



Department of Computer Engineering
College of Engineering
Polytechnic University of the Philippines Sta. Mesa



CMPE 40153: Big Data Analysis Final Requirement
Exploratory Data Analysis of World GDP by Country 1960-2022

Submitted by:

DELA PEÑA. JEBE F.

BS CpE 3-2

Submitted to:

EDCEL B. ARTIFICIO

I. Introduction and Purpose of the Analysis

Description:

The project's main objective is to analyze data on the global gross domestic product (GDP) that spans more than six decades, from 1960 to 2022. GDP is a significant economic measure which illustrates a nation's economic performance and condition. It incorporates the complete monetary worth of all commodities and services produced inside a nation's borders over a certain period of time. Global GDP data trends and patterns can be used to get important insights on the stability, development, and growth of economies around the world.

Project Objectives:

1. Total Global GDP Analysis: Analyze the trajectory of world economic development in order to identify the broad patterns influencing it.
2. Annual Global GDP Growth Rates: Discover the yearly variations in the global economy's growth, identifying critical junctures of expansion or decline.
3. Top 5 Countries' Economic Snapshot: Examine the economic performance of the top 5 nations with the largest GDP during the previous 5 years, emphasizing leaders in the global economic scene.
4. Global GDP Distribution (2022): Identify the economic powerhouses and rising markets by examining the GDP distribution among the top 10 countries, the Philippines, and other countries in 2022.
5. Philippine Economic Growth (1960–2022): Examine the GDP growth rates for the Philippines over a period of six decades, identifying prevailing historical economic trends and their underlying causes.
6. Philippine GDP During Presidencies: Investigate the GDP growth rates during several Philippine presidential periods to learn more about how the country's economic policies have affected its prosperity.

Key Questions:

1. Global Economic Trends: From 1960 through 2022, how has the global GDP evolved?

2. Annual Global Growth Fluctuations: What were the important economic events that affected the rates of global GDP growth? How do current happenings match up with past events?
3. Top 5 Global Economies: What economic variables caused certain nations to constantly rank among the top 5 economies worldwide? Are their growth approaches comparable to one another?
4. GDP Distribution in 2022: How will the top ten countries, the Philippines, and other countries be distributed in terms of global GDP in 2022? What conclusions concerning the economic dynamics of emerging markets may be drawn?
5. Philippine Economic Scene: How has the GDP of the Philippines increased from 1960 to 2022? What were the main factors promoting growth, and were there any noticeable changes in the economy?
6. Presidential Impact on Philippine GDP: How did the GDP of the Philippines change under the several presidents who governed on that period? How can these insights help us comprehend good economic governance, and which policies or events are associated with economic development or decline?

II. Data Dictionary

Overview of the Dataset:

This dataset offers an in-depth analysis of Gross Domestic Product (GDP) data from 1960 through 2022 for various nations and classifications. GDP figures are expressed in current U.S. dollars and represent the entire gross value added by domestic producers to the national economy of each nation.

Dataset Specifications:

Number of Rows: 266 Rows

The dataset consists of one row for each represented country and classification, resulting in a total count equivalent to the number of unique countries and classifications featured in the dataset.

Number of Columns: 65 columns, including:

1. Country: The name of the respective country or classification, providing a comprehensive overview of GDP for diverse entities.
Data Type: String
Example: "United States," "China," "Low Income," "High Income"
2. Country Code: A unique 3-letter code assigned to each country or classification, ensuring international data standardization.
Data Type: String
Example: "USA," "CHN," "LIC" (Low Income Country)
3. 1960 to 2022: Columns representing the GDP of the respective country or classification for each year.
Data Type: Float (GDP values in current U.S. dollars)
Example: 330442815.3

III. Analysis Process

Overview of Exploratory Data Analysis (EDA) Approach:

The methodology used for the World GDP by Country dataset (1960–2022)'s Exploratory Data Analysis (EDA) was established with the goal of ensuring an in-depth understanding of the data, efficient data cleaning, effective data visualization, and insightful insights and analysis. The following crucial steps were part of the EDA process:

Step 1: Understanding the Data:

Data Familiarization: Having a thorough understanding of the rows, columns and data types which composed the dataset.

Domain knowledge: Applying knowledge of global economics to assess the relevance of GDP numbers, and economic trends.

df.head()

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv("C:/Users/Jebe Dela Peña/Downloads/GDP.csv")
```

```
In [2]: #UNDERSTANDING THE DATA
```

```
df.head()
```

```
Out[2]:
```

	Country	Country Code	1960	1961	1962	1963	1964	1965	1966	1967	...	2013	
0	Aruba	ABW	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	2.727933e+09	2
1	Africa Eastern and Southern	AFE	2.112502e+10	2.161623e+10	2.350628e+10	2.804836e+10	2.592067e+10	2.947210e+10	3.201437e+10	3.326951e+10	...	9.859870e+11	1
2	Afghanistan	AFG	5.377778e+08	5.488889e+08	5.466667e+08	7.511112e+08	8.000000e+08	1.006667e+09	1.400000e+09	1.673333e+09	...	2.056449e+10	2
3	Africa Western and Central	AFW	1.044764e+10	1.117321e+10	1.199053e+10	1.272769e+10	1.389811e+10	1.492979e+10	1.591084e+10	1.451058e+10	...	8.339480e+11	8
4	Angola	AGO	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1.334020e+11	1

5 rows x 65 columns

df.tail()

```
In [3]: df.tail()
```

```
Out[3]:
```

	Country	Country Code	1960	1961	1962	1963	1964	1965	1966	1967	...	2013	
261	Kosovo	XKX	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	6.735328e+09	1
262	Yemen, Rep.	YEM	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	4.041523e+10	1
263	South Africa	ZAF	8.748597e+09	9.225996e+09	9.813996e+09	1.085420e+10	1.195600e+10	1.306899e+10	1.421139e+10	1.582139e+10	...	4.008860e+11	1
264	Zambia	ZMB	7.130000e+08	6.962857e+08	6.931429e+08	7.187143e+08	8.394286e+08	1.082857e+09	1.264286e+09	1.368000e+09	...	2.803724e+10	1
265	Zimbabwe	ZWE	1.052990e+09	1.096647e+09	1.117602e+09	1.159512e+09	1.217138e+09	1.311436e+09	1.281750e+09	1.397002e+09	...	1.909102e+10	1

5 rows x 65 columns

df.shape

