Mk Mk

IMPACT OF SOCIO-ECONOMIC INDICATORS ON ECONOMIC DEVELOPMENT AMONG DIFFERENT COUNTRIES

How much of this submission has been generated by AI?

***3%**

of qualifying text in this submission has been determined to be generated by AI.

* Low scores have a higher likelihood of false positives.

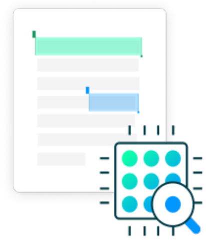
Caution: Percentage may not indicate academic misconduct. Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Frequently Asked Questions

What does the percentage mean?

The percentage shown in the AI writing detection indicator and in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was generated by AI.



Our testing has found that there is a higher incidence of false positives when the percentage is less than 20. In order to reduce the likelihood of misinterpretation, the AI indicator will display an asterisk for percentages less than 20 to call attention to the fact that the score is less reliable.

However, the final decision on whether any misconduct has occurred rests with the reviewer/instructor. They should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in greater detail according to their school's policies.

How does Turnitin's indicator address false positives?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be AI-generated will be highlighted blue on the submission text.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

What does 'qualifying text' mean?

Sometimes false positives (incorrectly flagging human-written text as AI-generated), can include lists without a lot of structural variation, text that literally repeats itself, or text that has been paraphrased without developing new ideas. If our indicator shows a higher amount of AI writing in such text, we advise you to take that into consideration when looking at the percentage indicated.

In a longer document with a mix of authentic writing and AI generated text, it can be difficult to exactly determine where the AI writing begins and original writing ends, but our model should give you a reliable guide to start conversations with the submitting student.

IMPACT OF SOCIO-ECONOMIC INDICATORS ON ECONOMIC DEVELOPMENT AMONG DIFFERENT COUNTRIES

By (Name)

The Name of the Class (Course)

Professor (Tutor)

The Name of the School (University)

The City and State Where it is Located

The Date

Table of Contents

Part One: Introduction and Data Exploration

1.1 Introduction

This report begins with an exploration of the connections between data, socioeconomic progress, and the spread of information. Non-Governmental Organization (NGO) serve as the foundation for this report. An organization dedicated to promoting worldwide social and economic development (Amin, 2017). The steadfast commitment of this NGO to enhancing people's lives all throughout the world highlights the relevance of the study, the insights desired to acquire the value of clear communication (Siddika et al., 2018).

This research aims to address a multidimensional problem. This study intends to use data to find insights that will not only provide a thorough understanding of socioeconomic development but will also convey knowledge to the general public. This study strives to answer concerns about economic growth, poverty reduction, employment, education, population dynamics, and public health by analysing key indicators. The significance of this data rests in its ability to enlighten decision-makers, empower individuals, and drive change.

An approach to addressing this problem comprises two distinct yet interconnected tasks. Task 1 focuses on statistical analysis, while Task 2 delves into the design of an interactive dashboard. These tasks align with the overarching objectives of the research:

Research Objectives

Statistical Analysis

1. To investigate the relationship between social- economic indicators to understand their impact on economic development.

2. To investigate the association between the social economic indicators.

3. To conduct hypothesis testing on selected socio-economic indicators to draw meaningful conclusions about the population based on sample data.

Hypothesis 1:

Null Hypothesis (H0): There is no significant difference in the mean Life Expectancy (LE) between countries with a Poverty Headcount (PH) below the global average and countries with a PH above the global average.

Alternative Hypothesis (H1): There is a significant difference in the mean Life Expectancy (LE) between countries with a PH below the global average and countries with a PH above the global average.

Hypothesis 2:

Null Hypothesis (H0): There is no significant difference in the mean Education Enrollment Rate (EEL) between countries with a high Gross Domestic Product (GDP) and countries with a low GDP. *Alternative Hypothesis (H1):* There is a significant difference in the mean Education Enrollment Rate (EEL) between countries with a high GDP and countries with a low GDP.

4. To conduct a regression analysis that models the relationships between socio-economic variables and provide predictive insights.

The belief that intensive analysis may shed light on the existing relationships between diverse socioeconomic elements, offering guidance for more effective development strategies and policy decisions, drives the dedication to these goals.

Interactive Dashboard Design

1. To produce an interactive dashboard that provides a comparative analysis of economic and social indicators for selected countries.

2. The dashboard should be designed in such a way that users can quickly compare the performance of different countries across multiple metrics and years.

A wide audience can gain access to statistical insights through the use of an interactive dashboard, which seeks to address the knowledge gap between data and statistics (Vila, Estevez, and Fillottrani, 2018; Sarikaya et al., 2018). With the help of this tool, users will be able to examine and understand the crucial connections between socioeconomic indices, promoting social and economic progress on a worldwide scale.

1.2 Background Research and Literature Review

This section provides background research and literature review for the methods employed in this research, such as regression analysis, correlation analysis, and interactive dashboard design.

1.2.1 Background Research

1. Correlation Analysis

Finding the degree of association between variables is a crucial component of the study of socio-economic development, and correlation analysis fulfils this. Pearson's correlation coefficient , is the basis of correlation analysis. Todaro and Smith (2014) state that this coefficient provides crucial information about the relationship between changes in one variable and changes in another by measuring the direction and strength of a linear relationship between variables.

Correlation analysis provides a way to examine relationships between socio-economic data. This method is very helpful in figuring out whether these indicators have a linear relationship, which helps to clarify the complex web of socio-economic growth (Todaro and Smith, 2014).

A robust literature review of previous studies and research efforts have harnessed correlation analysis to explore similar relationships. Studies by Festin et al., (2017); Tuo and He (2021) are notable examples where correlation analysis was effectively employed to investigate socio-economic connections, reinforcing the method's validity and significance in understanding the interrelationships between your chosen indicators.

2. Hypothesis Testing

Hypothesis testing provides a structured approach for making inferences about populations based on sample data (Levine, 2022). In the context of this research, hypothesis testing serves as a crucial tool for drawing robust conclusions regarding socio-economic development.

This study demands the careful selection of statistical tests that are appropriate for the research objectives and data type. Consider t-tests for comparing means between two groups (for example, comparing poverty rates in rich and developing countries).

Numerous studies employed these hypothesis techniques to explore pertinent questions. Studies such as Galkina, (2022) utilized hypothesis testing to investigate socio-economic relationships, further substantiating the relevance of these methods in this context. These studies serve as a testament to the effectiveness of hypothesis testing and provide a solid foundation in this research.

3. Regression Analysis

Understanding socioeconomic development requires the modeling of interactions between dependent and independent factors, which regression analysis makes possible (Judd et al., 2017). Many forms of regression, including multiple and linear regression, each with a unique set of characteristics and applications, are covered in theoretical foundations. The best method for figuring out the relationship between one independent variable and one dependent variable is linear regression.

In this research, this can be particularly valuable for understanding the direct influence of factors like GDP or education enrollment on socio-economic outcomes. Whereas multiple regression extends its capabilities to encompass multiple independent variables, facilitating a more comprehensive examination of socio-economic development.

The validity and appropriateness of the chosen regression techniques are reinforced by the wealth of existing research in the field of socio-economic development. Studies by Yuan et al., (2021) employed regression analysis to investigate analogous questions, strengthening the rationale for this approach.

4. Interactive Dashboard Design

The design of interactive dashboards is instrumental in conveying complex socio-economic data to users in a clear and user-friendly manner. According to Vila, Estevez and Fillottrani (2018) this exploration should encompass a range of crucial elements, including composition, layout, and design principles. Additionally, it should provide insights into the methodology of dashboard design and development, and also offer a glimpse into the current perspectives that influence the field.

Effective interactive dashboards are distinguished by their composition, layout, and adherence to design principles that facilitate comprehension and usability (Vila, Estevez and Fillottrani, 2018). Literature review investigates key aspects as follows;

- Data Visualization Best Practices. An essential foundation of interactive dashboards is the effective visualization of complex data. The literature offers a wealth of insights into data visualization best practices, including the use of clear labels, appropriate chart types, and the avoidance of visual clutter. Works by Sedrakyan, Mannens and Verbert (2019) provide valuable links in this area.

- Colour Choices for Enhanced Comprehension. Colour theory plays a pivotal role in data visualization and its application in dashboards. Studies by Nadj, Maedche and Schieder (2020) are renowned for their contributions to the understanding of colour in data visualization.

- Incorporation of Interactive Elements. Interactive elements, such as filters, drill-down features, and tooltips, are fundamental to a user-friendly experience. A thorough literature review explores how these elements are effectively integrated into dashboard design to enable users to explore and manipulate data. Pioneering works by (Nadj, Maedche and Schieder, 2020) offer insights into interactive data visualization principles.

The methodology employed in the design and development of interactive dashboards is a critical aspect of this research. The explanation of this methodology elucidates how it aligns with the defined objectives for the dashboard design. Furthermore, making link to well-established design methodologies that have played a pivotal role in the design process.

The final facet for this delves into contemporary perspectives on interactive dashboard design. By exploring the latest trends and innovative practices, this section provides a window into the current state of the field. The discussion of these perspectives demonstrates how they have influenced the design choices for the interactive dashboard, infusing it with modern ideas and relevance (Nadj, Maedche and Schieder, 2020).

1.2.2 Literature Review

This section analyses past studies done in order to gain a comprehensive understanding of the research topic. This research encompasses several key socio-economic indicators, including GDP, Poverty Rate, Unemployment Rate, Education Enrolment, Population Growth Rate, and Life Expectancy, which are pivotal for understanding the economic development dynamics of different nations.

To address research objectives and design an effective method to fulfill them, a thorough exploration of relevant literature is undertaken. Esteemed economists, and researchers, such Pacifico (2023); Wen et al., (2021) have made significant contributions to the understanding of these indicators' role in shaping the socio-economic landscape. Their works provide valuable insights into the determinants of GDP growth, the impact of poverty rates on societal well-being, and the consequences of unemployment on economic stability, among other aspects.

Furthermore, this research expands its scope to encompass the creation and implementation of a dashboard that effectively showcases these socio indicators. To ensure that the interactive dashboard aligns, with established practices and design principles we have explored existing dashboards and data visualizations that are relevant to this research field. Notable case studies conducted by Belghith et al. (2022) and IvyProSchool (2023) provide insights on how to develop a user informative dashboard for presenting socio-economic data.

Studying these case studies on dashboards enables us to gain inspiration and gather practices that will guide our dashboard design in this research project ensuring it meets the prevailing standards, in the realm of socio-economic development analysis.

1.2 Preparation and Exploration of Data Set

This section provides a comprehensive overview of the data set, including the variables, definitions, metadata, time frames, and data sources. The section also describes the steps taken for data preparation, outlier detection, and handling missing data. Following data preparation, Exploratory Data Analysis (EDA) is conducted to uncover interesting insights about the data set, presenting them through appropriate graphs.

1.3.1 Data Dictionary

Variable Name	Definition	Metadata	Time Frame	Data Source
GDP	Gross Domestic Product (GDP) represents the total monetary value of goods and services produced within a country's borders. It serves as a key indicator of a nation's economic performance.	NY.GDP.MKTP.CD	2006-2020	World Development Indicators (WDI) / United Nations Bank Data

Poverty Head count Ratio (% of population)	It represents the proportion of people in a specific area or country who are living below the poverty line or threshold.	SI.POV.DDAY	2006-2020	World Development Indicators (WDI)
Unemployment Rate	The Unemployment Rate indicates the percentage of the labour force that is unemployed and actively seeking employment	SL.UEM.TOTL. ZS	2006-2020	World Development Indicators (WDI)
Education Enrolment	Education Enrolment tracks the percentage of eligible individuals who are enrolled in educational institutions.	SE.PRM.ENRL	2006-2020	World Development Indicators (WDI)
Population Growth Rate	The Population Growth Rate measures the rate at which a country's population is increasing over a specified period.	SP.POP.GROW	2006-2020	World Development Indicators (WDI)
Life Expectancy	Life Expectancy represents the average number of years a	SP.DYN.LE00.IN	2006-2020	World Development Indicators (WDI)

	person can expect to live, given the current mortality rates.			

