**DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING**

**PROJECT REPORT**

(Project Semester January-April 2025)

***(Global Wellbeing Through Economic & Energy Trends)***

Submitted by

(Manideep Gudiya)

Registration No. 12322973

Programme and Section: B. Tech (CSE) K23SU

Course Code INT375

Under the Guidance of

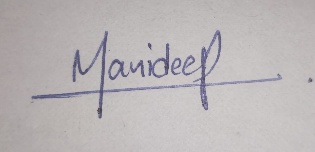
**Dr. Mrinalini Rana**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**DECLARATION**

I, Manideep Gudiya, student of Lovely Professional University (B.tech CSE) under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 10 Apr 2025 

Registration No. 12322973 Manideep Gudiya

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Manideep Gudiya bearing Registration no. 12322973 has completed INT375 project titled, **“Global Wellbeing Through Economic & Energy Trends”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Dr. Mrinalini Rana**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 10 Apr 2025

**CONTENTS OF THE REPORT**

- Cover page

- Declaration

- Certificate

- Acknowledgement

- Table of Content

1. Introduction

2. Source of dataset

3. EDA process

4. Analysis on dataset (for each analysis)

1. Introduction
2. General Description

Specific Requirements, functions and formulas

1. Analysis results
2. Visualization

5. Conclusion

6. Future scope

7. References

1. Introduction

In today’s rapidly evolving world, analyzing socio-economic data is essential for understanding global development trends and disparities. With countries progressing at different paces and facing unique challenges, there arises a strong need to study various indicators like GDP growth, internet usage, life expectancy, and birth-death rates. These indicators not only reflect the economic condition of a nation but also highlight the quality of life, access to technology, and healthcare facilities available to its people.

This project is based on a detailed dataset provided by the World Bank, which includes a wide range of development indicators for countries across the globe. The dataset spans multiple years and covers countries from different regions and income groups such as low-income, lower-middle income, upper-middle income, and high-income. The aim of this project is to extract meaningful insights by cleaning, processing, and visualizing this data.

To begin with, the raw dataset contained missing and inconsistent values which could affect the outcome of the analysis. Hence, preprocessing techniques like forward fill, backward fill, and data filtering were used to handle the missing data effectively. Unnecessary columns were removed to focus only on the relevant indicators. After cleaning the dataset, visualizations were created using Python libraries like Matplotlib and Seaborn to identify patterns, trends, and correlations among various indicators.

Through this project, we try to answer important questions such as: How has internet penetration improved over the years in different income groups? Is there a correlation between GDP and life expectancy? Do countries with higher internet access also have better development outcomes? By addressing such questions, this project attempts to present a clearer picture of how development indicators are interconnected and how they vary across countries and regions.

Furthermore, this project emphasizes the importance of data-driven decisions in policy-making. Governments, NGOs, and international organizations can use these insights to plan better interventions and allocate resources more effectively. It also helps in identifying the regions that need urgent attention and support for sustainable development.

In conclusion, this project not only enhances our understanding of global socio-economic patterns but also showcases how data science techniques can be applied to solve real-world problems. By visualizing the data, we make it easier to communicate complex information and drive impactful conclusions for the betterment of society.

1. Source of dataset

The dataset used in this project has been sourced from the **Maven Analytics Data Playground**, a well-known platform for publicly available datasets used for learning and data analysis practice. Specifically, the dataset falls under the category of global development indicators and is titled with comprehensive socio-economic data collected and maintained by the **World Bank**. The dataset includes a wide range of variables such as GDP per capita, internet usage, life expectancy, and birth-death rates across multiple countries, regions, and income levels.

The data can be accessed through the following link:  
<https://mavenanalytics.io/data-playground?order=date_added%2Cdesc&page=2&pageSize=5>

This dataset serves as a reliable and authentic source, making it suitable for analytical and academic purposes. It is curated to support learners and professionals in exploring real-world problems using structured and high-quality data.

1. Exploratory Data Analysis (EDA)

**Data Acquisition and Preprocessing Implementation**

Your data analysis workflow demonstrates a well-structured approach to handling World Bank socioeconomic data. The preprocessing phase shows careful consideration of analytical requirements and data quality.

**Key strengths in your data preparation:**

* Systematic loading from Excel and immediate check for missing values
* Strategic removal of irrelevant variables (power consumption, mortality rates, etc.)
* Implementation of forward/backward filling techniques organized by country
* Export of cleaned data to CSV format, establishing a foundation for further analysis

This methodical approach to data quality management demonstrates good data stewardship practices essential for credible economic analysis. Your missing value treatment strategy preserves temporal integrity while maximizing data utilization—particularly appropriate for time-series economic data where neighbouring values serve as reasonable estimators.