```
In [4]: df.shape
```

```
Out[4]: (266, 65)
```

df.columns

```
In [5]: df.columns
```

```
Out[5]: Index(['Country', 'Country Code', '1960', '1961', '1962', '1963', '1964',
               '1965', '1966', '1967', '1968', '1969', '1970', '1971', '1972', '1973',
               '1974', '1975', '1976', '1977', '1978', '1979', '1980', '1981', '1982',
               '1983', '1984', '1985', '1986', '1987', '1988', '1989', '1990', '1991',
               '1992', '1993', '1994', '1995', '1996', '1997', '1998', '1999', '2000',
               '2001', '2002', '2003', '2004', '2005', '2006', '2007', '2008', '2009',
               '2010', '2011', '2012', '2013', '2014', '2015', '2016', '2017', '2018',
               '2019', '2020', '2021', '2022'],
              dtype='object')
```

`df.nunique()`

```
In [6]: df.nunique()

Out[6]: Country          266
Country Code          266
1960                   130
1961                   132
1962                   134
...
2018                   256
2019                   255
2020                   255
2021                   249
2022                   231
Length: 65, dtype: int64
```

`df['2022'].unique()`

```
In [7]: df['2022'].unique()

Out[7]: array([          nan, 1.16948000e+12, 8.77863000e+11, 1.06714000e+11,
 1.88820955e+10, 3.35203274e+09, 3.55756000e+12, 5.07535000e+11,
 6.32770000e+11, 1.95027840e+10, 1.75760370e+09, 1.67542000e+12,
 4.71400000e+11, 7.87210588e+10, 3.07341468e+09, 5.78604000e+11,
 1.74017463e+10, 1.88846196e+10, 4.60201000e+11, 8.90403984e+10,
 4.43908205e+10, 1.28974000e+10, 2.45275073e+10, 7.27934576e+10,
 2.82408184e+09, 7.55050000e+09, 4.30688857e+10, 1.92010000e+12,
 5.63791452e+09, 1.66815316e+10, 2.03523222e+10, 2.38261862e+09])
```

Step 2: Data Cleaning:

1. Identifying Columns with Missing Values:

Objective: Identify columns in the dataset where data is missing.

Rationale: The purpose of this phase is to make sure that any data gaps are known, enabling knowledgeable choices about how to deal with missing values.

```
In [8]: # DATA CLEANING

# Identify columns with missing values
df.isnull().sum()
```

```
Out[8]: Country          0
Country Code          0
1960                 134
1961                 132
1962                 130
...
2018                  8
2019                  9
2020                  9
2021                 15
2022                 33
Length: 65, dtype: int64
```

2. Identifying and Removing Duplicate Rows:

Objective: Identify rows that are exact duplicates and remove them.

Rationale: Eliminating redundant entries maintains data integrity and prevents skewed analysis caused by repeated data.

```
In [9]: # Identify and remove duplicate rows
df = df.drop_duplicates()
```

3. Dropping Rows with All Missing Year Values:

Objective: Remove rows where all GDP values from 1960 to 2022 are missing. Rationale:

There is insufficient data for analysis in rows with missing GDP numbers for all years. A clean dataset for precise insights is ensured by removing these rows.

```
In [10]: # Drop rows where all values in year columns are missing
years = [str(year) for year in range(1960, 2023)]

df = df.dropna(subset=years, how='all')
```

4. Removing Non-Country Entries:

Objective: Exclude non-country entries, such as continents and other classifications.

Rationale: Focusing solely on country-specific data is essential for meaningful analysis. Removing non-country entries ensures the analysis pertains directly to individual nations, providing relevant insights.

```
In [12]: # Remove all non-country entries (Continents and Other Classifications)
col_to_remove = ['World', 'Africa Eastern and Southern', 'Africa Western and Central', 'Arab World', 'Central Europe and the Balti
'Channel Islands', 'Caribbean small states', 'East Asia & Pacific (excluding high income)', 'Early-demographic di
'East Asia & Pacific', 'Europe & Central Asia (excluding high income)', 'Europe & Central Asia', 'Arab Rep.', 'Eu
'European Union', 'Fragile and conflict affected situations', 'High income', 'Heavily indebted poor countries (H)
'IBRD only', 'IDA & IBRD total', 'IDA total', 'IDA blend', 'IDA only', 'Not classified', 'Latin America & Caribbe
'Latin America & Caribbean', 'Least developed countries: UN classification', 'Low income', 'Lower middle income',
'Late-demographic dividend', 'Middle East & North Africa', 'Middle income', 'Middle East & North Africa (excludir
'North America', 'OECD members', 'Other small states', 'Pre-demographic dividend', 'West Bank and Gaza', 'Pacific
'Post-demographic dividend', 'South Asia', 'Sub-Saharan Africa (excluding high income)', 'Sub-Saharan Africa', 'S
'East Asia & Pacific (IDA & IBRD countries)', 'Europe & Central Asia (IDA & IBRD countries)', 'Latin America & th
'Middle East & North Africa (IDA & IBRD countries)', 'South Asia (IDA & IBRD)', 'Sub-Saharan Africa (IDA & IBRD <