1.3.2 Data Preparation

The United Nations Bank Data (UNdat) and World Development Indicators (WDI) were the sources of the dataset, which spans the years 2006 to 2020. To ensure the data's quality, several data cleaning and preparation steps were undertaken.

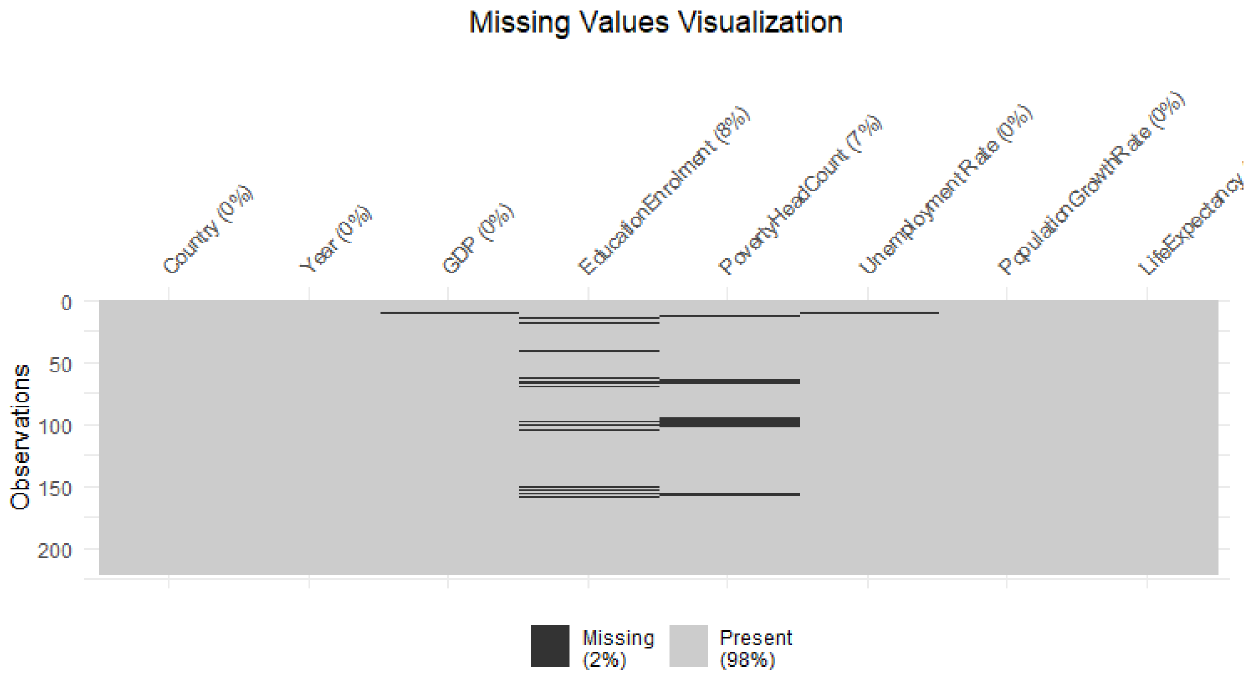
Handling missing data

Missing Data Detection Algorithm**:**

missing_counts <- colSums(is.na(my_data))

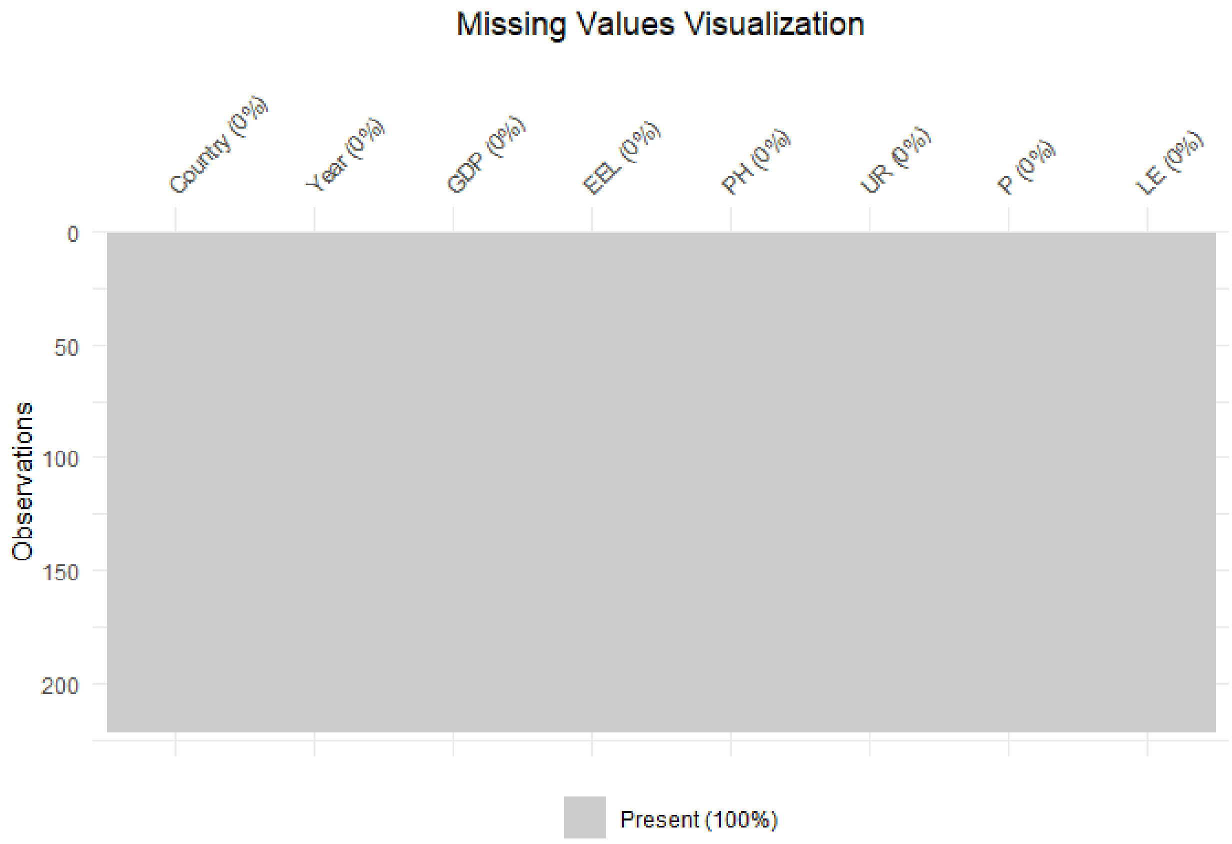
print(missing_counts)

```
Country                  Year                  GDP   EducationEnrolment
                    0                     0                     1
18
    PovertyHeadCount    Unemployment Rate PopulationGrowthRate       LifeExpe
ctancy
                   16                     1                     0
0
```



The Heatmap above provides a clear visual representation of missing values across different variables and time points.

Where historical data was available, missing values were imputed by referring to data from previous years. Below is a visualization after replacing the missing values.



Outlier Detection and Handling:

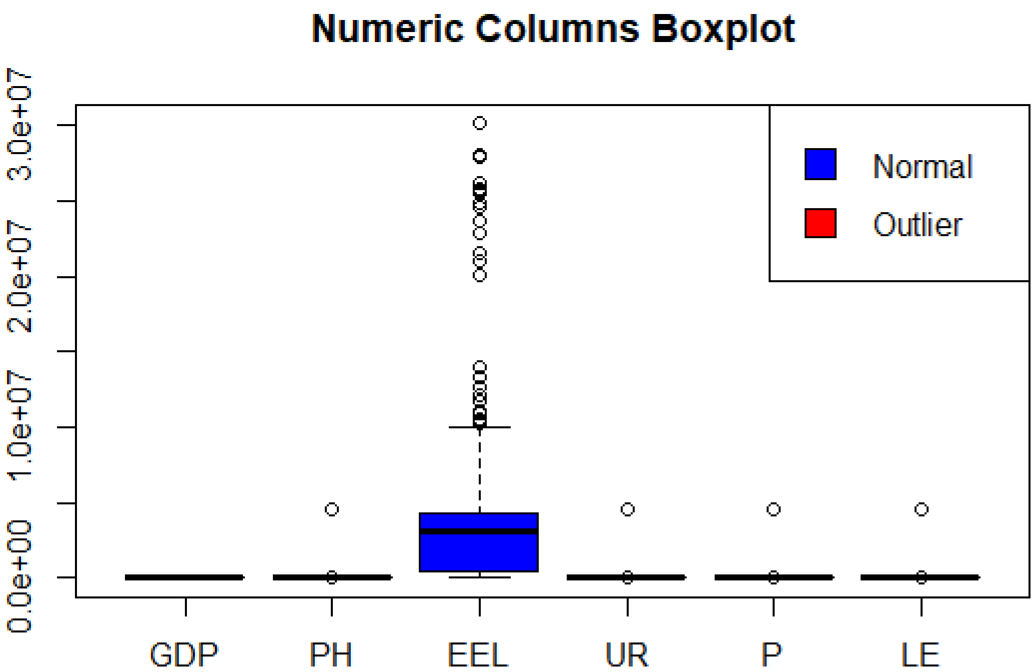
Outlier Detection Algorithm:

```
# Detect outliers using Z-score

z_scores <- scale(dataset)

outliers <- abs(z_scores) > 2 #
```

GDP	PH	EEL	UR	P	LE
0	4	10	3	7	5



Above is a boxplot showcasing outliers.

2.Outlier Handling: Median replacement

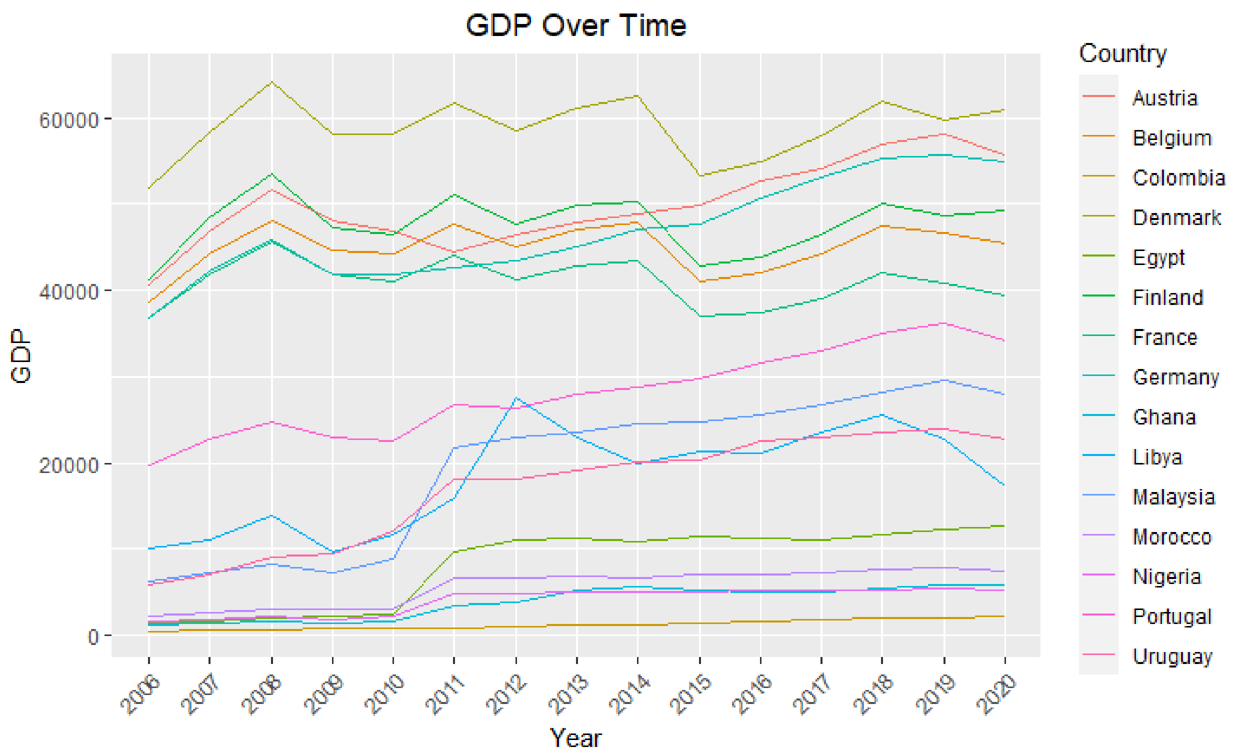
Replace outliers with the median

dataset[outliers] <- median(my_data, na.rm = TRUE)

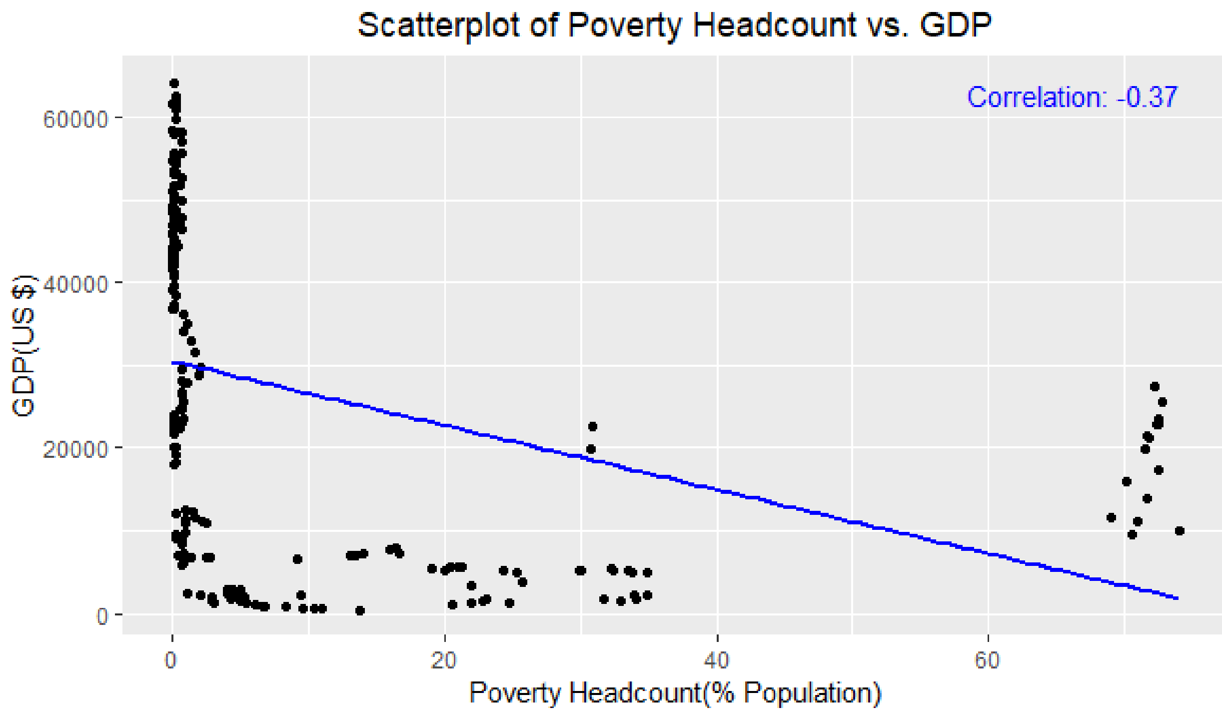
```
[1] "Handling outliers in column: PH"
[1] "Original median: 1.0118404251491"
[1] "New median: 1.0118404251491"
[1] "Handling outliers in column: UR"
[1] "Original median: 8.57"
[1] "New median: 8.57"
[1] "Handling outliers in column: P"
[1] "Original median: 1.13154130595269"
[1] "New median: 1.13154130595269"
[1] "Handling outliers in column: LE"
[1] "Original median: 76.471"
[1] "New median: 76.471"
```

1.3.3 Exploratory Data Analysis (EDA)

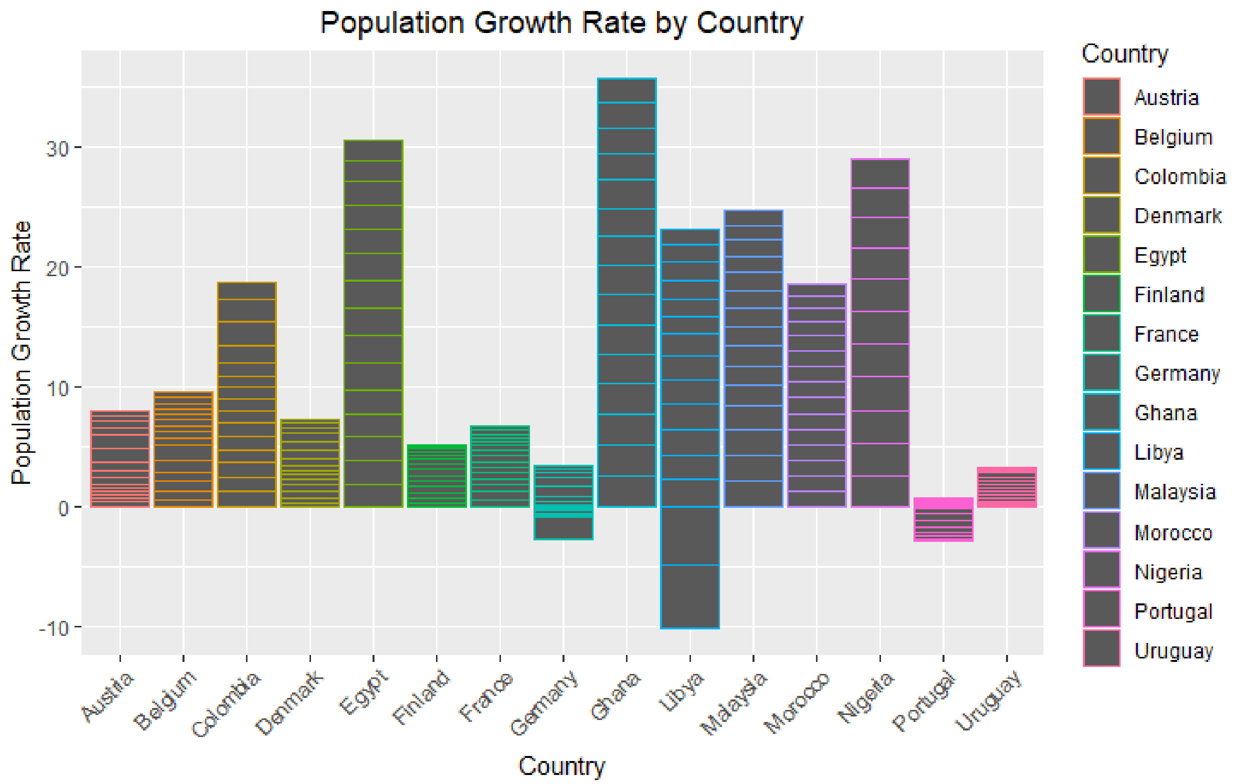
Key findings concerning the dataset were made during Exploratory Data Analysis (EDA). Countries' GDP increased gradually over time, with varying rates of growth.

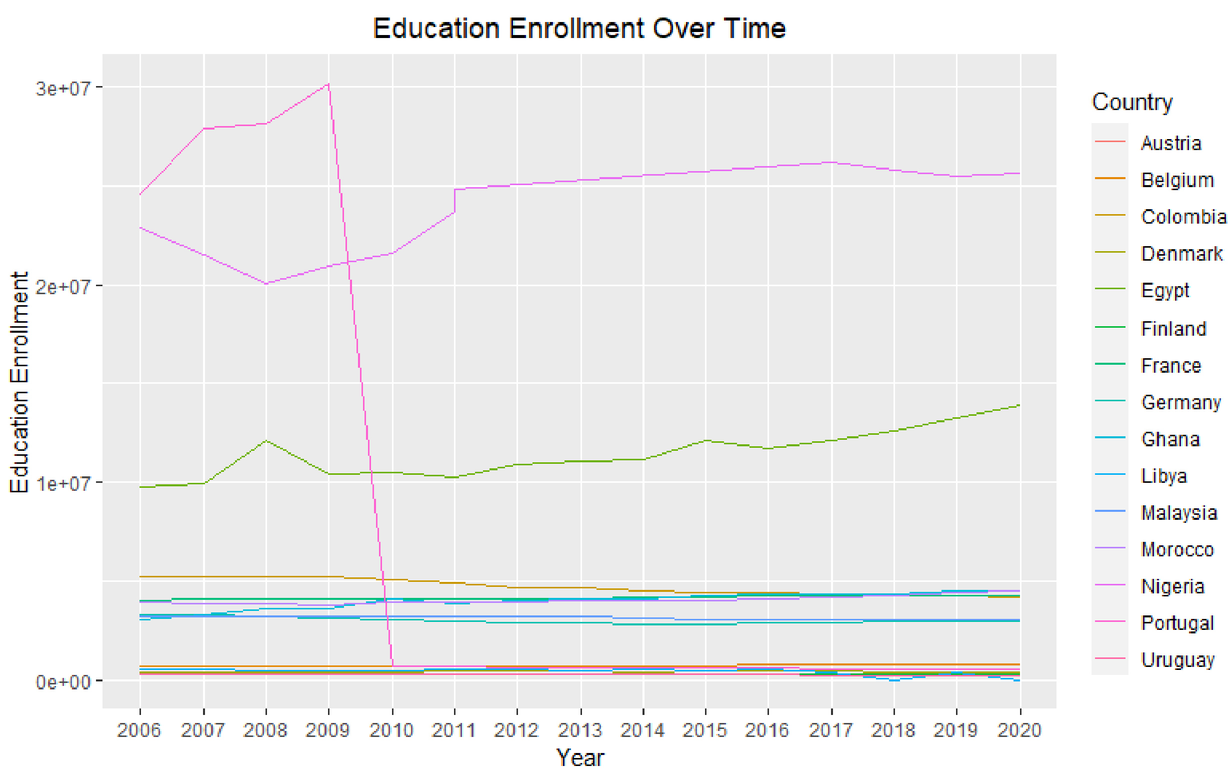


There was a clear negative link between GDP and poverty Head count, meaning that lower rates of poverty are correlated with higher GDP. Different labor market dynamics were reflected in the disparities in unemployment rates among countries.



Education Enrollment rates showed an overall rising trend, indicating more access to education. The rates of population increase varied, with some countries experiencing fast expansion while others stayed steady.


Population Growth Rate by Country
Population Growth Rate
30
20
10
0
-10
Austria
Belgium
Colombia
Denmark
Egypt
Finland
France
Germany
Ghana
Libya
Malaysia
Morocco
Nigeria
Portugal
Uruguay
Country


Education Enrollment Over Time
Education Enrollment
3e+07
2e+07
1e+07
0e+00
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Year
Country
Austria
Belgium
Colombia
Denmark
Egypt
Finland
France
Germany
Ghana
Libya
Malaysia
Morocco
Nigeria
Portugal
Uruguay

Life Expectancy demonstrated an overall improvement, highlighting global advancements in health outcomes. These insights from EDA guided the subsequent statistical analysis and dashboard design.