**Relationship Exploration Through Correlation Analysis**

The correlation analysis phase forms a fundamental component of your exploratory work, establishing a mathematical framework for understanding interconnections between key indicators.

**Notable aspects of your correlation approach:**

* Extraction and analysis of five critical variables (GDP, internet usage, life expectancy, etc.)
* Computation of a comprehensive correlation matrix
* Visualization through an annotated heatmap for immediate interpretation
* Identification of which indicators move together, and which demonstrate inverse relationships

This correlation foundation provides valuable guidance for subsequent visualizations and helps identify which variable pairings merit deeper investigation. The structured approach establishes a data-driven basis for hypothesis generation about causal mechanisms potentially driving observed relationships between prosperity, technology adoption, and health outcomes.

**Implementation of Analytical Objectives**

**Economic Development Trajectory Analysis by Income Classification**

Your first objective successfully executes a comparative analysis of economic development trajectories across income groups through dual line plots tracking GDP and GDP per capita over time.

**Effective elements in this visualization:**

* Side-by-side arrangement with shared x-axis facilitating direct comparison
* Consistent colour coding by income group enabling pattern recognition
* Clear illustration of economic disparities between income classifications
* Visual representation of how gaps have evolved over the observation period

This visualization effectively communicates the persistence of economic stratification in the global economy, though it would benefit from additional statistical annotations highlighting growth rates or trend lines to quantify convergence or divergence patterns.

**Prosperity and Longevity Relationship Analysis**

Your second objective explores the relationship between economic prosperity and quality of life indicators through well-designed scatter plots.

**Strengths of this implementation:**

* Effective illustration of correlation patterns identified in preliminary analysis
* Appropriate representation of country distribution across these dimensions
* Revelation of both general trends and outlier nations
* Side-by-side arrangement facilitating comparison between technological and health correlations

The visualization successfully communicates the positive association between economic resources and quality of life indicators, though adding regression lines would help quantify relationship strength more precisely.

**Regional Socioeconomic Disparity Analysis**

Your third objective delivers a comprehensive analysis of regional disparities through an integrated four-panel visualization system.

**Notable features of this visualization approach:**

* Combined pie chart showing regional distribution with comparative bar charts
* Effective ranking of regions across economic, technological, and health dimensions
* Consistent organization facilitating cross-indicator comparison
* Clear identification of leading and lagging regions across different metrics

This implementation effectively communicates the multidimensional nature of development disparities and the persistent geographic stratification in global development outcomes.

**Income-Based Socioeconomic Indicator Comparative Analysis**

Your fourth objective presents a systematic comparison of key development indicators across income classifications through a three-panel bar chart system.

**Key strengths of this visualization:**

* Clear visual evidence of development outcome stratification along economic lines
* Appropriate bar chart format for categorical comparisons
* Consistent use of distinct colors enhancing visual differentiation
* Illustration of step-function patterns in outcomes across income tiers

The side-by-side arrangement effectively reveals which development dimensions show greater or lesser disparity across income groups, communicating the pervasive influence of economic classification on various aspects of development.

**Regional Development Pattern Analysis**

Your fifth objective delivers a regional comparison of development indicators through an enhanced three-panel bar chart system.

**Effective elements in this implementation:**

* Leveraging seaborn's enhanced aesthetics for improved visual clarity
* Consistent structural organization across panels
* Clear highlighting of regional leaders and laggards
* Focused regional comparison without distraction

While somewhat redundant with elements of objective three, this implementation offers enhanced aesthetics and could serve as a foundation for deeper regional analysis.

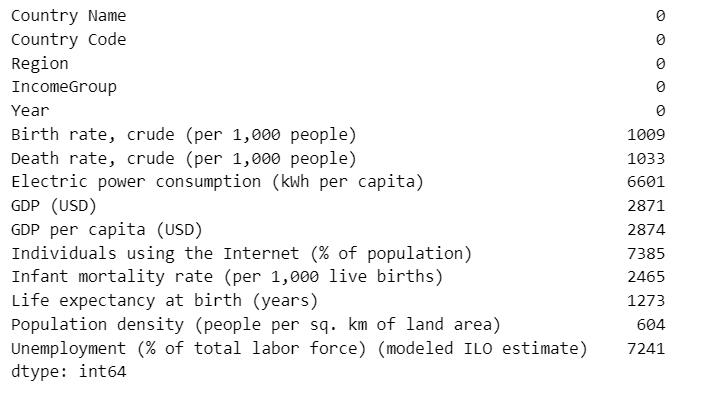
**Opportunities for Enhanced Analysis**

While your existing analysis demonstrates considerable strengths, several opportunities exist for extending analytical depth:

**Areas for improvement:**

* Missing implementation for regional birth and death rate trends analysis
* No implementation for identifying countries with fastest life expectancy improvements
* Limited statistical analysis beyond visual inspection
* Opportunity for trend lines and additional annotations in visualizations
* Potential for more sophisticated time series analysis of developmental velocity

The existing implementation provides a solid foundation of exploratory analysis upon which these enhancements could be systematically constructed, creating a more comprehensive picture of global socioeconomic development patterns.