df_countries = df[~df['Country'].isin(col_to_remove)]
```

Step 3: Data Visualization:

1.Total Global GDP Trend: A comprehensive picture of the global economy is provided by using line charts to illustrate the trend in total global GDP over time.

2.Global GDP yearly Growth Rate: By creating line charts to show the global GDP's yearly growth rate, it is possible to spot trends in the direction of economic development.

3.Top 5 Countries with greatest GDP: Bar charts are used to display the top 5 nations with the greatest GDP over the past five years, allowing for simple comparison.

4.GDP Distribution in 2022: A pie chart illustrating the GDP distribution in 2022 was created, showing the contributions of the top nations, the Philippines, and other nations.

5.Annual Growth Rate of GDP in the Philippines: By displaying the growth rate of the country's GDP over time, it is possible to gain insights into changing economic conditions.

6.Presidential Term Highlights: Using a bar chart to illustrate the GDP growth rate throughout several terms of presidency allowed for a thorough examination of how the economy fared under various administrations.

Step 4: Analysis and Insights

Approach: To identify significant economic movements, the study concentrates on Annual Growth Rates (AGR). In-depth examination of certain years with substantial growth or decrease sheds information on historical settings and underlying variables influencing economic stability and development.

Rationale: Recognizing economic changes provides crucial historical perspective. Investigating the factors that led to these changes yields insightful information that can be used to predict future economic trends and design efficient solutions.

Statistical Techniques:

1. Converting GDP to Billions:

Objective: Enhance readability by presenting GDP values in billions of USD.

Technique: Simple mathematical operation to divide GDP values by 1,000,000,000.

2. Calculating Annual Growth Rate (AGR):

Objective: Understand the growth trends of global GDP and GDP in the Philippines over time.

Technique: AGR calculation using the formula:

$$AGR = \frac{\text{Current Year GDP} - \text{Previous Year GDP}}{\text{Previous Year GDP}} \times 100$$

3. Aggregating Top Countries and Others:

Objective: Analyze the GDP distribution focusing on top countries and the rest of the world.

Technique: Aggregating top countries and calculating GDP for other countries.

Graphical Techniques:

1. Line Chart:

Objective: Visualize trends in global GDP, annual growth rates, and Philippines' GDP over the years.

Rationale: Line charts are effective for showing trends and variations over continuous time periods.

2. Bar Chart:

Objective: Compare GDP values across different countries and presidents' terms in the Philippines.

Rationale: Bar charts provide clear comparisons between categories.

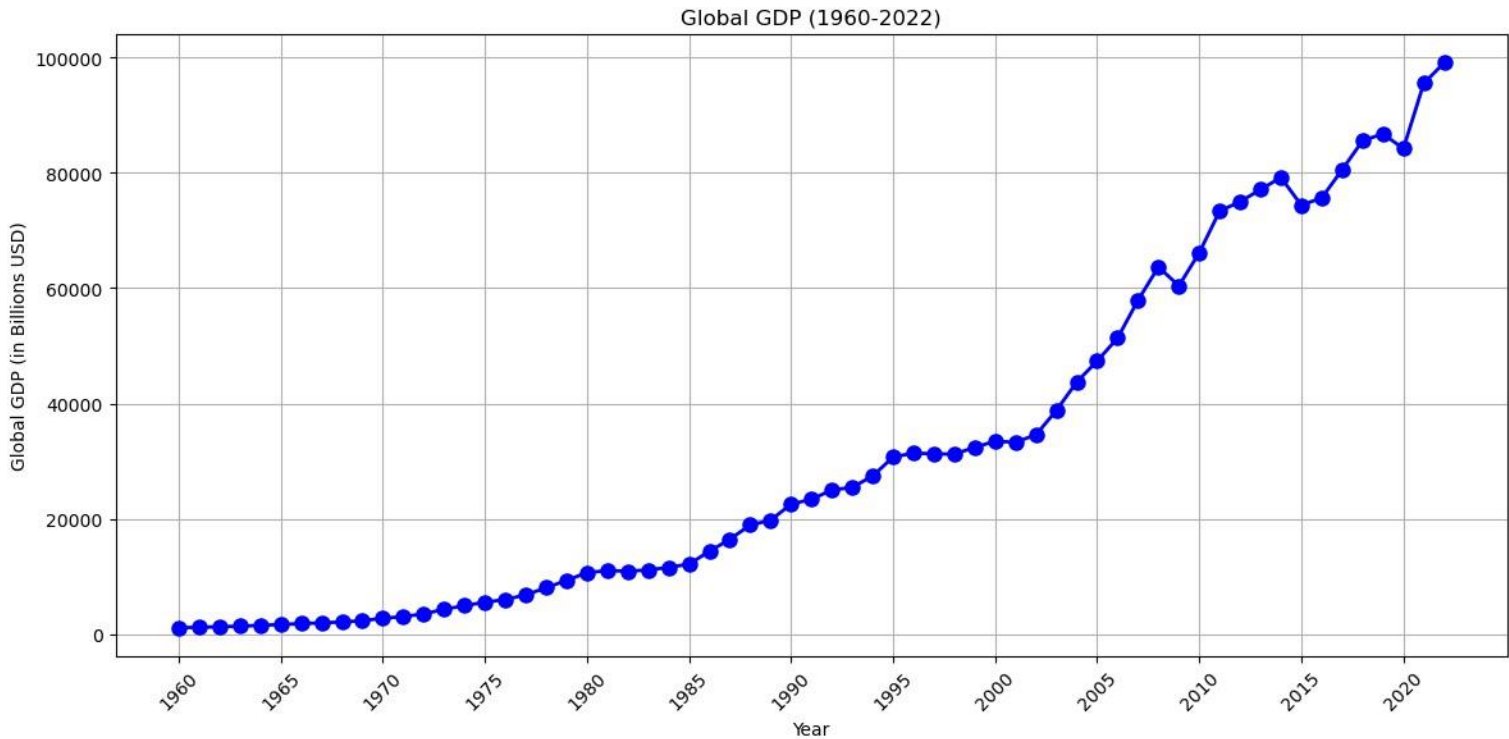
3. Pie Chart:

Objective: Illustrate the proportion of GDP distribution in 2022 among top countries, the Philippines, and other countries.

Rationale: Pie charts display parts of a whole and are useful for showing proportions.

IV. Analysis and Insights

1 - Global GDP over the years



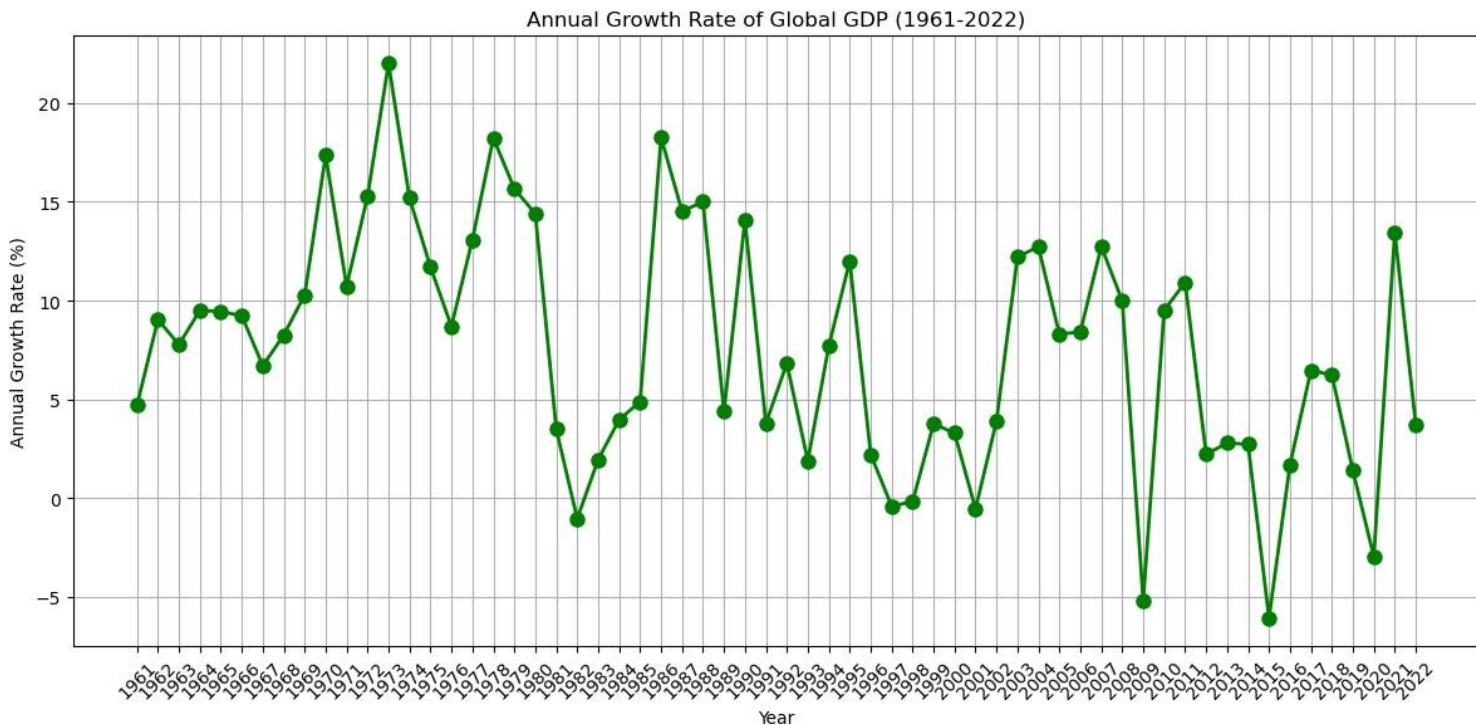
Analysis

- The chart showing global GDP from 1960 to 2022 tells a fascinating description of how the economy has developed through the years.
- The steady upward trajectory, which highlights the overall expansion of the global economy, is a notable aspect.
- In a normally increasing trend, periods of drop stand out as minor setbacks, particularly during economic recessions and global crises.

Insights

- *Continuous Economic Growth:* The steady upward movement represents the resilience of the world economy, illustrating its capacity to overcome obstacles and adjust to ever-shifting circumstances.
- *Cyclical Nature:* Economic downturns, while concerning, are part of the cyclical nature of economies. For policymakers to execute targeted actions during recessions and take advantage of periods of fast expansion, they must have a thorough understanding of these cycles.
- *Socioeconomic Progress:* The overall rising trend is driven by advances in technology, education, and healthcare as well as economic growth, signaling broader advancements in society.

