*Please delete all the cursive text before submission. It is here just for your reference*.

*Further: data set – DS, research question – RQ*

*The mark (****x words****) after each subchapter states the word count limit. This indicates the expected amount of information which you can exceed by 10% without losing the mark.*

7COM1079-0901-2024 - Team Research and Development Project

Final report title: (*the topic of your research.)*

Group ID: A117

Dataset number: DS008

Prepared by: Sravani Kumari Palem - 22067322

Chaithanya Krishna Nuthalapati - 23067825

Sai Ganesh Aavula - 24000424

Gopala Krishna Reddy Avuluri - 23036787

Gopala Krishna Pothumarthi - 23080478

*[Name and ID of submitting student first],*

*[Name and ID of other group members]*

***Please make sure*** *the document spelled correctly (including image labels, section headings, and table of contents). Please use correct punctuation.*

*Make sure your report is grammatically correct.*

University of Hertfordshire

Hatfield, 2024

Table of Contents *add page numbers here*

1. Introduction
   1. Problem statement and research motivation
   2. The data set
   3. Research question
   4. Null hypothesis and alternative hypothesis (H0/H1)

1. Background research
   1. Research papers (at least 3 relevant to your topic / DS)
   2. Why RQ is of interest (research gap and future directions according to the literature)

1. Visualisation
   1. Appropriate plot for the RQ *output of an R script (NOT a screenshot)*
   2. Additional information relating to understanding the data (optional)
   3. Useful information for the data understanding

1. Analysis
   1. Statistical test used to test the hypotheses and output
   2. The null hypothesis is rejected /not rejected based on the p-value

1. Evaluation – group’s experience at 7COM1079
   1. What went well
   2. Points for improvement
   3. Group’s time management
   4. Project’s overall judgement
   5. Comment on GitHub log output

1. Conclusions
   1. Results explained.
   2. Interpretation of the results
   3. Reasons and/or implications for future work, limitations of your stud

1. Reference list

Harvard (author, date) format.

1. Appendices
2. R code used for analysis and visualisation.
3. GitHub log output.

*The list below outlines the chapter/subchapter numbers, names, word count limits, and explanations of what to write in each section.*

1. Introduction

* 1. **Problem statement and research motivation** **(100 words)**

Happiness is a mix of many factors, including social, economic, and cultural influences. Understanding what makes people happy is important for governments and policymakers, as it will helps them create better programs to improve people’s lives. One big question is how much a country’s economy affects happiness. This study looks at GDP per capita, a common way to measure a country’s economy, and how it connects to happiness scores. Research (Helliwell et al., 2017) shows that economy does play a crucial role in happiness, but it’s not clear how strong this connection is. That’s why we’re exploring this topic further.

* *What is the problem in the area we want to learn more about (motivation for study).*
* *Use at least one citation from related literature for top marks.*
  1. **The data set** **(75 words)**

The dataset is from the 2017 World Happiness Report, which ranks 155 countries by their happiness levels. Happiness scores are based on a survey where people rate their lives. It includes data on factors like economy GDP per capita, social support, life expectancy, family and freedom. These factors help explain why some countries are happier than other countries and provide a good starting point for studying the connection between a country’s economy and happiness score.

* 1. **Research question (50 words).**

Is there a correlation between economy GDP per capita and happiness scores across countries?

* Independent variable (Interval/Measurement data): Economy GDP per capita
* Dependent variable(Interval/Measurement data): Happiness Score

To find out, we use visualization plots and apply Pearson’s correlation test. This will help us to measure the strength and direction of the relationship between GDP per capita and happiness scores.

* 1. **Null hypothesis and alternative hypothesis (H0/H1)** **(100 words)**

**Null Hypothesis (H0):** There is no significant correlation between GDP per capita and happiness scores across countries. Any observed relationship is purely due to chance.

**Alternative Hypothesis (H1):** There is a significant correlation between GDP per capita and happiness scores across countries. This suggests that changes in GDP per capita are associated with variations in happiness levels.

Pearson’s correlation test will be used to evaluate these hypotheses. If the p-value is greater than 0.05, we fail to reject the null hypothesis, indicating no significant relationship between the variables. Otherwise, we reject the null hypothesis, supporting a correlation.

1. **Background research**
   1. **Research papers (at least 3 relevant to your topic / DS) (200 words)**
   2. **Why RQ is of interest** **(research gap and future directions according to the literature)** **(100 word**s)

This research looks at the connection between GDP per capita and happiness scores, a topic of growing interest in economics and social studies. Studies like Stevenson and Wolfers (2013) find a positive link, but others suggest that happiness stops increasing as incomes rise and depends more on trust and good governance. This shows that happiness is complex and needs to be studied from different perspectives like family, health etc. Using data from the 2017 World Happiness Report, this study re-examines the relationship. Better understanding these factors can help policymakers create balanced strategies to improve well-being for people across the world.