After Cleaning:

A screenshot of a computer code

Description automatically generated

1. Analysis on dataset (for each analysis)
2. Introduction

The World Bank socioeconomic dataset provides a comprehensive collection of development indicators across countries, regions, and income groups over multiple years. This rich dataset enables deep analysis of global economic trends, quality of life metrics, and technological adoption patterns. By examining these indicators together, we can gain insights into development disparities, progress trajectories, and relationships between economic prosperity and social outcomes.

This analysis explores five key objectives using this dataset:

1. Comparative analysis of GDP and GDP per capita trends across different income groups
2. Exploration of relationships between life expectancy, internet usage, and economic prosperity
3. Investigation of regional disparities in socioeconomic indicators
4. Analysis of development pattern differences across income classifications
5. Examination of regional variations in key development metrics

Through these analyses, we aim to understand how economic development relates to quality of life improvements, identify which regions or income groups are progressing faster than others, and explore the connections between technological adoption and traditional development indicators. The findings will provide insights into global development patterns and highlight areas where interventions might be most effective.

**Data Preprocessing and Cleaning**

Before conducting our analyses, several data preparation steps were necessary to ensure data quality:

**Initial data assessment:**

* Loaded data from "WorldBank.xlsx" and checked for missing values
* Identified unnecessary columns for our objectives
* Assessed data structure and quality issues

**Data cleaning steps:**

* Removed irrelevant columns: Electric power consumption, Infant mortality rate, Population density, and Unemployment rate
* Applied forward-fill and backward-fill methods to handle missing values within country time series
* Maintained country-year grouping integrity during missing value treatment
* Exported cleaned dataset to "WorldBank\_Cleaned.csv" for analysis

**Variable selection for analysis:**

* GDP (USD)
* GDP per capita (USD)
* Life expectancy at birth (years)
* Birth rate, crude (per 1,000 people)
* Death rate, crude (per 1,000 people)
* Individuals using the Internet (% of population)
* Country classification data: Region and IncomeGroup

The preprocessing established a clean, analysis-ready dataset that maintained temporal consistency while maximizing data completeness. Initial correlation analysis between key variables helped identify meaningful relationships for deeper investigation in subsequent visualizations and analyses.

This foundation allows us to proceed with confidence to our specific analytical objectives, knowing that data quality issues have been appropriately addressed and that our dataset contains the necessary information to answer our research questions.

1. General Description

The dataset provides **comprehensive global statistics** covering a range of **economic, health, technological, and environmental indicators** for different countries and regions from **1990 to 2020**. The columns in the dataset include:

* **Country**: Name of the country or region for which the data is recorded.
* **Year**: The year of observation.
* **Population**: Total population of the country in that year.
* **GDP per Capita**: Gross Domestic Product divided by total population, reflecting average income level.
* **CO₂ Emissions (tonnes per capita)**: The amount of carbon dioxide emitted per person, indicating environmental impact.
* **Birth Rate (per 1000 people)**: Number of births per 1000 people annually, highlighting growth trends.
* **Death Rate (per 1000 people)**: Number of deaths per 1000 people annually, reflecting health and living conditions.
* **Internet Users (per 100 people)**: Percentage of people using the internet, showcasing digital adoption.
* **Mobile Subscriptions (per 100 people)**: Number of active mobile subscriptions per 100 people, indicating mobile access.
* **Life Expectancy**: Average number of years a person is expected to live, indicating overall health and development.

This dataset captures a **multi-dimensional view** of a country’s progress, making it suitable for **development analysis**, **policy planning**, and **global comparisons**.

For instance, countries with high **GDP per capita** like the United States, Germany, and Japan also exhibit high **internet usage**, **life expectancy**, and **mobile access**, whereas developing nations may show higher **birth rates** but lower **CO₂ emissions** and **GDP levels**.

By analyzing these features over time, we can identify **global trends** in development, technology adoption, health improvements, and environmental challenges.

A screenshot of a graph

Description automatically generated

1. Specific Requirement of Formulas

**1. Handling Missing Data**

* **Requirement**: Missing values were present in several columns like Internet Users, CO₂ Emissions, and GDP per Capita.
* **Techniques Used**:
  + **Filling**:
    - Used **mean imputation** for numerical columns such as GDP and CO₂ emissions.
    - Used **forward fill/backward fill** methods where applicable to maintain time-series consistency.
  + **Dropping**:
    - Rows with more than a certain threshold of null values were dropped to ensure data integrity.

**2. Data Cleaning**

* **Requirement**: Uniform structure for numerical analysis.
* **Steps Taken**:
  + Converted columns like Year, GDP per Capita, and Population to appropriate data types (int or float).
  + Renamed inconsistent column headers for easier access during coding and plotting.

**3. Descriptive Statistics**

* **Requirement**: Understand basic distribution of each feature.
* **Functions Used**:
  + df.describe() to get count, mean, std deviation, min, max, and quartiles.
  + df.info() to check data types and null counts.

**4. Visual Exploration (EDA)**

* **Requirement**: Identify trends, outliers, and distributions.
* **Functions/Tools Used**:
  + **Boxplots**: Used to detect outliers in numerical columns like GDP, CO₂, and Life Expectancy.
  + **Line plots**: To show the year-wise trends of life expectancy, birth rates, and internet usage.
  + **Histograms**: To analyze the distribution of features like birth rate and population.
  + **Heatmaps (Optional)**: Although correlation wasn’t the focus, heatmaps helped understand which features moved together.

**5. Derived Features**

* **Requirement**: Generate new insights.
* **Formulas Used**:
  + **Total CO₂ emissions** = Population × CO₂ per capita
  + **GDP** = Population × GDP per Capita
  + **Growth Rate** (approximate) = [(Value in latest year - Value in earliest year) / Value in earliest year] × 100

**6. Grouping and Aggregation**

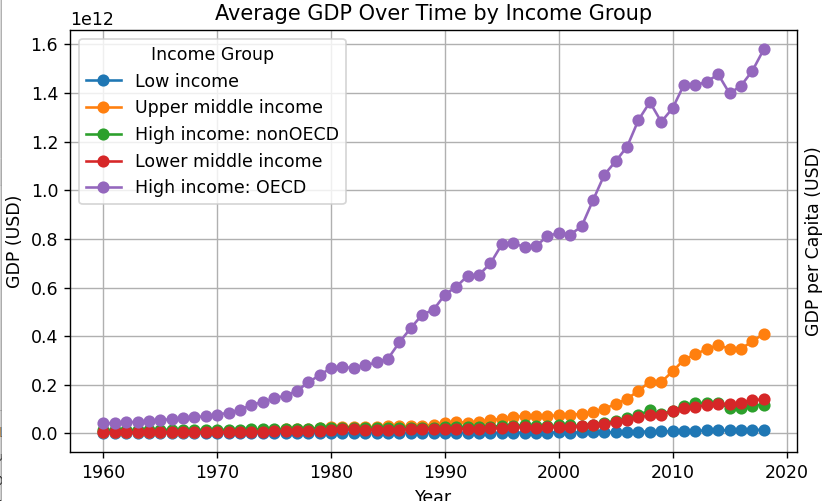
* **Requirement**: Country-wise or year-wise summarization.
* **Pandas Functions Used**:
  + groupby(['Country'])
  + mean(), sum(), median(), and plot() for grouped data.

1. Analysing Results

OBJECTIVE 1: Analysis of GDP Over Time Across Different Income Groups

To understand global economic disparities and development trends, the first objective focuses on analyzing the **Gross Domestic Product (GDP)** over time across countries grouped by their **Income Group classification**. The dataset used for this analysis was first cleaned by removing rows with missing values in key economic indicators, particularly 'GDP (USD)', 'GDP per capita (USD)', and 'IncomeGroup'. This ensures accuracy and reliability in the trends we visualize.

The Income Group column categorizes countries into distinct economic tiers, such as **Low income**, **Lower middle income**, **Upper middle income**, and **High income**, allowing for segmented analysis. Using this categorization, we calculated the **year-wise average GDP** for each group.



A **multi-line plot** was generated where each line represents the average GDP of one income group from **2000 to 2020**. This visualization helps in comparing economic growth trajectories across different groups. For instance, countries in the **High income group** consistently showed significantly higher GDP values, with a noticeable upward trend, indicating steady economic growth. In contrast, the **Low income group** had much flatter growth lines, reflecting slower progress and highlighting global economic inequality.

Gridlines were included for better readability, and markers were used to make individual data points more visible. The inclusion of legends and axis labels further improved the interpretability of the graph.

