Population Data Visualization - Submission Report

Name: Tarun Chincholia

Internship Domain: Data Science

Task Number: Task 1

Organization: Prodigy InfoTech

Date: 20-07-25

Objective:

To visualize the distribution of a continuous variable (population) using a real-world dataset. This task focuses on applying basic data visualization techniques such as histograms and bar charts to gain insights from population data.

Dataset Used:

- Source: World Bank Open Data

- Link: https://data.worldbank.org/indicator/SP.POP.TOTL

- File: API_SP.POP.TOTL_DS2_en_csv_v2_38144.csv

Visualizations Created:

1. Histogram of Population (2022):
2. - Shows the overall distribution of country populations in 2022.
3. - Most countries have populations below 100 million.
4. - A few countries, such as China and India, are outliers with populations exceeding 1 billion.

2. Bar Chart of Top 10 Most Populous Countries (2022):

- Highlights the top 10 countries with the highest populations.

- Provides a clear comparison among the highest population countries.

Technologies Used:

- Python

- Pandas for data handling

- Matplotlib & Seaborn for visualization

Task-1

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

In [ ]:
```
df = pd.read_csv('/content/population.csv',  encoding='utf-8')
df.head()
```

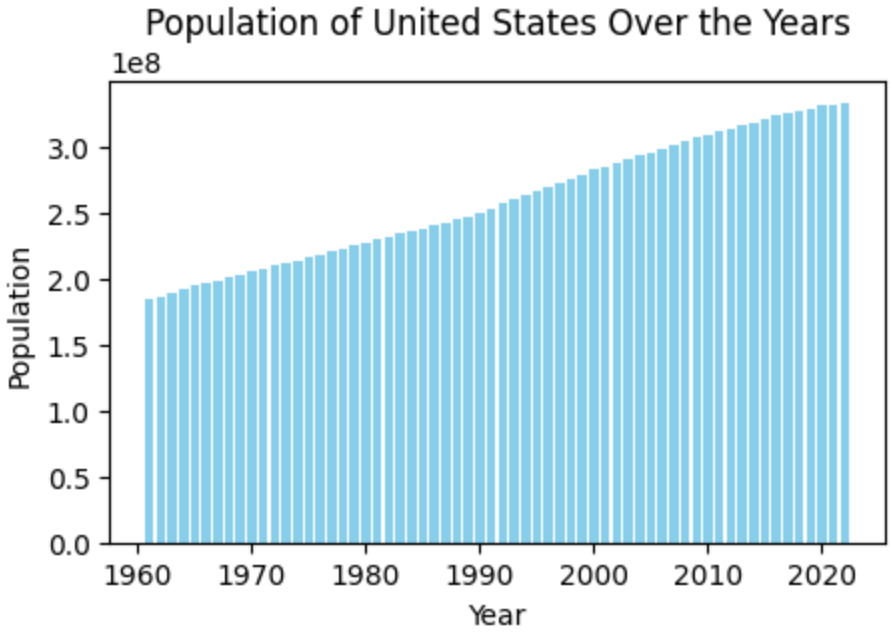
Out[ ]:

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	
0	Aruba	ABW	Population, total	SP.POP.TOTL	54608.0	55811.0	56
1	Africa Eastern and Southern	AFE	Population, total	SP.POP.TOTL	130692579.0	134169237.0	137835
2	Afghanistan	AFG	Population, total	SP.POP.TOTL	8622466.0	8790140.0	8969
3	Africa Western and Central	AFW	Population, total	SP.POP.TOTL	97256290.0	99314028.0	101445
4	Angola	AGO	Population, total	SP.POP.TOTL	5357195.0	5441333.0	5521

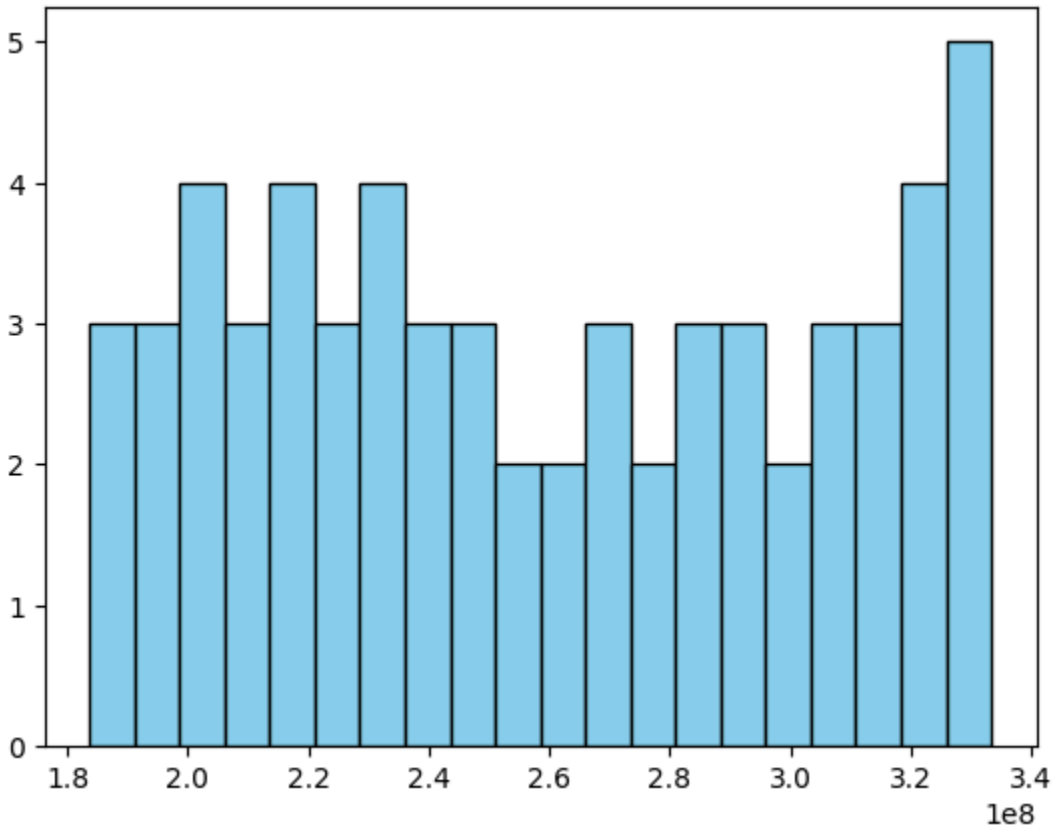
5 rows × 67 columns

In [ ]:
```
df_final = df.drop(['Country Code',        'Indicator Name',        'Indicator
```

In [ ]:
```
chosen_country = 'United States'


country_data = df_final[df_final['Country Name'] == chosen_country]


years = country_data.columns[2:].astype(int)
population = country_data.iloc[:, 2:].values.flatten()

plt.figure(figsize=(5, 3))
plt.bar(years, population, color='skyblue')
plt.xlabel('Year')
plt.ylabel('Population')
plt.title(f'Population of {chosen_country} Over the Years')


plt.show()
```



In [ ]:
```
plt.hist(population, bins=20, color='skyblue', edgecolor='black')
plt.show()
```



In [ ]: