A logo for a university

Description automatically generated

**ASSIGNMENT COVER SHEET**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PROGRAMME | : | Master of Business Analytics | | |
| SUBJECT CODE AND TITLE | : | BAA5043 Business Intelligence | | |
| ASSIGNMENT TITLE | : | The Role of Social Inclusion in the Pursuit of Happiness: A Panel Data Analysis | | |
|  |  |  | | |
| LECTURER | : | Dr Mubbasher Munir | ASSIGNMENT DUE DATE: | 15/10/2024 |

STUDENT’S DECLARATION

1. I hereby declare that this assignment is based on my own work except where acknowledgement of sources is made.
2. I also declare that this work has not been previously submitted or concurrently submitted for any other courses in Sunway University/College or other institutions.

[ Submit “Turn-it-in” report (please tick √): Yes \_\_\_\_\_ No \_\_√\_\_\_ ]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| NO. | NAME | STUDENT ID NO. | SIGNATURE | DATE |
| 1. | Abdul Hakim Bin Kamalur Rahman | 24015257 | Abdul Hakim | 14/10/2024 |
| 2. | Annabel Ching Ke Xin | 24002685 | Annabel | 14/10/2024 |
| 3. | Divani A/P Arumugam | 19058908 | Divani | 14/10/2024 |
| 4. | Harresh A/L Ragunathan | 19076090 | Harresh | 14/10/2024 |
| 5. | Harimanjato Fanaja Rambonimanana | 24036014 | Fanaja | 14/10/2024 |
| 6. |  |  |  |  |

E-mail Address / Addresses (according to the order of names above):

|  |  |
| --- | --- |
| 1. 24015257@imail.sunway.edu.my | 4. 19076090@imail.sunway.edu.my |
| 2. 24002685@imail.sunway.edu.my | 5. 24036014@imail.sunway.edu.my |
| 3. 19058908@imail.sunway.edu.my | 6. |

APPROVAL FOR LATE SUBMISSION OF ASSIGNMENT (If applicable)

IF extension is granted, what is the revised due date? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature of Lecturer: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |
| --- |
| Marker’s Comments: |

Marks and / or Grade Awarded: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ADDENDUM**

**USE OF ARTIFICAL INTELLIGENCE (A.I.) DECLARATION**

Students are allowed to use AI to support completion of assessments. However, students are reminded to do so ethically and transparently. This is so that (a) submissions can be fairly and accurately marked; and (b) feedback can be provided on the content that reflects student ability, in order to help with future submissions. Students are also reminded that in accordance with the University’s Academic Malpractice Policy, Item 4.11.2, “*… the representation of work: written, visual, practical or otherwise, of any other person, including another student or* ***anonymous web-based material*** *[emphasis added], or any institution, as the candidate’s own*” is considered malpractice.

**Declaration**

[ ] I / We used the following A.I. tools to produce content in this submission:

|  |  |  |  |
| --- | --- | --- | --- |
| **Tool** | **Purpose** | **Prompts** | **Sections where AI output was used / Outcome(s) in the submission** |
| ChatGPT | Generating coding reference | “How to merge data, visualize the data distribution, manage with missing value in Python?”  “How to use R for panel data analysis?” | The coding part in Python and R. All the answers provided by ChatGPT are for references. |
| Grammarly | Correcting grammar and spelling, improving sentence structure | N/A | Grammarly suggestions were used for all sections of the essay |
|  |  |  |  |

*Note: Add additional rows if necessary.*

**OR**

[ ] I / We did not use any A.I. tools to produce any of the content in this submission.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| NO. | NAME | STUDENT ID NO. | SIGNATURE | DATE |
| 1. |  |  |  |  |
| 2. |  |  |  |  |
| 3. |  |  |  |  |
| 4. |  |  |  |  |
| 5. |  |  |  |  |
| 6. |  |  |  |  |

E-mail Address / Addresses (according to the order of names above):

|  |  |
| --- | --- |
| 1. | 4. |
| 2. | 5. |
| 3. | 6. |

Table of Contents

[**List of Figure 4**](#_Toc179815501)

[**List of Table 4**](#_Toc179815502)

[**Abstract 5**](#_Toc179815503)

[**1.0 Introduction 6**](#_Toc179815504)

[**1.1 Research Background 6**](#_Toc179815505)

[**1.2 Problem Statement 6**](#_Toc179815506)

[**1.3 Research Questions 6**](#_Toc179815507)

[**1.4 Research Objectives 7**](#_Toc179815508)

[**2.0 Literature Review 7**](#_Toc179815509)

[**3.0 Methodology 10**](#_Toc179815510)

[**3.1 Nature of variables 10**](#_Toc179815511)

[**3.2 Data Pre-processing 11**](#_Toc179815512)

[**3.3 Exploratory Data Analysis (EDA) 14**](#_Toc179815513)

[**3.3.1 Description of data 14**](#_Toc179815514)

[**3.3.2 Descriptive statistics 14**](#_Toc179815515)

[**3.3.3 Missing value and outliers 15**](#_Toc179815516)

[**3.3.4 Distribution 17**](#_Toc179815517)

[**3.3.5 Model Selection for Predictive Analytics 18**](#_Toc179815518)

[**4.0 Result and analysis 19**](#_Toc179815519)

[**4.1 Descriptive analytics using PowerBI 19**](#_Toc179815520)

[**4.2 Predictive analytics using R and R studio 27**](#_Toc179815521)

[**4.2.1 Model Used 27**](#_Toc179815522)

[**4.2.2 Model training 32**](#_Toc179815523)

[**5.0 Conclusion 34**](#_Toc179815524)

[**Following questions 36**](#_Toc179815525)

[**References 39**](#_Toc179815526)

[**Appendix 43**](#_Toc179815527)

# List of Figure

[Figure 3.1: The Flowchart of Working Process 11](#_Toc179815194)

[Figure 3.2: Data Description 14](#_Toc179815195)

[Figure 3.3: Missing Values 15](#_Toc179815196)

[Figure 3.4: Outliers 15](#_Toc179815197)

[Figure 3.5: Outliers (Continued) 16](#_Toc179815198)

[Figure 3.6: Outliers (Continued) 16](#_Toc179815199)

[Figure 3.7: Data Distribution 17](#_Toc179815200)

[Figure 3.8: Model Selection Process 18](#_Toc179815201)

[Figure 4.1: Dashboard 1 19](#_Toc179815202)

[Figure 4.2: Dashboard 2 20](#_Toc179815203)

[Figure 4.3: Dashboard 3 23](#_Toc179815204)

[Figure 4.4: Dashboard 4 25](#_Toc179815205)

[Figure 4.5: Result of Time-Fixed Effects Model 28](#_Toc179815206)

[Figure 4.6: Result based on Driscoll and Kraay Standard Errors 30](#_Toc179815207)

[Figure 4.7: Coefficient Plot 31](#_Toc179815208)

[Figure 4.8: Model Training 32](#_Toc179815209)

[Figure 4.9: Result of Model Training 33](#_Toc179815210)

[Figure 4.10: Line Chart Comparing Predicted and Actual Result 33](#_Toc179815211)

# List of Table

[Table 3.1: Variables Description 10](#_Toc179815212)

[Table 3.2: Variables Description (Continued) 11](#_Toc179815213)

# Abstract

This research investigates the impact of social inclusion on happiness index, focusing on key indicators across different Human Development Index (HDI) levels. Social inclusion plays a pivotal role in enhancing the well-being of vulnerable groups such as the elderly, people with disabilities, and those living in lower HDI countries. Despite its benefits, inequalities persist between socio-economic groups, limiting access to essential resources, opportunities, and community participation.

This study analyzes social inclusive indicators such as nutrition and medical care, water and sanitation, education, and personal safety, alongside opportunity factors like personal rights, freedom, and inclusiveness, to assess their influence on happiness across 148 countries from 2011 to 2022. Descriptive analytics using Power BI and predictive modeling with R were employed to evaluate these relationships. The findings highlight significant correlations between basic needs, foundations of well-being, and opportunity with happiness index that measured through the Life Ladder score. High-income countries typically exhibit higher satisfaction across most indicators, while low-income countries struggle with access to basic services, reflecting lower happiness levels.

The study’s predictive analytics was conducted using a time-fixed effects model, identifies Personal Freedom & Choice and Health & Wellness as strong positive predictors of happiness index. Interestingly, some variables such as nutrition and medical care, environmental quality, education and personal rights showed negative correlations, suggesting that there is some complex dynamics. The research contributes to understanding how inclusive policies can foster societal well-being, aiding decision-makers in prioritizing interventions that reduce inequality and promote a fairer society.

# Introduction

## **1.1 Research Background**

Social Inclusion is vital for improving overall well-being, especially for vulnerable groups such as older people, individuals with disabilities and those in countries with low Human Development Index (HDI) scores. Gyasi et al. (2023) emphasizes that social inclusion positively affects happiness among older adults in a positive manner by reducing anxiety, depression and sleep problems. Yamb and Ekomane (2019) reveal that financial resources are important for well-being, especially in countries with low HDI, where economic growth can aid in satisfying basic needs and increasing happiness. Therefore, ensuring everyone has equal access to resources, support and economic opportunities is important for cultivating a fair and happy society.

## **1.2 Problem Statement**

Even though social inclusion improves well-being, there are still differences and inequalities between socio-economic groups. Vulnerable groups often face challenges that prevent them from accessing important resources and opportunities to participate in their communities which eventually lower their happiness and quality of life. Guillemot et al. (2024) found that children with disabilities face challenges in being included at school due to factors such as inadequate support from schools and poor relationships between parents and teachers. Therefore, there is a need to understand how social inclusion can help create more happiness for everyone.

## **1.3 Research Questions**

1. What roles do basic needs that includes nutrition and medical care, water and sanitation, housing and safety play in enhancing the happiness index?
2. How does the foundation of well-being such as basic education, information and communications, health and environmental quality affect happiness?
3. In what ways do aspects of opportunity, such as rights and voice, freedom and choice, inclusive society and advanced education impact happiness?
4. Which social inclusion indicators significantly affect nations with different Human Development Index (HDI) levels?
5. How does social inclusion influence the life ladder scores or happiness index across countries over time?

## **1.4 Research Objectives**

1. To assess the roles of basic needs, which include nutrition and medical care, water and sanitation, housing and safety in enhancing happiness.
2. To examine the impact of well-being foundations such as basic education, information and communication, health and environmental quality on happiness.
3. To investigate how opportunities’ aspects such as rights and voice, freedom and choice, an inclusive society and access to advanced education influence happiness.
4. To identify which social inclusion indicators significantly affect nations with different Human Development Index (HDI) levels.
5. To analyse the relationship between social inclusion and happiness using the life ladder scores across countries over time.

# Literature Review

**Social Inclusion**

Social inclusion is a multidimensional concept that aims to guarantee fair and equal participation across social, economic, and political domains for all individuals, regardless of their backgrounds. Mares et al. (2024) conclude that achieving social inclusion in education requires a systemic transformation to ensure equal access, particularly for marginalized groups, through collaboration at all educational levels. Extending this framework, Ibda et al. (2024) stipulate the importance of an inclusive model based on Gender Equality, Disability, and Social Inclusion (GEDSI), which strives to dismantle segregation and mainstream inclusive practices to address structural inequities. Burlea-Schiopoiu and Popovici (2024) emphasize that social inclusion significantly affects entrepreneurial intentions among Generation Z by enhancing their sense of belonging and boosting self-efficacy, thereby supporting cohesive behaviour. In contrast, Curto-Millet and Cañibano (2023) argue that social inclusion and exclusion are interconnected rather than being opposites, and that the design of information systems can either promote inclusion or deteriorate exclusion. Moreover, Gopalakrishnan (2024) affirms the critical role of microinsurance in advancing social inclusion by providing accessible financial protection to vulnerable populations, thus enabling them to manage risks effectively and participate in economic activities. Bringing together these perspectives, it is evident that social inclusion is a complex process that requires systemic and policy transformations across various sectors to create an inclusive and fair society where everyone can succeed.

**Basic Human Need**

Addressing basic human needs such as nutrition, healthcare, water, sanitation, and housing is vital for creating a fair and inclusive society. Mahlert (2023) emphasizes that the satisfaction of basic needs remains a fundamental issue in global social policy, requiring collaboration across different fields to ensure these needs are addressed comprehensively. In the context of urban Bangladesh, Anne et al. (2022) identified significant gaps in the quality of nutrition services provided by public facilities, which mostly affect vulnerable populations and increase health inequalities which eventually hinder social inclusion. Brown et al. (2023) asserts that issues like systemic racism and lack of infrastructure are major barriers to ensuring everyone has access to safe water and sanitation, even in high-income countries. These problems demonstrate a lack of focus on helping marginalized communities, which ultimately leads to social exclusion. Similarly, Omerov et al. (2020) claim that homelessness is associated with significant health problems as individuals must focus on finding shelter before addressing their health needs, resulting in chronic diseases and higher mortality rates compared to those with homes. As a whole, these studies underscore that meeting basic human needs like nutrition and medical care, water and sanitation, housing and safety require fairness, inclusiveness and support for those who need it most. Ensuring access to these basic needs not only improves quality of life but also encourages social inclusion by allowing all individuals to participate and contribute to society.

**Foundation of Wellbeing**

The foundations of wellbeing include basic education, information and communication, health and environmental quality which are vital to enhance social inclusion..Ngozi Elems-Ikwegbu (2023) demonstrates that Singapore’s emphasis on effective teaching methods and high academic standards has made its education system a global exemplar, while (Cajilog & Pabalan (2023) stresses on the significance of making more international to make sure students are well equipped with skills to face the global economy. Access to information and communication can be improved via ICT as it helps individuals and communities become more engaged in education, thereby nurturing social inclusion (Rasool & Naidoo, 2024). The next aspect to investigate is the foundation of wellbeing is health. In order to improve health, (Sturge et al., 2023) found that socially sustainable housing plays a crucial role. It provides a supportive environment that helps people stay physically and mentally fit, especially older adults, by allowing them to stay connected and engaged within their community. Besides, environmental quality is also important for wellbeing. Cavalli et al. (2024) argues for policies to reduce pollution while Eijkelenboom and Bluyssen (2020) claim that indoor environment quality enhances the comfort and health of people. Overall, all these studies portray that access to basic education, information and communication, health and environmental quality is instrumental in creating a community that is socially inclusive.

**Opportunity**

Opportunity includes, rights and voice, freedom and choice, an inclusive society and advanced education access. It is vital in fostering social inclusion. Appleby et al. (2023) emphasize the importance of giving marginalized communities a say in government decisions to ensure effective and equal participation which in turn strengthens democratic engagement. Freedom of choice can be defined as the ability to select one’s desired course of action and it plays an important role in nurturing a sense of autonomy and personal control, which are necessary for well-being and social inclusion (Carter, 2004). To support an inclusive society, Issahaku and Adam (2022) describe how community groups and non-profit organizations help young people feel included by giving them equal opportunities to participate and grow. Furthermore, (Meyer et al., 2024) conduct a systematic review on providing fair access to advanced education to reduce gaps and ensure everyone has the opportunity to thrive and succeed. Collectively, these studies depict that creating more opportunities, giving a voice to marginalized groups, providing access to education, freedom and choice are key factors in building a socially inclusive society.

**Life Ladder**

The Life Ladder, which is also known as happiness index, is influenced by different socio-economic factors like income, education and health. Sihombing (2022) discovered that in Indonesia, factors like income inequality, unemployment and spending levels affect the happiness index, where improved economic growth and higher Human Development Index scores result in higher levels of happiness. Similarly, Senasu et al. (2019) found that in Thailand, happiness levels vary widely due to differences in economic sufficiency, family relations and community engagement. Akgun et al. (2023) claim that in European countries, financial inclusion and economic stability are key factors in boosting happiness, indicating that equal access to resources improves overall well-being. These studies suggest that increasing economic opportunities and reducing inequality can improve overall happiness and encourage social inclusion.

# Methodology

## **3.1 Nature of variables**

The data that has been used for this report is the level of life ladder (life satisfaction) rated from 1 to 10 among 148 countries around the world and has been observed from 2011 to 2022. Hence the data has 1,776 observations. The data contains country name as the entity variable, year as time variable, and the rating of life ladder as dependent variable. The life ladder is the index of happiness that measures the degree of happiness which residents of the country are experiencing (Helliwell et al., 2022). The independent variables are all social inclusion indicators that under basic human need, fundamental of well-being and opportunity. According to the methodology from original source (Stern et al., 2024), all independent variables have been normalized, aggregated, weighted and eventually presenting in an index score that fit between 0 and 100. All dependent and independent variables are recorded in numerical datatype.

|  |  |
| --- | --- |
| **Variables** | **Description** |
| Nutrition and medical care | Do people have enough food to eat and are they receiving basic medical care? |
| Water and sanitation | Can people drink water and keep themselves clean without getting sick? |
| Housing | Do people have adequate housing with basic utilities? |
| Personal safety | Do people feel safe? |
| Access to basic knowledge | Do people have access to an educational foundation? |
| Access to information and communications | Can people freely access ideas and information from anywhere in the world? |
| Health | Do people live long and healthy lives? |
| Environmental quality | Does the environment support societal well-being? |
| Personal rights | Are people’s rights as individuals protected? |
| Personal freedom and choice | Are people free to make their own life choices? |
| Inclusiveness | Is no one excluded from the opportunity to be a contributing member of society? |
| Access to advanced education | Do people have access to the world’s most advanced knowledge? |
| Life Ladder | Happiness index/ life satisfaction rated from 1 to 10 |

Table .: Variables Description

Another variable we have utilized is the Human Development Index (HDI). In our descriptive analytics using Power BI, we employed this variable to categorize countries into four clusters which are low, lower-middle, upper-middle, and high-income countries. The United Nations’ Human Development Index (HDI), as defined by the World Development Indicators, aims to measure a nation’s prosperity by considering both economic and non-economic factors. Non-economic elements include life expectancy and educational achievements, while the economic aspect is represented by gross national income (GNI) per capita.

|  |  |
| --- | --- |
| **Variables** | **Description** |
| Human Development Index (HDI) | a statistical tool used to measure a country’s overall achievement in its social and economic dimensions. |

Table .: Variables Description (Continued)

The data has been merged and preprocessed using python. PowerBI is the business intelligence tool that has been used for descriptive analytics and R and R studio has been employed for predictive analytics.

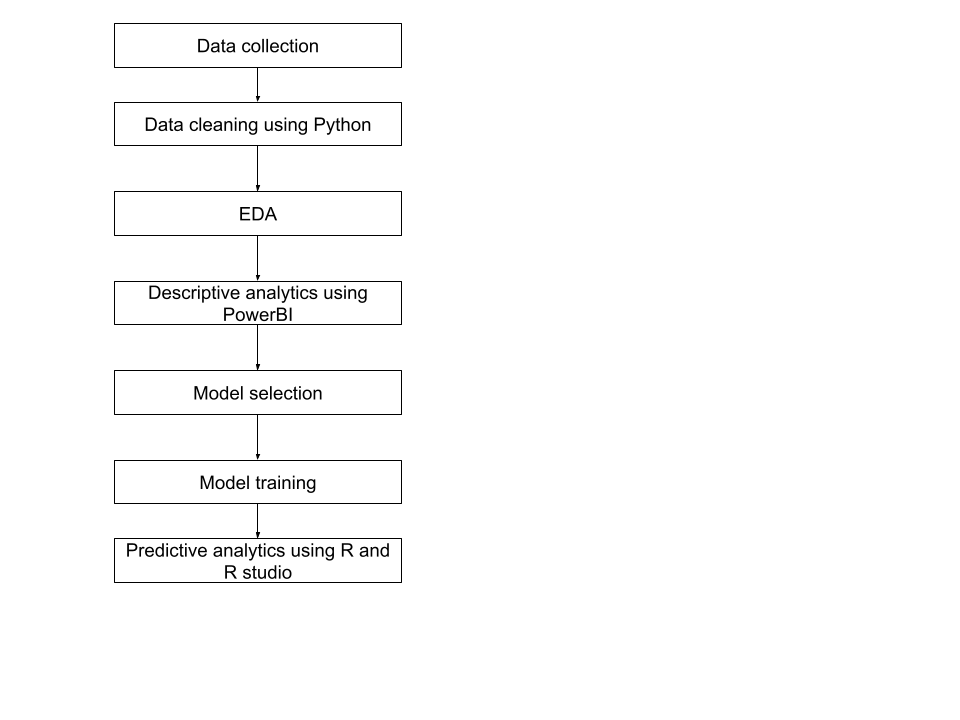


Figure .: The Flowchart of Working Process

## **3.2 Data Pre-processing**

In order to proceed with the descriptive and predictive analytics, the data must first be pre-processed. During this pre-processing stage, the main objective is to provide an accurate representation of the data in order to produce high quality visualizations and prediction models. Through the use of Python using Jupyter Notebook, this was achieved by merging the retrieved data, eliminating unneeded columns, and handling missing values.

Firstly, the name of the countries in the ‘Country name’ columns across the two datasets, ‘SocialInclusiondata’ and ‘Happinessdata’, were standardized in a way where the country names are consistent. This step is necessary to prevent any conflict during the merging process. Once the country names have been corrected, a ‘crossid’ column is created for both datasets, as it is essential in merging the two datasets. The ‘crossid’ column is generated for both datasets by combining the ‘Country name’ and ‘year’ (eg. “Afghanistan\_2011”).

After the ‘crossid’ column has been created, the two datasets may be merged using the left join. Left join was used to combine the two datasets without removing rows from the ‘SocialInclusiondata’ dataset. Due to the nature of using left join, there will be missing values present. This is because the right-side data may not have entries that correspond with the left-side data (eg. no entry for life ladder for Afghanistan in 2018, but there are values for Afghanistan in 2018 in the left-side dataset). As a result of merging the data, it may be observed that there are duplicates in the data which were then identified and removed. By observing the merged dataset, it may be noticed that the majority of the data is within the range of the years 2011 to 2022. Hence, the data was filtered to only output the rows within this timeframe.

After the merging process is complete, the missing values need to be identified and handled. However, due to the nature of the data, there are various conditions that will modify the method of handling the missing values.

Missing value conditions:

1. There are a few missing values for both dependent and independent variables for a given country.
   1. Median imputation is used to fill these values.
2. The values for all independent variables for a country is entirely missing
   1. All entries for that country are removed.
3. The values for a country’s dependent variable are entirely missing
   1. All entries for that country are removed.
4. The values for a country’s independent variables are partially complete (eg. one variable has all values, one variable is partially filled, and/or some variables are empty)
   1. The missing values are left alone
5. There is only one entry for a country’s dependent variable
   1. The missing values are filled with that one value

Aside from cleaning the data with the social inclusion indicators and happiness index, the dataset regarding each country’s HDI needed to be pre-processed as well. This is because the HDI may be used during the visualization process to perform clustering, providing more digestible and comprehensible visualizations through Power BI. The initial process of cleaning the HDI dataset was done by standardizing the country names, filtering the time period to only display values from 2011 to 2022, and dropping rows with missing values in the ‘HDI’ column.

In order to proceed with the clustering process, a data frame that contains the medians of the ‘HDI’ column for each country is created and standardized. Then k-means clustering is performed, using sklearn, grouping the medians into 4 clusters, Low income’, ‘Lower-medium income’, ‘Upper-medium income’, and ‘High income’ (Metreau et al., 2024). The centroids for these clusters are found using both the scaled and original HDI values. These centroids are then sorted in ascending order, where the lowest centroid is labeled as “low income”, while the highest centroid is labeled as “high income”.

## **3.3 Exploratory Data Analysis (EDA)**

### **3.3.1 Description of data**

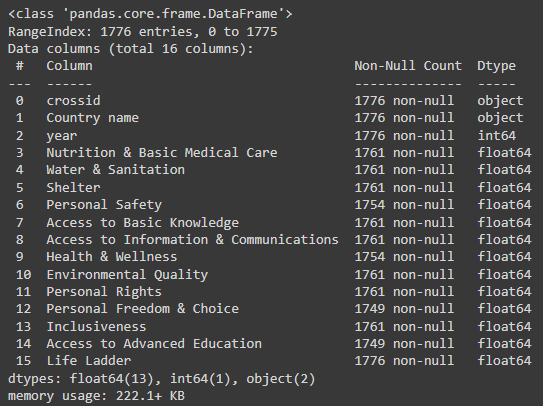


Figure .: Data Description

The figure above displays basic information of the dataset such as the number of rows and columns, the names of each column, and whether they are numerical or categorical (object, integer, or float). It may be observed that there are a total of 1776 rows with 16 columns, 14 columns being numerical, while 2 columns being categorical.

### **3.3.2 Descriptive statistics**

Based on the summary statistics, the results indicate relatively high scores for “Nutrition & Basic Medical Care” and “Access to Basic Knowledge”, with means of 80.39 and 73.99 respectively, highlighting good access and satisfaction from a period spanning 2011 through 2022. Nonetheless, notable fluctuations are observed in Water & Sanitation, with a mean of 76.43, emphasizing significant difficulties in certain regions. There is potential for improvement in Personal Safety (mean: 61.92), Health & Wellness (mean: 57.79), the areas of Environmental Quality (mean: 58.79) and Access to Advanced Education (mean: 49.18) scores. Whereas the average Life Ladder rating of 5.43 represents overall life satisfaction, it also shows that many people aim to raise their overall standard of living.

### **3.3.3 Missing value and outliers**

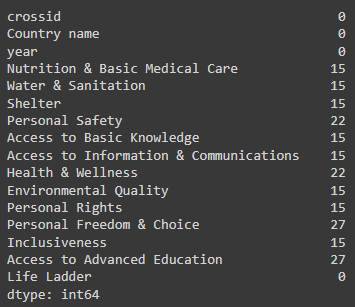


Figure .: Missing Values

The figure above shows the number of missing values for each column, ranging from 15 to 27. It may be observed that the “Personal Freedom & Choice” and “Access to Advanced Education” columns have the most missing values.

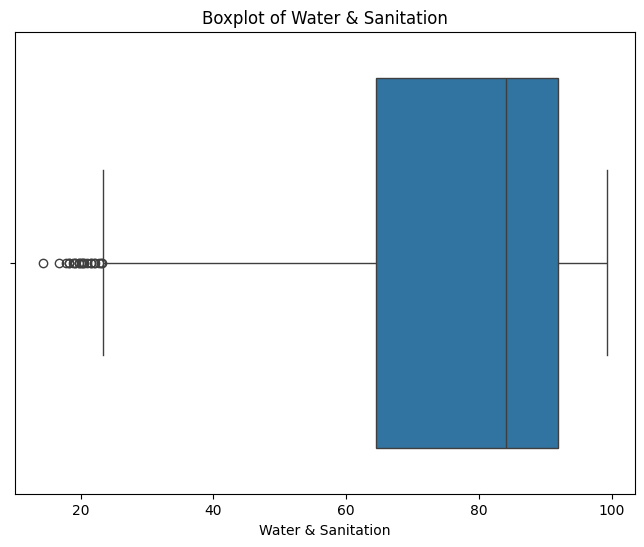


Figure .: Outliers

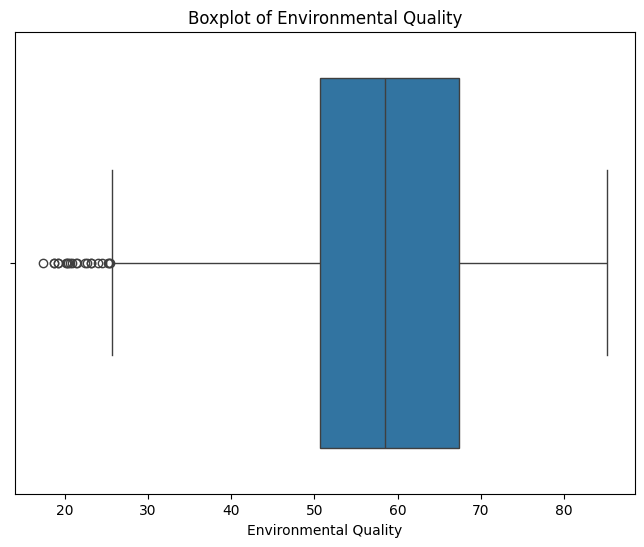


Figure .: Outliers (Continued)

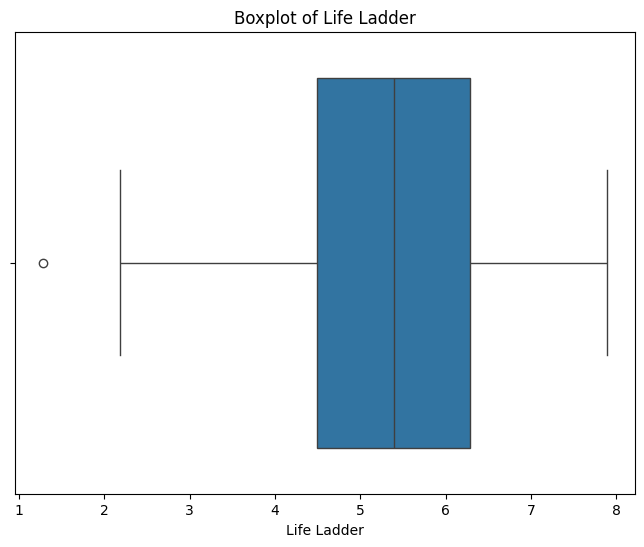


Figure .: Outliers (Continued)

Based on the box plots for the Environmental Quality, Water & Sanitation, and Life Ladder columns, it may be observed that there are outliers present in this dataset. Most of these values are considered outliers due to them falling below each column’s lower fence. Due to the nature of the data, these outliers have not been eliminated. Since these variables are indicators of social inclusion, it would not make sense to remove these outliers, as they represent real-world phenomena.

