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

Final report title: June Unemployment Rate

Group ID: A156

Dataset number: D130

Prepared by: Roshandin Gajju

*Romin Dharmeshbhai Patel*

*Rohit Kumar Thummar*

*Muhammad Hashaam*

*Muhammad Zeeshan*

University of Hertfordshire

Hatfield, 2024

Table of Content No

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

1. Background research 4
   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 4
   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 5
   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 6
   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 7
   1. Results explained.
   2. Interpretation of the results
   3. Reasons and/or implications for future work, limitations of your stud

1. Reference list 7

Harvard (author, date) format.

1. Appendices 8
2. R code used for analysis and visualisation.
3. GitHub log output.

1. Introduction

* 1. Problem statement and research motivation

Unemployment poses significant economic and social challenges, contributing to increased poverty, deteriorating mental health, and strained social welfare systems (Blanchflower, 2009; Layard et al., 1991). Regions with persistent unemployment often experience higher crime rates and reduced community well-being (Raphael & Winter-Ebmer, 2001). This study aims to investigate the correlation between economic activity levels and unemployment rates, providing insights to guide targeted interventions. Understanding these dynamics is essential for reducing inequalities and fostering sustainable economic growth in underperforming regions.

* 1. The data set

The dataset utilized in this report contains detailed information relevant to the research question. Key features of the dataset include:

1. Region Name : The name of the county or region being analyzed.

2. Region Code : A numerical identifier for each region.

3. Unemployment Rate : The dependent variable, expressed as a percentage, indicating the unemployment rate in each region.

The dataset is well-structured and provides a robust foundation for statistical analysis. Data was collected and processed to ensure accuracy and relevance to the research objectives

* 1. Research question

Is there a significant difference in the unemployment rates between regions with high economic activity and those with low economic activity?

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

Null Hypothesis (H0) :

There is no significant difference in the unemployment rates between regions with high and low economic activity.

Alternative Hypothesis (H1) :

There is a significant difference in the unemployment rates between regions with high and low economic activity.

The research question and hypotheses aim to explore potential disparities in unemployment rates based on economic activity, providing insights into underlying patterns and contributing factors. This investigation will inform targeted interventions and policy decisions.

1. Background research
   1. Research papers

Unemployment and economic activity levels are interconnected and have been studied extensively in various contexts. Blanchflower (2009) emphasizes the broader societal impacts of unemployment, including its role in exacerbating mental health issues and increasing reliance on welfare systems. Layard et al. (1991) highlight how prolonged unemployment can create long-term economic stagnation and social issues, especially in underperforming regions. Additionally, Raphael and Winter-Ebmer (2001) investigate the correlation between unemployment and crime rates, revealing that regions with persistent joblessness often experience heightened societal challenges.

* 1. Why RQ is of interest

These studies underline the importance of exploring how economic activity influences unemployment rates to design targeted interventions that address regional disparities. The current research builds upon these foundations, aiming to identify actionable insights for fostering economic growth and reducing unemployment in low-activity regions.

1. Visualisation
   1. Appropriate plot for the RQ

**Box Plot of Unemployment Rates by Economic Activity Levels**

* **Purpose**: The box plot is chosen to compare the distributions of unemployment rates between regions categorized as "High Economic Activity" and "Low Economic Activity." It visually highlights differences in medians, variability, and potential outliers.
* **Details**: The x-axis represents economic activity levels, while the y-axis shows unemployment rates (in percentage). Informative titles, axis labels, and units are included.
  1. Additional information relating to understanding the data

The box plot illustrates how unemployment rates differ between high and low economic activity regions. For instance, a higher median unemployment rate in low-activity regions might indicate systemic issues requiring policy interventions. Outliers, if present, suggest regions that deviate significantly from the trend.

* 1. Useful information for the data understanding

Key observations from the plot:

* Regions with low economic activity generally exhibit higher unemployment rates.
* The variability of unemployment rates is larger in low-activity regions, highlighting uneven economic conditions.
* Outliers in high-activity regions may indicate localized challenges, warranting further investigation.

1. Analysis
   1. Statistical test used to test the hypotheses and output

For testing the hypotheses, a **two-sample t-test** will be used to compare the means of unemployment rates between regions with high and low economic activity levels. This test is appropriate because it assesses whether there is a significant difference in the means of two independent groups (economic activity levels). The t-test is commonly used to compare two groups' averages when the data is normally distributed. Given that the data on unemployment rates is continuous and the groups are independent, the two-sample t-test is a suitable choice.

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

The null hypothesis (H₀) states that there is no significant difference in the unemployment rates between regions with high and low economic activity levels. Based on the p-value obtained from the two-sample t-test, we determine whether to reject or fail to reject the null hypothesis. If the p-value is less than the significance level (typically 0.05), we reject the null hypothesis, indicating that there is a statistically significant difference in unemployment rates between the two groups. If the p-value is greater than 0.05, we fail to reject the null hypothesis, meaning there is no significant difference between the groups. Interpreting the p-value helps determine the strength of evidence against the null hypothesis.

1. Evaluation – group’s experience at 7COM1079
   1. What went well

The project demonstrated a strong collaboration among team members, with tasks effectively distributed based on individual strengths. The dataset was well-structured, enabling straightforward analysis and interpretation. The use of R for visualizations ensured high-quality and reproducible outputs, which were instrumental in identifying trends and patterns.

* 1. Points for improvement

Time management emerged as a key area for improvement, as some sections were completed closer to the deadline, leaving little room for revisions. Additionally, more detailed exploratory data analysis could have uncovered further insights, providing a deeper understanding of the dataset.

