1. Solution Overview

Objective:

Our goal is to examine the relationship between life expectancy and key socioeconomic indicators such as GDP per capita, government healthcare expenditure per capita, and government spending on education per capita. By analyzing these factors, we aim to uncover meaningful insights into global development trends and the interplay between economic resources and population well-being.

Datasets:

Relational Database (RDBMS): World Database, hosted locally on a MySQL server. Contains information about countries, including population, geographical details, population, **life expectancy** and GNP.

API 1 – GDP Data: World Bank API, provides the **Gross Domestic Product (GDP)** in current US dollars for each country for the year 2021.

API 2 – Healthcare Expenditure Data: World Bank API, provides **healthcare expenditure per capita** by governments for each country for the year 2021.

CSV File – Education Spending: Contains data on **government spending on education as a percentage of GDP**, broken down by country and year.

2. Technical Choices Overview

Using Knime, the following workflow is used:

Loading the data:

(i) Imported world database into MySQL workbench, connected MySQL to Knime, extracted the entire country table using query and filtered the country table to relevant columns using DB query reader and Column filter node.

(ii) Loaded CSV file from local system using CSV Reader node and filtered rows to ‘2021’ using Row Filter node. 2021 was used as it had the most data.

Joined both the datasets using ‘Country’ as common column, using Joiner node.

Importing Data from APIs: Filtered both the World Bank APIs to include data for all countries and for the year 2021 using Postman. Then, imported the APIs using a GET request and processed the data with the JSON Reader node in Knime.

Data Transformation:

Data processing: Using Column and Row filter nodes, filtered the columns of the unified dataset to relevant columns and removed missing values, reducing rows from 239 to 150 values, as education spending by the government for some countries was missing.

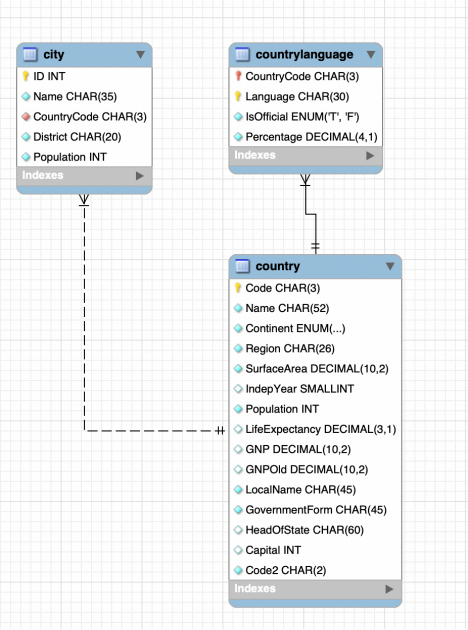
Feature Engineering: Using Math formula nodes, made the following transformations and rounded the numbers using the Number Rounder node:

(i) GDP **>** GDP per Capita (GDP / Population)

(ii) Public spending on education as % of GDP **>** Education spending per capita (Public spending on education as a share of GDP / 100 * GDP / Population)

3. Data Model Overview

We have primarily worked with the ***Country*** table as it contains all relevant information about countries.



Attributes Breakdown:

`Code` (Primary Key): Unique identifier for each country, an ISO 3166-1 alpha-3 code.
`Name`: Official name of the country.
`Continent`: Continent a country belongs to (e.g., Asia, Europe).
`Region`: Describes the specific region within a continent.
`SurfaceArea`: Total land area of the country in square kilometers.
`IndepYear`: The year the country gained independence (if applicable).
`Population`: Total number of people living in the country.
`LifeExpectancy`: Average life expectancy of the population (in years).
`GNP & GNPOld`: Current and historical Gross National Product values.
`LocalName`: The name of the country in its native language.
`GovernmentForm`: Description of the country's governance system (e.g., Republic, Monarchy).
`HeadOfState`: Name of the country's current head of state.
`Capital` (Foreign Key): Links to the **City** table to specify the capital city.
`Code2`: Secondary country code (alpha-2).

4. Visualization Overview

Life Expectancy vs GDP Per Capita Scatter Plot:

- For lower GDP per capita (e.g., 10,000), there is a large variance in life expectancy. Interestingly, some countries achieve higher life expectancy with lower GDP per capita, which probably suggests that economic wealth alone does not guarantee better health outcomes.

Correlation Matrix:

- **GDP Per Capita** and **Education Expenditure Per Capita** (0.93). Very high correlation. As expected, countries with higher GDP Per Capita tend to spend more resources on Education.
- **Healthcare**. Strongly correlated with Education Expenditure and GDP per capita (0.86 and 0.80, respectively). Again, wealthier countries are more likely to spend more on education and healthcare.
- **Life Expectancy**. Moderate correlation with Healthcare expenditure per capita (0.58), GDP per capita (0.54), and Education expenditure per capita (0.54). Higher expenditure in all of these indicators contributes to better health outcomes.