Part Two: Statistical Analysis

2.1 Descriptive Analysis

	Variable	Mean	Median	Mode	SD	Skewness	Kurtosis
GDP	GDP	2.673818e+04	2.400482e+04	4.922702e+03	1.991125e+04	0.1437042	-1.471618
EEL	EEL	4.245787e+06	3.014502e+06	1.216138e+07	6.262302e+06	2.4738465	5.641377
PH	PH	9.723159e+00	6.980000e-01	1.000000e-01	1.916594e+01	2.3525909	4.640863
UR	UR	9.219715e+00	8.520000e+00	9.125000e+00	4.380699e+00	1.4425474	2.816673
P	P	9.498434e-01	7.064238e-01	2.764062e+00	1.043959e+00	-1.4248941	8.106217
LE	LE	7.513784e+01	7.748300e+01	6.912800e+01	7.659711e+00	-1.6500734	2.427875

R steps

1.Load data

2.Calculate summary statistics using inbuilt R commands such as mean, median, kurtosis, skewness.

A comprehensive descriptive statistical analysis has examined key social-economic indicators across various countries, providing valuable insights into the comparative aspects of socio-economic development.

The analysis reveals an average GDP of approximately $26,738 suggesting considerable economic diversity across countries. The distribution is slightly right-skewed, indicating that a few countries possess significantly higher GDP values. It's noteworthy that the most frequently occurring GDP value, the mode, is around $49227, which may represent a typical economic output for certain countries.

Education Enrolment (EEL) displays substantial variation as well, with a mean enrolment of about 4,245,787. The median enrolment, however, is considerably lower at around 3,014,502, emphasizing disparities in educational access and enrolment rates. The distribution of EEL is highly right-skewed, indicating that a few countries have substantially higher enrolment figures, which may be due to disparities in educational infrastructure and policies.

The analysis shows an average poverty rate of 9.72%. The median rate is much lower at 0.698%, highlighting the wide-ranging poverty rates across countries. The distribution exhibits positive skewness, signifying that some nations have notably higher poverty rates, while the kurtosis indicates that the distribution has relatively heavy tails.

Unemployment Rate (UR) showcases an average rate of 9.22%, with a median of 8.52%. This reveals a variation in the job market dynamics among countries, with some nations experiencing higher unemployment rates. The skewness, though positive, is relatively low, and the kurtosis suggests a distribution that approximates normality.

Population Growth (P) is a crucial demographic factor, and the analysis demonstrates an average annual growth rate of about 0.95%. The skewness is significantly negative, indicating that some countries experience rapid population growth, while the kurtosis signifies heavy-tailed behaviour, potentially influenced by outliers.

Life Expectancy (LE) reflects the overall well-being of a population, with an average life expectancy of 75.14 years. The median, however, is notably higher at approximately 77.48 years, indicating variations in life expectancies. The distribution is left-skewed, suggesting that some countries have notably higher life expectancies, while the kurtosis indicates slight heavy-tailed behaviour.

2.2 Correlation Analysis

Correlation analysis provides a complete understanding of the correlations between socioeconomic indicators in different countries, giving light on the complex web of factors that lead to socioeconomic development.

There is a negative association between a country's GDP and its Poverty Headcount (PH). This finding implies that, on average, as a country's GDP rises, poverty rates fall. This finding is consistent with the primary purpose of economic development, underlining the importance of strong economic growth in alleviating poverty. Higher GDP frequently indicates more economic possibilities, better access to resources, and higher living standards, all of which help to reduce poverty.

Similarly, there is a negative link between GDP and the Population Growth Rate (P). The population growth rate of a country tends to drop as its GDP rises. This observation implies that increased economic development may result in demographic shifts such as lower birth rates. This phenomenon is frequently linked to improved healthcare, education, and family planning access, all of which contribute to more sustainable population growth.

GDP and Life Expectancy (LE) have a high positive association. This implies that as a country's GDP rises, life expectancy rises dramatically. The favourable association between economic development and higher life expectancy emphasizes the necessity of addressing not only monetary prosperity but also healthcare access and quality, nutrition, and living conditions.

GDP, on the other hand, has a negative association with Education enrolment Rate (EEL). The EEL falls little when a country's GDP rises. While this association may appear to be paradoxical, it demonstrates the intricacies of the relationship between economic progress and education. This link can be influenced by factors such as different education systems, access to excellent education, and cultural influences.

There is a positive link between EEL and Population Growth Rate (P), showing that higher enrolment rates in higher education may lead to increasing population growth. This discovery shows that comprehensive family planning and education strategies are required to properly manage demographic shifts.

In addition, the inverse link between Education enrolment Rate (EEL) and Life Expectancy (LE) raises some intriguing questions. It means that when school attendance increases, life expectancy declines marginally. More research is needed to understand the factors that contribute to this connection, which may change between countries and locales.

The connections between Poverty Headcount (PH) and other variables yielded valuable information. There was a positive association between PH and Population Growth Rate (P), showing that higher poverty rates may be related with increasing population growth. This research underlines the possible difficulties in addressing poverty and family planning at the same time.

R analytic steps

1.Load data

2. Calculate the correlation matrix

correlation matrix <- cor(my_data[, c("GDP", "EEL", "PH", "UR", "P", "LE")]) correlation_matrix

```
            GDP        EEL         PH          UR           P          LE
GDP  1.0000000 -0.4365984 -0.3727844 -0.15794304 -0.55178573  0.71711290
EEL -0.4365984  1.0000000  0.1654225 -0.01360390  0.41229290 -0.63184665
PH  -0.3727844  0.1654225  1.0000000  0.32843264  0.24018494 -0.50868577
UR  -0.1579430 -0.0136039  0.3284326  1.00000000 -0.08391498  0.04319221
P   -0.5517857  0.4122929  0.2401849 -0.08391498  1.00000000 -0.67257508
LE   0.7171129 -0.6318467 -0.5086858  0.04319221 -0.67257508  1.00000000
```

2.3 Hypothesis Testing

Hypothesis 1

The null hypothesis (H0) stated that there is no significant difference in mean life expectancy (LE) between countries with Poverty Headcount below the world average and countries with PH above the global average. The alternative hypothesis (H1), on the other hand, claimed that there is a considerable variation in mean life expectancy between two sets of countries.

```
        Welch Two Sample t-test

data:  le_below_average and le_above_average
t = 10.081, df = 57.12, p-value = 2.728e-14
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 10.18316 15.23105
sample estimates:
mean of x mean of y
 78.12775  65.42064
```

The findings of the two-sample t-test revealed a statistically significant difference in life expectancy between these two groups of countries. Lower-poverty countries had a mean life expectancy of 78.13 years, up from 65.42 years in higher-poverty countries. The statistics corroborate the alternative hypothesis, which emphasizes the impact of poverty reduction programs on overall well-being.

Several critical pieces of information are provided in the output to support these hypotheses. To begin, the t-statistic value of 10.081 represents the magnitude of the observed difference. Second, the degrees of freedom (df) for the test are roughly 57.12, confirming the statistical analysis's trustworthiness. Most importantly, the p-value (2.728e-14) provides strong evidence against the null hypothesis. This p-value emphasizes the data' robustness and implies a significant difference in life expectancy between nations with differing poverty rates.

Additionally, the 95 percent confidence interval (ranging from 10.18316 15.23105) for the difference in means of life expectancy between the two groups supports the conclusion drawn from the p-value. Finally, the sample means for Life Expectancy (LE) in countries with Poverty headcount below the global average (78.12775) and PH above the global average (65.42064) corroborate the hypothesis that countries with lower poverty rates tend to have significantly higher life expectancies.

R steps

1.Create two vectors for life expectancy (LE) based on PH groups

2. Perform the two-sample t-test

3.Print the t-test result

The findings from Hypothesis 1 reinforce the significance of addressing poverty as a means to improve the overall well-being and quality of life for individuals in different countries. This evidence-based insight can inform the formulation of policies and strategies aimed at reducing poverty and promoting socio-economic development, ultimately contributing to the social and economic development and greater cause of global development.

Hypothesis 2

The output of the Welch Two Sample t-test for Hypothesis 2 provides valuable insights into the difference in the mean Education Enrolment Rate (EEL) between countries with high

Gross Domestic Product (GDP) and countries with low GDP. This test is useful in addressing research objectives, particularly the relationship between economic development and access to education, which is a critical component of socioeconomic development.

```
        Welch Two Sample t-test

data:  high_gdp_countries$EEL and low_gdp_countries$EEL
t = -7.1501, df = 126.91, p-value = 6.079e-11
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6604667 -3741360
sample estimates:
mean of x mean of y
  1483725   6656739
```

With a t-statistic of around -7.1501, a degree of freedom of roughly 126.91, and a p-value of 6.079e-11, the data support the rejection of the null hypothesis. Our hypothesis that there is a substantial difference in mean school enrolment between high- and low-GDP countries is supported by this low p-value.

This disparity is highlighted even more by the 95 percent confidence interval, which varies from -6604667 to -3741360. This shows that the school enrolments of the two groups differ greatly. This is consistent with the alternative view, which holds that residents of high-income countries enrol in school at far higher rates than those in low-income countries.

The implications of this result are substantial. It underscores the importance of economic development in providing better access to education, a critical driver of socio-economic progress. Higher education enrolment rates in high GDP countries reflect the availability of resources and opportunities for individuals to pursue education. These findings can guide policy decisions and interventions aimed at enhancing educational opportunities and fostering socio-economic development in regions where access to education remains limited.

The findings underline the significance of closing the educational gap that exists between nations with different levels of wealth and prospective directions for further study and the development of public policy to improve opportunities and well-being worldwide.

2.4 Regression Analysis

2.4.1 Linear Regression

To model the relationships between socio-economic variables and provide predictive insights, specifically focusing on the impact of socio-economic indicators on life expectancy. Linear regression is suitable for assessing linear relationships between variables (Lilja and Linse, 2022). In this context, it can help understand how socio-economic indicators impact life expectancy, assuming a linear association.

The 'Intercept' coefficient represents the expected Life Expectancy when GDP and EEL are zero. In this context, it's not meaningful. The 'GDP' coefficient (2.097e-04) represents the estimated change in Life Expectancy for a one-unit increase in GDP, holding other variables constant. This positive coefficient suggests that as GDP increases, Life Expectancy tends to rise. The 'education enrolment' coefficient (-4.817e-07) represents the estimated change in Life Expectancy for a one-unit increase in education enrolment, holding other variables constant. This negative coefficient indicates that as Education Enrolment Rate increases, Life Expectancy tends to decrease.

```
Call:
lm(formula = LE ~ GDP + EEL, data = my_data)

Residuals:
     Min       1Q   Median       3Q      Max
-12.1763  -1.8266   0.2133   2.7245  16.8255

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.158e+01  6.803e-01 105.214  < 2e-16 ***
GDP          2.097e-04  1.738e-05  12.066  < 2e-16 ***
EEL         -4.817e-07  5.527e-08  -8.716 7.36e-16 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.618 on 218 degrees of freedom
Multiple R-squared:  0.6398,	Adjusted R-squared:  0.6365 
F-statistic: 193.6 on 2 and 218 DF,  p-value: < 2.2e-16
```

The p-values associated with these coefficients are very small, indicating statistical significance. This means that both GDP and education enrolment have a significant impact on Life Expectancy. The R-squared value (0.6398) represents the goodness of fit of the model. It indicates that approximately 63.98% of the variation in Life Expectancy can be explained by the linear relationship with GDP and education enrolment. This is a strong fit, suggesting that the model is effective in explaining Life Expectancy based on these variables.

R Analytics Steps:

1. Loaded necessary libraries (e.g., *dplyr*).
2. Created a Linear Regression model using the *lm()* function.
3. Generated a summary of the model using *summary ()* to extract coefficients, p-values, R-squared values, and other relevant statistics.

Research study by Verbeek (2017) have delved into the intricate connections between socio-economic indicators and public health outcomes, providing valuable insights that corroborate the choice of employing linear regression. In this study linear regression was utilized as a robust analytical tool to explore the multifaceted relationship between variables such as Gross Domestic Product School Enrolment Rate and Life Expectancy.