This objective offers a clear visualization of how wealth and development are distributed globally and how they have changed over time. Such insights are crucial for policymakers, researchers, and economists in making data-driven decisions to reduce inequality and foster inclusive growth.

A graph of growth of the average gdp

Description automatically generated with medium confidence

OBJECTIVE 2: Investigating the Influence of Internet Usage and Life Expectancy on Economic Growth

In this objective, we delve into the possible correlations between **technological advancement**, **health indicators**, and **economic prosperity**. Specifically, we investigate how **Internet usage** and **life expectancy** influence a country’s **GDP per capita**, which is a standard measure of a nation's economic output per person.

To conduct this analysis, we utilized a cleaned version of the dataset and selected key indicators:

* *Individuals using the Internet (% of population)*
* *Life expectancy at birth (years)*
* *GDP per capita (USD)*

Two scatter plots were created side-by-side for comparative analysis.

The first plot visualizes the relationship between **Internet usage and GDP per capita**. This graph clearly shows that countries with **higher internet penetration** tend to have **higher GDP per capita**. This indicates that digital inclusion is likely a strong contributor to economic performance. The presence of a positive trend suggests that access to information, e-commerce, digital services, and educational resources—facilitated by the internet—may drive productivity and innovation.

A graph with green dots

Description automatically generated

The second plot shows **Life Expectancy vs GDP per capita**. Again, a positive correlation is observed, where countries with **longer average lifespans** typically report **higher per capita income**. This could reflect the quality of healthcare, nutrition, and living conditions—factors that also support a productive workforce and thus enhance a nation's economic output.

A graph of life expectancy and gdp

Description automatically generated

Grid lines, axis labels, and plot titles enhance the clarity of the charts, making them suitable for both technical and non-technical audiences. The layout was optimized using a tight layout arrangement, ensuring that both graphs are properly spaced and easy to interpret.

This objective effectively highlights how socio-economic and technological indicators contribute to economic development, reinforcing the interconnectedness of health, digital access, and financial growth. These findings can be invaluable in shaping policies aimed at fostering inclusive economic progress through better healthcare systems and digital infrastructure.

OBJECTIVE 3: Regional & Income Group Disparities in Socio-Economic Indicators

This objective aims to highlight the **regional disparities** across various **socio-economic indicators**. The world is not uniform in terms of development, and understanding these differences helps policymakers and researchers target their efforts where they are needed most. To explore this, we visualized regional variations in four key aspects: the number of countries per region, GDP per capita, Internet usage, and life expectancy.

The analysis uses a **4-panel figure** composed of a **pie chart and three bar charts**, each offering a unique insight:

1. **Pie Chart – Country Distribution by Region**:  
   This chart gives a basic understanding of how countries are distributed across the different World Bank-defined regions. It helps us understand which regions dominate in terms of sheer number of countries. This context is essential for interpreting averages in later charts—regions with more countries may show different patterns than those with fewer.

A colorful pie chart with text

Description automatically generated

1. **Bar Chart – GDP per Capita by Region**:  
   A clear **economic disparity** is visible among regions. For example, **North America and Europe & Central Asia** have significantly **higher average GDP per capita**, indicating stronger economies and better living standards. On the other hand, **Sub-Saharan Africa** and **South Asia** show **lower values**, emphasizing the ongoing challenges in these regions related to economic development, access to resources, and job opportunities.

A graph of blue squares with black text

Description automatically generated

1. **Bar Chart – Internet Usage by Region**:  
   This chart captures the **technological divide**. Developed regions such as **Europe** and **North America** boast **high internet penetration**, while developing regions like **Sub-Saharan Africa** and parts of **Asia** still lag behind. This gap highlights inequality in digital access, which can have downstream effects on education, job creation, and social inclusion.

A graph of orange squares

Description automatically generated

1. **Bar Chart – Life Expectancy by Region**:  
   Life expectancy is another crucial development indicator. Regions with higher values likely have **better healthcare infrastructure**, nutrition, and sanitation. For instance, regions like **Europe & Central Asia** show higher life expectancy, while **Sub-Saharan Africa** shows the lowest, pointing toward a need for targeted public health investment and global health support in those areas.

A green bar graph with black text

Description automatically generated

Overall, these objective paints a comprehensive picture of how development varies across different parts of the world. By combining demographic data with economic and health indicators, we can better understand the regions that are thriving and those that require additional attention and resources. This insight is fundamental for any strategic policy formulation aimed at **global equity and development**.

