A green shield with gold leaves

Description automatically generated

**Data Analysis and Visualization**

**Project Title:**

**“***Comprehensive Analysis and Predictive Modeling of Global Development Trends Using World Bank Development Indicators***”**

**Submitted By:**

**Muhammad Faisal**

**BSCS202110**

**Submitted To:**

**Muhammad Bilal Hidral**

2024-06-12

**World Bank Development Indicators Exploratory Analysis**

This report summarizes the initial exploration of a World Bank development indicators dataset loaded from Google Drive. The script focuses on understanding the data's structure, identifying missing values, and visualizing key characteristics.

**Data Loading and Selection**

The script utilizes pandas to load the CSV file located at /content/drive/MyDrive/Colab Notebooks/world\_bank\_development\_indicators.csv and filters the data to include these specific columns:

* country
* date
* GDP\_current\_US
* population
* CO2\_emisions
* life\_expectancy\_at\_birth
* forest\_land%
* access\_to\_electricity%
* individuals\_using\_internet%
* inflation\_annual%
* rule\_of\_law\_estimate

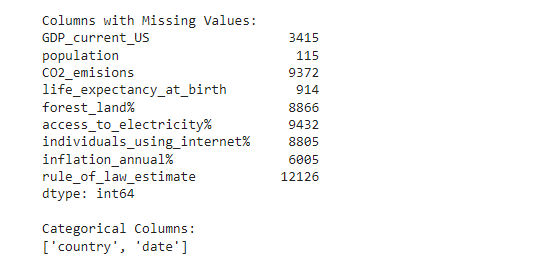
**Data Description**

* The script displays the first five rows to provide a glimpse of the data.
* Column names and data types are listed.
* Basic information about the dataset is presented using df.info().
* Summary statistics are generated using df.describe().

A screenshot of a computer code

Description automatically generated

* Missing value counts are identified for each column.
* Categorical columns are listed.



**Handling Missing Values**

* The script addresses missing values by imputing the mean of each numeric column within each country group.

**Data Visualization**

* Histograms are generated to visualize the distribution of numeric features.

