Vivega Task1

IMPORT NECESSARY LIBRARIES

In [59]:
```
import pandas as pd
import numpy  as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import warnings
warnings.filterwarnings("ignore")
```

LOAD DATASET

In [60]:
```
train = pd.read_csv('worldpopulationdata.csv')
```

Checking the first 5 rows

In [61]:
```
train.head(5)
```

Vivega Task1

Out[61]:

	Series Name	Series Code	Country Name	Country Code	2022	2021	2020	2019	2018	2017	...	2010	
0	Population, total	SP.POP.TOTL	Afghanistan	AFG	41128771.0	40099462.0	38972230.0	37769499.0	36686784.0	35643418.0	...	28189672.0	273
1	Population, total	SP.POP.TOTL	Albania	ALB	2775634.0	2811666.0	2837849.0	2854191.0	2866376.0	2873457.0	...	2913021.0	29
2	Population, total	SP.POP.TOTL	Algeria	DZA	44903225.0	44177969.0	43451666.0	42705368.0	41927007.0	41136546.0	...	35856344.0	351
3	Population, total	SP.POP.TOTL	American Samoa	ASM	44273.0	45035.0	46189.0	47321.0	48424.0	49463.0	...	54849.0	
4	Population, total	SP.POP.TOTL	Andorra	AND	79824.0	79034.0	77700.0	76343.0	75013.0	73837.0	...	71519.0	

5 rows × 26 columns

Checking the last 5 rows

In [62]:
```
train.tail(5)
```

Vivega Task1

Out[62]:

	Series Name	Series Code	Country Name	Country Code	2022	2021	2020	2019	2018	2017	...	2010	
1080	Population, male (% of total population)	SP.POP.TOTL.MA.ZS	Virgin Islands (U.S.)	VIR	46.613382	46.764444	46.914637	47.057307	47.185912	47.314214	...	47.801059	47
1081	Population, male (% of total population)	SP.POP.TOTL.MA.ZS	West Bank and Gaza	PSE	49.893678	49.877839	49.858957	49.835542	49.811374	49.785969	...	49.876336	49
1082	Population, male (% of total population)	SP.POP.TOTL.MA.ZS	Yemen, Rep.	YEM	50.519031	50.538516	50.554317	50.571320	50.596614	50.616964	...	50.594170	50
1083	Population, male (% of total population)	SP.POP.TOTL.MA.ZS	Zambia	ZMB	49.344602	49.344951	49.338301	49.326233	49.309087	49.288400	...	49.056379	48
1084	Population, male (% of total population)	SP.POP.TOTL.MA.ZS	Zimbabwe	ZWE	47.214139	47.167153	47.130679	47.099796	47.076238	47.051613	...	46.995893	47

5 rows × 26 columns

Checking the shape of the dataset

In [63]:
```
train.shape
```

Out[63]:
```
(1085, 26)
```

Checking the columns of the dataset

In [64]:
```
train.columns
```

Out[64]:
```
Index(['Series Name', 'Series Code', 'Country Name', 'Country Code', '2022',
       '2021', '2020', '2019', '2018', '2017', '2016', '2015', '2014', '2013',
       '2012', '2011', '2010', '2009', '2008', '2007', '2006', '2005', '2004',
       '2003', '2002', '2001'],
      dtype='object')
```

Checking for missing values

In [65]:
```
train.isnull().sum()
```

Out[65]:
```
Series Name     0
Series Code     0
Country Name    0
Country Code    0
2022            0
2021            0
2020            0
2019            0
2018            0
2017            0
2016            0
2015            0
2014            0
2013            0
2012            0
2011            0
2010            0
2009            0
2008            0
2007            0
2006            0
2005            0
2004            0
2003            0
2002            0
2001            0
dtype: int64
```