### **3.3.4 Distribution**

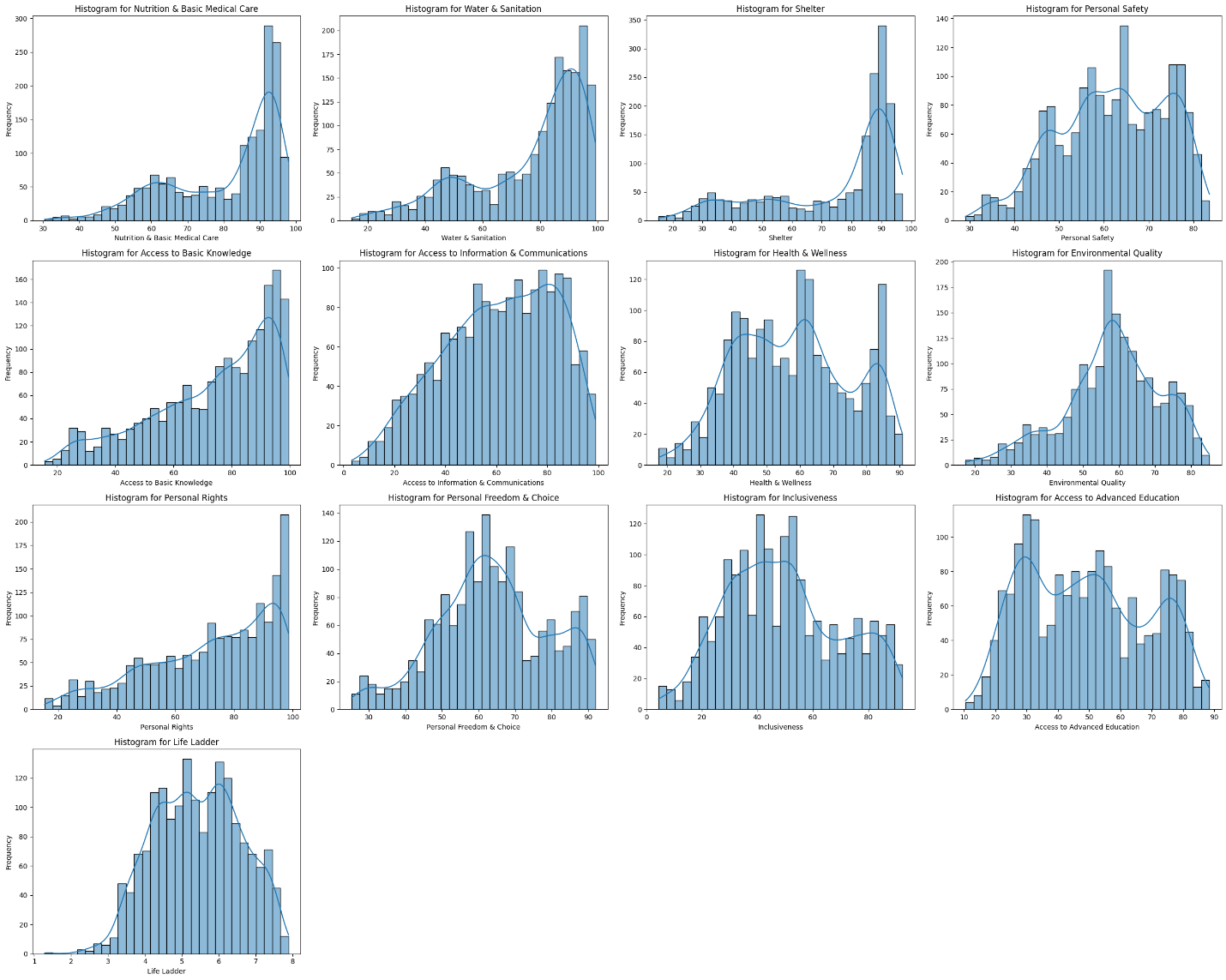


Figure .: Data Distribution

The histograms above represent how the numerical columns in the data set are distributed. Based on the histograms above, it may be seen that the distributions of the “Access to Basic Knowledge”, “Water & Sanitization”, “Personal Rights”, “Nutrition & Basic Medical Care”, and “Shelter” histograms are significantly left-skewed, due to its longer tail on the left side and peak at the right side (higher end of the scale).

Looking further into the Nutrition & Basic Medical Care and Shelter histograms, it may be observed that the data points descend sharply after the peak. This is an indication that the majority of countries have high indicators for these columns.

Furthermore, the rest of the histograms reveal a significant concentration around the middle of the distribution in which Access to Information & Communication is most remarkable.

### **3.3.5 Model Selection for Predictive Analytics**

The figure below outlines the model selection process for predictive analytics using panel data. We begin with the pooled Ordinary Least Squares (OLS) model, which identifies significant predictors of life satisfaction but ignores country-specific effects. To address this, we test the random effects model, and the Breusch-Pagan Lagrangian Multiplier test confirms that it is preferred over the pooled OLS model due to country-specific variations.

Next, we compare the random and fixed effects models using the Hausman test, which suggests that the fixed effects model is more suitable due to a correlation between country-specific effects and the regressors. An F-test comparing the fixed effects model with the pooled OLS model further confirms that the fixed effects model is more appropriate. Thus, we select the fixed effects model for the panel data analysis. Details of this selection process are in the appendix.

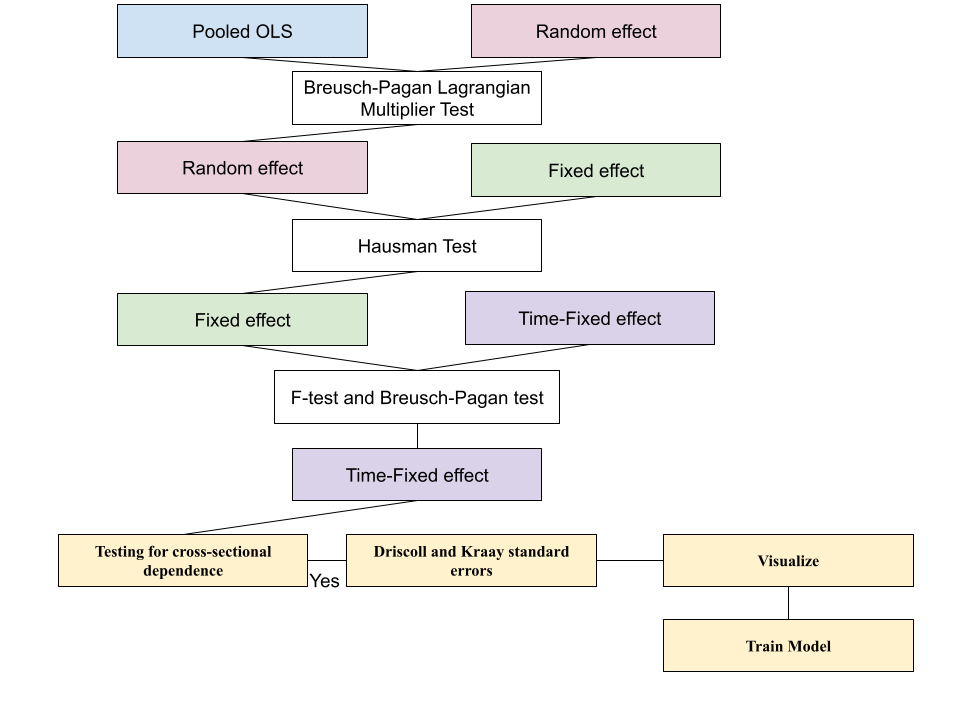


Figure .: Model Selection Process

# Result and analysis

## **4.1 Descriptive analytics using PowerBI**

**Dashboard 1**

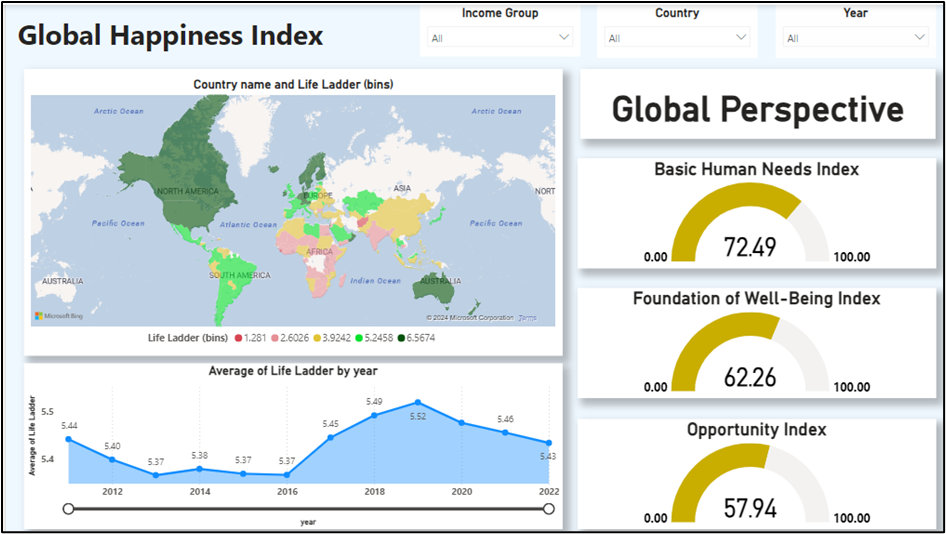


Figure .: Dashboard 1

The diagram above shows the overall overview of the dashboard that analyse happiness level across three metrices which are Basic Human Needs, Foundation of Well-being and Opportunity across countries over time.

The Map Visualization shows the countries that are color coded according to their Life Ladder score. A legend below the map shows the color coding for Life Ladder scores, with bins ranging from the lowest, 1.281 (represented by a reddish hue) to the highest, 6.574 (represented by a greenish hue). The countries which score in between these two figures are shaded with yellow and lighter green color.

The line chart which shows the global average Life Ladder scores from years 2011 to 2022. The Life Ladder scores are represented by y-axis and the years are represented by x-axis. Based on the graph, there is an inconsistent pattern in global well-being over time. In 2011, the score is at 5.44 but slightly dips in 2012 and shows variation until reaching peak of 5.52 in 2019. After 2019, there is a decreasing trend with the scores ends at 5.43 in 2022. The decreasing pattern is seen after 2019 as the world is facing global challenges which had a major impact on life satisfaction. The most notable events that happened during the period is COVID-19 which put the world to a halt. This pandemic has caused many people to lose their jobs which led to an economic downturn. This has caused people to become less happy as they are stressed and depressed with the things that happened hence the Life Ladder score decreases.

On the right-hand side of the dashboard, there are three-gauge charts which are Basic Human Needs Index, Foundation of Well-Being Index and Opportunity Index, each reflecting a different dimension of social inclusion. The Human Needs Index measures the extent to which basic human needs like food, medical, shelter and safety are met. These metrics show the score of necessities that meet across countries or globally. The Foundation of Well-Being Index assesses border well-being factors beyond the basic needs of humans such as health, education, good environment and access to information and communication. The Opportunity Index measures the opportunities available to individuals including economic prospects, advanced education, social mobility and freedom.

**Dashboard 2**

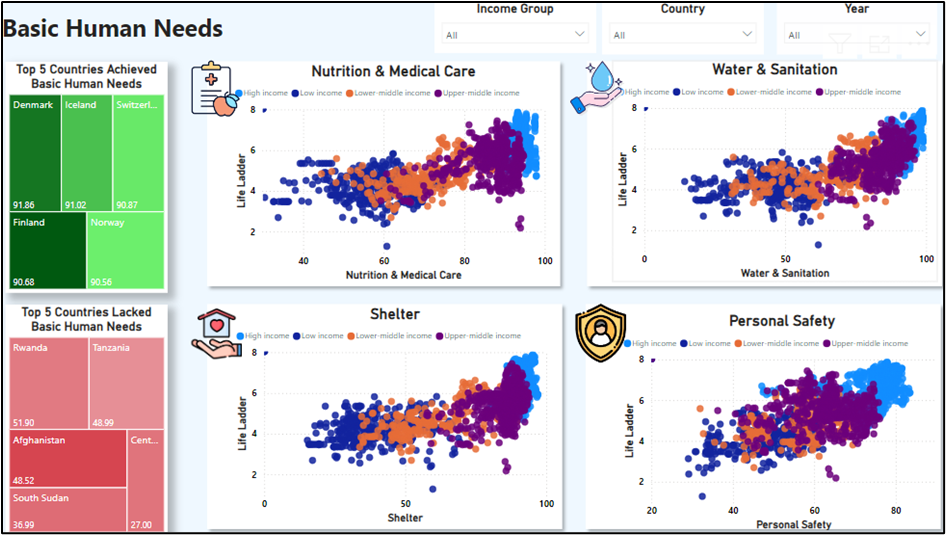


Figure .: Dashboard 2

In this dashboard, it shows the relationship between essential living conditions and people’s happiness. The top 5 achievers’ countries like Denmark, Iceland and Switzerland have of score over 90, indicating that these countries are performed well in meeting basic human needs. In contrast, countries like Rwanda, Tanzania and Afghanistan have a score around 50 and below. These countries struggle to meet the basic needs of the people, which reflect the severity of their socioeconomics.

There are four scatter plots which highlighted the relationship between the Life Ladder Score and independent variables that highlight the specific human needs which are Nutrition & Basic Medical Care, Water & Sanitation, Shelter and Personal Safety. These countries are clustered into four income groups based on HDI which are represented by different colors which give in depth on how wealth distribution effect the basic needs of the populations.

The scatter plot shows a positive correlation between the independent variables for Basic Human Needs and the Life Ladder scores across countries. High-income countries are clustered at the higher of both x and y axis which show the best access to healthcare and nutrition which have a range of score between 80 to 100 with the highest Life Ladder score (between 6 to 8). Countries with upper-middle income are scattered across the mid-range of graph with 60 to 90 on the x-axis and 5 to 7 on the Life Ladder score. With the score to the nutrition and medical access range between 50 to 70, the cluster of the low-income countries are spread across the quadrant with the Life Ladder scores ranging from 4 to 6. Towards the end left of the quadrant, low-income countries can be seen in this range, with scores for nutrition and medical between 40 to 60 and the lowest Life Ladder score among other income groups, mostly between 3 and 5. Overall, the scatter plot shows the impact of income levels on well-being, with countries that have higher income experiencing better access to the basic needs hence increased the level of happiness. In contrast, low-income countries face more challenges in these areas hence degrading the level of satisfaction.

For Water & Sanitation, the scatter plot shows a positive correlation with better access to these leading to higher satisfaction levels. With scores between 80 to 100 in the x-axis and Life Ladder scores between 6 to 8, the high-income countries are concentrated in this region. Upper-middle income countries are dominating in the range of 60 to 85 on the x-axis and Life Ladder scores between 5 and 7. Lower-middle income countries show lower access between 40 to 70 in access to water and sanitation with corresponding Life Ladder scores of 4 to 6. In the end spectrum, the low-income countries have the lowest access (below 60 on the x-axis) and the lowest Life Ladder scores mostly between 3 and 5. This shows that nations with better access to water and sanitation significantly boosts happiness and it is represented by the wealthy nations in the datasets. Low-income countries which face challenges with these resources are experiencing lower satisfaction levels.

The scatter plot for Shelter and Life Ladder shows a positive correlation across different income groups, where higher subjective well-being corresponds to better access to shelter. In the top right, high-income countries are clustered in this range with the scores between 80 to 100 in x-axis and the highest Life Ladder scores, between 6 to 8, indicating greater well-being. Upper-middle income countries have moderate-to high access with the scores of 60 to 90 in x-axis and Life Ladder scores between 5 and 7. In the mid-range of scatter plot, the lower-middle class incomes are scattered in this range with the score of shelter access of 40 to 70 and satisfaction levels between 4 to 6. In the end of the left quadrant, the low-income countries are dominating this area with limited shelter access, below 60 and lower Life Ladder (scores between 3 to 5). The trend that is shown in this graph highlights the critical impact of adequate shelter on improving the well-being with high income nations benefiting the most and low-income nations facing challenges that affected their happiness levels.

Personal Safety shows high Life Ladder scores for countries in the high-income bracket with the scores of 70 to 90 in the x-axis and 6 to 8 in the y-axis, reflecting both high safety and well-being. However, there are some outliers that can be seen for the high-income group, which shows relatively high levels of personal safety (between 70 to 80), yet the Life Ladder scores are unexpectedly low, around 4 to 5. This shows that despite having strong personal safety, the overall well-being of the countries is lower than expected. Upper-middle income countries have moderate personal safety scores between 50 to 80 with the satisfaction level between 5 and 7 while lower-middle income countries fall in the range of 40 to 60 in the x-axis and happiness level between 4 and 6. Low-income countries dominate the bottom left quadrant, with scores below 60 for personal safety and Life Ladder scores, mostly between 3 and 5. This trend emphasis that high level of personal safety is strongly liked to higher level of satisfaction.