2 – Annual Growth Rate of Global GDP



Analysis

- The Annual Growth Rate of Global GDP chart shows significant changes in the rate of global economic expansion over time.
- Significant rises can be seen in 1973, 1978, 1986, and 2001.
- Notable decreases in 1981, 2009, 2015, and 2020 signify periods of economic contraction.

Insights

2009 Economic Downturn - The Great Recession:

- Housing Market Bubble Burst: The crisis of 2008–2009, also referred to as "The Great Recession," was brought on by the American housing market bubble's collapse. Financial institutions had made significant investments in securities backed by mortgages that contained loans with high risk. Financial institutions suffered extraordinary losses as a significant portion of these loans went bad, which led to a series of collapses.

2015 Economic Decline - "Demographic Cliff" and Global GDP Fall:

- Demographic Factors: In 2015, the global economy was significantly hampered by demographic changes. In many nations, aging populations resulted in lower labor force participation and higher dependence ratios, which had an effect on economic production.

Cyclical Nature: Economic downturns, while concerning, are part of the cyclical nature of economies. For policymakers to execute targeted actions during recessions and take advantage of periods of fast expansion, they must have a thorough understanding of these cycles.

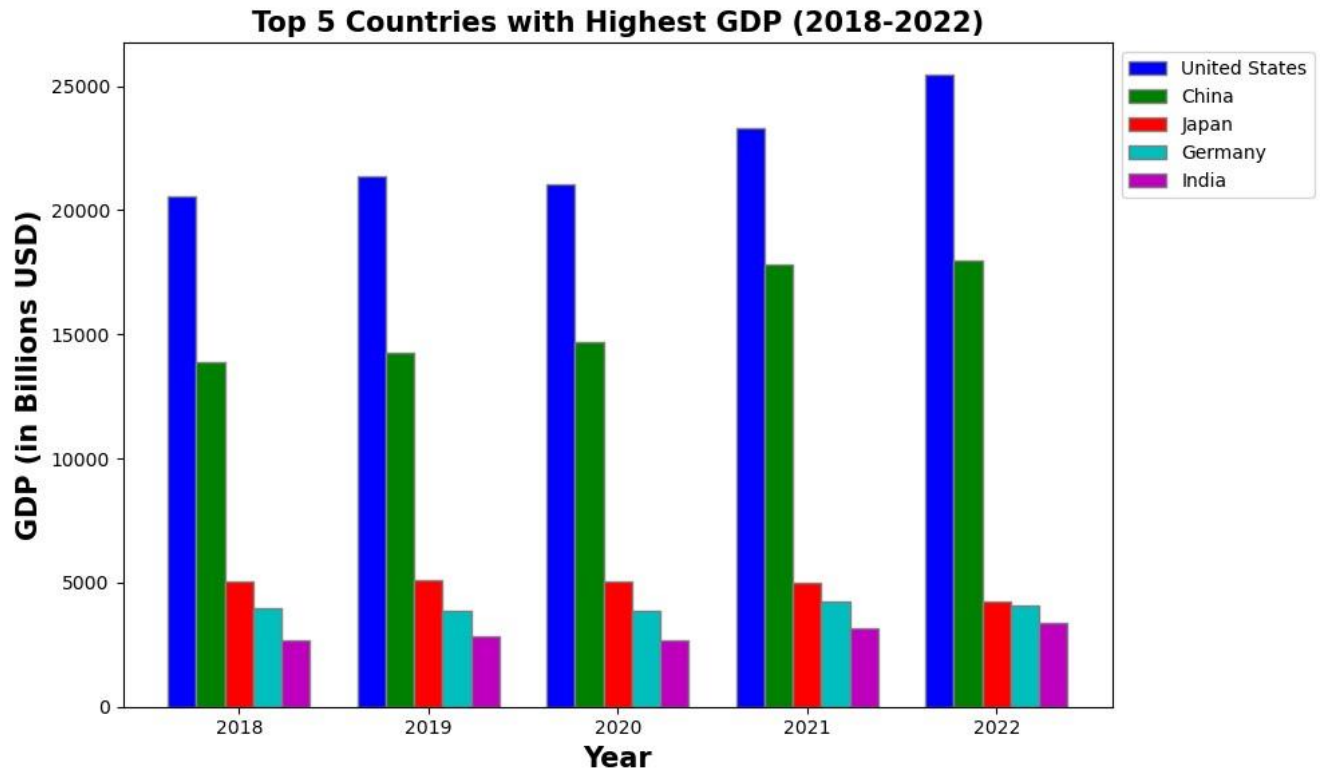
2020 - The COVID-19 Pandemic's Impacts:

- Unprecedented Disruptions: The COVID-19 pandemic caused unmatched disruptions to the world economy in 2020. Lockdowns, travel restrictions, and supply chain disruptions had a significant negative impact on economic activity in all sectors. Widespread GDP declines contributed to one of the most difficult economic downturns in recent history.

2021- Vaccination – Economic Rebound:

- Despite the difficulties, aggressive initiatives including stimulus packages and immunization campaigns helped the economy significantly rebound in 2021. In countries that progressively restarted economic activity and upped vaccination rates in response to the new normal, GDP estimates recovered.

3 - Top 5 Countries with Highest GDP for the last 5 years



Analysis

- The graph showing the GDP of the top five nations over the past five years—the United States, China, Japan, Germany, and India—displays severe economic differences.

Insights

United States:

- The United States has a varied economy that is supported by technological innovation and entrepreneurship. Some of the biggest technology corporations in the world call it home, greatly boosting its GDP. Economic growth is fueled by the existence of Silicon Valley and thriving financial markets.

China:

- **Manufacturing and Export Powerhouse:** The world's largest manufacturing sector is the foundation of China's economic success. It has a substantial industrial base and a vast network of facilities, making it a hub for exports on a global scale. China's major investments in technology and infrastructure have accelerated its economic development.

Japan:

- Technological Innovations and an Export-Oriented Economy: Japan is well known for its technological innovations, particularly in the fields of electronics and autos. Japan retains a strong export-oriented economy despite an aging population, with its industrial and technology sectors considerably contributing to its GDP.

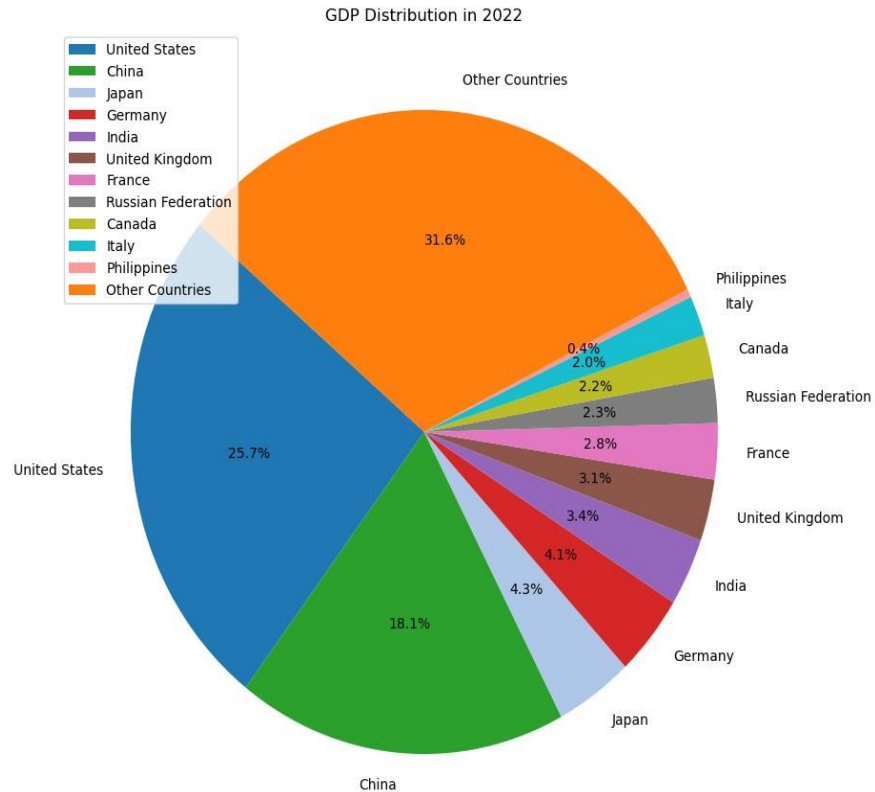
Germany:

- Exports and Engineering Excellence: Germany is known for its superior manufacturing and engineering capabilities. It leads the world in the export of vehicles, equipment, and business supplies. The nation's economy is driven by a talented workforce and a significant emphasis on research and development.

India:

- Service-Oriented Economy and IT Hub: India's service sector, notably information technology (IT) and software services, is the country's economic growth engine. The nation provides IT services to large international organizations and is a popular outsourcing location. The size of India's domestic market also contributes to the expansion of its GDP.

4 - GDP Distribution in 2022 - Top 10 Countries, Philippines and the entire world



Analysis