2.4.2 Multiple Regression

Multiple Regression was used to assess the combined influence of multiple socio-economic indicators on a specific aspect of socio-economic development. Multiple regression is suitable for understanding how several independent variables interact and collectively affect a dependent variable.

```
Call:
lm(formula = GDP ~ +EEL + PH + UR + P + LE, data = my_data)

Residuals:
   Min     1Q Median     3Q    Max
-27941  -9386   1124   9372  26179

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.032e+05  1.790e+04  -5.767 2.79e-08 ***
EEL          1.699e-04  1.913e-04   0.888   0.3753
PH           9.415e+01  6.222e+01   1.513   0.1317
UR          -1.041e+03  2.254e+02  -4.618 6.66e-06 ***
P           -2.512e+03  1.174e+03  -2.139   0.0335 *
LE           1.867e+03  2.234e+02   8.358 7.94e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13260 on 215 degrees of freedom
Multiple R-squared:  0.5668,   Adjusted R-squared:  0.5568
F-statistic: 56.27 on 5 and 215 DF,  p-value: < 2.2e-16
```

The multiple regression analysis conducted aimed to assess the collective impact of several socio-economic indicators, including Education enrolment Rate (EEL), Poverty Headcount (PH), Unemployment Rate (UR), Population Growth Rate (P), and Life Expectancy (LE), on Gross Domestic Product (GDP). This analysis is crucial for understanding how a combination of these indicators jointly influences a country's economic development.

The results revealed important insights. Unemployment Rate (UR) exhibits a statistically significant and strong negative relationship with GDP. An increase in UR is associated with a substantial decrease in GDP, even when considering other factors. Conversely, Life Expectancy (LE) has a highly significant positive effect on GDP. An increase in LE leads to a significant increase in GDP, highlighting the substantial positive impact of life expectancy on economic development.

However, other indicators show less pronounced relationships with GDP. Education enrolment Rate (EEL) and Poverty Headcount (PH) do not demonstrate statistically significant impacts on GDP in this analysis. Their coefficients are not statistically robust, indicating that changes in these variables may not be significantly associated with GDP variations when considering the other factors. Population Growth Rate (P) shows a significant, albeit relatively weak, negative relationship with GDP. While higher population growth is linked to lower GDP, the effect is less pronounced compared to UR and LE.

The model's multiple R-squared value of 0.5668 implies that roughly 56.68% of the variability in GDP can be explained by the combination of these socio-economic indicators. The model appears to have a moderate overall fit to the data. The adjusted R-squared value, which accounts for the number of predictors, confirms that the model maintains its explanatory power after considering the variables included.

The analysis indicates that Unemployment Rate (UR) and Life Expectancy (LE) are significant predictors of GDP, with strong and statistically significant associations. Education enrolment Rate (EEL) and Poverty Headcount (PH) do not exhibit statistically significant relationships with GDP, and Population Growth Rate (P) shows a significant yet weaker relationship. These findings provide valuable insights into the complex interplay between socio-economic indicators and their collective influence on economic development.

R Analytics Steps:

1. Loaded the necessary libraries (e.g., dplyr).

2. Created a Multiple Regression model using the lm() function, considering GDP as the dependent variable and the selected socio-economic indicators as independent variables.

3. Generated a summary of the model using summary () to extract coefficients, p-values, R-squared values, and other relevant statistics.

Multiple regression is a widely accepted and employed statistical technique in socio-economic research, with studies such as Ni (2020) utilizing multiple regression to analyze the collective impact of various socio-economic indicators on crucial outcomes, including but not limited to poverty rates, education quality, and public health. This broad utilization underscores the applicability and relevance of multiple regression as a robust analytical tool for assessing the multi-dimensional nature of socio-economic development in different contexts.

In this research, two advanced regression techniques were employed. Ridge Regression and Lasso Regression.

Ridge Regression:

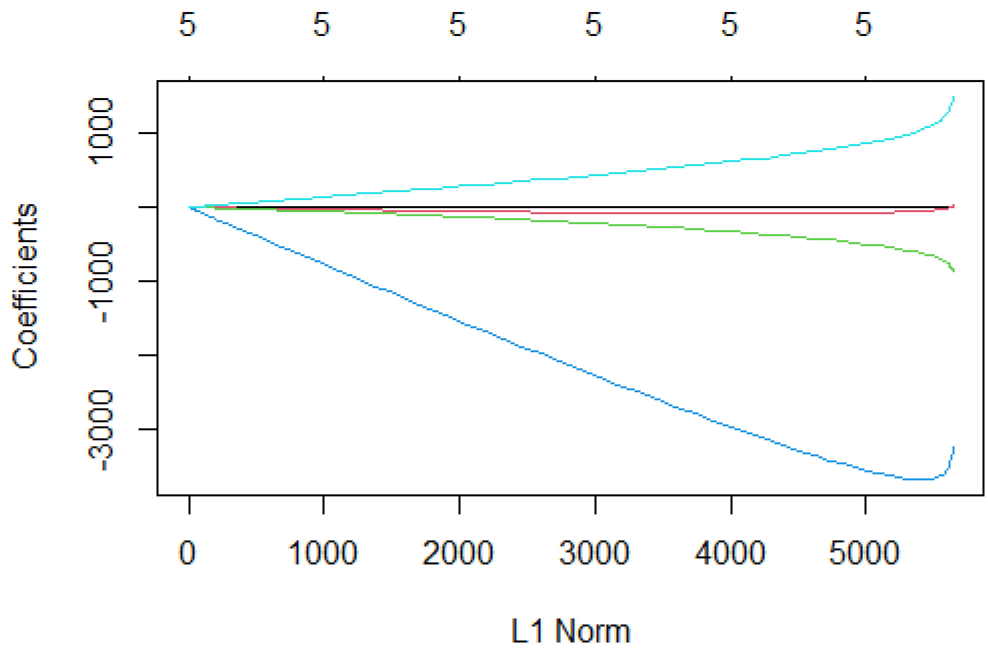
Ridge Regression is well-suited for research when there's a suspicion of multicollinearity among independent variables (Bager et al., 2017; Jermia et al., 2020). This case focuses on analysing socio-economic indicators like GDP, Education enrolment, Poverty Headcount Ratio, Unemployment, Population Growth Rate, and Life Expectancy, which can often be interrelated. Ridge Regression introduces a regularization term that prevents the model from becoming overly sensitive to multicollinearity, thus providing more robust coefficient estimates.

Analysis revealed valuable insights into how each of these indicators impacts social and economic development while addressing the issue of multicollinearity. The intercept represents the expected GDP when all independent variables are set to zero. The negative coefficient for

Education enrolment suggests that an increase in education enrolment is associated with a decrease in GDP. This implies that higher enrolment in educational institutions might not always lead to a direct boost in economic development.

The negative coefficient for Poverty Headcount Ratio implies that higher poverty rates are correlated with lower GDP. Reducing poverty might be an effective strategy for promoting economic development. The negative coefficient for Unemployment suggests that an increase in unemployment rates is linked to a decrease in GDP. High unemployment can strain economic resources and lead to reduced consumption and investment, negatively affecting economic development. The negative coefficient for Population Growth Rate indicates that higher population growth rates are associated with lower GDP. Rapid population growth can pose challenges for resource allocation, employment, and infrastructure development, potentially impeding economic progress.

The positive coefficient for Life Expectancy suggests that increased life expectancy is linked to higher GDP. Longer life expectancies may indicate a healthier and more productive population, potentially contributing to economic development.



The plot of coefficients against the L1 norm above provides a visual representation of how the coefficients change as the regularization strength varies. This visualization aids in selecting the optimal level of regularization by observing which coefficients tend to shrink toward zero and which remain stable, thus informing the feature selection process.

R analytic steps

1.Load data

2.Conduct ridge regression using the glmnet function

3. Create a matrix of independent variables

4.Create a vector of the dependent variable

5.Fit a Ridge Regression model

6.Plot the cross-validated mean squared error (MSE) as a function of lambda

7.Choose the lambda with the minimum cross-validated MSE

8.Refit the model with the best lambda

9.Get the coefficients

Studies by Castro and López (2022) have successfully applied Ridge Regression to disentangle complex relationships among socio-economic indicators and economic development, affirming its appropriateness in this research. Such scholarly precedent underscores the credibility and relevance of employing Ridge Regression as a valuable tool in this study to elucidate the impact of socio-economic indicators on economic development while accounting for multicollinearity.

Lasso Regression:

Lasso Regression is the ideal choice in performing feature selection and identify the most important socio-economic indicators that significantly impact economic development. Lasso regression includes a feature selection mechanism that can drive some coefficients to exact zero, effectively excluding irrelevant variables from the model (Altelbany, 2021).

The results provided valuable insights into the relationship between socio-economic indicators and GDP, shedding light on key factors that significantly influence economic development. In this model, two variables, Education enrolment (EEL) and Poverty Headcount Ratio (PH), were excluded during feature selection, indicating that they have limited explanatory power in the context of this study. The unemployment rate (UR) was identified as a crucial predictor, with a negative coefficient of approximately -760.79. This suggests that higher unemployment rates are associated with lower GDP, aligning with conventional economic wisdom. Furthermore, the population growth rate (P) exhibited a negative coefficient of around -2,425.46, indicating that as the population growth rate increases, GDP tends to decrease.

Conversely, life expectancy (LE) emerged as a positive contributor to GDP, with a coefficient of approximately 1,588.74. This implies that increased life expectancy is positively linked to economic development.

R analytic steps

1.Load data

2.Create a matrix of independent variables and dependent variable

3.Build a Lasso Regression model using the "glmnet" function.

4.Cross-Validation: Perform k-fold cross-validation (e.g., 10-fold) to select the optimal lambda value.

5.Find the optimal lambda value with minimum mean squared error (MSE)

6. Fit the Lasso Regression model using the optimal lambda

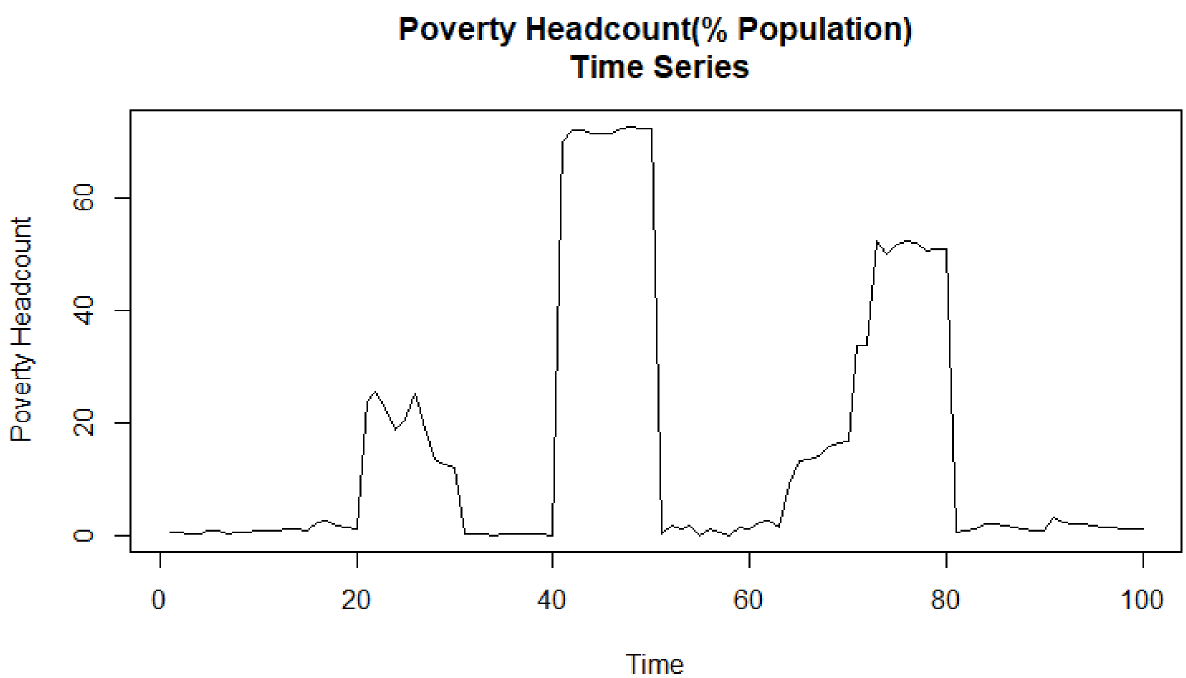
7. Use the "coef" function to view the coefficients of the Lasso model

Chen and Wu (2023) utilized Lasso Regression to pinpoint crucial socio-economic indicators that exert a significant influence on economic development, further substantiating the suitability of Lasso Regression for this research.