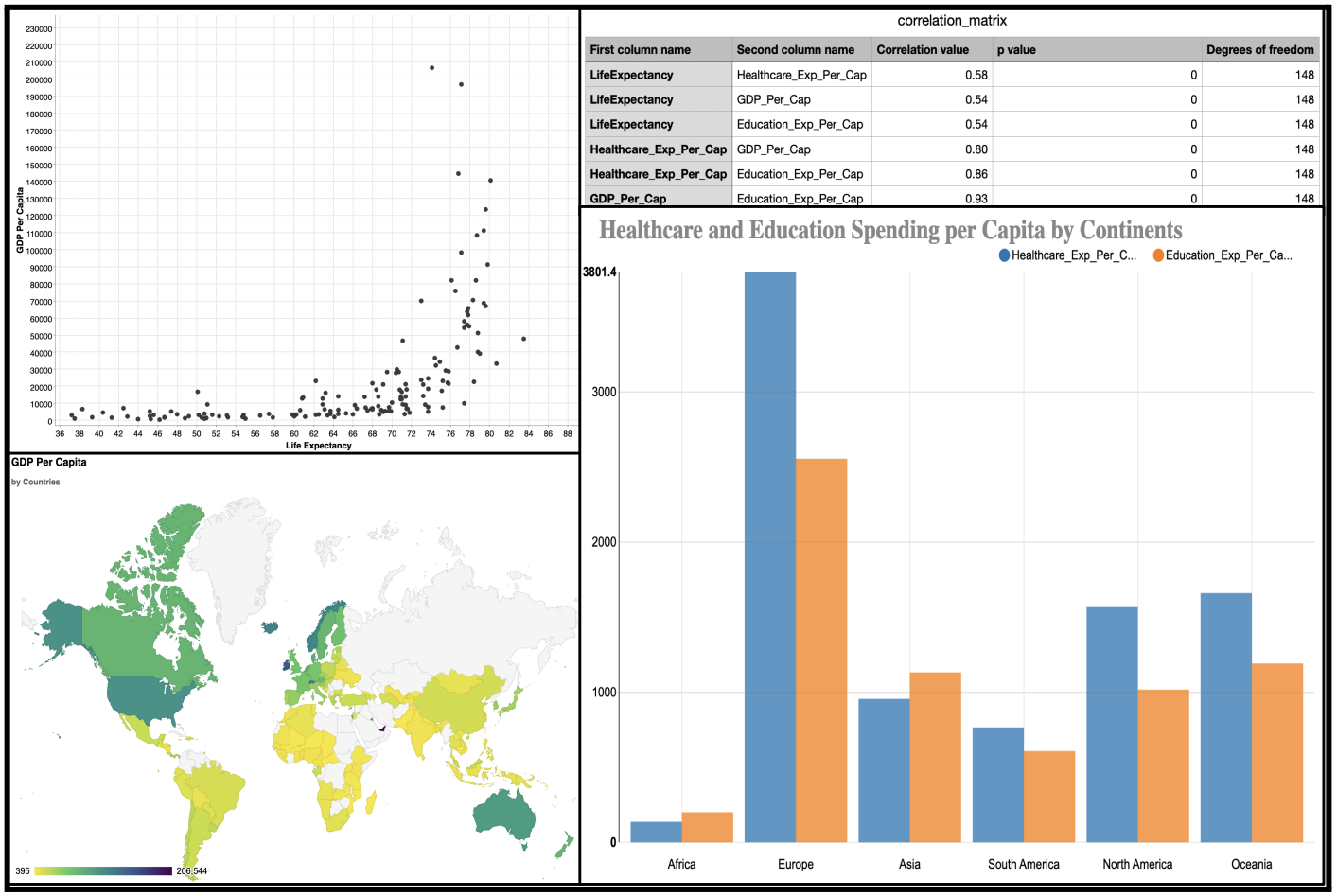
GDP Per Capita Map:

- There are high GDP per capita regions like **North America, Western Europe, Australia, and parts of the Middle East**. There are low GDP per capita regions like **South America, Africa, and parts of Asia**.

(Average) Healthcare and Education Spending by Continent:

- Europe has the highest healthcare and education expenditure per capita among all continents.

Appendix



correlation_matrix

First column name	Second column name	Correlation value	p value	Degrees of freedom
LifeExpectancy	Healthcare_Exp_Per_Cap	0.58	0	148
LifeExpectancy	GDP_Per_Cap	0.54	0	148
LifeExpectancy	Education_Exp_Per_Cap	0.54	0	148
Healthcare_Exp_Per_Cap	GDP_Per_Cap	0.80	0	148
Healthcare_Exp_Per_Cap	Education_Exp_Per_Cap	0.86	0	148
GDP_Per_Cap	Education_Exp_Per_Cap	0.93	0	148

Note: To use the Choropleth Map, you need to download and put it into your workflow.

References

World Bank. (2024). *GDP (current US$)*. Retrieved from https://genderdata.worldbank.org/en/indicator/ny-gdp-mktp-cd.

World Bank. (2024). *Current health expenditure per capita (current US$)*. Retrieved from https://data.worldbank.org/indicator/SH.XPD.CHEX.PC.CD?locations=1W.

World Bank. (2024). *Public spending on education as a share of GDP*. Retrieved from https://ourworldindata.org/grapher/total-government-expenditure-on-education-gdp.

Oracle Corporation. (n.d.). *Other MySQL Documentation*. Retrieved from https://dev.mysql.com/doc/index-other.html.