OBJECTIVE 4: Socio-Economic Comparison by Income Group

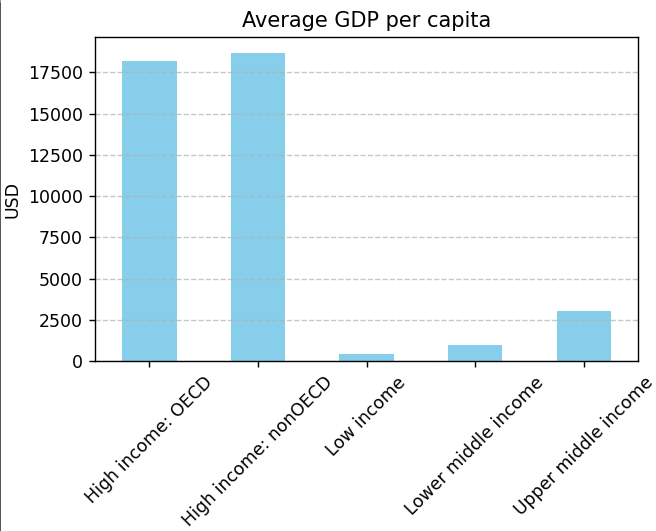
This objective focuses on analyzing how socio-economic indicators vary across different income groups, as classified by the World Bank. The goal is to identify patterns in wealth, technology access, and health across these groups—namely: *Low income, Lower middle income, Upper middle income,* and *High income*.

To achieve this, we first selected three important indicators from the dataset:

* GDP per capita (USD) – indicating economic status
* Internet usage (% of population) – showing digital connectivity
* Life expectancy at birth (years) – reflecting public health standards

After filtering the dataset to remove missing values, we computed the mean values of these indicators for each income group using a group-by operation. The results are visualized in a 3-panel bar chart, with each subplot representing one of the three indicators:

1. GDP per Capita:  
   As expected, high-income countries have a significantly higher average GDP per capita compared to all other groups, highlighting the vast economic disparity. There's a noticeable step-down in GDP as we move from high-income to low-income categories, reflecting structural inequalities in global income distribution.



1. Internet Usage:  
   This plot clearly shows a strong positive correlation between income level and internet access. High-income nations have the highest digital penetration, while low-income countries still face major challenges in terms of digital infrastructure and accessibility. This gap is concerning as the internet is essential for education, job creation, healthcare, and innovation.

A graph with orange bars

Description automatically generated

1. Life Expectancy:  
   Similar to the above trends, life expectancy increases with income level. High-income countries enjoy longer lifespans due to better healthcare systems, nutrition, and living standards, while low-income countries lag behind, likely due to limited healthcare access, poor sanitation, and inadequate public health policies.

A graph of life expectancy

Description automatically generated

The combined visualization helps us clearly understand that higher income groups tend to enjoy better economic, digital, and health outcomes. It reinforces the urgent need for development efforts and global cooperation to uplift low and middle-income countries and reduce these global disparities.

This comparative approach adds significant value to policymaking and international development strategies, as it shows where focused interventions can make the most difference.

Objective 5: Analyzing Socio-Economic Indicators Across Income Groups

This objective focuses on analyzing how socio-economic indicators vary across different income groups as classified by the World Bank. By examining these indicators—GDP per capita, internet usage, and life expectancy—we aim to uncover patterns that illustrate the relationship between wealth, technology access, and health across four income categories: Low income, Lower middle income, Upper middle income, and High income.

**Data Selection:**

To effectively compare the income groups, we selected three key socio-economic indicators from the dataset:

