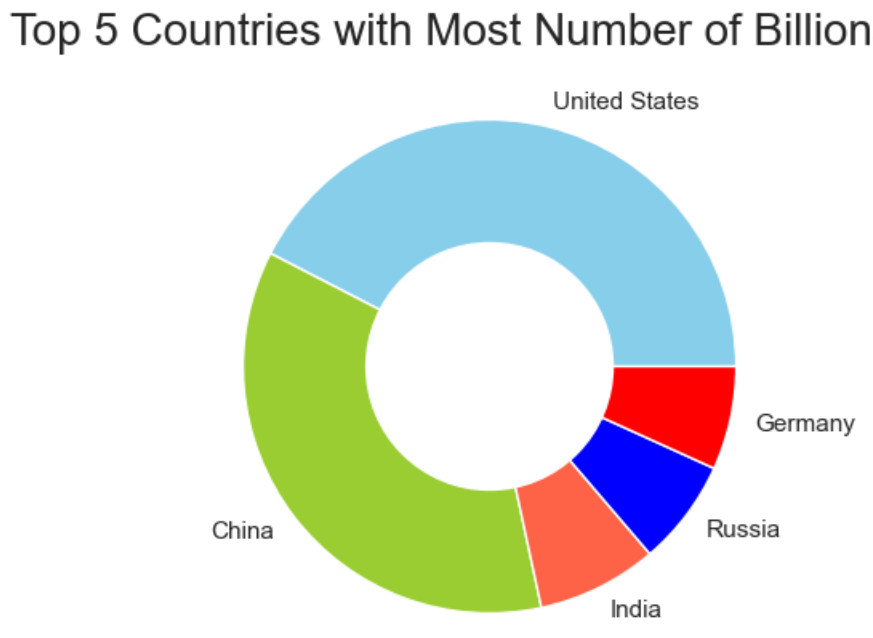
In [37]:
```
a = data["Industry"].value_counts().head()
index = a.index
industries = a.values
custom_colors = ["skyblue", "yellowgreen", 'tomato', "blue", "red"]
plt.figure(figsize=(5, 5))
plt.pie(industries, labels=index, colors=custom_colors)
central_circle = plt.Circle((0, 0), 0.5, color='white')
```

```
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 [38]:
```
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()
```



The visualization above shows that the United States and China are the countries from which most people become billionaires. So that means the business environment and the startup success rate is really good in the US and China

compared to the rest of the world.

THANK YOU!

GitHub Link: https://github.com/anujtiwari21?tab=repositories