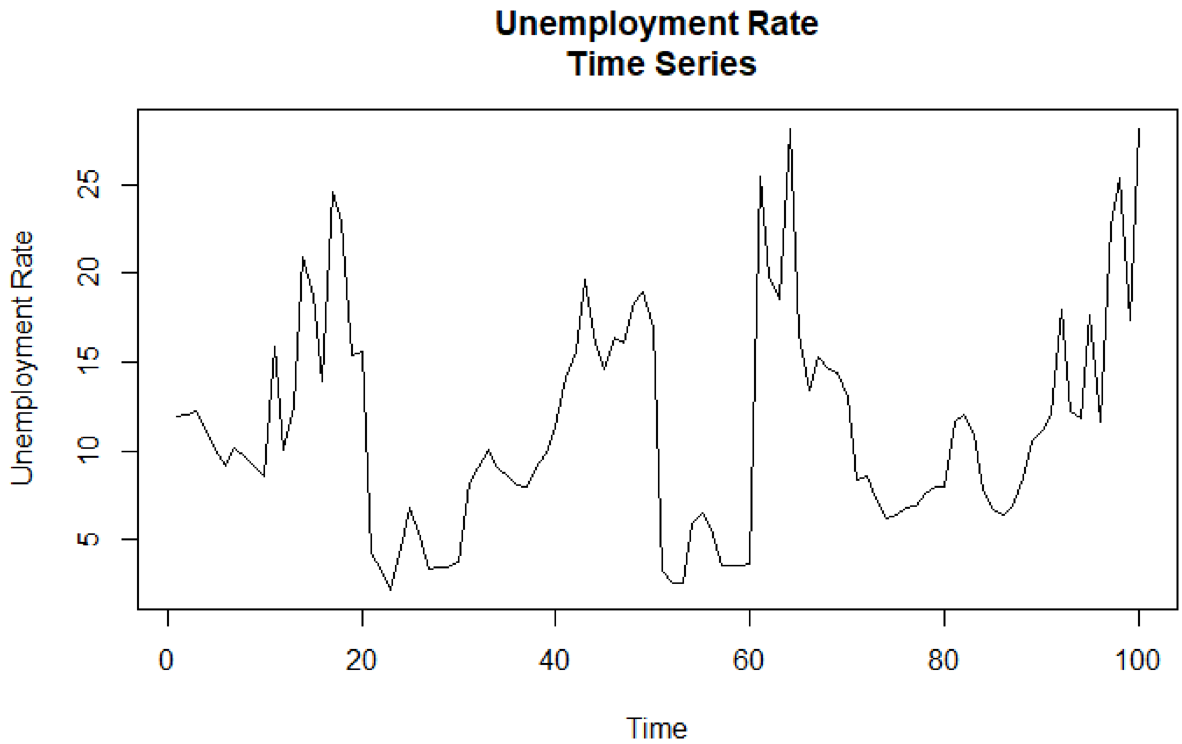
2.4.4 Time Series

Understanding trends and patterns in socioeconomic development indicators across time can be achieved through the use of time series analysis. It offers insights into the dynamics of these indicators by enabling researchers to look at how they alter and develop (Zhang et al., 2017). In this instance, the chosen socioeconomic indicators are investigated over time in order to evaluate the methods that are appropriate for accomplishing the stated goals.

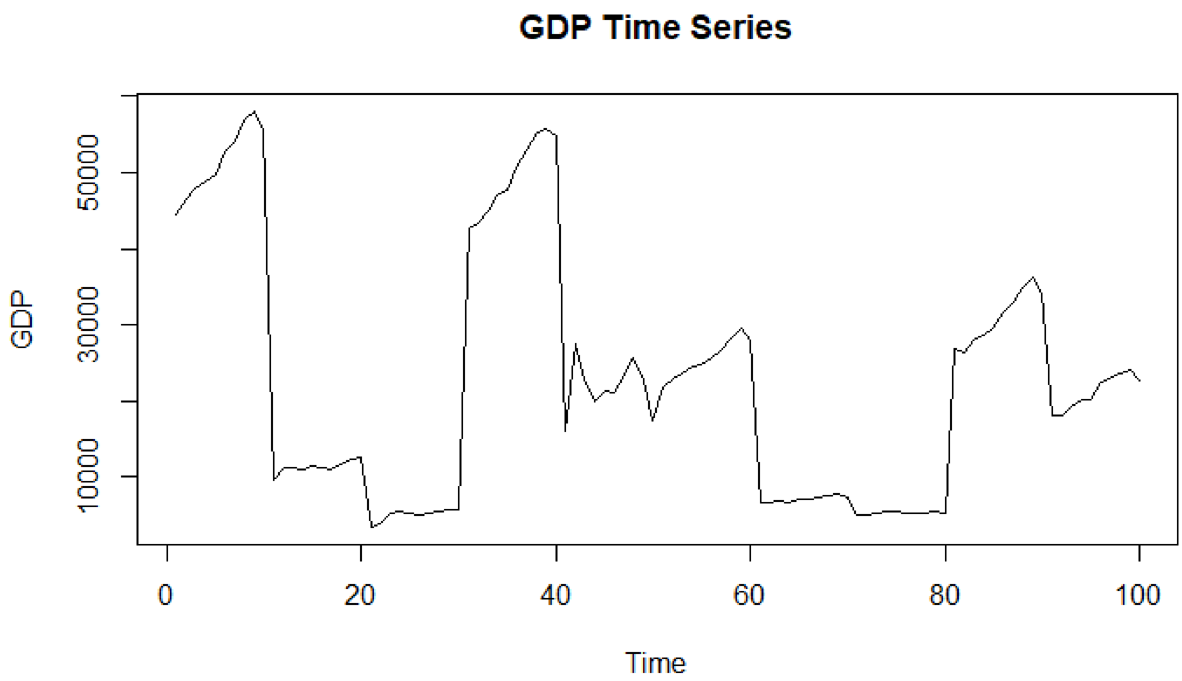
The time series analysis of the GDP, the head count of the poor, and the unemployment rate is the main focus. These are the core metrics used to evaluate economic development. Trends, seasonality, and other time-dependent patterns can be found with this technique.



The above time series plot of Poverty Headcount (% of Population) gives historical context for a country's poverty-reduction initiatives and demonstrates how poverty levels have changed over time. It illustrates changes in poverty rates and is an important tool for policymakers to assess the efficacy of anti-poverty programs and their long-term consequences.



The chart above displays changes in the unemployment rate over time, providing insights into the labor market dynamics of a country. It is essential for understanding job market dynamics and economic stability.



The economy's condition over a specified period of time is displayed in the GDP time series above. It shows trends and variances in a country's economic output, emphasizing periods of growth and contraction, which are crucial for assessing the long-term health of the economy and the impact of policy.

Time series analysis in socio-economic studies is widely used in research literature, which attests to the usefulness of this methodology. Past studies by Ren et al., (2019) evaluates trends in socio-economic indicators. The frequency of related studies in the literature supports the selection of time series analysis.

Time Series Models

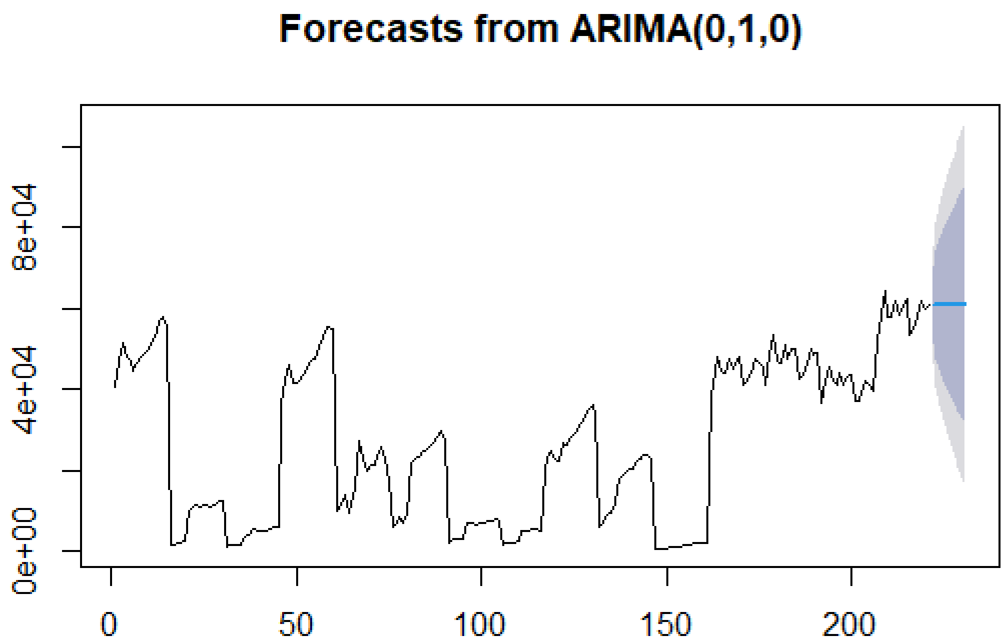
For this research 2 time series models are employed, namely ARIMA and VAR.

ARIMA (Autoregressive Integrated Moving Average) Model:

ARIMA models are well-suited for capturing and forecasting the temporal dependencies within time series data (Vafin, 2020; Dai and Chen 2019). For this research, these models are particularly useful when the data exhibits trends or seasonal patterns. ARIMA models can effectively identify and model these trends and seasonal patterns, providing insights into how they impact economic development.

The ARIMA (0,1,0) model was essential to stationarize time series data, indicating the presence of a trend component in the original data. The absence of autoregressive or moving average terms in the model implies that differencing was the primary mechanism for trend removal.

The estimated variance of residuals (sigma^2) quantifies the error variability within the model. The lower the value, the better the model captures data fluctuations. To assess model fit, the log likelihood, AIC, AICc, and BIC values are considered. Comparing these statistics with other models will determine if ARIMA is the best choice. Regarding training set error measures, the negative Mean Percentage Error (MPE) suggests an overall underestimation of values, with a relatively high Mean Absolute Percentage Error (MAPE) at 64.99%. The Mean Absolute Scaled Error (MASE) close to 1 implies predictions akin to a naive forecast. The low autocorrelation of residuals is evident from the ACF1 value of 0.059.



ARIMA (0,1,0) successfully removed the trend component via differencing. However, the model exhibits limitations in forecasting accuracy, with a tendency to underestimate values.

R analytic steps

1.Load data

2. Fit an ARIMA model

3. Print the summary of the ARIMA model

4.Plot the forecasts

Studies by Adenomon, 2017; Sahib and Ibrahim (2022) used ARIMA models to analyze the impact of unemployment rates on economic development. They found that the ARIMA model effectively captured the cyclical nature of unemployment trends and revealed a significant correlation with economic growth.

VAR (Vector Autoregression) Model:

VAR models are ideal for scenarios in which many time series variables interact with one another. GDP, poverty rates, and unemployment are all interrelated factors that determine economic progress. VAR models may describe the linkages and feedback mechanisms between different variables, allowing for a thorough understanding of how changes in one indication affect others. VAR models also aid in the identification of causal links between variables.