- The top 10 economies, the Philippines, and the rest of the world are clearly illustrated by the GDP distribution in 2022, which is represented by a pie chart.
- United States Dominance: With a 25.7 percent share of the global economy, the United States continues to be the world's economic leader.
- China's Significant Role: China comes in second place with a share of 18.1%, showing its significant economic influence as a global trading and manufacturing hub.
- Diverse European Contributions: With a combined 12.2 percent contribution, European nations including Germany, the United Kingdom, France, and Italy highlight the continent's economic diversity.
- Philippines Contribution: Despite being considerably less significant, the Philippines' 0.4% contribution still reflects its position within the global economic system.
- Other Countries: The category "Other Countries" accounts for 31.6 percent of the total GDP, highlighting the economic output of several countries besides the top 10 and the Philippines.

Insights

Global Economic Powerhouses:

- US-China Hegemony: With a combined share of 43.8 percent, the United States and China are once again the world's two most powerful economies. Their economic decisions, trade relations, and policies have a big impact on international markets and global economic trends.
- European Influence: The significant contributions from Germany, the United Kingdom, France, and Italy demonstrate the region's economic sway. The European Union continues to be a crucial economic force that influences international trade and financial markets.

5 - Philippines Annual Growth Rate of GDP (1960-2022)



Analysis

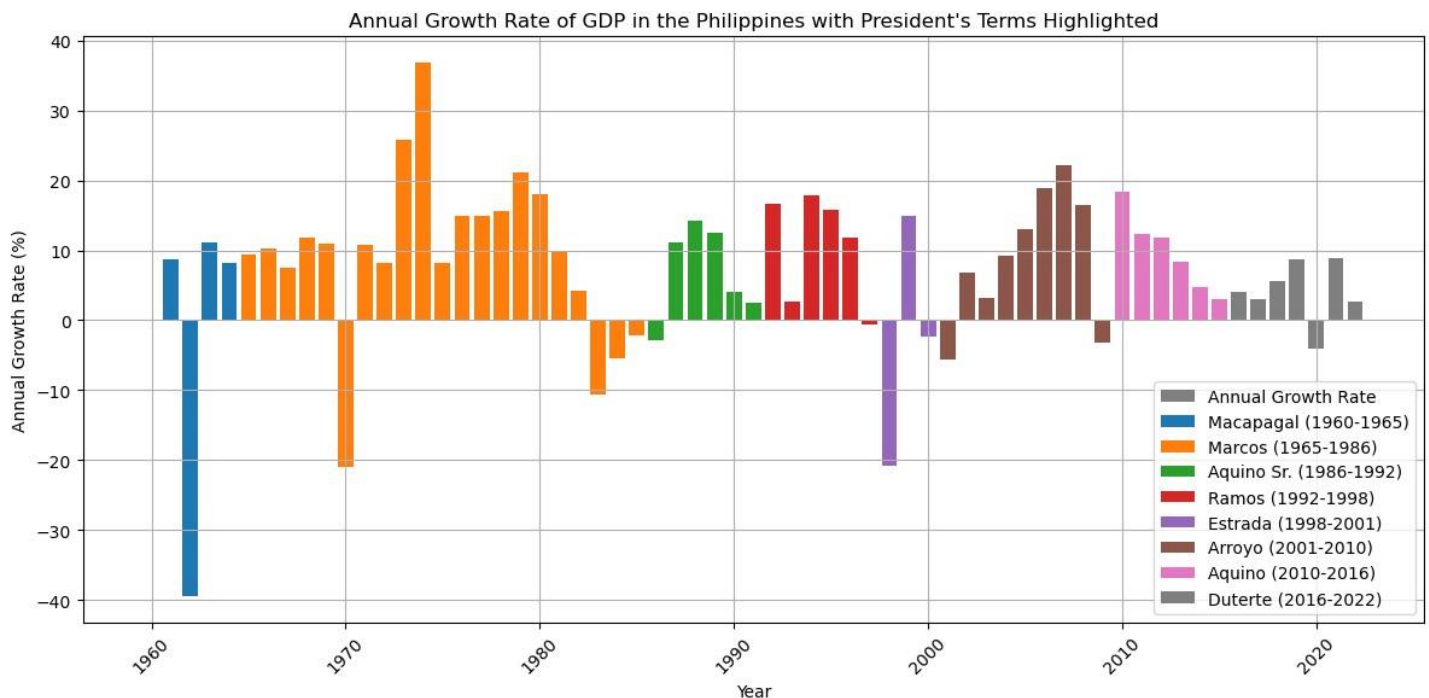
- The Annual Growth Rate of GDP in the Philippines from 1960 to 2022 exhibits distinctive trends, with notable swings reflecting both internal and global economic forces.
- Resilience in the Face of Global Challenges: Notably, in 2015, despite a substantial global economic slowdown, the Philippines showed resilience by maintaining a positive GDP growth rate. This resiliency can be linked to strong domestic demand, more government spending, and the expansion of service-oriented businesses, which protected the economy against shocks from outside sources.
- Strategic Economic Booms: The year 1974 saw an enormous jump in GDP growth, indicating a time of economic prosperity. Strategic economic policies, increased industrialization, and advantageous global trade circumstances can all be attributed to this expansion, which in turn has increased investments and economic activity.
- Economic expansionary phases are indicated by upward trends in 1979 and 2007. These phases could be ascribed to strong leadership, business-friendly regulations, and rising investment levels that foster thriving economies.
- Setbacks & Obstacles: In contrast, significant downturns in 1962, 1970, 1998, and 2009 reflect difficult times. Economic contractions during these years may be

attributed to natural disasters, policy concerns, or worldwide economic recessions, which could have a detrimental effect on consumer and investor sentiment.

Insights

- Economic Diversification: Phases of strong growth correspond to times when the economy is diversifying. During difficult global economic times, policies supporting varied industries like business process outsourcing (BPO), tourism, and manufacturing significantly boosted economic resilience and growth.
- Global and regional economic developments are closely related to the economic performance of the Philippines. The trajectory of the Philippine economy is shaped by external factors such as regional economic stability, global financial and economic crises, and investor confidence.

6 - Annual Growth Rate of GDP in the Philippines each President's term



Analysis:

Macapagal Administration (1960-1965):

- Economic difficulties were experienced during President Macapagal's term, particularly in 1961. There are a number of causes for the current economic difficulties, including changes in the global economy and domestic policy worries.

Marcos Administration (1965-1986):

- The Marcos era was marked by mix of economic growth and difficulties. It went through significant downturns, most notably the one in 1970. The final three years of the Marcos administration saw a drop in GDP growth.