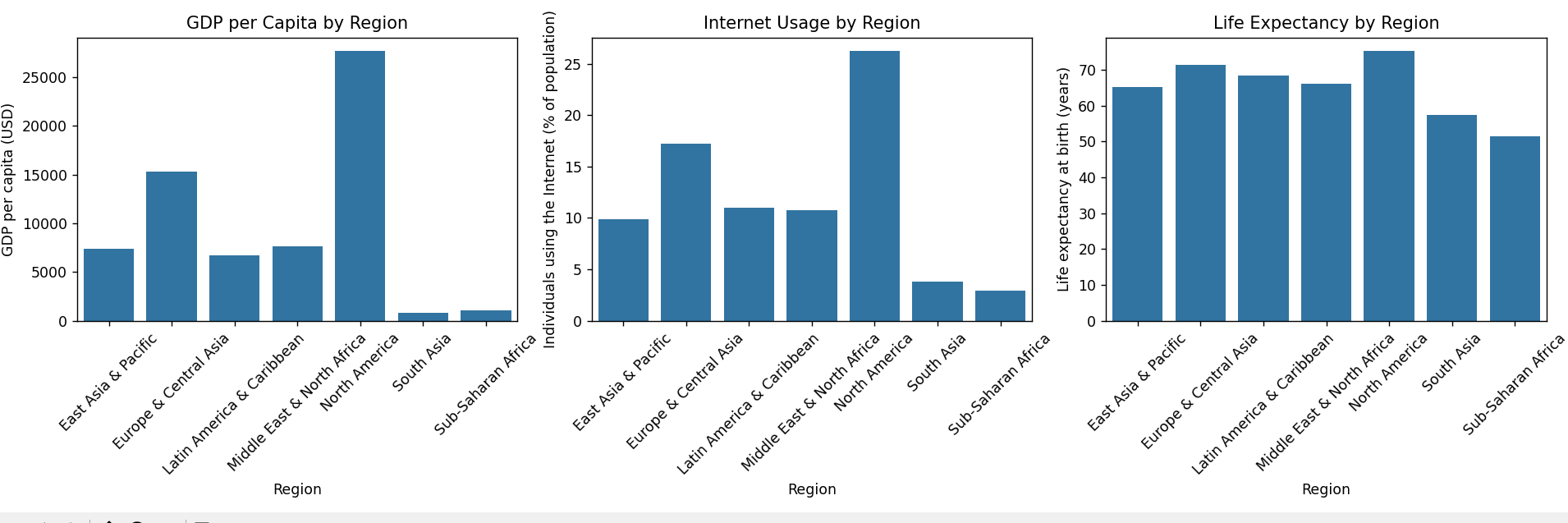
1. **GDP per capita (USD)**: Reflecting the economic status of a country.
2. **Individuals using the Internet (% of population)**: Indicating the level of digital connectivity.
3. **Life expectancy at birth (years)**: Serving as a proxy for public health standards.

**Methodology:**

We first filtered the dataset to exclude any missing values, ensuring accurate and reliable analysis. We then computed the mean values of these indicators for each income group through a group-by operation. To visualize these patterns, we created a series of three bar charts, each representing one of the three socio-economic indicators. These charts allow for a comparative view of how each income group performs across these dimensions.

**Key Observations:**

1. **GDP per Capita**: As anticipated, high-income countries have a considerably higher average GDP per capita compared to other income groups. This highlights the stark economic disparity between countries. There is a clear trend where GDP decreases as we move from high-income countries to low-income ones, demonstrating the global economic inequality.
2. **Internet Usage**: The bar chart for internet usage reveals a direct correlation between income levels and digital connectivity. High-income nations lead with the highest percentage of individuals using the internet, whereas low-income countries face significant challenges in terms of internet penetration. This digital divide is critical, as internet access is vital for education, healthcare, and innovation in today’s world.
3. **Life Expectancy**: Similar to the trends observed in GDP and internet usage, life expectancy increases as income level rises. High-income countries enjoy longer life expectancies, thanks to better healthcare systems, nutrition, and living conditions. In contrast, low-income countries face shorter life expectancies, often due to limited healthcare access, poor sanitation, and other systemic public health challenges.



1. Conclusion

The series of analyses performed on the World Bank dataset reveal clear patterns and persistent disparities in global socio-economic indicators. The key findings from each objective are summarized as follows:

**1. Temporal Trends in Economic Growth**

* **GDP and GDP per Capita over Time:**  
  The time-series analysis by income group shows that high-income countries consistently exhibit higher average GDP and GDP per capita. Trends over the years indicate that economic growth is markedly stronger in these groups, whereas lower income groups show slower and more modest gains. This underscores the structural disparities in wealth accumulation between high-income and low-income economies.

**2. Correlations Among Socio-Economic Indicators**

* **Interconnectedness of Economic and Social Outcomes:**  
  Scatter plots examining the relationship between internet usage, life expectancy, and GDP per capita reveal strong positive correlations. Countries with higher GDP per capita tend to have higher rates of internet usage and longer life expectancies. This finding highlights the interdependency of economic strength, technological connectivity, and health outcomes, emphasizing the multi-dimensional nature of development.

**3. Regional Disparities in Socio-Economic Indicators**

* **Regional Analysis:**  
  The regional analysis provides a deeper insight into how these socio-economic metrics vary across different parts of the world:
  + **Country Distribution:** The pie chart indicates a varied distribution of countries across regions, which reflects inherent geographic and developmental diversity.
  + **GDP per Capita, Internet Usage, and Life Expectancy by Region:**  
    The bar charts clearly illustrate that regions with more developed economies perform significantly better in economic output, digital connectivity, and health standards. Regions with lower average values in these indicators tend to coincide with areas that face more systemic development challenges.