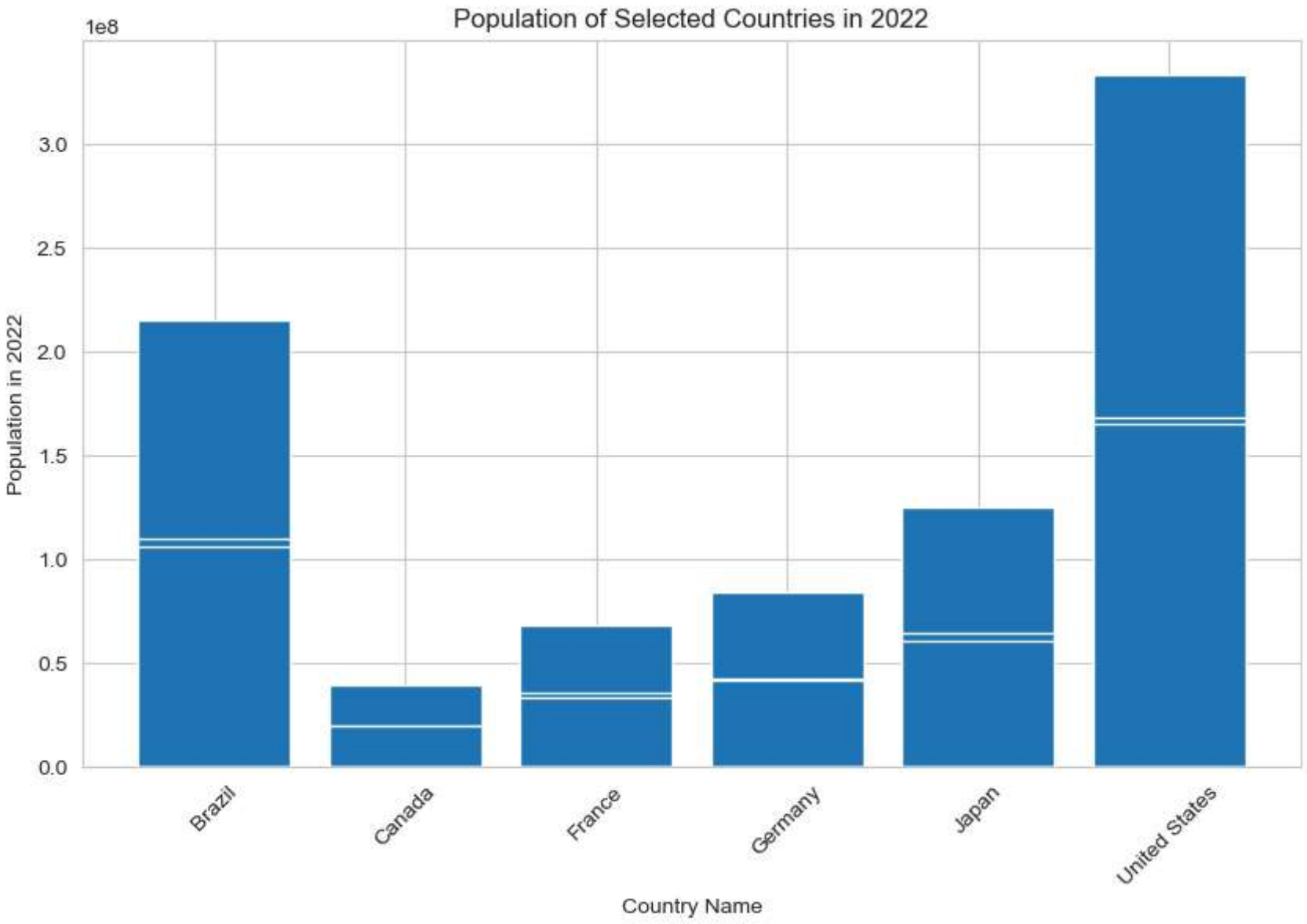
`Observation:`

- There are no missing values in the dataset

Population of Selected Countries in 2022

In [66]:
```
countries_to_plot = ['United States', 'Canada', 'Brazil', 'Japan', 'Germany', 'France']
filtered_data = train[train['Country Name'].isin(countries_to_plot)]
plt.figure(figsize=(10, 6))
plt.bar(filtered_data['Country Name'], filtered_data['2022'])
plt.xlabel('Country Name')
plt.ylabel('Population in 2022')
plt.title('Population of Selected Countries in 2022')
plt.xticks(rotation=45)
plt.show()
```