- Massive lending from commercial banks, accounting for about 62% percent of external debt, allowed the GDP of the Philippines to rise during martial law.

Aquino Sr. Administration (1986-1992):

- President Aquino Sr.'s first year in office in 1986 was characterized by political and economic uncertainty as the country transitioned from the previous government. The economic fallout from the Marcos government caused this initial recession.
- Positive GDP growth in succeeding years pointed to a steady economic recovery. Reforms in governance and initiatives to regain investor trust contributed to economic stability and consistently favorable growth rates.

Ramos Administration (1992-1998):

- The Ramos administration got off to a good start economically, showing stability and recuperation from prior difficulties. Ramos' final year in office saw a minor economic decline.

Estrada Administration (1998-2001):

- The Estrada administration had economic instability, which was characterized by major downturns particularly in 1998.

Arroyo Administration (2001-2010):

- The economic performance of the Arroyo government was mixed. There were

difficulties in the first and last year, but also economic development that was evident in the years that followed.

Aquino Administration (2010-2016):

- The time of the Aquino administration was characterized by sustained economic growth without significant downturns. Various reasons, such as increased foreign direct investment, infrastructure spending, and governance reforms, were the causes for the period's economic stability.
- The GDP performance reflects how pro-business policies, anti-corruption programs, and strategic investments helped to create an environment that was favorable for economic growth.

Duterte Administration (2016-2022):

- The economic performance of the Duterte government maintained to be lacking of notable GDP changes. The global COVID-19 pandemic caused a dramatic decline in 2020. Supply chains were disrupted and strict lockdowns caused the economy to decline.

V. Conclusion and Recommendation

My extensive study of the World GDP dataset from 1960 to 2022 through exploratory data analysis (EDA) has given me a profound insight of global economic trends. I discovered significant insights on economic movements, policy impacts, and growth trends through a thorough process of data cleaning and insightful visualization. The analysis gave a lens in shedding light on historical economic occurrences. The emphasis I placed on the Philippines enabled a comprehensive analysis of this country's economic development, providing insightful insights for both analysts and policymakers.

It is essential to comprehend the specifics of economic changes. During downturns, specific economic issues should be addressed, and during upswings, growth opportunities should be taken advantage of. Lastly, the knowledge obtained through monitoring GDP growth in the Philippines during presidential terms highlights the necessity of consistent economic policies and their long-term effects.

Future investigations might yield richer insights by going deeper into the regional economic differences within nations and comprehending the social repercussions of economic changes. The project provides a solid framework for more focused research,

the development of policy, and strategic planning, leading to a better understanding of our interconnected global economy.

VI. References

Anna. (2023). World GDP by Country 1960-2022

<https://www.kaggle.com/datasets/annafabris/world-gdp-by-country-1960-2022/code>

Andrew Loo. 2008 – 2009 Global Financial Crisis.

<https://corporatefinanceinstitute.com/resources/economics/2008-2009-globalfinancial-crisis/>

Hodges, Paul. (2016). Global GDP sees record fall in 2015 as world hits “demographic cliff.

<https://www.icis.com/chemicals-and-the-economy/2016/01/global-gdp-seesrecord-fall-in-2015-as-world-hits-demographic-cliff/>

Sicat, Gerardo. (2015). Philippine economic history since 1930’s in perspective

<https://econ.upd.edu.ph/perse/?p=4671>

VII. Annex

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv("C:/Users/Jebe Dela Peña/Downloads/GDP.csv")

#UNDERSTANDING THE DATA

df.head()

df.tail()

df.shape

df.columns

df.nunique()

df['2022'].unique()
```