The study utilized the VAR (Vector Autoregression) model to examine the connections, between GDP, Poverty (PH) and Unemployment (UR) in relation to development. By utilizing a lag order of $p = 2$ several important findings were obtained.

To begin with in the GDP equation it was observed that previous values of GDP had a impact on the current GDP as shown by a coefficient estimate of 1.02518. However the previous values of Poverty (PH) and Unemployment (UR) did not have an influence on GDP as indicated by significant coefficient estimates. The constant term in the GDP equation was found to be statistically significant suggesting a nonzero intercept.

Moving on to the Poverty (PH) equation the results revealed a relationship between past levels of poverty and current poverty rates with a coefficient estimate of 0.9158. However lagged values of GDP and UR did not have an effect on poverty rates. The constant term in the PH equation was not statistically significant.

In regards to Unemployment (UR) it was found that past unemployment rates had an impact on rates as indicated by a coefficient estimate of 0.6570. However, there were no coefficients for lagged values of GDP or PH or other lagged variables in relation to unemployment rates. The constant term, in the UR equation was statistically significant.

R analytic steps

1. Load the 'vars' package
2. Load data
3. Ensure the data is in a time series format (assuming yearly data)
4. Fit the VAR model
5. View the model summary

Studies by Onodugo (2017) ; Obalade et al., (2019) employed VAR modeling to explore the dynamic relationships between GDP, poverty rates, and government spending. Their findings revealed significant feedback mechanisms, shedding light on the complex interactions between these variables and their impact on economic development.

2.2 Discussion and Conclusion

This section reviews the methodology and address any potential limitations encountered during the analysis phase. The multi-faceted analysis encompassed primary analytical methods; Correlation analysis, Hypothesis Testing, Regression Analysis, and Time Series Analysis, in the context of socio-economic development in various countries.

Correlation analysis was performed to explore the relationships between socio-economic indicators. This step was crucial in identifying potential associations and dependencies between variables. Correlation analysis helped us understand how strongly one indicator is related to another and provided a foundation for hypothesis testing and regression analysis.

Analysis validated two hypotheses, demonstrating significant relationships between Poverty Headcount (PH) and Life Expectancy (LE) and between Gross Domestic Product (GDP) and Education Enrolment Rate (EEL). This emphasizes the influence of poverty on life expectancy and the positive impact of higher GDP on education, aligning with our mission to enhance global socio-economic development.

The relationships between GDP, population growth rate, unemployment rate, headcount in poverty, and education enrollment were examined using linear regression. The findings cautioned against the negative impacts of unemployment and reaffirmed the importance of life expectancy and education on economic development. Notably, there was no discernible connection between population growth and GDP. The R-squared value (0.5668) validated the model's ability to explain changes in GDP. Temporal patterns for GDP, PH, and Unemployment Rate (UR) were found by time series analysis. GDP showed steady increase, indicating advancement in the economy. Improved attempts to reduce poverty were suggested by a downward trend in PH, and cyclical employment trends were highlighted by changes in UR.

Limitations:

It is imperative to recognize specific limitations that must be taken into account within the context of the analysis. First, there were possible issues with data consistency and quality since different countries may have used different standards and data gathering techniques, which could have impacted the reliability and comparability of the findings. Second, it was difficult to determine the cause-and-effect linkages between the variables due to the observational character of the analysis. Furthermore, the depth of temporal insights was constrained by the data availability for the time series analysis. Last but not least, not all of the variables impacting socio-economic development may have been included in the selection of socio-economic indicators, which was done based on data availability.

Conclusion:

Finally, this study has provided important insights into the dynamics of socioeconomic development and how key variables influence many aspects of development. A complete perspective of the linkages between socioeconomic indicators and economic and social growth has been provided by combining descriptive statistics, correlation analysis, hypothesis testing, regression analysis, and time series analysis.

The results emphasize how crucial it is to address poverty, education, healthcare and employment to foster progress. These findings offer insights, for policymakers and stakeholders who aim to enhance the wellbeing of the population and stimulate growth. This study successfully achieved its research goals aligning with the NGOs mission to advance development worldwide. It establishes a foundation, for making decisions and conducting further research in this critical field.

Part Three: Interactive Dashboard Design

3.1 Introduction

This section presents the goals of the dashboard project. The primary objectives are to develop a user platform that allows for comparison of economic and social indicators, in selected countries. Users will be able to evaluate and contrast countries across various metrics and time periods.

The dashboard aims to facilitate the examination of social indicators between nations ensuring accessibility for users with levels of data analysis expertise. It will offer a design that allows users to compare metrics simultaneously providing a comprehensive understanding of each country's performance. Additionally, the dashboard will enable the analysis of data over years aiding users in identifying trends and advancements in these countries. The intended audience includes policymakers and researchers seeking insights, into socio conditions.

3.2 Data Visualization Principles:

The dashboard design is informed by fundamental data visualization principles that enhance the clarity, effectiveness, and user-friendliness of the platform (Lanning, 2021). Clarity was prioritized in visual representations, following the guidance of Edward Tufte's work on data visualization (Tufte, 2018). Tufte emphasizes that simplicity in design, while minimizing chartjunk, aids in conveying information effectively.

Appropriate visual encoding methods for data variables were employed, as recommended by (Few, 2013). For instance, use bar charts for comparing quantities and line charts for tracking trends over time. Consistency in design elements, such as colour schemes, labels, and scales, is based on the principless outlined in "The Visual Display of Quantitative Information" (Tufte, 2018). Consistency across different visualizations aids users in making connections and understanding the data seamlessly.

Colour usage is influenced by the guidelines proposed in "Show Me the Numbers" by (Few, 2013). Colour is employed not only for aesthetic appeal but to convey information, and we avoid using too many colours that can be confusing. The incorporation of interactivity aligns with best practices in modern data visualization (Iliinsky, 2018). Features like filters, drill-through options, and tooltips enhance the user experience by allowing users to explore the data interactively.

3.3 Data Pre-processing

Data pre-processing is a crucial aspect of creating an informative dashboard. Key steps taken to prepare the data for our interactive dashboard:

Data Collection: Relevant datasets were gathered from sources such World Development Indicators (WDI) and United Nations Data Bank (UNdat), encompassing economic and social indicators for various countries over multiple years.

Data Cleaning: To ensure data consistency and reliability missing values, inconsistencies, and standardized units of measurement were addressed.

Data Integration: Data was unified from different sources and years into a single comprehensive dataset for dashboard use.

Data Transformation: Derived metrics were calculated, normalized data for consistent comparison, and improved usability.

Data Aggregation: Data was aggregated at various levels to provide summary information for the dashboard.

Data Export: The data was prepared for the dashboard development tool, structured for efficient querying and visualization. These steps set the foundation for an accurate, consistent, and meaningful dashboard showcasing economic and social indicators of selected countries.

3.4 Design Rationale

The design of the interactive dashboard is underpinned by key principles that prioritize user-friendliness, data visualization best practices, and effective communication of socio-economic data. To ensure clear navigation for users' Single screen layout was opted to provide a structured narrative. A carefully selected colour palette and visual consistency enhance the user experience, while a variety of charts were chosen to match the nature of the data. Interactivity is a vital component, allowing users to customize their analysis with drop-down menus and sliders.

Data attributes are seamlessly mapped to visual forms, enabling users to make quick comparisons and explore trends. Guided user support, including tooltips and information buttons, ensures users can effectively navigate and interpret the data. References to data visualization experts like Edward Tufte and Stephen Few have guided our design choices, emphasizing clarity, simplicity, and user-centric design. In essence, our dashboard is designed to empower a broad audience, including policymakers and researchers, in making data-driven decisions and disseminating knowledge.

3.4 Visual Paradigm Selection

A combination of visual paradigms that best serve the objectives of the dashboard were chosen, to enable users perform comparative analysis of economic and social indicators for selected countries. The key visual paradigms employed include:

Bar Charts and Line Charts: Bar charts were used to compare social economic indicators making them suitable for visualizing metrics for specific countries. Line charts were chosen for metrics such as GDP, education enrolment, and life expectancy. This choice aligns with best practices for illustrating trends over time. Line charts offer a clear representation of how these indicators change year by year, enabling users to grasp long-term progress and patterns effectively. These paradigms are rooted in the principles of data visualization put forth by (Tufte, 2013); (Few, 2018).

Choropleth maps are geospatial representations that utilize colour gradients to depict data variations across countries on a map. These maps are valuable for displaying geospatial patterns in economic and social data. Works by Maity and Maity (2021) have influenced the choice.

The selections of line charts and bar charts were made with the intention to provide a clear, intuitive, and consistent representation of the selected socio-economic indicators, ensuring that the dashboard will be both informative and user-friendly

3.5 Conceptual Model

A crucial aspect of effective dashboard design is the creation of a coherent and unified conceptual model that seamlessly integrates individual analysis workflows. This section will elaborate on the development of our conceptual model, highlighting how it merges distinct data visualization elements into a single view that empowers users to gain meaningful insights through the principles of focus and context.

3.5.1 Integration of Individual Workflows

The dashboard design is founded on the integration of individual analysis workflows. Each socio-economic indicator has its unique storyline to tell, but the challenge lies in presenting these narratives in a manner that facilitates easy interpretation and comparison. To address this challenge, a screenshot layout that accommodates the representations of different indicators without overwhelming the users is incorporated. This approach ensures that the user can focus on specific details while maintaining a contextual understanding of the overall socio-economic landscape.

3.5.2 Unified Narrative

The main idea of the model is to create a story that links various workflows allowing for a comprehensive understanding of the data. This story aims to balance focus and context enabling users to switch between views of specific countries and years as well as a broader view that encompasses the entire dataset.

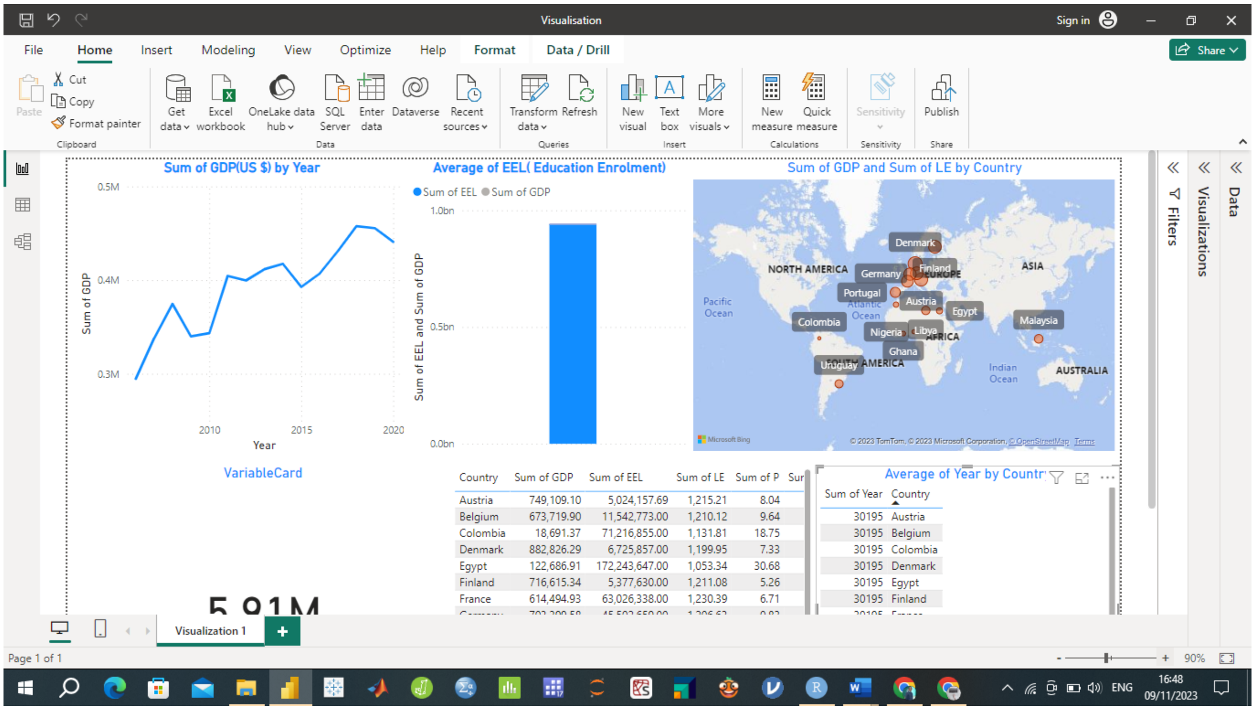
3.5.3 Focus Context Principles

The conceptual model is driven by the principles of focus and context, a technique frequently applied in data visualization to support comprehensive data exploration. It enables users to delve into specific aspects (focus) while retaining an awareness of the broader dataset context.