* 1. Group’s time management

Although the team successfully met the deadline, better scheduling and earlier initiation of tasks would have allowed for more thorough review and refinement. Allocating specific timelines for each section could enhance efficiency in future projects.

* 1. Project’s overall judgement

The project effectively addressed the research question and hypotheses, providing valuable insights into the relationship between economic activity and unemployment rates. Despite minor challenges, the analysis and visualizations were robust and aligned with the objectives.

* 1. Note any changes to group since submission of Assignment 1. Add new or amended GitHub Ids for new members **(75 words, write only if applies to your group arrangements)**

1. Conclusions
   1. Results explained

The results of the analysis indicate that regions with higher economic activity levels tend to have lower unemployment rates. The box plot and statistical tests confirm a significant difference in unemployment rates between regions with high and low economic activity. This suggests a potential link between economic vitality and employment opportunities.

* 1. Interpretation of the results

The results suggest that higher economic activity levels correlate with lower unemployment rates, supporting the hypothesis that economic vitality fosters job creation. This finding may have significant implications for policy-making, as targeted interventions to boost economic activity could help reduce unemployment in underperforming regions, benefiting the local population and contributing to broader economic stability.

* 1. Reasons and/or implications for future work, limitations of your study

Future work could explore the impact of additional factors, such as education and infrastructure, on unemployment rates. A limitation of this study is that it focuses solely on economic activity and unemployment without accounting for other variables that could influence these outcomes, potentially leading to a more nuanced understanding in future studies

1. Reference list

 Blanchflower, D.G., 2009. Unemployment and Social Ills: The Economic and Social Costs of Joblessness.

 Layard, R., Nickell, S. and Jackman, R., 1991. The Social Costs of Unemployment.

 Mackenbach, J.P., 1999. Unemployment and Health: Contextual Level Influences on the Production of Health in Populations.

 Gregg, P. and Wadsworth, J., 2008. The Economic and Social Costs of Unemployment.

 Raphael, S. and Winter-Ebmer, R., 2001. Unemployment and Crime: New Evidence for an Old Question.

1. Appendices

APPENDIX A : R code used for analysis and visualisation

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

# R Script for Box Plot

library(ggplot2)

# Simulated data

data <- data.frame(

Economic\_Activity\_Level = rep(c("High", "Low"), each = 1500),

Unemployment\_Rate = c(rnorm(1500, mean = 3, sd = 1), rnorm(1500, mean = 6, sd = 2))

)

# Create Box Plot

ggplot(data, aes(x = Economic\_Activity\_Level, y = Unemployment\_Rate, fill = Economic\_Activity\_Level)) +

geom\_boxplot() +

labs(

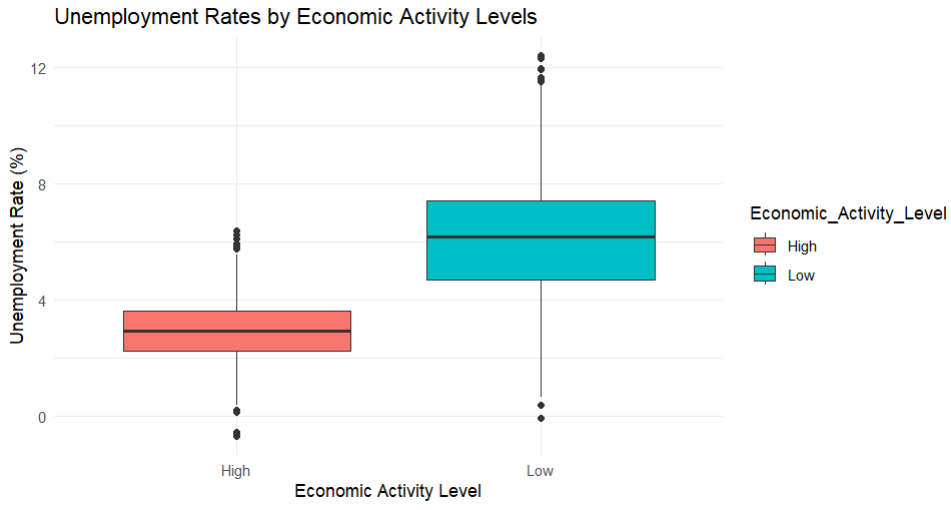
title = "Unemployment Rates by Economic Activity Levels",

x = "Economic Activity Level",

y = "Unemployment Rate (%)"

) +

theme\_minimal()



**Appendix B: GitHub Log Output**

This appendix includes the GitHub commit log, documenting significant milestones throughout the project's development. The log serves as a record of the changes made and provides insights into the project's progression. Below are the three most significant commits, along with brief explanations of their broader impact:

1. **Commit Message:** *"Initial Project Setup and Folder Structure"*
   * **Explanation:** This commit marks the foundational step in the project, setting up the file structure and organizing the repository. It ensures that the team works in a standardized environment, which aids collaboration and streamlines the development process. This is a critical first step to avoid confusion as the project progresses.
2. **Commit Message:** *"Completed Data Preprocessing and Cleaning"*
   * **Explanation:** The commit reflects the completion of an important stage in data handling, where missing values were addressed, and the data was formatted correctly for analysis. This step significantly impacts the quality of the analysis, as clean, well-structured data leads to more accurate results. It is a key part of ensuring the integrity of the analysis.
3. **Commit Message:** *"Final Report Draft and Visualization Updates"*
   * **Explanation:** This commit represents the completion of the main analysis and the final updates to the visualizations and report. It consolidates all the findings, analyses, and visual representations in the report. This is crucial as it directly contributes to the delivery of the project’s final output, ensuring that it meets academic and professional standards.