MILESTONE 2

GLOBAL ECONOMIC & DEMOGRAPHIC TRENDS ANALYSIS  
KISHORE VASUDEVAN  
26/05/2025  
DATA ANALYTICS & DATA SCIENCE  
FEB 25’

PROJECT OVERVIEW:

The dataset represents the world population and GDP from 1960 to 2016. It is provided in three formats: SQL, Excel, and CSV. The SQL file contains details such as country name, country code, population, GDP per capita, infant mortality, death rate, birth rate, phones per 1000 people, literacy rate, sector contributions (agriculture, industry, service), population density, arable land percentage, other land percentage, and climate. The Excel and CSV files include population and GDP data from 1960 to 2016. The Excel file contains data for 256 countries, while the SQL file contains data for 225 countries.

OBJECTIVE:

To analyze the trends in GDP and population over the years, and identify the key factors that influence GDP and population growth.

TOOLS USED:

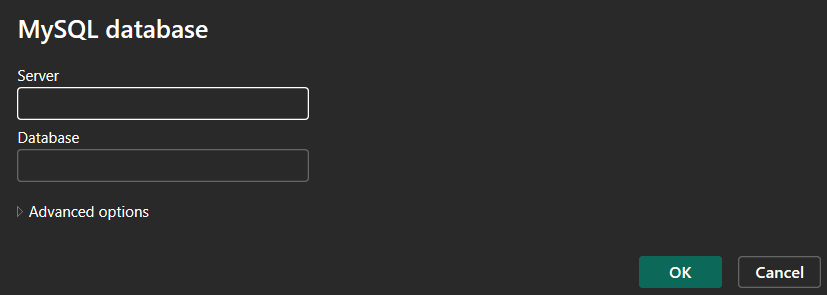
* Excel
* SQL
* Power Query
* Power BI

SQL:

* Created the database using given SQL data.
* Export the country names and their corresponding country codes as a reference for reliable countries.

DATA CLEANING AND PREPARARTION:

* Using the exported country data, each country was verified using Excel’s online geography feature to check for reliability.
* All tables were appended in Excel, and the country names and country codes were checked for duplicates. It was observed that some country names had different country codes. These inconsistencies were identified, and the data was cleaned to ensure uniformity.
* The cleaned list of countries was imported into Power BI, followed by the step-by-step import of all Excel and CSV files.
* Click transform data, directs to power query. The population table was cleaned first, as it contained no null values. The top three empty rows were removed, and the first row was promoted to headers.
* Next, the GDP table was cleaned by removing the top three empty rows and promoting the first row to headers. Some countries had null values, likely because they were not recognized as independent countries during that time period. These null values were replaced with zeros.
* Both the GDP and population tables were merged using the country name and country code. The year columns were unpivoted to transform them into rows, allowing the values to be used effectively in filters and on the X-axis for visualizations.
* With the cleaned country table, the metadata table was first merged to add the income group information to the dataset.
* Then, import the SQL database using the Get Data feature. You need to enter the server’s name and database name in the columns below.



* Then, enter the MySQL username and password, and click **Connect** to transform the database.
* Now, merge this data with the cleaned country data using an inner join on the matching fields: country name and country code. Then, expand the table to include all fields from the joined table.
* Just filter the reliable countries by merging with the cleaned country, population, and GDP tables to get the reliable data.
* Now, we have 170 cleaned records from the Population & GDP table and the Countries table.
* Just right-click on the table in Power Query and uncheck **Enable Load** to prevent overloading Power BI.
* Just click on **Close & Apply** to load the data into Power BI.
* Now, click the **Model** tab to model the tables. Match the country code to create a one-to-many relationship between the tables, and set the cross-filter direction to **Both** to enable filtering from both sides.

FINDINGS:

* Overall population of the dataset is 259.96bn and GDP is 1370.35trillion.
* Population has consistently increased from 2.7 billion in 1960 to 6.6 billion in 2016. Significant growth acceleration is observed after 1980.
* GDP has consistently increased from 1 trillion (T) in 1960 to 72 trillion (T) in 2016. There's a noticeable acceleration in GDP growth from the 1990s onwards.
* There's a clear positive correlation between increasing total population and increasing total GDP over the years. As total population grows, total GDP also shows a significant upward trend.
* **Total Population by Income Group:** Lower-middle income group accounts for the largest share at **38.12%** & Upper-middle income group is **36.31%** & High-income group is **19.86%** & Low-income group is the smallest at **5.71%**.
* The **United States has the highest GDP among the listed countries at 386 trillion**. Following the U.S., **Japan ranks second with 152trillion**, and **China is third with 102trillion**. Germany and France round out the top five, with GDPs of **89trillion** and **66trillion** respectively.
* Top Population: China with the highest figure **61 billion**, followed by the United States **14 billion**, Japan **7 billion**, Germany **5 billion**, and France **3 billion**.
* The Top 5 Average of Agriculture, Service, and Industry countries: Liberia, Somalia, Guinea-Bissau, Central African Republic, and Sierra Leone. **Liberia:** Has the highest average of agriculture (0.77), lowest average of industry (0.18), and relatively low average of service (0.05).
* The literacy rate has increased over time, reaching approximately 83.40 by 2016 from a starting point around 83.20 in 1960. The increase appears gradual.
* Net migration hovered around -0.20 to -0.21 for much of the period from 1960 to the late 1980s. There was a significant drop to around -0.29 in the early 1990s, after which it recovered slightly but remained negative, around -0.21.
* The "Top 5 by Phones per 1000" data indicates that **Cyprus leads significantly with 1,490 phones per 1000 people (1.49K)**. Monaco follows as second with 1,040 phones per 1000 (1.04K). The United States, Gibraltar, and Bermuda complete the top five, with 900, 880, and 850 phones per 1000 people, respectively.

INSIGHTS:

* In Tuvalu, the GDP (0.58bn) is the lowest among countries, but it still has an upper-middle-income population. Both population and GDP are growing, but it ranks last due to its very low population size.
* The United States has the highest GDP (386T) due to tremendous growth in both population and GDP over time. It clearly shows a strong relationship between GDP and population, and most people fall under the high-income group.
* For Low-Income, Ethiopia (0.67T) is the first country and least is Guinea-Bissau.
* In 2016, Tuvalu had the lowest population and GDP, while the United States remained in first place for GDP with $19T, followed by China with $11T. Interestingly, China had the highest population at 1.4 billion, followed by the United States with 0.3 billion.
* The country with the highest birth rate (50.73) is Niger, with an infant mortality rate of 12.17 and a death rate of 12.17, making it relatively balanced.
* The countries with the lowest GDP per capita, around 500, are Somalia and Sierra Leone, while the highest is Luxembourg with 55,100.
* There is a downward trend in infant mortality over the years. If we look at the graph, the distribution appears fairly even, resembling a platykurtic distribution.
* Mongolia has the lowest population density at 1.80, while Monaco has the highest at 16,271.50.

CONCLUSION:

From 1960 to 2016, both global population and GDP steadily increased, with notable acceleration after 1980 and 1990 respectively. There is a strong positive correlation between population growth and GDP rise. Lower-middle income countries hold the largest share of the population, while the United States leads in total GDP and Luxembourg in GDP per capita. Tuvalu, despite having the lowest GDP, maintains an upper-middle income status but its growth is limited by its small population. China has the largest population, whereas the U.S. remains the economic leader. Over time, literacy rates have improved and infant mortality rates have declined. Population density varies widely, with Mongolia being the lowest and Monaco the highest. Overall, while global disparities persist, the general trend shows consistent development and economic growth worldwide.