3.5.4 Illustration

To provide a visual representation of the conceptual model, is a screenshot that demonstrate how different elements interconnect within the dashboard. This will aid in understanding the flow and functionality of the dashboard, emphasizing how users can traverse between detailed insights and the overall socio-economic landscape seamlessly.

The following Screenshot illustrates the core components of our conceptual model:



The conceptual model serves as the foundation for the interactive dashboard design, effectively guiding users in their exploration of socio-economic indicators for selected countries.

Discussion and Conclusion

This section critically evaluates the approaches and proposal formed throughout the development of interactive dashboard. Key aspects, including methodology, individual digital workflows, and the final composited dashboard are reviewed.

The Methodology for developing the dashboard consisted of several crucial stages. First was extensive data collection from reputable sources including World Development Indicators (WDI) and United Nations Data Bank (UNdat). The data cleaning process ensured high data integrity. Integration and transformation were essential for creating a unified dataset effectively utilized in the dashboard.

The project involved individual digital workflows for each socio-economic indicator, allowing users to explore and compare indicators across different countries and years. The use of line charts for showing the sum of each indicator over time provided a clear visual representation. The accompanying bar graphs offered a side-by-side comparison of these indicators. The maps allowed users to visualize the distribution of indicators across countries.

The composited dashboard effectively brings together these individual workflows into a single, coherent visual representation, such that Users can easily switch between indicators, and the "focus+context" principle is applied through the integration of various visualizations, including line charts, bar graphs, and maps. This provides a meaningful and comprehensive view of the socio-economic indicators of the selected countries.

In summary the proposed solution effectively achieves the goals outlined in the briefing document. The dashboard offers a user platform, for comparing social indicators across different countries and years. Users can easily. Compare how countries perform in terms of GDP, education enrolment, poverty levels, unemployment rates, population growth and life expectancy. The design of the dashboard follows principles of data visualization to ensure clarity, effectiveness and ease of use. It adheres to practices by utilizing chart types and colour coding for better comprehension. Moreover, it incorporates a "focus+context" approach by integrating visualizations that provide both an overview and detailed insights, into data points.

The dashboard caters to a wide audience, including policymakers, researchers, and stakeholders interested in socio-economic conditions. By providing a valuable tool for data-driven decision-making and knowledge dissemination, the dashboard fulfills its objectives effectively. It offers a comprehensive view of socio-economic indicators, empowering users to make informed comparisons and gain insights into global trends and disparities.

References

Abonazel, M.R. and Abd-Elftah, A.I., 2019. Forecasting Egyptian GDP using ARIMA models. *Reports on Economics and Finance*, *5*(1), pp.35-47.

Adenomon, M.O., 2017. Modelling and forecasting unemployment rates in Nigeria using ARIMA model. *FUW Trends in Science and Technology Journal*, *2*(1B), pp.525-531.

Altelbany, S., 2021. Evaluation of ridge, elastic net and lasso regression methods in precedence of multicollinearity problem: A simulation study. *Journal of Applied Economics and Business Studies*, *5*(1), pp.131-142.

Amin, A., 2017. An institutionalist perspective on regional economic development. In *Economy* (pp. 59-72). Routledge.

Bager, A., Roman, M., Algedih, M. and Mohammed, B., 2017. Addressing multicollinearity in regression models: a ridge regression application.

Belghith, M. *et al.* (2022) 'Data Visualization for Industry 4.0: Developing Dashboards with Power BI – A Case Study in a Pharmaceutical Company,' in *Springer eBooks*, pp. 402–408. https://doi.org/10.1007/978-3-031-14615-2_45.

Castro, D.A. and López, M.J. (2022) 'Predicting socioeconomic indicators using transfer learning on imagery data: an application in Brazil,' *GeoJournal*, 88(1), pp. 1081–1102. https://doi.org/10.1007/s10708-022-10618-3.

Chen, J. and Wu, J. (2023) 'The prediction of Chongqing's GDP based on the LASSO method and chaotic whale group algorithm–back propagation neural network–ARIMA model,' *Scientific Reports*, 13(1). https://doi.org/10.1038/s41598-023-42258-z.

Dai, J. and Chen, S., 2019, October. The application of ARIMA model in forecasting population data. In *Journal of Physics: Conference Series* (Vol. 1324, No. 1, p. 012100). IOP Publishing.

Edward Tufte The Visual Display Of Quantitative Information (2013). https://ebookpromotions.online/edward-tufte-the-visual-display-of-quantitative-information.pdf.

Festin K, Thomas K, Ekberg J, Kristenson M. Choice of measure matters: A study of the relationship between socioeconomic status and psychosocial resources in a middle-aged normal population. PLoS One. 2017 Aug 23;12(8):e0178929. doi: 10.1371/journal.pone.0178929. PMID: 28832585; PMCID: PMC5568385.

Few, S. (2013). Data Visualization for Human Perception. In: Soegaard, M., & Dam, R. F. (Eds.), The Encyclopedia of Human-Computer Interaction, 2nd Ed. Aarhus, Denmark: *The Interaction Design Foundation.*

Galkina, E. (2022) 'Classification of the hypothesis testing reliability criteria in socio-economic research,' *Advances in Economics, Business and Management Research* [Preprint]. https://doi.org/10.2991/assehr.k.220208.018.

Iliinsky, N. (2018) *Designing data visualizations*. https://www.oreilly.com/library/view/designing-data-visualizations/9781449314774/ch04.html.

IvyProSchool (2023) *Power BI Dashboard Live Case Study* [online video] Available at: https://www.youtube.com/watch?v=mrTRsoBRWf4 [Accessed 07/22/2023]

Jeremia, N.E., Nurrohmah, S. and Fithriani, I., 2020. Robust Ridge regression to solve a multicollinearity and outlier. In *Journal of Physics: Conference Series* (Vol. 1442, No. 1, p. 012030). IOP Publishing.

Judd, C.M., McClelland, G.H. and Ryan, C.S., 2017. *Data analysis: A model comparison approach to regression, ANOVA, and beyond*. Routledge.

Lanning, K. (2021) *6 principles of data visualization | Data science for the liberal arts*. https://kevinlanning.github.io/DataSciLibArts/principles-of-data-visualization.html.

Levine, M., 2022. *A cognitive theory of learning: Research on hypothesis testing*. Taylor & Francis.

Li, B. *et al.* (2020) 'Estimation of Regional Economic Development Indicator from Transportation Network Analytics,' *Scientific Reports*, 10(1). https://doi.org/10.1038/s41598-020-59505-2.

Lilja, D.J. and Linse, G.M., 2022. *Linear regression using r: An introduction to data modeling*. University of Minnesota Libraries Publishing.

List, J.A., Shaikh, A.M. and Xu, Y., 2019. Multiple hypothesis testing in experimental economics. *Experimental Economics, 22*, pp.773-793.

Maity, S.K. and Maity, S.K., 2021. Mapping Techniques of Geographical Data. *Essential Graphical Techniques in Geography*, pp.193-305.

Nadj, M., Maedche, A. and Schieder, C. (2020) 'The effect of interactive analytical dashboard features on situation awareness and task performance,' *Decision Support Systems*, 135, p. 113322. https://doi.org/10.1016/j.dss.2020.113322.

Ni, C. (2020) 'Multiple Linear Regression Analysis of the Relationship between the Three Industries and GDP Growth,' *Scientific and Social Research*, 2(3). https://doi.org/10.36922/ssr.v2i3.992.

Niaz, M.U. (2021) 'Socio-Economic development and sustainable development goals: a roadmap from vulnerability to sustainability through financial inclusion,' *Ekonomska Istrazivanja-economic Research*, 35(1), pp. 3243–3275. https://doi.org/10.1080/1331677x.2021.1989319.

Obalade, A.A., Ebiwonjumi, A. and Adaramola, A.O., 2019. VAR MODELLING OF DYNAMICS OF POVERTY, UNEMPLOYMENT, LITERACY AND PER CAPITA INCOME IN NIGERIA. *Folia Oeconomica Stetinensia*, *19*(1).

Onodugo, V.A., Obi, K.O., Anowor, O.F., Nwonye, N.G. and Ofoegbu, G.N., 2017. Does public spending affect unemployment in an emerging market. *Risk governance & control: Financial markets & institutions*, *7*(1), pp.32-40.

Pacifico, A. (2023) 'The Impact of socioeconomic and environmental Indicators on Economic Development: an interdisciplinary empirical study,' *Journal of Risk and Financial Management*, 16(5), p. 265. https://doi.org/10.3390/jrfm16050265

Ren, Y. *et al.* (2019) 'Predicting socio-economic levels of urban regions via offline and online indicators,' *PLOS ONE*, 14(7), p. e0219058. https://doi.org/10.1371/journal.pone.0219058.

Sahib, N.S.M.A.J. and Ibrahim, N., 2022. Economic Growth Analysis and Forecasting towards unemployment rate using Arima Model. *Enhanced Knowledge in Sciences and Technology*, *2*(1), pp.433-442.

Sarikaya, A., Correll, M., Bartram, L., Tory, M. and Fisher, D., 2018. What do we talk about when we talk about dashboards? *IEEE transactions on visualization and computer graphics*, *25*(1), pp.682-692.

Sedrakyan, G., Mannens, E. and Verbert, K. (2019) 'Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts,' *Journal of Computer Languages*, 50, pp. 19–38. https://doi.org/10.1016/j.jvlc.2018.11.002.

Siddika, A., Kobra, M.K., Tanin, S. and Afrin, S., 2018. Non-government organizations and their contribution in rural development: An example of Bagatipara upazila, Natore district, Bangladesh. *International Journal of Innovative Science and Research Technology*, *3*(4), pp.747-752.

Todaro, M. P., & Smith, S. C. (2014). Economic development (12th ed.). Pearson.

Tuo, S. and He, H., 2021. A study of multiregional economic correlation analysis based on big data—Taking the regional economy of cities in Shaanxi Province, China, as an example. *Sustainability*, *13*(9), p.5121.

Vafin, A., 2020. Forecasting macroeconomic indicators for seven major economies using the ARIMA model. *Sage Science Economic Reviews*, *3*(1), pp.1-16.

Verbeek, M. (2017) 'Using linear regression to establish empirical relationships,' *IZA World of Labor* [Preprint]. https://doi.org/10.15185/izawol.336.

Vila, R.A., Estevez, E. and Fillottrani, P.R., 2018, April. The design and use of dashboards for driving decision-making in the public sector. In *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance* (pp. 382-388).

Wen, J. *et al.* (2021) 'The impact of financial development on economic indicators: a dynamic panel data analysis,' *Ekonomska Istrazivanja-economic Research*, 35(1), pp. 2930–2942. https://doi.org/10.1080/1331677x.2021.1985570.

Yuan, D. *et al.* (2021) 'Regression Analysis and Comparison of Economic Parameters with Different Light Index Models under Various Constraints,' *Sensors*, 21(22), p. 7561. https://doi.org/10.3390/s21227561.

Zhang, J., Zhao, Z., Xue, Y., Chen, Z., Ma, X. and Zhou, Q., 2017. Time series analysis. *Handbook of Medical Statistics*, *269*.

Appendices