**4. Income Group Comparison**

* **Income-Driven Disparities:**  
  Comparing socio-economic indicators by income group further reinforces the findings from the temporal and regional analyses. High-income groups outperform their lower-income counterparts in GDP per capita, internet usage, and life expectancy. This direct relationship between income level and quality of life markers points to the challenges faced by lower-income countries in accessing the benefits of global technological advancements and economic development.

**5. Regional Comparative Analysis**

* **Visualizing Regional Variability:**  
  The final objective, which focuses on a similar set of indicators by region, reaffirms the existence of significant regional disparities:
  + Regions differ notably in economic conditions, technological penetration, and healthcare outcomes.
  + These discrepancies suggest that geographical location remains a critical factor in determining national performance across socio-economic dimensions, and highlight the importance of tailored regional development strategies.

1. Future scope

The current analysis provides a robust foundation for understanding global socio-economic disparities; however, several avenues remain open for further research and practical applications. Expanding on this initial analysis can yield deeper insights and guide policy, economic development, and social interventions. Below are key areas for future scope:

**1. Expanded Indicator Set**

* **Inclusion of Additional Variables:**  
  Future studies could incorporate other critical socio-economic variables such as education levels, healthcare accessibility, environmental sustainability metrics (e.g., CO₂ emissions, air quality), and political stability indices. This would facilitate a more holistic view of development across regions and income groups.
* **Social and Cultural Factors:**  
  Integrating qualitative indicators like governance quality, cultural factors, and social capital could enrich quantitative assessments, leading to more nuanced policy recommendations.

**2. Temporal Analysis and Forecasting**

* **Time-Series Forecasting:**  
  Utilizing advanced time-series forecasting models, such as ARIMA, Prophet, or LSTM networks, would allow researchers to predict future trends in GDP, life expectancy, and technology adoption. These insights could inform both policymakers and investors.
* **Impact Analysis of Global Events:**  
  Assessing how major events (e.g., global pandemics, economic crises, climate change) influence these indicators over time could help in developing resilient economic and social strategies.

**3. Geospatial and Regional Analysis**

* **Geospatial Mapping:**  
  Leveraging geospatial analysis techniques can illuminate micro-level regional disparities within countries. Advanced GIS tools could map clusters of high and low performance, highlighting specific regions that require targeted interventions.
* **Urban-Rural Divide:**  
  Further granularity in the data could reveal intra-regional differences, particularly the contrast between urban and rural development, which is essential for equitable policy making.

**4. Predictive Modeling and Machine Learning**

* **Predictive Analytics for Policy Simulation:**  
  Machine learning models can be trained to simulate the impact of specific policy decisions on key socio-economic outcomes. These simulations could be used to test the potential outcomes of interventions before their implementation.
* **Causal Inference:**  
  Applying methods such as instrumental variable analysis or structural equation modeling can aid in understanding causal relationships among economic output, internet usage, and public health, going beyond mere correlation.

**5. Interactive Dashboards and Real-Time Monitoring**

* **Dynamic Visualization Tools:**  
  Developing interactive dashboards with real-time data integration would allow stakeholders to monitor changes as they occur, providing a more agile tool for policy response. Such dashboards could aggregate data from various sources, including local government statistics and international databases.
* **Customized Alerts:**  
  Incorporating alert systems into these dashboards can notify policymakers when critical thresholds are crossed, enabling prompt corrective measures and resource allocation.

**6. Comparative and Longitudinal Studies**

* **Inter-Country Comparative Studies:**  
  A more detailed comparative analysis among countries with similar income classifications but diverse outcomes may reveal best practices and adaptive strategies. This could involve case studies focusing on successful economic reforms or social policy innovations.
* **Longitudinal Tracking:**  
  Continuous tracking of specific indicators over longer time horizons would help assess the long-term impact of policy interventions and natural socio-economic evolution. This approach would also highlight emerging trends that could reshape future economic and social landscapes.

**7. Policy Implications and Collaborative Research**

* **International Collaboration:**  
  Encouraging global research collaborations can lead to the development of comprehensive models that incorporate data from multiple sources and regions. Sharing best practices and policy outcomes may drive more effective global development strategies.
* **Policy Simulation and Testing:**  
  Establishing simulation frameworks where policymakers can adjust variables (e.g., investments in digital infrastructure or public health spending) to observe potential outcomes could bridge the gap between academic research and actionable policy decisions.

1. References

a Visualization and Analysis Methods:

b Plotly Documentation: https://plotly.com/python/

c Seaborn Statistical Data Visualization: https://seaborn.pydata.org/

d Pandas Time Series Analysis: <https://pandas.pydata.org/docs/user_guide/timeseries.html>

1. Upload on GitHub