**Dashboard 3**

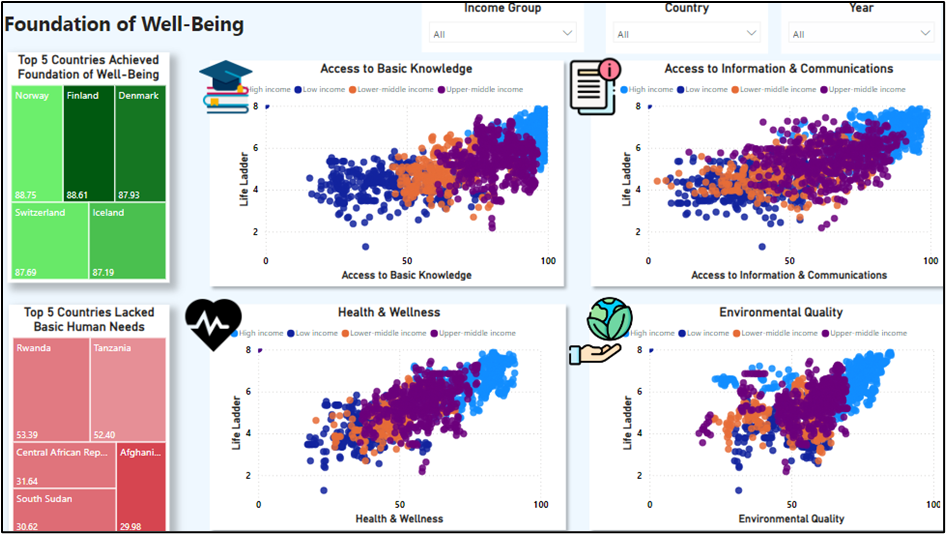


Figure .: Dashboard 3

Foundation of Well-being explores the key dimensions of well-being that go beyond the basic human needs. Similar to the second dashboard, the layout also provides the same parameters using tree maps and scatter plot. Once again, countries like Norway, Finland and Denmark score the highest which prove that these countries provide strong foundational system for their citizens to access education, information, healthcare and healthy environment. In contrast, countries in the red gradient map tree are among the most struggling in this area with low scores indicating serious deficiencies in their infrastructure.

Four scatter plots show the relationship between well-being and Life Ladder scores. The first scatter plot which is Access to Basic Knowledge and Life Ladder shows a positive correlation which indicates education can improve happiness scores. At the upper-right quadrant, high-income countries are scattered around the area with the score of Access to Basic Knowledge ranges between 80 to 100 with the Life Ladder score between 6 and 8, which explaining the strong relationship between education and well-being. Upper-middle-income countries are showing moderate scores in the scatter plot with 60 to 90 in the x-axis and Life Ladder scores between 5 and 7. With Access to Basic Knowledge scores ranging from 40 and 70 with the Life Ladder scores between 4 and 6, the lower-middle income countries are dominating in this quadrant. At the very bottom left of the scatter plot, the low-income countries are situated most in this area with access below 60 and lower happiness score (3 to 5). This trend shows the important role of having access to basic education as it plays a crucial role in the satisfaction level.

Access to Information & Communication also shows a similar pattern where wealthy nations exhibit a higher level of satisfaction. The wealthy nations are clustered at the very top left of the scatter plot with the range between 80 to 100 for the x-axis and 6 to 8 for the Life Ladder scores. The upper-middle income countries which surround the area of 60 to 90 for the x-axis and corresponding Life Ladder scores between 5 and 7. Lower-middle income countries range from 40 to 70 for the Access to Information & Communication scores and Life Ladder scores between 4 and 6. In the bottom left of the graph shows the low-income nations which have both lowest for the x-axis and y-axis scores with range below 60 and below 5 respectively. This pattern highlights that access to information and communication technologies play an important role in enhancing well-being, with the high-income nations benefiting the most and lower-income nations facing greater challenges.

For Health & Wellness, the same trends can be also seen here with the high-income nations dominating the top left chart with the scores of 80 to 100 for the x-axis and Life Ladder scores between 6 and 8. The upper-middle income nations clustered at the area of 60 to 90 for the x-axis and 5 to 7 for the y-axis. For the lower-middle income countries, their access scores to these resources range from 40 and 70 with the Life Ladder scores mostly between 4 and 6. Low-income countries are concentrated in the lower end left, with scores below 60 for x-axis and Life Ladder scores between 3 and 5. This pattern has shown the important role of healthcare in improving the well-being of the nations, with wealthier countries benefiting the most while lower-income countries facing greater challenges in accessing the basic healthcare system.

Environmental Quality has a positive correlation on the Life Ladder, indicating that better environmental conditions are associated with higher well-being across income group. In the upper right, the high-income countries are showing higher scores ranging between 60 and 100 in the x-axis and Life Ladder scores between 6 and 8. However, there are high-income countries that show relatively high scores in environmental quality, but their Life Ladder scores are lower than expected, between 4 and 5, which contrast with the most high-income countries. This pattern suggests that despite having a better environment quality, these countries report lower subjective well-being.

In the same scatter plot, the upper-middle income nations show moderate environmental quality ranging between 50 to 80 and satisfaction level of 5 and 7 while lower-middle income nations clustered within 40 to 70 range for environmental quality scores and reported Life Ladder scores mostly between 4 and 6. Countries with lower-income populations are concentrated in the lower left with environmental quality below 60 and Life Ladder scores between 3 and 5. This trend suggests that the majority of wealthy nations are experiencing better environmental quality in their countries hence improving their satisfaction while poorer environmental conditions in lower income countries correlate with lower happiness.

**Dashboard 4**

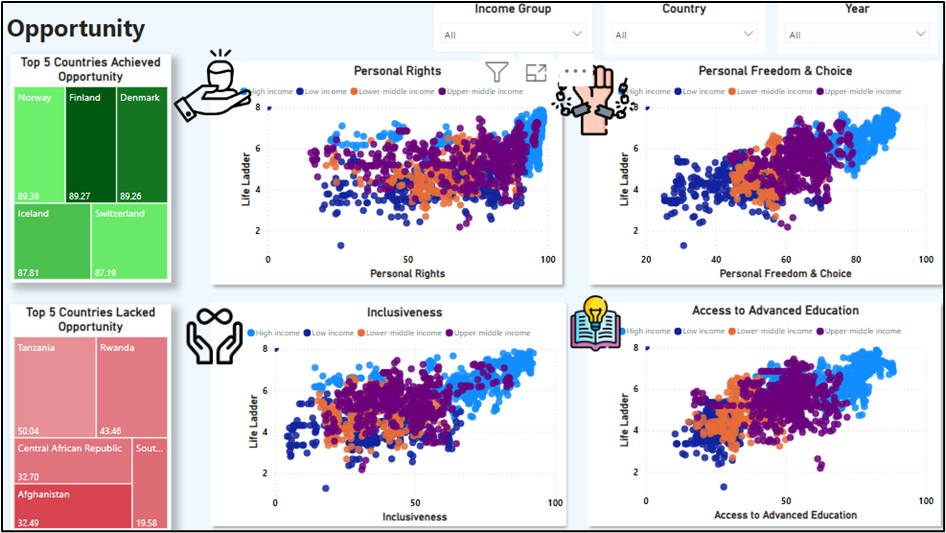


Figure .: Dashboard 4

This dashboard shows how various opportunities such as personal rights, freedom, inclusiveness and access to advanced education affect the Life Ladder scores across different countries. The two tree maps on the left-hand side show the top 5 countries with the most and least countries in the Opportunity spectrum. Countries like Norway, Finland, Denmark, Iceland and Switzerland scores the highest for the opportunity index. In contrast, countries like Tanzania, Rwanda and the Central African Republic are at the bottom of the spectrum. The tree maps provide a clear representation of countries with both tree maps showing countries at opposite ends of the happiness spectrum.

The first scatter plots which is Personal Rights and Life Ladder shows a positive relationship. High-income countries are grouped at the upper-right, with high personal rights, scores between 80 and 100 and Life Ladder scores between 6 and 8. The outliers in this graph can be seen with the high-income nations with some of them are scattered between 4 and 5 in Life Ladder scores despite having high Personal Right scores. Upper-middle income countries fall between 60 and 90 in the Personal Right scores and Life Ladder scores between 5 and 7 while, lower-middle income countries show moderate access with 40 to 70 in x-axis and satisfaction level between 4 and 6. The low-income countries score the lowest with Personal Rights below 60 and Life Ladder scores between 3 and 5.

For Personal Freedom & Choice indicates the importance of having strong personal autonomy which improves the nations’ well-being. The scatter plot shows the positive correlation with high-income countries dominating in the top left, with personal freedom scores between 80 and 100 with Life Ladder scores between 6 and 8. Upper-middle income countries show moderate scores of 60 to 90 in personal rights corresponding the satisfaction level scores between 5 and 7 while lower-middle income countries have personal freedom levels between 40 and 70 with Life Ladder scores mostly between 4 and 6. At the bottom left of the graph, the low-income countries concentrate in these regions with personal freedom below 60 and Life Ladder scores between 3 and 5.

In inclusiveness, the pattern distributed in the graph is positive correlated with higher income scattered at the upper right with the inclusiveness scores of 80 to 100 and Life Ladder score between 6 and 8 which shows that social inclusion brings more happiness to the people. However, there are some outliers in the high-income group which are at the upper left, indicating that there are some others social economic factors that play a role in their well-being. Upper-middle income countries score between 60 to 90 for the inclusiveness and corresponding Life Ladder scores between 5 and 7 while lower-middle income countries show lower inclusiveness score level between 40 and 70 and Life Ladder scores mostly scattered around 4 and 6. For the lower-income countries, they have the lowest score for both axis where the inclusiveness scores are below 60 and the Life Ladder scores between 3 and 5. By looking at the pattern, it shows that countries with high-income excel in providing inclusiveness to their people, although there are some outliers that may suggest other factors than this variable that bring them happiness. Lower-income nations feel less inclusive thus reporting lower happiness scores.

The final scatter plot shows the positive correlation between Access to Advanced Education and the Life Ladder across the income group. On the upper left of the scatter plot, the high-income nations are plotted around the axis with the scores of 80 to 100 in the x-axis and 6 to 8 in the y-axis. However, there are some outliers in the income group that may suggest that education alone does not contribute to happiness. Upper-middle income and low-middle income countries are scattered below the higher income group with a moderate score of 80 to 40 for the education and 7 to 4 for the Life Ladder. At the bottom-left of the chart, the low-income groups are scattered around the area with the score below 60 for the access to advanced education and 4 to 6 for the Life Ladder scores.

In summary, nations with high levels of incomes tend to have a higher score in the Life Ladder which proves that the populations there are having a higher satisfaction level across the variables measured. Conversely, the nations in low-income brackets are dominating the lower left quadrant of the scatter plot graph which shows that the people living in these countries have experienced lower satisfaction level for all the variables measured.

## **4.2 Predictive analytics using R and R studio**

Before we select the predictive model, we have handle the missing value and time-fixed effect model have been selected. The handle missing value process and the whole model selection process has been included in appendix.

### **4.2.1 Model Used**

**Focusing on fixed effects model**

Since we run a lot of tests to figure out that fixed effects model is suitable to analyse our panel data, we will discuss the result of fixed effect model in this section. The whole process of this model selection can be found in the appendix.

**Time-fixed effect**

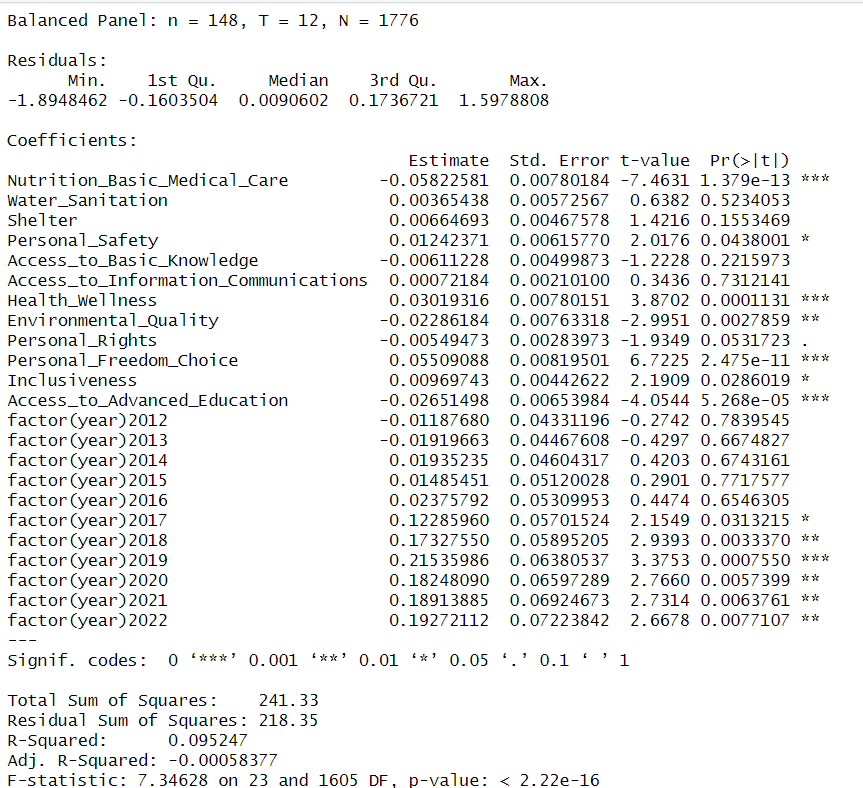


Figure .: Result of Time-Fixed Effects Model

In the time-fixed effects model, several interesting patterns emerge when controlling for time-specific shocks that might affect the Life Ladder. The variable Nutrition & Basic Medical Care retains a significant negative effect on life satisfaction, with an even larger negative coefficient than in previous models. This indicates that improvements in nutrition and basic medical care are associated with lower life satisfaction, which continues to challenge conventional expectations.

Health & Wellness and Personal Freedom & Choice maintain their positive and significant effects on life satisfaction, reinforcing their critical roles in enhancing well-being. Personal Safety also continues to show a significant positive effect, suggesting that a sense of safety enhances life satisfaction.

Environmental Quality has now become statistically significant (p = 0.0216), demonstrating a negative effect on life satisfaction. This counterintuitive finding implies that as environmental quality improves, life satisfaction decreases, possibly reflecting trade-offs or challenges individuals face due to increased environmental regulations or changes.

The variable Personal Rights shows a significant negative effect, indicating that improvements in personal rights are linked to lower life satisfaction in this model—another surprising outcome. Similarly, Access to Advanced Education becomes significant with a negative effect, suggesting that higher levels of education are associated with reduced life satisfaction. This unexpected result may reflect educational stress, unmet expectations, or competition in environments where advanced education is more accessible.

Inclusiveness demonstrates a weak positive relationship with the Life Ladder, suggesting that higher levels of inclusiveness correlate with greater life satisfaction. This implies that societies that are more inclusive and equitable may foster overall well-being, though the effect is not strongly significant in this model. Other variables, including Water & Sanitation, Shelter, Access to Basic Knowledge, and Access to Information & Communications, are not significant in this analysis.

Regarding the time variables, Significant positive effects are noted for several years (2018-2022), indicating overall improvements in life satisfaction during these years compared to the baseline year. This aligns with broader social or economic developments during these years, possibly reflecting recovery or progress from global disruptions caused by the pandemic. The inclusion of time effects accounts for temporal shocks and emphasizes how societal and global changes over time interact with individual life satisfaction.

**Driscoll and Kraay standard errors**

After we run the test for cross-sectional dependence or contemporaneous correlation using Breusch-Pagan LM test of independence and Pesaran CD test, it showed that there is presence of cross-sectional dependence, Driscoll and Kraay standard errors that allows for both heteroskedasticity and autocorrelation were employed to accommodate cross-sectional dependence.

