# Web Scraping Project

# GDP per Capita Analysis

### Analyzing GDP per Capita In today's data-driven world, the ability to harness information from various online sources is invaluable, particularly when it comes to understanding economic indicators like GDP per capita. This web scraping project aims to systematically collect and analyze data related to GDP per capita across different countries.

**Prepared by: M. Dinesh Kumar**

**Role: Data Analyst Intern**

**Date:28/09/2024**

## Contents

* Project Overview
* Tools and Libraries Used
* Data Source
* Web Scraping Process
* Data Cleaning and Storage
* Data Visualization and Plotting
* Results and Insights
* Conclusion

## Project Overview

* **Objective:**

The main objective of this project was to collect economic data related to GDP per capita for different countries from the **Worldometers** website. The data gathered includes both the GDP (PPP) and nominal GDP per capita for the year 2022, along with a comparison to the world average. This information can be useful for analyzing the economic standing of various nations.

* **Importance of the Project:**

Understanding global economic indicators, such as GDP per capita, is crucial for financial analysts, policymakers, and economists. The data collected allows for a deeper understanding of the wealth distribution across countries, enabling analysts to make informed decisions or create reports that could influence economic policy or investment strategies.

## Tools and Libraries Used

In this project, the following tools and Python libraries were used:

* **Requests:**
* **BeautifulSoup:**
* **pandas:**

## Data Source

**Website:**

The data was scraped from the **Worldometers** website, specifically from the page listing GDP per capita for countries around the world. The webpage contains a table that displays the GDP (PPP) per capita, nominal GDP per capita for the year 2022, and a comparison of each country’s GDP to the global average.

**Data Structure:** The table on the website includes the following columns:

1. Country Ranking (#)
2. Country Name
3. GDP (PPP) per capita (2022)
4. GDP (Nominal) per capita (2022)
5. Percentage comparison with World GDP per capita

This data provides an insight into both the purchasing power parity and nominal economic output of countries across the globe. This dataset is crucial for comparing economies on a more equal footing by adjusting for the relative cost of living and inflation rates

## Web Scraping Process

* **Step 1: Fetching the Webpage Content**

The first step was to send an HTTP GET request to the target website using the requests library

* **Step 2: Parsing the HTML with BeautifulSoup**

Once the webpage was fetched, the next step was to parse the HTML content using **BeautifulSoup**

* **Step 3: Extracting the Table Data**

To locate the table containing the GDP data,

We utilized the find method to successfully locate the table tag within the HTML. Once identified, we were able to extract the headings and rows with precision. This meticulous process is greatly appreciated for its efficiency and accuracy.

* **Step 4: Storing the Data**

the extracted data was stored in a pandas DataFrame for further manipulation and analysis:

The data was organized into a structured format and then exported to a CSV file for later use.

## Data Cleaning and Storage

After scraping the raw data from the Worldometers website, several steps were taken to clean and store the data for further analysis.

* **Step 1: Removing Unnecessary Characters**

The raw data contained some unwanted characters like dollar signs ($), commas (,), and extra spaces. These were removed to make the data easier to manipulate and perform numerical operations on.

* **Step 2: Converting Data Types**

After removing the unwanted characters, the GDP columns, which were initially in string format, were converted to numerical data types for proper analysis

* **Step 3:** **Storing the Clean Data**

After cleaning, the final dataset was stored in a CSV file for easy access and sharing:

The cleaned dataset can be used for various types of analysis, and it’s stored in a structured format that is easy to load and use in future projects.

## Data Visualization and Plotting

After cleaning and storing the data, several visualizations were created to better understand the distribution of GDP per capita across different countries. These visualizations provide insights into economic disparities and highlight trends among top and lower-performing countries.

* **Step 1: Importing Visualization Libraries:**

The **Matplotlib** and **Seaborn** libraries were used to generate the plots. These libraries offer powerful tools to visualize the data in a simple and clear manner

* **Step 2: Plotting the Top 10 Countries by GDP (PPP) per Capita:**

To identify the top countries based on their GDP (PPP) per capita, **a bar plot** was generated. This plot helps visualize which countries have the highest purchasing power per capita.

* **Step 3: Comparing GDP (Nominal) per Capita with World Average:**

A **scatter plot** was created to visualize the nominal GDP per capita of different countries and compare it with the world average. This highlights the economic disparities across different nations.

* **Step 4: Distribution of GDP (PPP) per Capita:**

A **histogram** was created to visualize the overall distribution of GDP (PPP) per capita across all countries. This helps understand how GDP is distributed globally.

* **Step 5: GDP Growth Over Time:**

The **line plot** would show the GDP growth trend over time for a specific country, in this case, Luxembourg.

* **Step 6: Frequency of Countries in GDP Ranges:**

**count plot** helps visualize the number of countries that fall within specific ranges of GDP per capita

* **Step 7 : GDP Comparison by Region:**

A **swarm plot** can be useful for visualizing the distribution of GDP values across different regions.

* **Step 8: Detailed Distribution of GDP:**

The **dis plot** (or distribution plot) provides a more detailed view of the distribution of GDP per capita values.

## Results and Insights

After successfully scraping and cleaning the data, several insights were obtained from the GDP per capita figures.

1. **Top Economies by GDP (PPP) per Capita:**

The top countries in terms of GDP (PPP) per capita were **Luxembourg**, **Singapore**, and **Ireland**, with Luxembourg leading the chart at $142,214 per capita.

1. **Significant Differences Between GDP (PPP) and Nominal GDP:**

In some cases, there were large discrepancies between a country's GDP (PPP) and nominal GDP, highlighting differences in the cost of living and local purchasing power.

1. **Global Comparisons:**

In some cases, there were large discrepancies between a country's GDP (PPP) and nominal GDP, highlighting differences in the cost of living and local purchasing power.

## Conclusion

This project demonstrated how to use web scraping to gather real-world data and perform meaningful analysis on global GDP per capita figures. The code and dataset used in this project are available for review and reuse in the following GitHub repository:

**GitHub Repository:**<https://github.com/dineshsibi/DataAnalyst_Internship_Project>