DATA CLEANING

Identify columns with missing values

```
df.isnull().sum()
```

Identify and remove duplicate rows

```
df = df.drop_duplicates()
```

Drop rows where all values in year columns are missing

```
years = [str(year) for year in range(1960, 2023)]
```

```
df = df.dropna(subset=years, how='all')
```

Remove all non-country entries (Continents and Other Classifications)

```
col_to_remove = ['World', 'Africa Eastern and Southern', 'Africa Western and Central', 'Arab World', 'Central Europe and the Baltics',  
                 'Channel Islands', 'Caribbean small states', 'East Asia & Pacific (excluding high income)', 'Early-demographic dividend',  
                 'East Asia & Pacific', 'Europe & Central Asia (excluding high income)', 'Europe & Central Asia', 'Arab Rep.', 'Euro area',  
                 'European Union', 'Fragile and conflict affected situations', 'High income', 'Heavily indebted poor countries (HIPC)',  
                 'IBRD only', 'IDA & IBRD total', 'IDA total', 'IDA blend', 'IDA only', 'Not classified', 'Latin America & Caribbean (excluding high income)',  
                 'Latin America & Caribbean', 'Least developed countries: UN classification', 'Low income', 'Lower middle income', 'Low & middle income',  
                 'Late-demographic dividend', 'Middle East & North Africa', 'Middle income', 'Middle East & North Africa (excluding high income)',  
                 'North America', 'OECD members', 'Other small states', 'Pre-demographic dividend', 'West Bank and Gaza', 'Pacific island small states',  
                 'Post-demographic dividend', 'South Asia', 'Sub-Saharan Africa (excluding high income)', 'Sub-Saharan Africa', 'Small states',  
                 'East Asia & Pacific (IDA & IBRD countries)', 'Europe & Central Asia (IDA & IBRD countries)', 'Latin America & the Caribbean (IDA & IBRD countries)',  
                 'Middle East & North Africa (IDA & IBRD countries)', 'South Asia (IDA & IBRD)', 'Sub-Saharan Africa (IDA & IBRD countries)', 'Upper middle income']
```

```
df_countries = df[~df['Country'].isin(col_to_remove)]
```

DATA VISUALIZATION

1 - Global GDP over the years

```

# Calculate global GDP by summing GDP values across all countries for each year
global_gdp = df_countries.loc[:, '1960':'2022'].sum()

# Extract years and global GDP values
years = global_gdp.index.astype(int)
gdp_values = global_gdp.values / 1_000_000_000 # Convert GDP to billions for better readability

# Create a line chart for global GDP growth by year
plt.figure(figsize=(12, 6))
plt.plot(years, gdp_values, marker='o', color='b', linestyle='-', linewidth=2, markersize=8)
plt.title('Global GDP (1960-2022)')
plt.xlabel('Year')
plt.ylabel('Global GDP (in Billions USD)')
plt.grid(True)

# Show only a subset of years on the x-axis (every 5 years, for example)
subset_years = years[::5] # Every 5 years
plt.xticks(subset_years, rotation=45) # Rotate labels for better visibility

plt.tight_layout()
plt.show()

# 2 - Annual Growth Rate of Global GDP

# Calculate Annual Growth Rate (AGR) of Global GDP
agr_values = []
years = df_countries.columns[2:] # Assuming the columns start from the third column
for i in range(1, len(years)):
    gdp_current_year = df_countries[years[i]].sum()
    gdp_previous_year = df_countries[years[i - 1]].sum()
    agr = ((gdp_current_year - gdp_previous_year) / gdp_previous_year) * 100
    agr_values.append(agr)

# Create a line chart for Annual Growth Rate of Global GDP
plt.figure(figsize=(12, 6))
plt.plot(years[1:], agr_values, marker='o', color='green', linestyle='-', linewidth=2, markersize=8)
plt.title('Annual Growth Rate of Global GDP (1961-2022)')
plt.xlabel('Year')
plt.ylabel('Annual Growth Rate (%)')
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()