A group of blue and white graphs

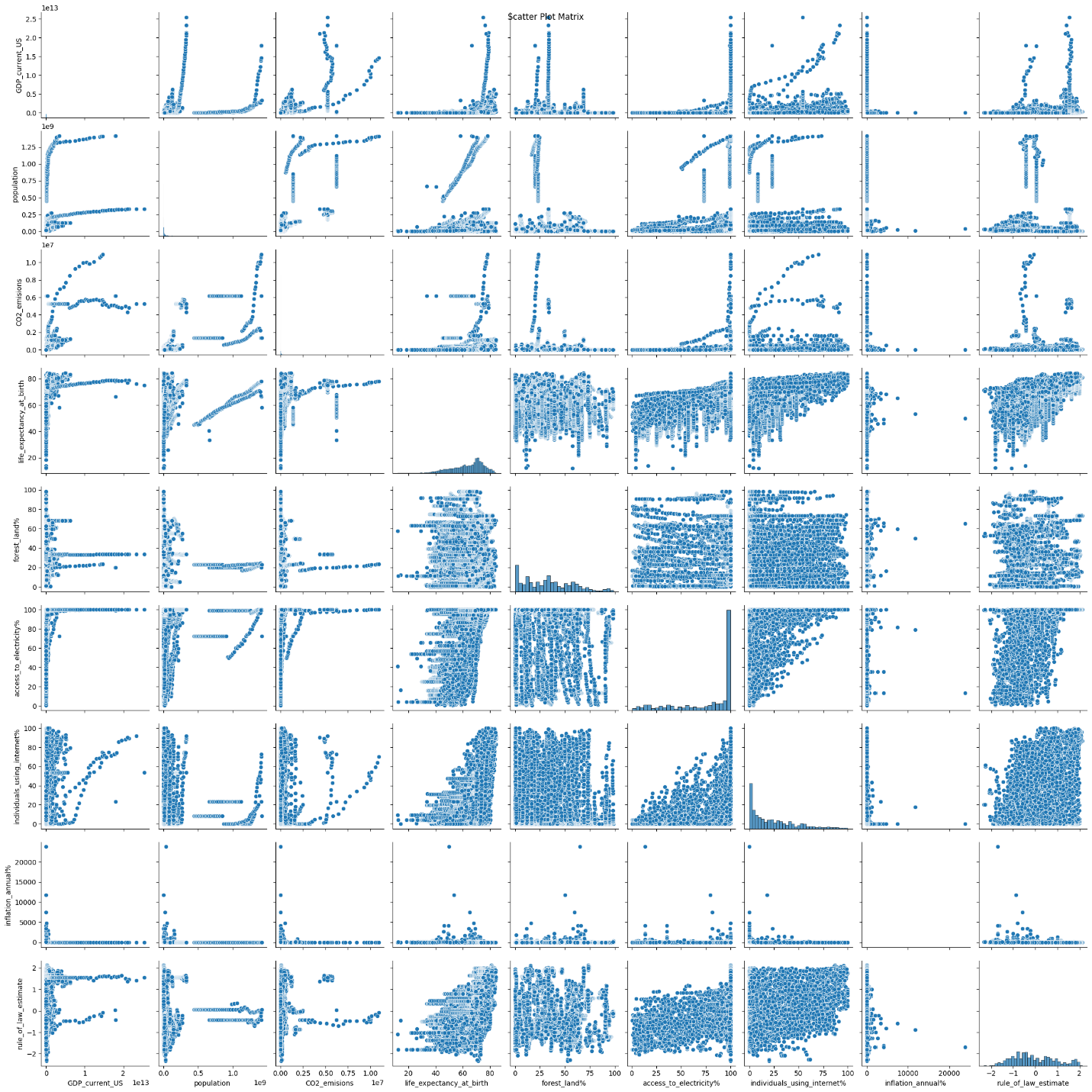
Description automatically generated

* Box plots are created to explore potential outliers and skewness.

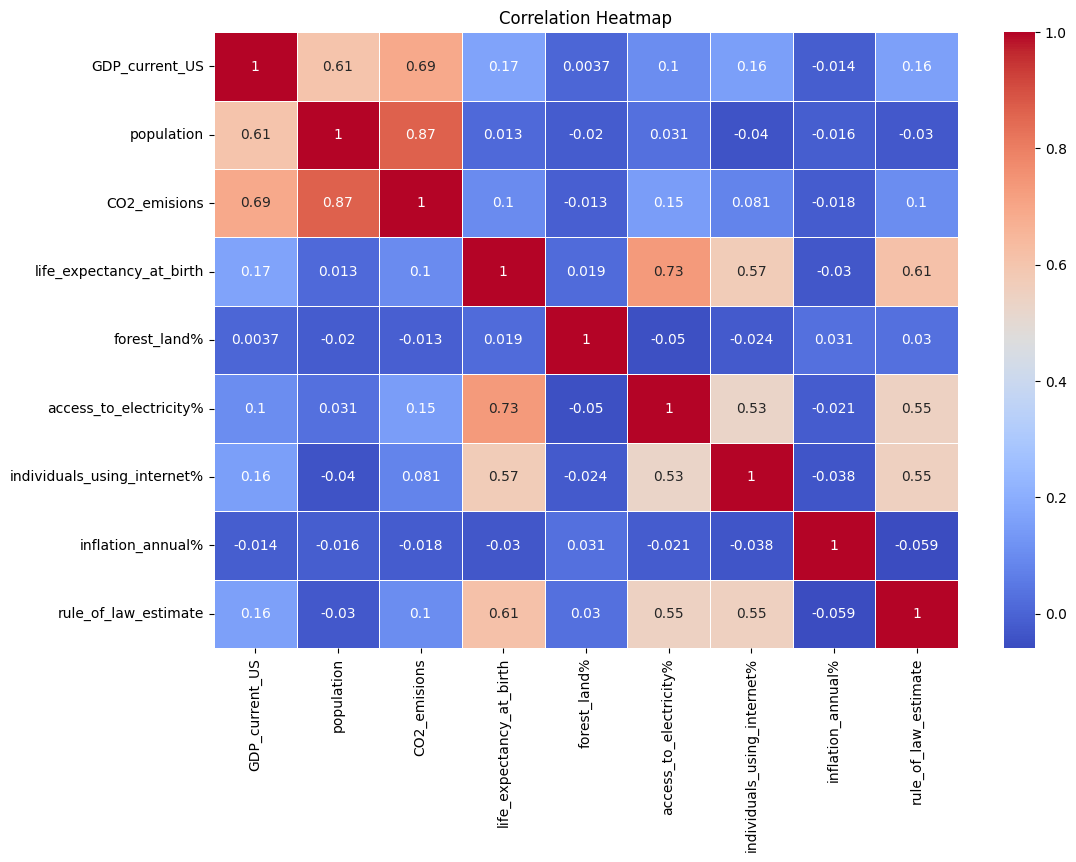
A screenshot of a graph

Description automatically generated

* A scatter plot matrix is used to identify potential relationships between numeric features.



* A correlation heatmap depicts the strength and direction of linear relationships between numeric features.



* Bar charts are produced to visualize the distribution of categorical data.

**Initial Findings**

* The dataset encompasses information on various development indicators across multiple countries and years.
* The data contains a mix of numeric and categorical features.
* Missing values were identified in some columns and addressed by imputation using the country-wise mean for numeric features.
* Exploratory visualizations reveal varying distributions, potential skewness in some numeric features, and potential correlations between variables.

This initial exploration provides a foundational understanding of the World Bank development indicators dataset. Further analysis can delve deeper into specific aspects of development based on the research questions and delve into the relationships between various indicators.