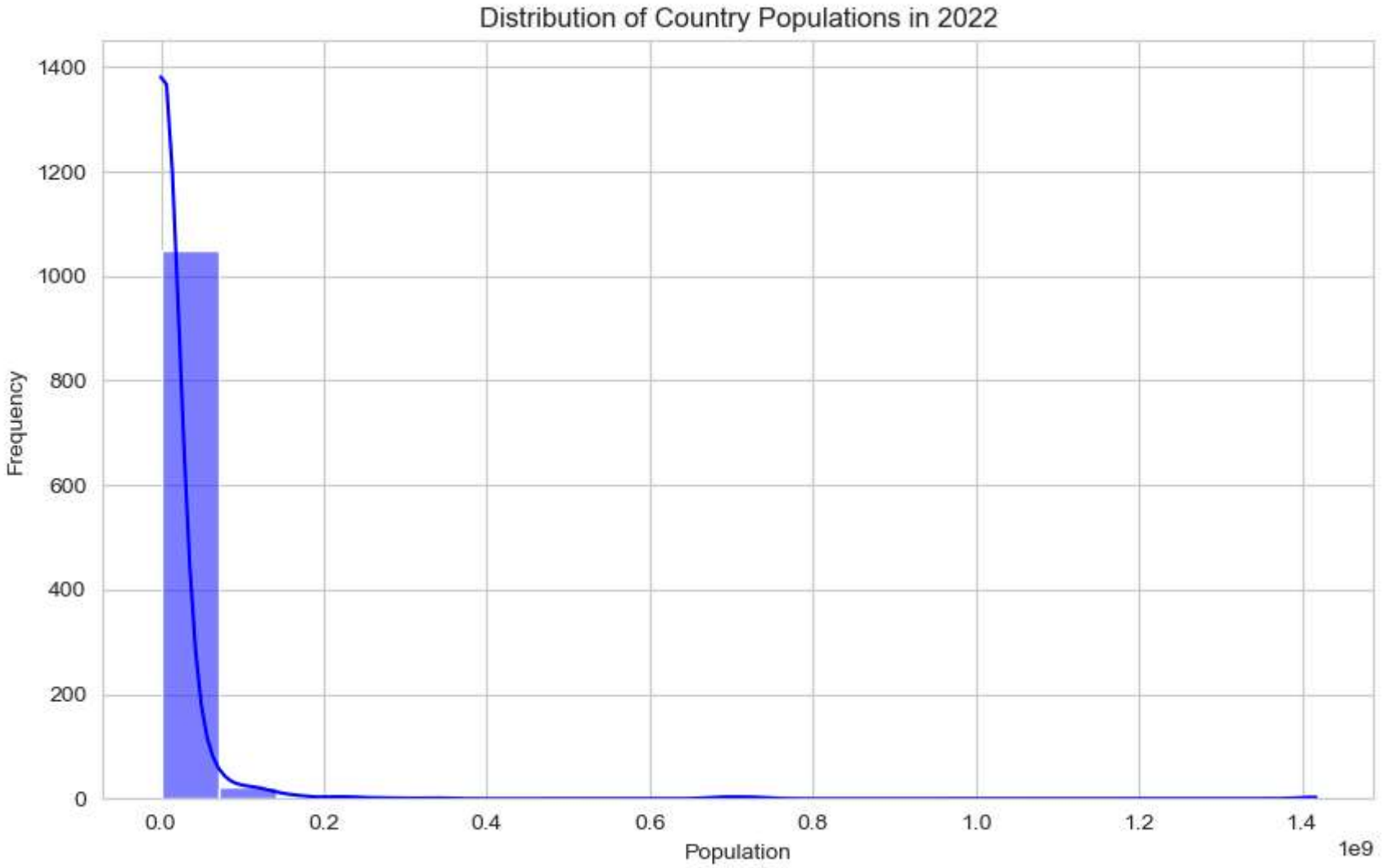
Vivega Task1



Histogram for population distribution in 2022

In [67]:
```
plt.figure(figsize=(10, 6))
sns.histplot(train['2022'].dropna(), bins=20, kde=True, color='blue')
plt.title('Distribution of Country Populations in 2022')
plt.xlabel('Population')
```

```
plt.ylabel('Frequency')
plt.show()
```



Bar chart for population distribution in 2022

In [68]:
```
import matplotlib.pyplot as plt
import seaborn as sns

# Set the aesthetic style of the plots
```

Vivega Task1

```
sns.set_style("whitegrid")

plt.figure(figsize=(15, 10))
sns.barplot(x='Country Name', y='2022', data=train.sort_values('2022', ascending=False).head(20), palette='viridis')
plt.xticks(rotation=90)
plt.title('Population of Different Countries in 2022 (Top 20)')
plt.xlabel('Country')
plt.ylabel('Population')
plt.show()
```

Vivega Task1