```

```

# Show the plot
plt.show()

# 3 - Top 5 Countries with Highest GDP for the last 5 years

# Select the top 5 countries with the highest GDP for the last 5 years (2018-2022)
top_countries = df_countries.nlargest(5, '2022')

# Extract years from the columns of the DataFrame
years = [str(year) for year in range(2018, 2023)]

# Extract country names and GDP values for the last 5 years
country_names = top_countries['Country']
gdp_values = top_countries[years].values / 1_000_000_000 # Convert GDP to billions

# Set the bar width for the chart
bar_width = 0.15

# Set the positions of bars on X-axis for better visibility
r1 = list(range(len(years)))
r2 = [x + bar_width for x in r1]
r3 = [x + bar_width for x in r2]
r4 = [x + bar_width for x in r3]
r5 = [x + bar_width for x in r4]

# Create the bar chart
plt.figure(figsize=(10, 6))
plt.bar(r1, gdp_values[0], color='b', width=bar_width, edgecolor='grey',
label=country_names.iloc[0])
plt.bar(r2, gdp_values[1], color='g', width=bar_width, edgecolor='grey',
label=country_names.iloc[1])
plt.bar(r3, gdp_values[2], color='r', width=bar_width, edgecolor='grey',
label=country_names.iloc[2])
plt.bar(r4, gdp_values[3], color='c', width=bar_width, edgecolor='grey',
label=country_names.iloc[3])
plt.bar(r5, gdp_values[4], color='m', width=bar_width, edgecolor='grey',
label=country_names.iloc[4])

# Customize the plot
plt.xlabel('Year', fontweight='bold', fontsize=15)
plt.ylabel('GDP (in Billions USD)', fontweight='bold', fontsize=15)
plt.xticks([r + 2 * bar_width for r in range(len(years))], years)
plt.title('Top 5 Countries with Highest GDP (2018-2022)', fontweight='bold', fontsize=15)
plt.legend(loc='upper left', bbox_to_anchor=(1,1))

```

```

# Show the plot
plt.tight_layout() # Ensures the legend fits within the plot area
plt.show()

# 4 - GDP Distribution in 2022 - Top 10 Countries, Philippines and the entire world

year = '2022'
top_countries_2022 = df_countries.nlargest(10, year) # Top 10 countries (GDP in 2022)

# Add the GDP data of the Philippines
ph_gdp = df_countries.loc[df_countries['Country'] == 'Philippines', year].values[0]
philippines_df = pd.DataFrame({'Country': ['Philippines'], year: [ph_gdp]})
top_countries_2022 = pd.concat([top_countries_2022, philippines_df], ignore_index=True)

# Calculate the total GDP of other countries (excluding top 10 and Philippines)
other_countries_gdp = df_countries[year].sum() - top_countries_2022[year].sum()

# Create a list of GDP values (in billions USD) including top 10 countries, Philippines, and
"Other Countries"
gdp_values = list(top_countries_2022[year] / 1_000_000_000) + [other_countries_gdp /
1_000_000_000]

# Create a list of country names including top 10 countries, Philippines, and "Other
Countries"
countries = list(top_countries_2022['Country']) + ['Other Countries']

# Define colors for each segment
colors = ['#1f77b4', '#2ca02c', '#aec7e8', '#d62728', '#9467bd', '#8c564b', '#e377c2',
'#7f7f7f', '#bcbd22', '#17becf', '#ff9896', '#ff7f0e']

plt.figure(figsize=(10, 10))
plt.pie(gdp_values, labels=countries, autopct='%1.1f%%', startangle=140, colors=colors)

# Add legends with country names
plt.legend(countries, loc='upper left')

plt.title(f'GDP Distribution in {year}')
plt.show()

# 5 - Philippines Annual Growth Rate of GDP (1960-2022)

# Calculate Annual Growth Rate (AGR) of GDP for the Philippines
agr_values_philippines = []
years = df.columns[2:] # Assuming the GDP data starts from the third column

```



```

for i in range(len(years)):
    gdp_current_year = df[df['Country'] == 'Philippines'][years[i]].values[0]
    if i > 0:
        gdp_previous_year = df[df['Country'] == 'Philippines'][years[i - 1]].values[0]
        agr = ((gdp_current_year - gdp_previous_year) / gdp_previous_year) * 100
        agr_values_philippines.append(agr)
    else:
        # For the first year (1960), set AGR to 0 as there is no previous year for comparison
        agr_values_philippines.append(0)

# Create a line chart for Annual Growth Rate of GDP in the Philippines including year 1960
plt.figure(figsize=(12, 6))
plt.plot(years, agr_values_philippines, marker='o', color='blue', linestyle='-', linewidth=2,
markersize=8)
plt.title('Annual Growth Rate of GDP in the Philippines (1960-2022)')
plt.xlabel('Year')
plt.ylabel('Annual Growth Rate (%)')
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()

# Show the plot
plt.show()

# 6 - Annual Growth Rate of GDP in the Philippines each President's term

# Define the years corresponding to each president's term
president_terms = {
    'Macapagal': (1960, 1965),
    'Marcos': (1965, 1986),
    'Aquino Sr.': (1986, 1992),
    'Ramos': (1992, 1998),
    'Estrada': (1998, 2001),
    'Arroyo': (2001, 2010),
    'Aquino': (2010, 2016),
    'Duterte': (2016, 2022)
}

# Calculate Annual Growth Rate (AGR) of GDP for the Philippines including year 1960
agr_values_philippines = []
years = df.columns[2:] # Assuming the GDP data starts from the third column
for i in range(len(years)):
    gdp_current_year = df[df['Country'] == 'Philippines'][years[i]].values[0]
    if i > 0:

```

```

    gdp_previous_year = df[df['Country'] == 'Philippines'][years[i - 1]].values[0]
    agr = ((gdp_current_year - gdp_previous_year) / gdp_previous_year) * 100
    agr_values_philippines.append(agr)
else:
    # For the first year (1960), set AGR to 0 as there is no previous year for comparison
    agr_values_philippines.append(0)

# Extract the years for x-axis labels
years = list(range(1960, 2023))

# Create a bar chart for Annual Growth Rate of GDP in the Philippines with president's terms
highlighted
plt.figure(figsize=(12, 6))
plt.bar(years, agr_values_philippines, color='gray', label='Annual Growth Rate')
plt.xlabel('Year')
plt.ylabel('Annual Growth Rate (%)')

# Highlight each president's term with different colors and add legends
for president, (start_year, end_year) in president_terms.items():
    plt.bar(range(start_year, end_year + 1),
            agr_values_philippines[start_year - 1960:end_year - 1960 + 1],
            label=f'{president} ({start_year}-{end_year})')

plt.title('Annual Growth Rate of GDP in the Philippines with President\'s Terms Highlighted')
plt.legend()
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()

# Show the plot
plt.show()

```