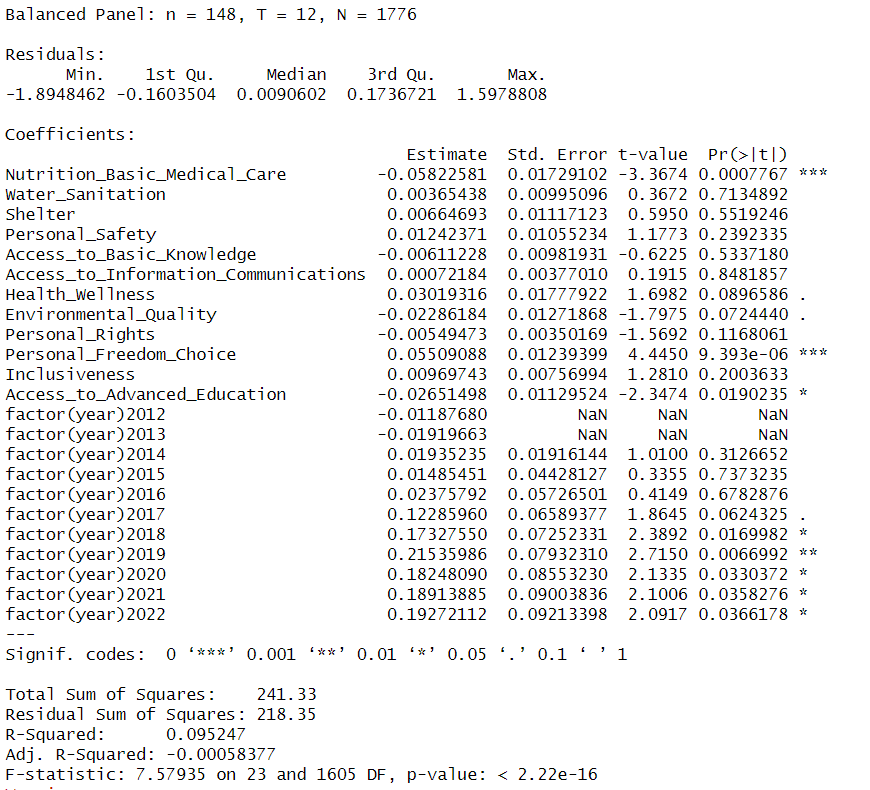


Figure .: Result based on Driscoll and Kraay Standard Errors

The results from the Driscoll and Kraay adjusted time-fixed effects model reveal several significant relationships between socio-economic variables and life satisfaction (Life Ladder). Notably, Nutrition & Basic Medical Care has a significant negative effect on Life Ladder, suggesting that increased access to basic medical care and nutrition may paradoxically be associated with lower life satisfaction, which warrants further investigation. Conversely, Personal Freedom & Choice shows a significant positive effect, indicating that individuals’ ability to make autonomous life decisions is crucial for enhancing life satisfaction. Additionally, the negative effect of Access to Advanced Education on life satisfaction is significant, implying that while education is vital for long-term development, it may not directly contribute to immediate life satisfaction, potentially due to the stresses associated with educational attainment. In terms of yearly trends, the year 2013 had a significant negative effect compared to the reference year, reflecting a possible downturn in life satisfaction during that period. Other factors, such as Water Sanitation, Shelter, and Personal Safety, did not show statistically significant effects, suggesting that while they are important for overall well-being, their variation within the dataset does not directly correlate with life satisfaction in a meaningful way. Overall, although the model explains only a modest portion of the variance in life satisfaction (R-squared = 9.52%), it highlights the complex and sometimes counterintuitive relationships between socio-economic factors and well-being, emphasizing the need for nuanced approaches in understanding these dynamics.

**Coefficient Plot**

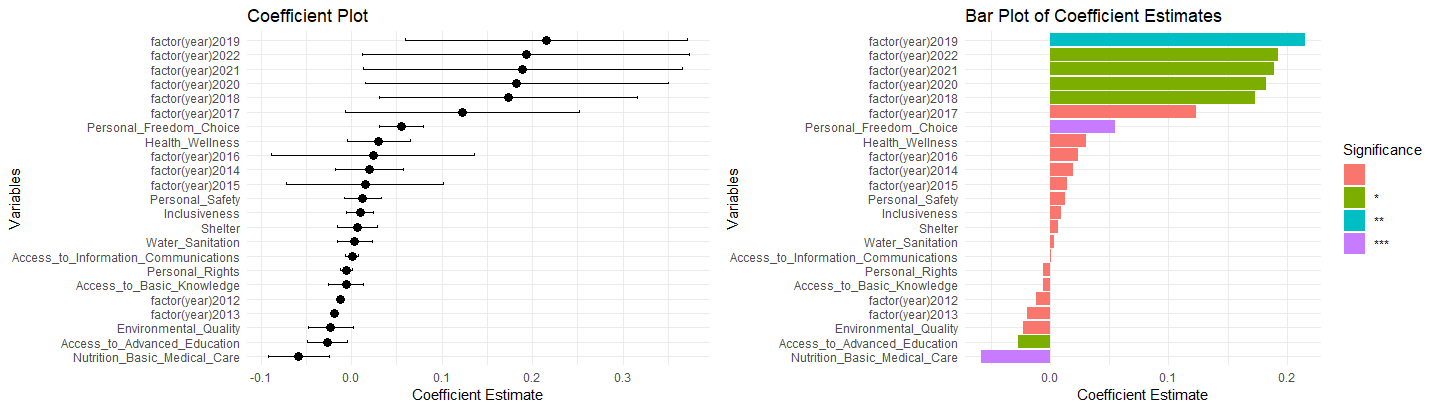


Figure .: Coefficient Plot

The chart above displays coefficient estimates. Variables such as Nutrition and Medical Care, Access to Advanced Knowledge, Environmental Quality, Access to Basic Knowledge as well as Personal Rights performed negative coefficients in the model, indicating the complex and counterintuitive relationships between various factors and life ladder.

Although access to nutrition and medical care is typically associated with enhancing life satisfaction, there may be situations where increased investment in these areas does not lead to immediate improvements in perceived life quality. For example, poor quality of care, inequalities in access or high costs associated with these services might lower the overall life satisfaction, especially in regions where people expect better service but do not receive it (Pemau et al., 2024; Wray et al., 2021). Based on the research from Möwisch et al. (2021) and Wilson Fadiji & Lomas (2024), while higher education generally improves individual capabilities and career prospects, it can also increase stress, competition, and pressure to succeed, which could reduce life satisfaction, especially in highly competitive environments. Other than that, good environmental quality is generally associated with higher life satisfaction. However, it may not be immediately perceived as impacting life satisfaction. For example, the efforts to improve environmental quality might lead to short-term disruptions, personal sacrifies or costs of industrial jobs that will negatively impact life satisfaction (Silva, De Keulenaer, & Johnstone, 2012). Increased access to basic education could theoretically boost satisfaction, but in some cases, the acquisition of knowledge may lead to dissatisfaction by exposing people to unmet opportunities, inequalities, or frustrations with the system. People might become more aware of their lack of opportunities or face competition and stress in pursuing further education or employment (Araki, 2022). According to Steckermeier (2021), the negative correlation between personal rights and the life ladder suggests that multiple factors, such as political stability and cultural context, could influence this relationship. Greater awareness of personal rights can make people more conscious of social inequalities or injustices, especially if they realize they are being denied certain rights. This awareness might lead to dissatisfaction, as people may feel more disappointed or upset about their situation.

### **4.2.2 Model training**

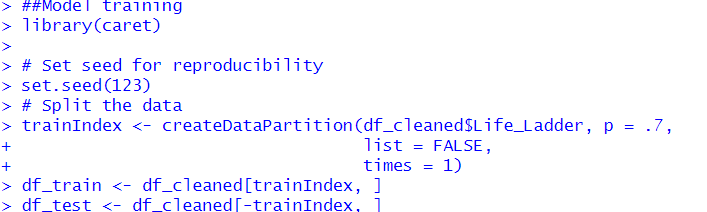


Figure .: Model Training

The analysis conducted using a fixed effects time model aimed to predict life satisfaction, represented by the variable Life\_Ladder, based on various socio-economic factors. The dataset was split into training and testing sets in a 70:30 ratio to facilitate model training and evaluation.

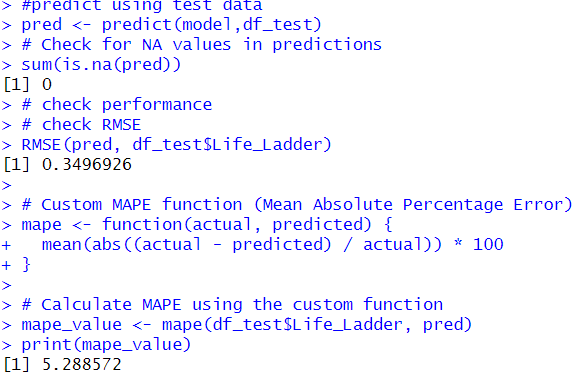


Figure .: Result of Model Training

Although the time fixed effect model explained only 9.52% of the variance in life satisfaction, indicating that other unmeasured factors likely influence this outcome, the model performed reasonably well in predicting Life Ladder values on the test set. The Root Mean Square Error (RMSE) was approximately 0.35, suggesting a modest fit, and the Mean Absolute Percentage Error (MAPE) of 5.29% indicated that the average prediction error was low relative to actual values. Key variables such as Personal Freedom and Choice and Access to Advanced Education were significant predictors of life satisfaction, providing valuable insights into how socio-economic factors impact people’s happiness.

**Predicted VS Actual Result**

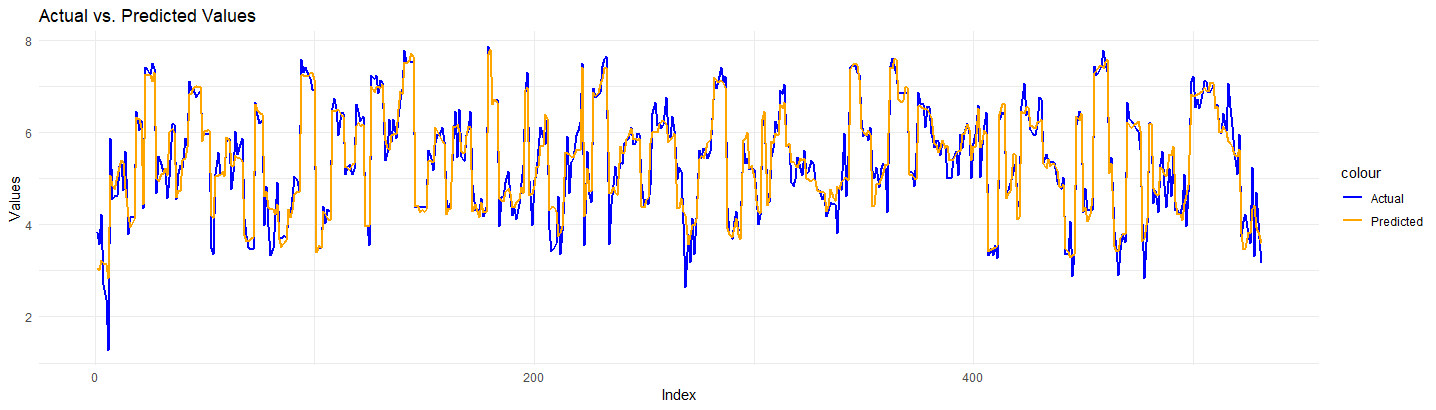


Figure .: Line Chart Comparing Predicted and Actual Result

The line chart above compares actual versus predicted values from a predictive model over an index of observations. The blue line represents the actual values, while the orange line shows the predicted values. The close alignment of the two lines indicates that the model performs reasonably well, as the predicted values track the actual values closely across the majority of the index.

# Conclusion

When it comes to analyzing a country’s happiness index, difficulties arise in identifying which social inclusion indicator has the most significant impact. In order to gauge a better understanding of how these social inclusion indicators affect happiness index, descriptive and predictive analytics have been performed. Preliminary descriptive analytics was performed through Exploratory Data Analysis using Python, while the main descriptive analytics was performed through dashboard development using Power BI. Descriptive analytics was performed with the intention of providing more comprehensible insights in the form of visualizations. In terms of predictive analytics, regression models were created and tested using R, based on the panel data regarding social inclusion indicators and happiness index. The best model will be selected based on various evaluation metrics and testing methods.

By using Power BI, dashboards were developed containing visualizations that provide insights on how the social inclusion indicators affect a country’s happiness index. This was achieved by creating dashboards that display the social inclusion indicators based on “Basic Human Needs”, “Foundations of Wellbeing”, and “Opportunity”. One dashboard was created to display the summarized information, while three additional dashboards were created to display the relationship of the individual indicators against the happiness index, producing scatter plots. Furthermore, the countries were also clustered based on different HDI levels. By clustering the countries based on HDI, the dashboard can display different perspectives on how happy a country is based on different levels of HDI and how the social indicators affect its happiness index. By observing these dashboards, nations can gain key insights on what social inclusion indicators they should aim to improve to increase their overall happiness index.

In terms of the regression models for predictive analytics, the best model has been identified and selected based on performing multiple tests comparing models. As a result, it was found that the best predictive model is “Time-fixed effect”. Furthermore, inspecting the Driscoll and Kraay standard errors, the evaluation metrics and able to be analyzed. It may be noted that the “Time-fixed effect” model performs the best for the social inclusion indicators and happiness index data due to its low RMSE and MAPE values.

Our journey through the implementation of business intelligence tools such as Power BI, Python and R Studio has been insightful. From the start, there were significant learning curves especially in the the processes of data cleaning, merging datasets and handling missing values in Python. This process required attention to detail, as discrepancies between datasets such as inconsistent country names had to be resolved before moving forward with the analysis. However, once the data was pre-processed and prepared, the workflow become more efficient and smoother.

Besides, using Power BI for descriptive analytics was instrumental in our learning experience. It enabled us to create interactive and visually appealing dashboards that provided valuable insights into global life satisfaction trends. Features like map visualizations and scatter plots made it easy to explore relationships between variables such as basic human needs and the Life Ladder score. Power BI’s user-friendly interface and flexibility allowed us to uncover key patterns in the data.

Moving on to predictive analytics, R Studio has its own set of challenges. Since this is our first time handling panel data, understanding the differences between pooled OLS, random effects, and fixed effects models was crucial for selecting the right approach. Through tests like the Breusch-Pagan and Hausman tests, we learned the importance of accounting for country-specific effects in panel data analysis. Despite the complexity, the ability to build predictive models and interpret their results was one of the most valuable learning experiences. It provided deeper insights into which socio-economic factors most strongly influence life satisfaction.

Overall, this business intelligence journey greatly enhanced our ability to approach data analytically and use advanced tools to obtain actionable insights. It was a comprehensive and enlightening learning experience that improved our technical skills and deepened our understanding of how data-driven decision-making can be applied to real-world problems.

**Recommendations for improvement**

There are some recommendations for improvement for this project.To improve the predictive model, incorporating variables such as GDP, Gini Index, employment rate and many other variables will enhance its explainability. These factors provide a clearer understanding of the socio-economic conditions that affect happiness levels. Incorporating these variables will make the model better at identifying what impacts happiness and lead to more accurate predictions. Additionally, to make the visualizations in Power BI more effecting, including a variety of chart types can offer deeper and better insights. By diversifying the chart types, the data can be presented more interactively which makes it easier for stakeholders to comprehend key trends and patterns found in the data.

## **Following questions**

1. **Why Microsoft Power BI is important to take decision?**

Microsoft Power BI is a powerful business analytics tool that is important for decision-making as it simplifies raw and complex data into understandable visualizations, including interactive dashboards that make insights accessible. It ensures that users have access to the most recent information by establishing real-time connections with several data sources. Additionally, we can create reports and dashboards without having to process advanced technical skills because the platform is simple to use due to its user-friendly interface. Power BI makes it easier for teams to share insights with one another and unifies data from various sources to deliver a comprehensive perspective, which promotes cooperation. This integration and ease of use enable organizations to make timely, data-driven decisions that enhance strategic planning and operational efficiency.

