**GDP Data Analysis**

# Context

Gross Domestic Product (GDP) is a measure of the size and health of an economy. It represents the total value of all goods and services produced within a country over a specific period of time, usually a year. GDP is considered a key indicator of a country's economic performance and is used to compare the economic output of different countries.

The GDP is one of the metrics to ensure self-sustained growth for any country. The total monetary value of goods and services flowing through an economy over time is measured by GDP. The goal of our project is to analyse the Patterns inside the taken dataset of multiple countries. This project involves applications of Data Analysis.

# About this project

Understanding of economy and growth of all countries is an essential for all citizens. As we are progressing towards machine learning and artificial intelligence era, it is helping us to understand complex problems having good amount of data. The data analysis and engineering are very essential and critical in modern day world. Data science can provide excellent insight of data using patterns and trends. The primary goal of this project is to investigate the dataset and to focus on the Country's GDP over the years.

# About this dataset

This dataset provides a comprehensive overview of GDP data across all the countries from the past 64 years (from 1960 to 2023), offering insights into analysis of GDP per capita of world and India. There are GDP data of 218 countries. Also contains GDP values for several regions and income groups as well as world bank groups.

Dataset is downloaded from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD.

Here's a brief description of each of the attributes or labels in the dataset:

**CSV File**: API\_NY.GDP.MKTP.CD\_DS2\_en\_csv\_v2\_2.csv

**Rows**: Around 260.

**Timeframe**: Data covers the years 1960 to 2023.

* **Country Name:** Name of the country.
* **Country Code:** Unique identifier of the country. Every record in this dataset starts with a Row ID which helps identifying rows uniquely.
* **Indicator Name and** **Indicator Code:** The unique name for the report category.
* **1960-2023:** The values of GDP in US Dollar is mentioned in the specific columns year-wise.

**CSV File**: Metadata\_Country\_API\_NY.GDP.MKTP.CD\_DS2\_en\_csv\_v2\_2.csv

**Rows**: Around 260.

* **Country Code:** Unique 3-characters identifier of the country. Every record in this dataset starts with a Row ID which helps identifying rows uniquely.
* **Region:** The name of the region is mentioned if the area is not identified as Country or a group of Country.
* **IncomeGroup:** The income groups are categorised as High income, Upper middle income, Lower middle income and Low income.
* **SpecialNotes:** The corresponding description.
* **TableName:** This column contains the list of the countries under a particular Region.

# Data Cleaning & Analysis

Tools Used: Python, Pandas, NumPy, Seaborn, Plotly, Matplotlib

The dataset underwent thorough cleaning and analysis to prepare it for further exploration. Data cleaning is also referred to as data preparation, is a vital step that comprises reformatting the data, making data corrections, and merging data sets to enhance the data.

On the other hand, the Data Exploration is used to explore and visualize data to derive insights from the start or identify patterns to dig deeper.

# Tasks performed

* Imported the csv data into a Panda dataframe for analysis
* Unpivoted the dataset to bring the years in row format (extended upto 17K) for further analysis
* Performed the basic data cleaning activities
* Checked datatypes of columns and dropping unnecessary columns
* Renamed the columns Country Name and Country Code as these two columns have some space in their names.
* Handled missing values
* Korea, Dem. People's Rep. , British Virgin Islands and Gibraltar do not have any GDP value from 1960 to 2023. Apart from that one junk row was identified as ‘Not Classified’.
* Checked for duplicate values and dropping them
* Performed general data exploration and validation checks on the GDP dataset
* Generated table summary statistics counted total number of rows and total number of columns
* Analysed distinct values and distributions for different columns like Income Group, Region etc
* Summarized GDP across the regions and years
* Used Python tools to analyse data for visualization

# Key Findings

* The dataset consists of GDP of 262 countries incl. regions, income groups, world bank groups etc.
* There are 214 unique countries are in the GDP dataset.
* We are handling 64 years of data ranging from 1960 up to 2023.
* There are countries where all the yearly data is not available.
* Number of countries per Region and Income Group –
  + Income Group wise : High income (83), Low income (25), Lower middle income (51), Upper middle income (54).
  + Region wise : East Asia & Pacific (36), Europe & Central Asia (57), Latin America & Caribbean (41), Middle East & North Africa (21), North America (3), South Asia (8), Sub-Saharan Africa (48)
* In 1960, World’s GDP was 1.37 trillion which has reached to 106 trillion in 2023 with a growth of 7638.76% compared to 1960’s GDP.
* The average GDP growth of the World is 7.1% per Anum.
* India’s GDP has recorded to 3.56 trillion in 2023 which was 37 billion in 1960.
* Average GDP growth of India is 7.75%.
* The %age of GDP growth of India in year 2009 during the Great Recession was 11.93.
* In the year of 2020, the GDP growth of the World falls by 2.71% because of COVID-19. After COVID-19 era, the GDP of the World takes a sharp turn and increases by 14.09%.
* For the last two decades (2003 - 2023), India's GDP has increased by 487.06%, whereas increment of World's GDP was 170.97% only.
* Top 10 Countries based on the average GDP growth are Iraq (25.03%), Azerbaijan (18.37%), United Arab Emirates (16.57%), Qatar (15.91%), Oman (15.91%), Brunei Darussalam (14.67%), Equatorial Guinea (13.60%), Saudi Arabia (12.87%), Maldives (12.82%), Viet Nam (12.48%).
* In 1960, top 10 countries based on GDP are United States, Germany, United Kingdom, France, China, Japan, Italy, Canada, India and Australia.
* In 2023, according to GDP, The United States continues to hold the 1st position followed by China. India comes at 5th. Whereas Tuvalu has the smallest GDP with 62.28 million.
* From 1960 to 2023, India has seen some drop in GDP in compared to previous years. Among them 1966 and 1991 are some big drops, though the Great Recession could not make severe impact on in India’s GDP growth.
* At present (in 2023), The United States holds almost 26.1% of world’s GDP, whereas China holds around 16.8%. India has 3.36% of world’s GDP.

# Purpose

As technology is used in every aspect of our lives, the country's economy is no exception. Data science deals with massive amounts of data using modern tools and techniques and enables better decision making, predictive analysis, and pattern discovery. Using data science in GDP analysis enables us to know the factors that are affecting the GDP per capita of various countries. This helps to focus on the areas that help to foster economic development.

# Conclusion

The main point of doing this project are -

* Creative thinking
* Understanding the GDP Dataset
* Use the best tools for working