1. **Visualisation**
   1. **Appropriate plot for the RQ** *output of an R script (NOT a screenshot)* (**50 words)**

* *Explain the choice of the plot.*
* *Anything on the plot from R is not counted towards word count limit*
* *Make sure that the plot is from output of an R script (NOT a screenshot)*
* *Make sure that the plot has a caption or title, X and Y-axis labels, legend if appropriate and units.*
* *Make sure the title or caption and axis labels are informative.*

The scatter plot is ideal for showing the relationship between GDP per capita (X-axis) and happiness scores (Y-axis). Each data point represents a country, while the linear trend line (red) highlights the positive correlation. The plot includes an informative title, axis labels, and units, clearly visualizing the strength and direction of the relationship.

* 1. **Additional information relating to understanding the data (optional) (50 words)**

**Scatter Plot**: The scatter plot shows the relationship between GDP per Capita and Happiness Score. Each point represents a country, with the blue dots highlighting the observed data. The red trendline indicates a positive correlation, suggesting that countries with higher GDP per Capita generally have higher happiness scores.

**Histogram:** The histogram depicts the distribution of Happiness Scores across all countries. It reveals a normal distribution, with most countries scoring between 4 and 6. This suggests that while happiness varies, it is generally centred around moderate levels. The overlaying bell curve confirms the suitability of the data for correlation analysis.

* 1. Useful information for the data understanding (**50 words)**
* *Summarise key observations from the plot.*

1. **Analysis**
   1. **Statistical test used to test the hypotheses and output** (**75 words)**

The Pearson's product-moment correlation test was selected to evaluate the relationship between GDP per capita and happiness scores because both variables are continuous and follow a linear pattern, as visualized in the scatter plot. This test measures the strength and direction of the linear association between the two variables. The test produced a correlation coefficient of 0.8125, indicating a strong positive relationship, with a highly significant p-value (< 2.2e-16), confirming the appropriateness of the test for the research question.

* 1. **The null hypothesis is rejected /not rejected based on the p-value** (**100 words)**

Based on the p-value (< 2.2e-16), which is far below the significance level of 0.05, we reject the null hypothesis. This means there is sufficient evidence to support the alternative hypothesis that GDP per capita and happiness scores are significantly correlated. The Pearson correlation coefficient of 0.8125 indicates a strong positive relationship between the two variables. Higher GDP per capita is generally associated with higher happiness scores, suggesting that economic performance is an important factor in national well-being. However, this result does not imply causation and should be interpreted alongside other factors influencing happiness

1. **Evaluation – group’s experience at 7COM1079**
   1. **What went well (75 words)**

Our group successfully analysed the dataset, ran statistical tests, and created visualization results to answer the research question. We worked well together, collaboration between the group members was smooth and efficient, dividing tasks effectively. Using GitHub, we maintained version control, which ensured seamless integration of our contributions. Using R for data analysis and visualization went smoothly. Clear communication among the team members enabled us to solve problems, stay focused on research goals, and achieved the project results.

* 1. **Points for improvement (75 words)**

While our project was successful and reached project goals, we identified a few areas for improvement. We could have conducted a more in-depth exploration of additional factors influencing happiness, such as family, health or governance. However earlier planning for the data visualization and analyzation has saved the time. Communication among the group members was effective, though we could have been improved by scheduling more meetings. Finally, we ensuring equal contribution from all members would make the workload more balanced.

* 1. **Group’s time management (50 words)**

Our group managed time effectively by setting clear deadlines and milestones for each task. Early planning allowed for smooth progress and clear results. Although, there were moments of delay due to overlapping personal schedules, we have managed smooth and effective communication among the group members to meet the project goals.

* 1. **Project’s overall judgement (50 words)**

Overall, the project was a success, as we met our objectives and gained valuable insights into the relationship between GDP per capita and happiness. The results were well-supported by statistical evidence, and the group’s collaborative effort was evident in the quality of the analysis and the clarity of the presentation.

* 1. Comment on the GitHub log output **(50 words)**

*Please comment on the GitHub log output, and refer to it as being placed into**Appendix B.*

*From your Git log, select the three most significant commits during this project and include the following for each:*

1. ***Commit Message:*** *[Insert Commit Message] Brief explanation of the broader impact of the change*
2. ***Commit Message:*** *[Insert Commit Message] Brief explanation of the broader impact of the change*
3. ***Commit Message:*** *[Insert Commit Message] Brief explanation of the broader impact of the change*

1. **Conclusions**
   1. **Results explained (75 words)**

The analysis showed a strong positive correlation (r = 0.8125, p < 0.001) between GDP per capita and happiness scores, as seen in the scatter plot, leading us to reject the null hypothesis. The histogram of happiness scores showed a normal distribution, with most countries having moderate happiness levels and fewer with very high or low scores. These results confirm that economic performance plays a significant role in happiness, but other factors like governance, social support, health, also matter.

* 1. **Interpretation of the results (75 words)**

1. Reference list ***(not included in the work count)***

Harvard (author, date) format.

1. Appendices
2. R code used for analysis and visualisation ***(not included in the word count)***

Analysis.R code with the appropriate statistics to test the hypotheses.

* ***No word count****, but ensure the code is without redundant lines, well-commented and produces the correct output.*
* *Make sure it runs (look in Rscript.log for output from a statistical test)*
* *It should compute appropriate statistics to test the hypotheses*

1. GitHub log output.