1. **Why Descriptive Analytics methods are important to take decisions?**

Descriptive analytics methods are vital for decision-making as they allow organizations to comprehend past and present performance by summarizing large amounts of data into meaningful patterns and trends. By analyzing historical data, these methods offer insights into what has occurred, helping businesses recognize key performance drivers, identify emerging trends, and detect any anomalies. Complex data becomes more accessible and easier to interpret through clear visualizations like reports, charts, and dashboards. Descriptive analytics supports informed decision-making by providing factual evidence, enabling organizations to make better decisions and improving strategic planning, resource allocation, and operational efficiency.

1. **Why Predictive Analytics methods are important to take decisions?**

Predictive Analytics methods are vital for decision-making as they estimate future outcomes and trends based on historical data, statistical algorithms and machine learning techniques. By identifying patterns and relationships within the data, predictive analytics help organizations to foresee changes and take proactive measures. They help with risk assessment by detecting potential negative events, improve consumer insights for targeted marketing and enhance resource allocation by forecasting demand. These techniques also improve operational effectiveness, assist long-term strategic planning, and generate a data-driven culture that values insights over instincts.

1. **Why Regression model is important to take decisions?**

A regression model is crucial for decision-making as it defines the relationships between variables, helping organizations understand how changes in one factor can influence outcomes. By analyzing past data and identifying patterns, regression models offer predictive insights that enhance resource allocation, risk management, and strategic planning. These models provide a data-driven approach, reducing reliance on intuition and enabling businesses to evaluate the impact of different solutions in complex scenarios with multiple influencing factors. By recognizing trends over time and estimating the effect of specific variables, regression models optimize decision-making processes and support long-term forecasting and performance improvement.

1. **What is the concept of data mining in business scenarios?**

Data mining in business involves analyzing large datasets to uncover hidden patterns, trends, and insights that support decision-making. Techniques such as classification, clustering, regression, and association enable businesses to derive valuable information from raw data to better understand customer behavior, enhance operations, and predict future outcomes. For example, market basket analysis reveals which products are frequently purchased together, helping companies refine marketing strategies, cross-selling, and product placement. In order to enhance proactive decision-making, predictive analytics forecasts future trends using previous data. Moreover, data mining is used in fraud detection to identify odd patterns that point to fraudulent activity. In risk management, data mining evaluates potential hazards, strengthening strategic responses. Then, sentiment analysis is another key component for data mining. It uses social media and consumer input to measure public opinion and enables prompt corrections. Lastly, by emphasizing market gaps and desirable features, data mining insight guide product development. Overall, data mining transform data into a strategic asset which fosters improved decision-making and competitive advantage.

1. **How Business Intelligence subject is aligned and important in your degree program with practical example.**

Business Intelligence plays a vital role in our degree program, equipping us with the analytical tools needed to evaluate trends and patterns in order to support strategic decision-making and foster business growth. A solid understanding of BI cultivates a data-driven mindset, enabling us to leverage data as a powerful tool to solve complex business challenges. The program emphasizes the development of critical analytical skills, which helps us transform raw data into meaningful insights. Through hands-on experience with tools like R and Power BI, we gain essential skills in advanced analysis and data visualization. For example, in our Business Intelligence course project, we were introduced to panel data analysis, using tools like Python for data preprocessing, Power BI for descriptive analytics and visualization, and R for predictive modeling. By analyzing social inclusion factors that influence the Happiness Index across countries across time, we managed to transform raw data into actionable insights and gain a deeper understanding of global happiness trends. This project provided us with hands-on experience addressing real-world social issues, teaching us the importance of precise data handling. Overall, this experience has sharpened our skills and provided valuable insights that will benefit both our academic growth and future careers in business analytics.

# References

Akgun, A. İ., Türkoğlu, S. P., & Erikli, S. (2023). Investigating the determinants of happiness index in EU-27 countries: a quantile regression approach. *International Journal of Sociology and Social Policy*, *43*(1–2), 156–177. https://doi.org/10.1108/IJSSP-01-2022-0005

Anne, F. I., Akter, S. M., Sheikh, S. P., Ireen, S., Escobar-DeMarco, J., Kappos, K., Ash, D., & Rasheed, S. (2022). Quality of nutrition services in primary health care facilities of Dhaka city: State of nutrition mainstreaming in urban Bangladesh. *PLoS ONE*, *17*(12 December). https://doi.org/10.1371/journal.pone.0278621

Appleby, G., Levy, R., & Whalan, H. (2023). *Voice Versus Rights: The First Nations Voice and The Australian Constitutional Legitimacy Crisis*. 1–321. https://doi.org/10.5040/9781509928958

Araki, S. (2022). Does education make people happy? Spotlighting the overlooked societal condition. Journal of Happiness Studies, 23(2), 587-629.

Brown, J., Acey, C. S., Anthonj, C., Barrington, D. J., Beal, C. D., Capone, D., Cumming, O., Pullen Fedinick, K., MacDonald Gibson, J., Hicks, B., Kozubik, M., Lakatosova, N., Linden, K. G., Love, N. G., Mattos, K. J., Murphy, H. M., & Winkler, I. T. (2023). The effects of racism, social exclusion, and discrimination on achieving universal safe water and sanitation in high-income countries. In *The Lancet Global Health* (Vol. 11, Issue 4, pp. e606–e614). Elsevier Ltd. <https://doi.org/10.1016/S2214-109X(23)00006-2>

Burlea-Schiopoiu, A., & Popovici, N. (2024). Social Inclusion: A Factor That Influences the Sustainable Entrepreneurial Behavior of Generation Z. *Administrative Sciences*, *14*(3). https://doi.org/10.3390/admsci14030059

Cajilog, A. M. L., & Pabalan, A. P. (2023). Basic Education Through the Lens of Internationalization: The Case of Bohol Province. *International Journal of Multidisciplinary: Applied Business and Education Research*, *4*(6), 2131–2165. https://doi.org/10.11594/ijmaber.04.06.35

Carter, I. (2004). Choice, freedom, and freedom of choice. In *Social Choice and Welfare* (Vol. 22, Issue 1, pp. 61–81). https://doi.org/10.1007/s00355-003-0277-z

Cavalli, F., Moretto, E., & Naimzada, A. (2024). Green transition and environmental quality: an evolutionary approach. *Annals of Operations Research*, *337*(3), 1009–1035. https://doi.org/10.1007/s10479-023-05449-5

Curto-Millet, D., & Cañibano, A. (2023). The Design of Social Inclusion Interventions: A Paradox Approach. *Journal of the Association for Information Systems*, *24*(5), 1271–1291. https://doi.org/10.17705/1jais.00795

Eijkelenboom, A., & Bluyssen, P. M. (2020). Profiling outpatient staff based on their self-reported comfort and preferences of indoor environmental quality and social comfort in six hospitals. *Building and Environment*, *184*. <https://doi.org/10.1016/j.buildenv.2020.107220>

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. Journal of Big Data, 8(1). <https://doi.org/10.1186/s40537-021-00516-9>

Gopalakrishnan, H. (2024). *Insurance as a Tool for Financial and Social Inclusion*.

Guillemot, F., Lacroix, F., & Nocus, I. (2024). Subjective well-being and social inclusion at school for students with a disability, according to their parents, in France. *Research in Developmental Disabilities*, *153*. https://doi.org/10.1016/j.ridd.2024.104814

Gyasi, R. M., Hajek, A., Asante, F., Accam, B. T., Osei-Tutu, S., Rahmati, M., Hervie, V. M., Abass, K., & Phillips, D. R. (2023). Ageing Happily in Ghana: How Does Social Inclusion Contribute? *Psychogeriatrics*, *23*(5), 821–830. <https://doi.org/10.1111/psyg.13004>

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2022). World Happiness Report 2022. New York: Sustainable Development Solutions Network.

Ibda, H., Gandi Wijanarko, A., Hilmi, M. N., Mabruroh, S. S., Anzakhi, A., & Fadhilah, T. D. (2024). Inclusive Education based on Gender Equality, Disability, and Social Inclusion (GEDSI) in Elementary School. *Pegem Journal of Education and Instruction*, *14*(3), 276–286. https://doi.org/10.47750/pegegog.14.03.26

Issahaku, P. A., & Adam, A. (2022). Young People in Newfoundland and Labrador: Community Connectedness and Opportunities for Social Inclusion. *SAGE Open*, *12*(3). https://doi.org/10.1177/21582440221113845

Mahlert, B. (2023). Discussing basic human needs: Insights into collective reflexivity in Global Social Policy. *Global Social Policy*, *23*(3), 499–515. https://doi.org/10.1177/14680181231195292

Mares, G., Cojocariu, V.-M., Cîrtiță-Buzoianu, C., Amălăncei, B.-M., & Mâță, L. (2024). Social Inclusion Attitude, an Insight among Teachers from Disadvantaged Areas. *Revista Romaneasca Pentru Educatie Multidimensionala*, *16*(1), 429–448. https://doi.org/10.18662/rrem/16.1/830

Metreau, E., Young, K. E., & Eapen, S. G. (2024, July 1). *World Bank country classifications by income level for 2024-2025*. World Bank Blogs. <https://blogs.worldbank.org/en/opendata/world-bank-country-classifications-by-income-level-for-2024-2025>

Meyer, M. S., Shen, Y., & Plucker, J. A. (2024). Reducing Excellence Gaps: A Systematic Review of Research on Equity in Advanced Education. *Review of Educational Research*, *94*(1), 33–72. https://doi.org/10.3102/00346543221148461

‌ Möwisch, D., Brose, A., & Schmiedek, F. (2021). Do higher educated people feel better in everyday life? Insights from a day reconstruction method study. Social Indicators Research, 153, 227-250.

Ngozi Elems-Ikwegbu, H. (2023). Evaluating Basic Education in Singapore: Trends and Dynamics. In *Syntax Admiration* (Vol. 4, Issue 3).

Omerov, P., Craftman, Å. G., Mattsson, E., & Klarare, A. (2020). Homeless persons’ experiences of health- and social care: A systematic integrative review. In *Health and Social Care in the Community* (Vol. 28, Issue 1, pp. 1–11). Blackwell Publishing Ltd. <https://doi.org/10.1111/hsc.12857>

Pemau, R. C., González-Palacios, P., & Kerr, K. W. (2024). How quality of life is measured in studies of nutritional intervention: a systematic review. Health and Quality of Life Outcomes, 22(1), 9.

Rasool, K. N., & Naidoo, P. (2024). The sustainability of information and communication technology (ICT) in previously disadvantaged public schools in Gauteng, South Africa. *South African Journal of Education*, *44*(1). https://doi.org/10.15700/saje.v44n1a2302

Senasu, K., Sakworawich, A., & Russ-Eft, D. F. (2019). Developing Thai Happiness Index. *Social Indicators Research*, *146*(3), 425–448. https://doi.org/10.1007/s11205-019-02136-6

Sihombing, P. R. (2022). Macro Socio-Economic Factors that Affect the Happiness Index in Indonesia Article Info. *Journal of Social Sciences and Humanities*, *12*(2), 2022.

Silva, J., De Keulenaer, F., & Johnstone, N. (2012). Environmental quality and life satisfaction: Evidence based on micro-data.

Sjåstad, H., Zhang, M., Masvie, A. E., & Baumeister, R. (2021). Social exclusion reduces happiness by creating expectations of future rejection. *Self and Identity*, *20*(1), 116–125. <https://doi.org/10.1080/15298868.2020.1779119>

Steckermeier, L. C. (2021). The value of autonomy for the good life. An empirical investigation of autonomy and life satisfaction in Europe. Social Indicators Research, 154(2), 693-723.

Stern, S., Harmacek, J., Krylova, P. & Htitich, M.: 2024 Social Progress Index Methodology Summary. Social Progress Imperative. Washington, DC. Available at: [www.socialprogress.org/global/methodology](http://www.socialprogress.org/global/methodology)

Sturge, J., Miedema, E., Elf, M., & Nordin, S. (2023). Socially sustainable housing and built environments to support the health and social inclusion of older adults: Protocol for a scoping review and stakeholder consultation. In *BMJ Open* (Vol. 13, Issue 12). BMJ Publishing Group. <https://doi.org/10.1136/bmjopen-2023-075447>

The world by income and region. WDI - The World by Income and Region. (n.d.). <https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html>

United Nations Development Programme. (2024, June 4). Human development index. Human Development Reports. <https://hdr.undp.org/data-center/human-development-index#/indicies/HDI>

Wilson Fadiji, A., & Lomas, T. (2024). Understanding the Association Between Education and Wellbeing: An Exploration of the Gallup World Poll. Applied Research in Quality of Life, 1-35.

Wray, C. M., Khare, M., & Keyhani, S. (2021). Access to care, cost of care, and satisfaction with care among adults with private and public health insurance in the US. JAMA Network Open, 4(6), e2110275-e2110275.

Yamb, B., & Ekomane, J. L. (2019). *Human Development: “Can money buy Happiness?”*

# Appendix

**Literature Review**

|  |  |  |  |
| --- | --- | --- | --- |
| Title of the Articles | Objective | Methodology | Findings |
| Developing Thai Happiness Index  (Senasu et al., 2019) | To develop a national happiness index to assess the happiness level of Thai people and explore related issues. | Data were collected across 13 Thai provinces using stratified multi-stage and simple random sampling, and the Alkire-Foster method was used to create the Thai Happiness Index ranging from 0 to 1. | The 2017 Thai Happiness Index (THaI0) was 0.506, with 14.6% of the population happy and 85.5% not yet happy. The index supports national well-being policies and can be adapted for other countries. |
| Macro Socio-Economic Factors that Affect the Happiness Index in Indonesia  (Sihombing, 2022) | To identify the macro socio-economic factors that influence the Happiness Index in Indonesia. | Panel data regression analysis was employed, with the fixed effects model being the best fit. The model was transformed into the white cross-section GLS model to address heteroscedasticity and autocorrelation issues. | Gini ratio, poverty, and unemployment negatively affect happiness, while HDI, per capita expenditure, and economic growth positively impact happiness. |
| Investigating the determinants of happiness index in EU-27 countries: a quantile  regression approach  (Akgun et al., 2023) | To examine the determinants of happiness in EU-27 countries, focusing on macroeconomic factors and labour market conditions. | The study uses panel data from the Eurostat, World Bank, and World Happiness Reports, employing ordinary least squares (OLS) and quantile regression models to analyse data from 27 EU countries across eight time points. | Taxes and inflation positively affect happiness, while employment has a negative impact. Financial inclusion’s effect varies by income, and GDP boosts happiness mainly in wealthier countries. |
| Social Inclusion Attitude, an Insight among Teachers from Disadvantaged Areas  (Mares et al., 2024) | To explore the perspectives of teachers in disadvantaged rural high schools and identify models and future strategies for support. | The research utilized a 23-item questionnaire administered online to teachers in the North-East region of Romania, with responses analysed using the SPSS software. | The findings show that most teachers are interested in social inclusion but lack practical experience and find working with families challenging. |
| Social exclusion reduces happiness by creating expectations of future rejection  (Sjåstad et al., 2021) | To understand how social exclusion lowers happiness by making people expect future rejection, and to see how a sense of belonging can reduce this effect. | Participants were assigned to social exclusion or inclusion scenarios via a guided imagination task and measured on exclusion, future rejection expectations, and happiness using self-reported scales. | Social exclusion has a long-term negative impact on happiness, largely due to the anticipation of further rejection which worsens mental health. |

**Predictive analytics using R and R studio**

**Unobserved Heterogeneity**

A graph with lines and dots

Description automatically generatedThe chart above shows the heterogeneity across countries that present the variation in Life Ladder scores across different countries. Each dot represents the Life Ladder score for a country, and the vertical lines indicate the heterogeneity within each country’s scores. The large number of countries, compressed on the x-axis, suggests substantial diversity in well-being across countries, as seen from the broad range of Life Ladder scores and the varied lengths of the error bars, indicating differences in consistency within each country’s scores.

