**📘 Case Study Title:**

**Global Socio-Economic Comparison Dashboard**

**📝 Objective:**

To build a dynamic dashboard that enables users to explore and compare various countries across regions in terms of their population, development indicators, economic structure, and infrastructure availability using a visual, filterable format. The project is aimed at government planners, NGOs, educators, and policy analysts.

**🌍 Dataset Overview:**

The dataset contains information on **multiple countries** and includes the following categories:

* **Demographics**: Population, Population Density, Birthrate, Deathrate, Net Migration
* **Economic Indicators**: GDP per Capita, Employment Sectors (Agriculture, Industry, Services)
* **Health and Education**: Infant Mortality, Literacy Rate
* **Infrastructure**: Phones per 1000 people, Arable land, Crop and Other land usage
* **Geography**: Area, Region, Coastline to area ratio
* **Climate Category**: Simplified climate type (1–4)

**🎯 Project Goals:**

1. Create **regional profiles** that summarize demographic and economic characteristics by region.
2. Identify countries with **high literacy but low GDP** or **high GDP but low infrastructure** (e.g., phone density).
3. Provide **sector-wise employment** visualization to understand development focus (agrarian vs industrial vs service-oriented).
4. Enable dynamic filtering to compare countries based on **climate, region, or population size**.
5. Present **policy implications** and **development opportunities** based on the visual insights.

**🧰 Tools & Technologies ():**

* **Microsoft Power BI / Tableau / Google Data Studio** for dashboard creation
* **Python** for data cleaning and preparation
* **Manual rules-based filters and calculated fields** for categorization and derived metrics

**📊 Dashboard Components:**

**1. Country Profile Card (Dynamic)**

* Country Name, Region
* GDP per Capita
* Literacy Rate
* Infant Mortality
* Employment Sector % (Pie Chart)
* Phones per 1000 people

**2. World Map Visualization**

* Color-coded based on GDP per Capita or Literacy Rate
* Tooltips showing population and climate

**3. Comparative Region Charts**

* Bar/Column charts comparing:
  + Average GDP by Region
  + Average Infant Mortality by Region
  + Average Net Migration by Region

**4. Climate-Based Grouping**

* Filter by climate type (1 to 4)
* Analyze patterns in development, agriculture, and health across climate zones

**5. Employment Sector Trends**

* Ternary or stacked bar chart of % in Agriculture, Industry, Service
* Drill-downs for viewing country-level detail

**6. Custom Filters:**

* **Slider for population size**
* **Dropdown for continent/region**
* **Checkboxes for climate type**
* **Range filter for GDP or Literacy**

**🧩 Insights & Use Cases (Non-ML):**

* **Policy Planning:** Governments can identify needs for infrastructure in high-literacy, low-infrastructure countries.
* **NGO Strategy:** Focus regions with poor infant mortality and low GDP.
* **Education:** Understanding socio-economic patterns across climates and geography.
* **Investment Decisions:** Identify countries transitioning from agriculture to services for potential economic opportunities.

**🔄 Project Workflow:**

1. **Data Cleaning** (Manual/Excel)
   * Fix formatting, handle missing values (e.g., blank climate), and unify decimal separators.
2. **Data Modeling** (in Power BI or Tableau)
   * Create calculated fields like “Total Employment % Check,” “Development Index (Composite),” etc.
3. **Dashboard Design**
   * Use consistent color schemes, interactive components, and layered filters.
4. **Insight Documentation**
   * Write a brief report summarizing key regional insights and disparities.

**📌 Potential Extensions (Still No ML):**

* Add **temporal trendlines** if time-series data is added.
* Allow **scenario analysis** (e.g., if GDP increases 10%, how might infrastructure indicators change).
* Incorporate **qualitative tags** such as "Tourist-heavy", "Oil-based economy", "Post-conflict recovery".

:

**🧾 General Data Overview**

1. What are the column names in the dataset?
2. How many rows and columns are present?
3. Are there any missing or null values in the dataset?
4. What are the data types of each column?
5. Which columns have numeric values and which are categorical?

**🌍 Geographic and Regional Analysis**

1. How many unique regions are represented in the dataset?
2. Which countries belong to each region?
3. Which region has the highest number of countries?
4. What is the total population per region?
5. Which region has the largest land area in total?

**🧑‍🤝‍🧑 Population and Density**

1. Which countries have the highest and lowest population?
2. Which countries have the highest and lowest population density?
3. Is there any country with high population but low density?
4. Are there countries with similar populations but very different densities?

**🐣 Birth, Death, and Migration**

1. Which countries have the highest birth rates?
2. Which countries have the highest death rates?
3. Which countries have the highest net migration rates?
4. Are there any countries with negative net migration?

**💸 Economy and Services**

1. What is the range of GDP per capita across countries?
2. Which countries have the highest and lowest GDP per capita?
3. Are there countries with high GDP but low literacy?
4. Which countries have agriculture as the main contributor to their economy?
5. Which countries are more service-oriented?

**👶 Health and Literacy**

1. Which countries have the highest infant mortality rates?
2. Are there countries with low GDP but high literacy rates?
3. Which countries have both high infant mortality and low literacy?
4. What are the literacy rates per region?

**📱 Technology and Communication**

1. Which countries have the highest phone penetration per 1000 people?
2. Are there any countries with very low phone penetration?
3. Is there a relationship between phone penetration and GDP?

**🌾 Land Usage and Climate**

1. Which countries have the highest percentage of arable land?
2. Are there any countries with 0% arable land?
3. What climate types exist in the dataset and how many countries are in each?
4. Do countries with more arable land tend to have higher birthrates?

**🧩 Data Quality and Consistency**

1. Are there any duplicate rows in the dataset?
2. Are there outliers in any numeric columns (e.g., very high population density)?
3. Are the percentage values in "Arable", "Crops", and "Other" columns adding up to 100%?
4. Are all country names unique and correctly spelled/formatted?