**1. Data Loading and Cleaning**

*data = pd.read\_csv(file\_path)*

*pop\_area\_data = data[['2022 Population', 'Area (km²)']].dropna()*

**Description:** The dataset is loaded using `pandas` from a CSV file, and two key columns, "2022 Population" and "Area (km²)", are extracted. Rows containing any missing values in these columns are dropped to ensure accurate calculations.

**2. Mean Calculation**

*mean\_population = pop\_area\_data['2022 Population'].mean()*

*mean\_area = pop\_area\_data['Area (km²)'].mean()*

Mean: The mean (or average) is calculated for both population and area. It is the sum of all values divided by the number of data points.

Purpose: It provides a central value to summarize the dataset.

Example: The mean population helps to understand the average population size across countries.

**3. Median Calculation**

*median\_population = pop\_area\_data['2022 Population'].median()*

*median\_area = pop\_area\_data['Area (km²)'].median()*

Median: The median is the middle value when the data is sorted in ascending order. If the dataset has an even number of values, the median is the average of the two middle numbers.

Purpose: The median gives a better sense of the "typical" value in cases where data is skewed by outliers.

**4. Variance Calculation**

*variance\_population = pop\_area\_data['2022 Population'].var()*

*variance\_area = pop\_area\_data['Area (km²)'].var()*

Variance: This measures how far the data points are spread out from the mean. It is calculated as the average of the squared differences from the mean.

Purpose: A higher variance means the data is more spread out, while a lower variance indicates that data points are close to the mean.

**5. Standard Deviation Calculation**

*std\_dev\_population = pop\_area\_data['2022 Population'].std()*

*std\_dev\_area = pop\_area\_data['Area (km²)'].std()*

Standard Deviation: This is the square root of the variance. It quantifies the amount of variation or dispersion in the dataset.

Purpose: Like variance, standard deviation provides insight into data spread, but in the same units as the data itself, making it easier to interpret.

**6. Covariance Calculation**

*covariance\_population\_area = pop\_area\_data['2022 Population'].cov(pop\_area\_data['Area (km²)'])*

Covariance: This statistic measures how two variables (in this case, population and area) change together. If both variables tend to increase together, the covariance is positive; if one increases while the other decreases, the covariance is negative.

Purpose: Covariance indicates the direction of the relationship between population and area but not the strength of the relationship.

**7. Correlation Calculation**

*correlation\_population\_area = pop\_area\_data['2022 Population'].corr(pop\_area\_data['Area (km²)'])*

Correlation: Unlike covariance, correlation not only measures the direction of the relationship between two variables but also its strength. It ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no correlation.

Purpose: Correlation helps assess how strongly population size is related to the area of the country.

**8. Visualization**

The then proceeds to generate several visualizations using `matplotlib` and `seaborn`:

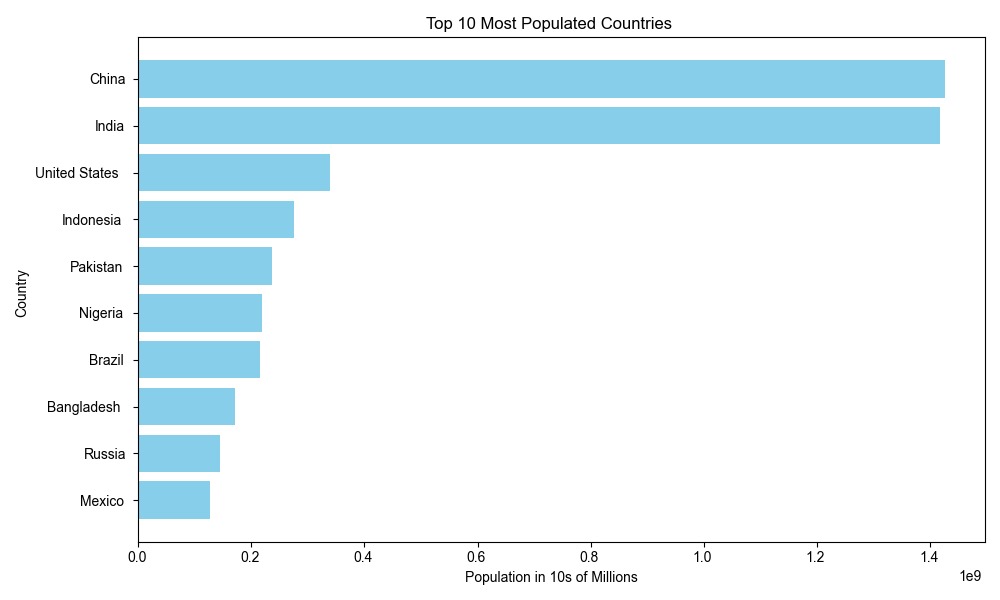
Table of Statistics: A table displays the computed mean, median, variance, standard deviation, covariance, and correlation.

A close-up of a number

Description automatically generated

Top 10 Most Populated Countries: A bar chart ranks countries by their population size.

Histograms: Separate histograms for population and area distributions show how these values are spread across countries.



Scatter Plot: A scatter plot visualizes the relationship between population and area, helping to spot trends or patterns between these two variables.

A graph of different sizes and colors

Description automatically generated with medium confidence

**Summary of Statistical Functions:**

* Mean: The average value of a dataset.
* Median: The middle value in the sorted dataset.
* Variance: The measure of how much the data varies from the mean.
* Standard Deviation: The spread of the data around the mean, in the same units as the data.
* Covariance: Shows how two variables change together (direction of the relationship).
* Correlation: Measures both the strength and direction of the relationship between two variables.