A graph with numbers and a line

Description automatically generatedFigure above shows the heterogeneity across time. It visualizes the variation in the Life Ladder from 2011 to 2022. Each point on the line represents the average Life Ladder score for a given year, while the vertical bars likely representing error bars or standard deviations depict the heterogeneity or variability in the Life Ladder scores within that year. The trend shows a slight decline from 2011 to 2014, followed by a relatively steady period until 2017, when an upward trend appears, peaking in 2019. After 2019, there is a modest decrease. The relatively wide error bars suggest significant variability in the Life Ladder scores across the dataset.

**Handle Missing Values for Predictive Analytics**

Several variables in the dataset contain missing values, which include key factors such as nutrition and basic medical care, water and sanitation, shelter, access to basic knowledge, access to information and communications, environmental quality, personal rights, and inclusiveness, all of which have 15 missing entries. Additionally, personal safety and health and wellness have 22 missing values, while personal freedom and choice, as well as access to advanced education, are missing data in 27 instances. To address this issue, we employed mean imputation to fill in the missing values before proceeding with the predictive modeling phase. This approach ensures that the analysis remains robust by minimizing data loss and maintaining consistency across the dataset, allowing for a more accurate and reliable prediction (Emmanuel et al., 2021).

A screenshot of a computer

Description automatically generated

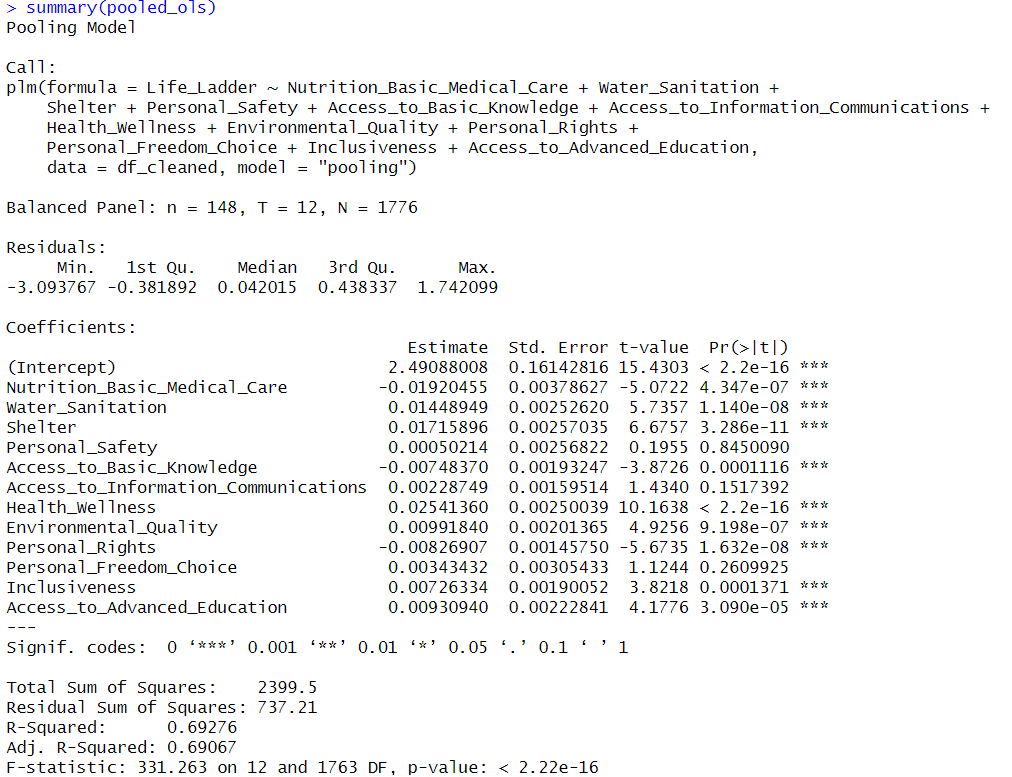
A screenshot of a computer

Description automatically generated

Hence, we handle the missing value before model selection and model training.

**Model Selection Process**

**Pooled Ordinary Least Square (OLS)**

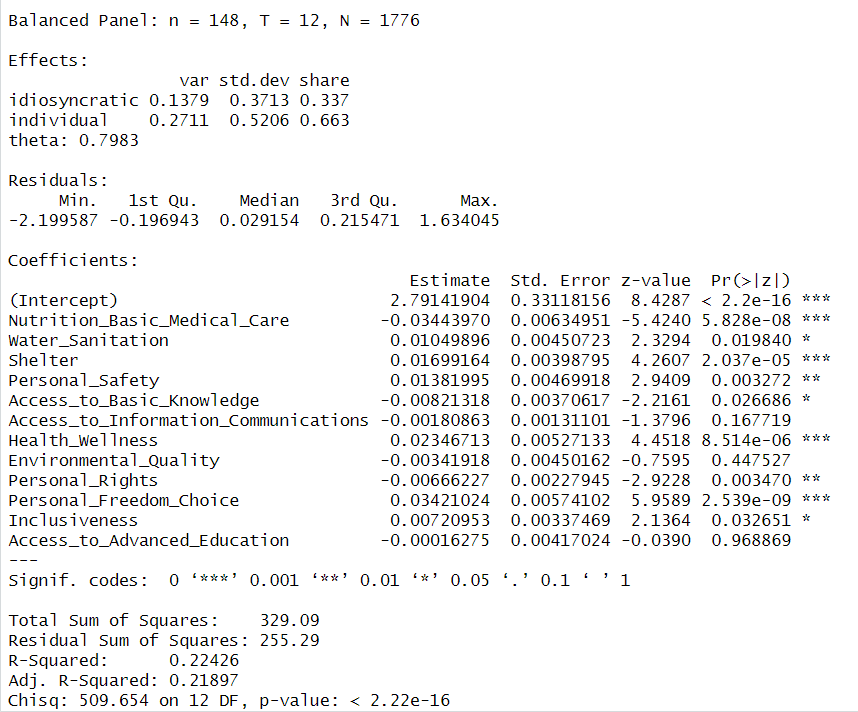
The pooled OLS assumes that there is no heterogeneity between countries so it will treat them as one large group. The results from the pooled OLS model indicate a strong relationship between the independent variables and the Life Ladder score, as evidenced by a high R-squared value of 0.693. This suggests that approximately 69.3% of the variance in the Life Ladder can be explained by the models

In the pooled OLS model analyzing the factors affecting the Life Ladder, several independent variables show significant effects. Nutrition & Basic Medical Care has a negative and significant impact on life satisfaction, indicating that higher provision of these services may coincide with lower subjective well-being, possibly due to issues like unequal access or poor quality. In contrast, Water & Sanitation, Shelter, and Health & Wellness exhibit positive and highly significant effects, suggesting that improvements in these areas are associated with higher life satisfaction. Environmental Quality, Inclusiveness, and Access to Advanced Education also positively influence the Life Ladder, highlighting the importance of better environmental conditions, social inclusiveness, and education in enhancing well-being. Conversely, Access to Basic Knowledge and Personal Rights have negative and significant coefficients, indicating complexities or inequalities in their provision that adversely affect life satisfaction. Access to Information & Communications shows a marginally significant small positive effect. Notably, Personal Safety and Personal Freedom & Choice are not statistically significant, suggesting their effects are minimal or inconsistent in this model.

Overall, while the pooled OLS model demonstrates significant predictors of life satisfaction, it does not account for individual-specific effects that may exist, which is a limitation compared to the random effect model and fixed effect model. The pooled model treats the data as a single cross-section, potentially overlooking important variations among countries over time.

**Random effect model**

The random effect model and fixed effect model can handle the entity variable for panel data.

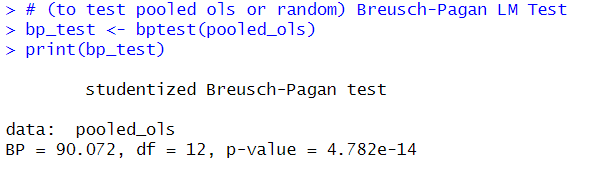


The random effects model shows significant relationships between several independent variables and the dependent variable, Life Ladder, in a balanced panel of 148 countries over 12 years. In the random effects model analyzing the factors influencing Life Ladder, several variables remain significant, though the effects vary compared to the pooled OLS results.

The analysis indicates that Nutrition & Basic Medical Care continues to exert a significant negative impact on life satisfaction, while Water & Sanitation, Shelter, and Health & Wellness exhibit positive and significant effects, confirming that improvements in these areas enhance well-being. Notably, unlike the pooled OLS model, Personal Safety is significant in this model, suggesting that feelings of security play an important role in life satisfaction. Personal Rights have a significant negative effect, implying that complexities or inequalities in the protection of rights may decrease life satisfaction. Personal Freedom & Choice is highly significant and positively correlated with well-being, indicating that greater freedom and autonomy are crucial for enhancing life satisfaction. Inclusiveness also remains significant, positively influencing well-being. Access to Basic Knowledge is marginally significant with a small negative effect, whereas Access to Information & Communications, Environmental Quality, and Access to Advanced Education do not show significant impacts, suggesting their effects on life satisfaction may be inconsistent or context-dependent.

The model has an R-squared of 0.224, indicating that approximately 22.4% of the variation in the Life Ladder is explained by the model. The theta value of 0.798 suggests that the individual effects of country differences are relatively important in explaining the Life Ladder.

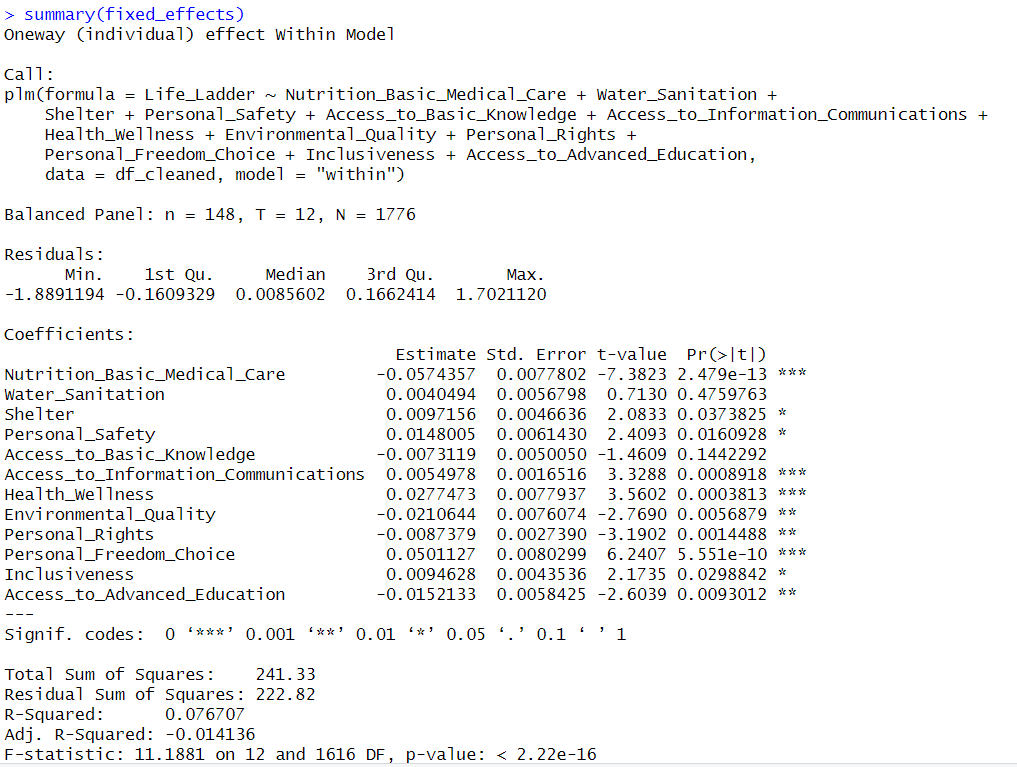
**Breush-Pagan Lagrangian Multiplier Test**



The Breusch-Pagan Lagrangian Multiplier test indicates a significant p-value (p < 0.05) for the pooled OLS model, with a BP statistic of 90.072 and 12 degrees of freedom. This result suggests that the random effects model is preferred over the pooled OLS model due to the presence of individual-specific effects.

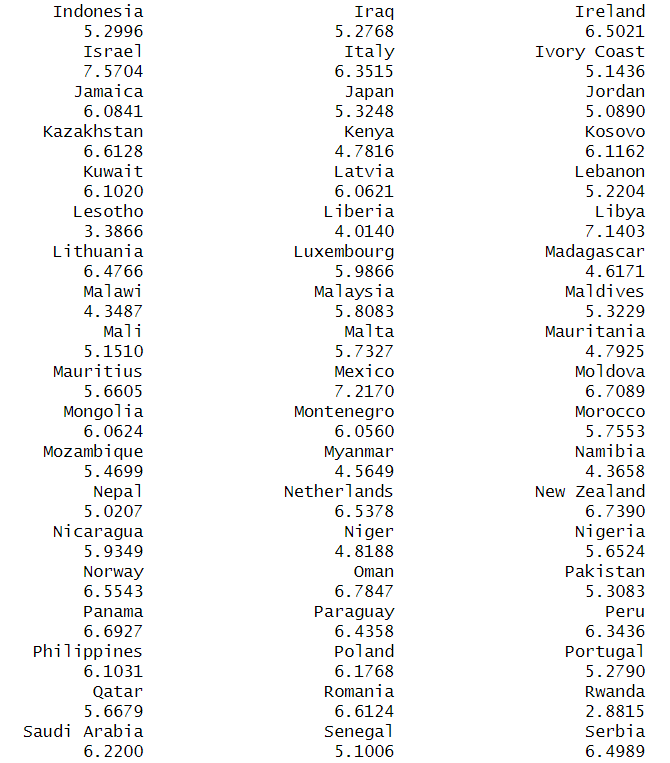
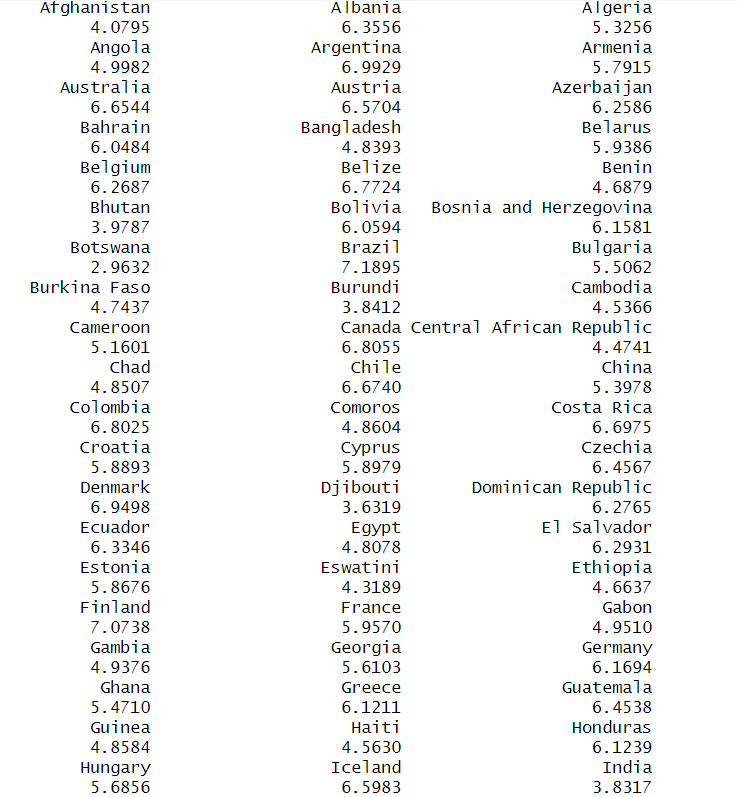
However, it remains uncertain whether these effects are fixed among the countries or merely random. Therefore, it is essential to conduct a regression analysis using the fixed effects model and subsequently compare the random effects model with the fixed effects model using the Hausman test to determine which model is more appropriate for this analysis.

**Fixed effect model**

The fixed effects model analyzing the impact of various independent variables on the Life Ladder reveals several significant relationships at the 5% significance level. The results indicate that while many variables continue to exhibit significant relationships, the effects differ compared to the pooled OLS and random effects models.

Nutrition & Basic Medical Care maintains a significant negative impact on life satisfaction, underscoring its critical role in determining well-being. Water & Sanitation loses significance, while Shelter shows marginal significance. Notably, Personal Safety emerges as significant, demonstrating a positive effect on life satisfaction and reinforcing the importance of security. Access to Basic Knowledge is found to be insignificant, indicating no discernible effect on the Life Ladder in this model. Access to Information & Communications has a significant positive effect, suggesting that improved access to communication tools enhances well-being. Health & Wellness remains a strong positive predictor of life satisfaction. Interestingly, Environmental Quality appears marginally significant with a negative impact, which contrasts with findings from previous models and suggests potential concerns regarding environmental degradation. Personal Rights again exhibit a significant negative relationship, while Personal Freedom & Choice has a highly significant positive effect, reinforcing the notion that greater autonomy significantly boosts life satisfaction. Inclusiveness is also positively significant, highlighting its importance in promoting well-being. Notably, Access to Advanced Education shows a negative and significant relationship, possibly indicating that the benefits of education may be countered by certain complexities or expectations.

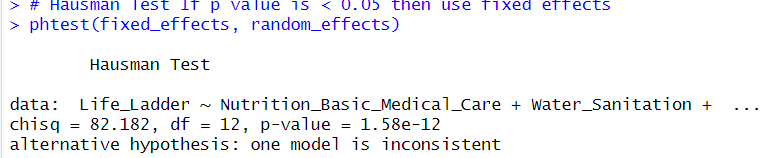
The model’s R-squared value of approximately 0.077 indicates that about 7.7% of the variation in Life Ladder scores can be explained by the independent variables included in the model, while the adjusted R-squared of -0.014 suggests that the model may not adequately capture the underlying data complexity. The overall model is statistically significant, as indicated by the F-statistic (11.188) with a p-value of less than 2.22e-16. Despite the relatively low R-squared, the significant coefficients suggest that specific factors are important for understanding variations in life satisfaction levels across individuals in this dataset.



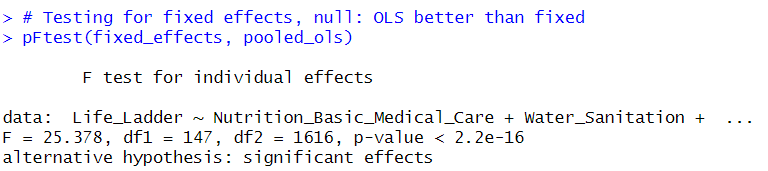
The fixed effects model results provide individual-specific intercepts for each country, reflecting their unique average Life Ladder scores when controlling for the included independent variables. For instance, Afghanistan has a fixed effect estimate of 4.0795, suggesting that, on average, its Life Ladder score is lower than the overall average when accounting for other factors. In contrast, Australia has a higher fixed effect estimate of 6.6544, indicating a relatively higher Life Ladder score.

This variability highlights the differences in life satisfaction across countries, with nations like Luxembourg (6.6299) and Canada (6.8055) also showing elevated scores, while countries such as Somalia (4.7705) and Rwanda (2.8815) demonstrate lower scores, pointing to significant disparities in perceived quality of life. These country-specific intercepts illustrate how contextual factors influence the Life Ladder, extending beyond the effects of the independent variables. This analysis reinforces the importance of considering fixed effects when studying cross-country variations in life satisfaction, as it accounts for unobserved characteristics that may impact the outcomes.

**Hausman Test**

The Hausman test was conducted to compare the fixed effects and random effects models in our panel data analysis of the Life Ladder as the dependent variable. The test produced a chi-squared statistic of 82.182 with 12 degrees of freedom and a p-value of 1.58e-12. Given this extremely low p-value, which is significantly below the conventional alpha level of 0.05, we reject the null hypothesis that the random effects estimator is consistent and efficient. This indicates that the random effects model may not be appropriate for our data, suggesting a correlation between the individual effects and the regressors. Therefore, we conclude that the fixed effects model should be utilized for our analysis to ensure consistent estimates.

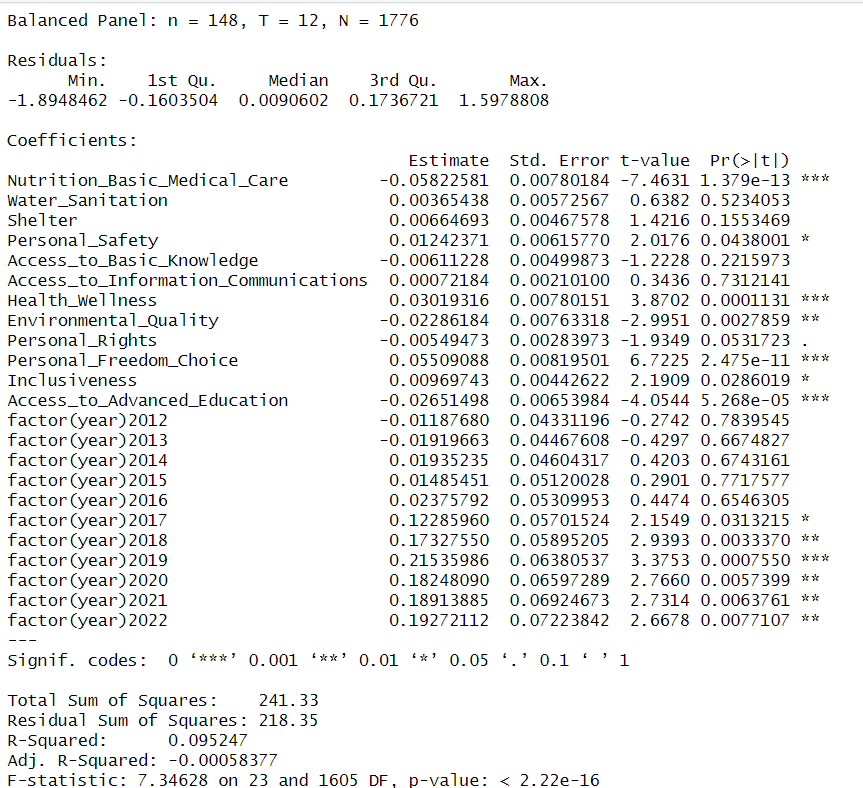
**PFtest for individual effects**

The F-test compares the fixed effects model to a pooled OLS model. The results of the F-test for individual effects indicate a strong rejection of the null hypothesis that there are no individual-specific effects in this panel data. The F-statistic is 25.378, with degrees of freedom df1 = 147 and df2 = 1616, yielding a p-value < 2.2e-16. This extremely low p-value suggests that there are significant individual effects present in the model, meaning that the fixed effects model is more appropriate than the pooled OLS model for analyzing the data. In other words, the differences across countries over time significantly contribute to the variation in the Life Ladder scores. Hence, we can confidently apply the fixed effects model for the analysis.

**Focusing on fixed effects model**

Since we run a lot of tests to figure out that fixed effects model is suitable to analyse our panel data, we will discuss the result of fixed effect model in this section. The whole process of this model selection can be found in the appendix.

**Time-fixed effect**

In the time-fixed effects model, several interesting patterns emerge when controlling for time-specific shocks that might affect the Life Ladder. The variable Nutrition & Basic Medical Care retains a significant negative effect on life satisfaction, with an even larger negative coefficient than in previous models. This indicates that improvements in nutrition and basic medical care are associated with lower life satisfaction, which continues to challenge conventional expectations.

Health & Wellness and Personal Freedom & Choice maintain their positive and significant effects on life satisfaction, reinforcing their critical roles in enhancing well-being. Personal Safety also continues to show a significant positive effect, suggesting that a sense of safety enhances life satisfaction.

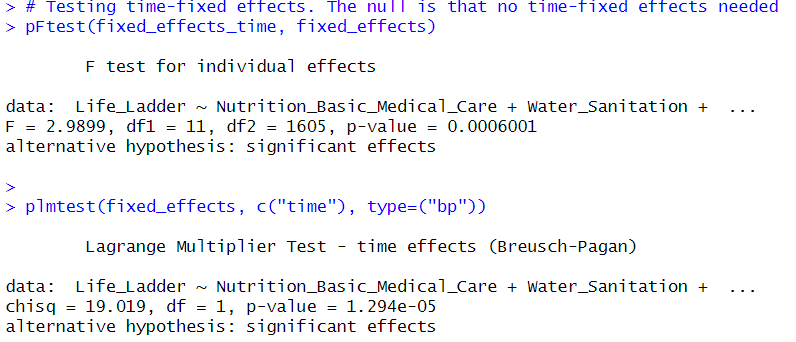
Environmental Quality has now become statistically significant (p = 0.0216), demonstrating a negative effect on life satisfaction. This counterintuitive finding implies that as environmental quality improves, life satisfaction decreases, possibly reflecting trade-offs or challenges individuals face due to increased environmental regulations or changes.

The variable Personal Rights shows a significant negative effect, indicating that improvements in personal rights are linked to lower life satisfaction in this model—another surprising outcome. Similarly, Access to Advanced Education becomes significant with a negative effect, suggesting that higher levels of education are associated with reduced life satisfaction. This unexpected result may reflect educational stress, unmet expectations, or competition in environments where advanced education is more accessible.

Inclusiveness demonstrates a weak positive relationship with the Life Ladder, suggesting that higher levels of inclusiveness correlate with greater life satisfaction. This implies that societies that are more inclusive and equitable may foster overall well-being, though the effect is not strongly significant in this model. Other variables, including Water & Sanitation, Shelter, Access to Basic Knowledge, and Access to Information & Communications, are not significant in this analysis.

Regarding the time variables, Significant positive effects are noted for several years (2018-2022), indicating overall improvements in life satisfaction during these years compared to the baseline year. This aligns with broader social or economic developments during these years, possibly reflecting recovery or progress from global disruptions caused by the pandemic. The inclusion of time effects accounts for temporal shocks and emphasizes how societal and global changes over time interact with individual life satisfaction.

**F-test and Breush-Pagan test for Time-Fixed effect**

The F test for individual effects yielded an F-statistic of 2.9899 with a p-value of 0.0006001, leading to the rejection of the null hypothesis that there are no time-fixed effects, thus suggesting that life satisfaction varies significantly over time. Similarly, the Lagrange Multiplier Test produced a chi-square statistic of 19.019 and a p-value of 1.294e-05, further supporting the inclusion of time-fixed effects to control for time-specific factors that impact the Life Ladder scores.

Therefore, these results suggest that factors changing over time such as global events or time-specific shocks have a significant influence on the Life Ladder score and should be included in this model.

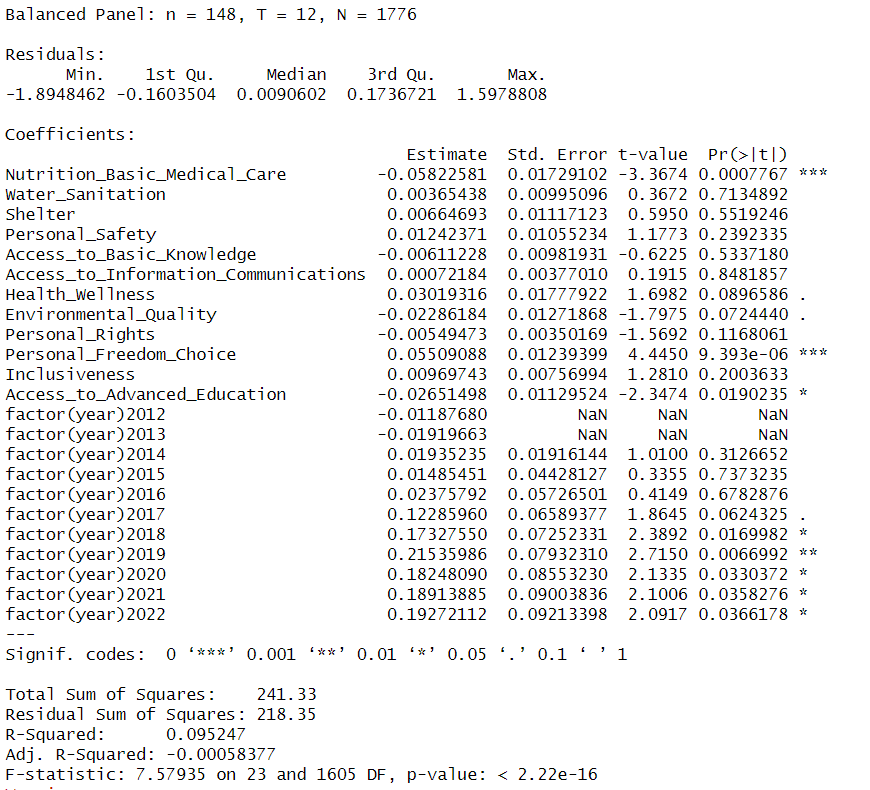
**Testing for cross-sectional dependence/contemporaneous correlation: using Breusch-Pagan LM test of independence and Pesaran CD test**

A close-up of a computer screen

Description automatically generatedThe results from the tests for cross-sectional dependence reveal significant insights regarding the relationships within the panel data. The Breusch-Pagan LM test produced a chi-square statistic of 21,319 with a p-value less than 2.2e-16, strongly indicating the presence of cross-sectional dependence among the units in the dataset; thus, the null hypothesis of independence is rejected. This suggests that the observations in different cross-sections may be correlated, which could impact the validity of standard errors in the model. Conversely, the Pesaran CD test yielded a z-score of -1.7831 and a p-value of 0.07457, suggesting marginal evidence of cross-sectional dependence at a 10% significance level but not conclusively rejecting the null hypothesis. Overall, the findings from these tests underscore the necessity of considering potential correlations between cross-sectional units in the analysis of factors affecting life satisfaction over time.

**Driscoll and Kraay standard errors**

Since it showed that there is presence of cross-sectional dependence, Driscoll and Kraay standard errors that allows for both heteroskedasticity and autocorrelation were employed to accommodate cross-sectional dependence.



The results from the Driscoll and Kraay adjusted time-fixed effects model reveal several significant relationships between socio-economic variables and life satisfaction (Life Ladder). Notably, Nutrition & Basic Medical Care has a significant negative effect on Life Ladder, suggesting that increased access to basic medical care and nutrition may paradoxically be associated with lower life satisfaction, which warrants further investigation. Conversely, Personal Freedom & Choice shows a significant positive effect, indicating that individuals’ ability to make autonomous life decisions is crucial for enhancing life satisfaction. Additionally, the negative effect of Access to Advanced Education on life satisfaction is significant, implying that while education is vital for long-term development, it may not directly contribute to immediate life satisfaction, potentially due to the stresses associated with educational attainment. In terms of yearly trends, the year 2013 had a significant negative effect compared to the reference year, reflecting a possible downturn in life satisfaction during that period. Other factors, such as Water Sanitation, Shelter, and Personal Safety, did not show statistically significant effects, suggesting that while they are important for overall well-being, their variation within the dataset does not directly correlate with life satisfaction in a meaningful way. Overall, although the model explains only a modest portion of the variance in life satisfaction (R-squared = 9.52%), it highlights the complex and sometimes counterintuitive relationships between socio-economic factors and well-being, emphasizing the need for nuanced approaches in understanding these dynamics.