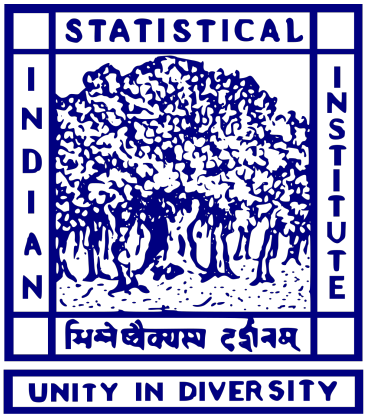

Indian Statistical Institute, Kolkata

Project for 'Statistical Methods -III'

Bachelor of Statistics (Hons.), 2023-26

Unemployment in India

Submitted By:
Unnati Mandal
Roll No: BS2355
Batch: 2023-26

Submitted To:
DR. AYANENDRANATH BASU
Professor, Interdisciplinary Statistical Research Unit

Submission Date:
November 8, 2024

Contents

1 Introduction

1.1 Abstract

Unemployment is a persistent problem in India, with significant social and economic consequences. This paper provides an overview of the state of unemployment in India before and after COVID-19, including the latest unemployment rate and the factors contributing to this issue. Indian government has taken various measures to address unemployment, such as skill development programs and the promotion of entrepreneurship and foreign investment. Despite these efforts, the challenge of creating adequate job opportunities for the growing population of India remains significant. This paper highlights the need for continued efforts to address unemployment in India and suggests potential solutions for improving the situation.

1.2 Problem Statement

Unemployment is measured by the unemployment rate, which is the number of people who are unemployed as a percentage of the total labor force. We have seen a sharp increase in the unemployment rate during COVID-19. This analysis intends to shed light on the socioeconomic consequences of the pandemic on India's workforce and labor market.

This dataset aids in comprehending the unemployment dynamics across India's states during the COVID-19 crisis. It offers valuable insights into how the unemployment rate, employment figures, and labor participation rates have been impacted across different regions in the country.

1.3 Data Overview

This dataset contains the unemployment rate of all the states in India. The dataset has 740 entries and 7 variables, which are as follows:

- Region
- Date
- Frequency
- Estimated Unemployment Rate
- Estimated Employed
- Estimated Labour Participation Rate
- Area

The link of the data is here:Unemployment in India
The link of the data source is here:Unemployment in India - Source link

1.4 Research Questions

- How does the unemployment rate vary in urban and rural places in India?
- How does the unemployment rate vary by region in India?
- How does the unemployment rate vary by both region and area (urban/rural)?
- How does the unemployment rate vary between the years 2019 and 2020?
- Which variable is strongly correlated with the unemployment rate ?

2 Analysis of Unemployment Rate in Different Areas of India

2.1 Performing t test

In this report, we analyze the unemployment rate in different areas (rural and urban) of India. The purpose is to determine whether there is a statistically significant difference in unemployment rates between rural and urban areas. To achieve this, we employ a two-sample t-test.

2.2 Methodology

The data for this analysis includes unemployment rates, categorized by rural and urban areas. We conducted a two-sample t-test to examine the differences between these groups. The null hypothesis (H_0) and alternative hypothesis (H_1) are defined as follows:

- H_0: There is no difference in unemployment rates between rural and urban areas.
- H_1: There is a difference in unemployment rates between rural and urban areas.

2.3 Data Analysis

The R code below was used to load the data, perform the t-test, and obtain summary statistics.

2.3.1 R Code

```r
# Load necessary libraries
library(dplyr)

# Load the dataset
data <-read.csv("C:\Users\UTTATI\Downloads\Unemployment in India.csv")

# Perform the t-test
t_test_result <- t.test(EstimatedUnemploymentRate ~ Area, data = data)

# Display the result
print(t_test_result)
```

2.4 Results

The results of the t-test are summarized below:

- **t-Statistic**: `-3.6448`
- **p-Value**: `0.0002865`
- **Degrees of Freedom**: `736.4`
- **Mean Unemployment Rate (Rural)**: `10.32479`
- **Mean Unemployment Rate (Urban)**: `13.16661`
- **95% confidence Interval** : `-4.372524 to -1.311122`

Based on the p-value obtained, we **reject** the null hypothesis, concluding that there **is** a significant difference in unemployment rates between rural and urban areas. The 95% confidence interval provides a range of values for the difference in means. Since both bounds are negative, it indicates that, on average, rural areas have a lower unemployment rate than urban areas.

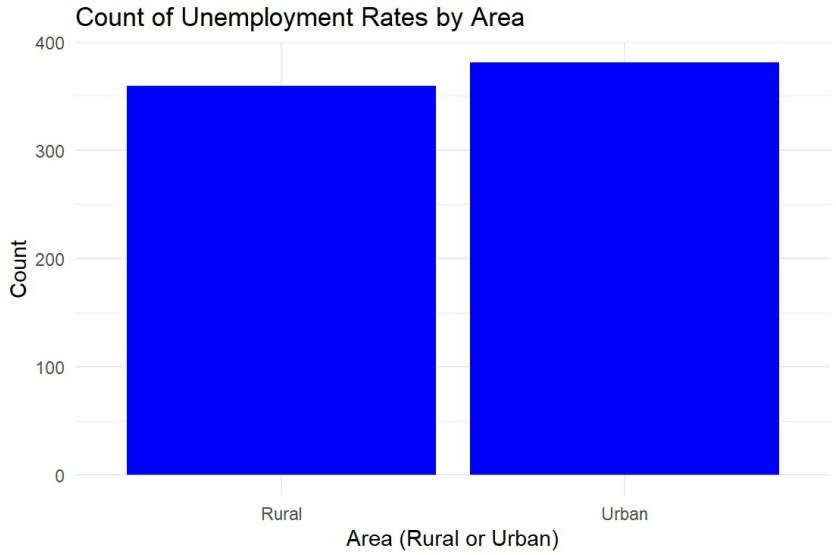

Figure 1: Count of Unemployment Rates by Area

2.5 Conclusion

This analysis provides insights into the distribution of unemployment rates in rural and urban areas. These findings can inform policy decisions focused on addressing area-specific unemployment issues.

3 Analysis of Unemployment Rates Across Different Regions of India

3.1 Performing One-way ANOVA

This report examines whether there is a statistically significant difference in unemployment rates across different regions (states) of India. To analyze this, we perform a one-way ANOVA test.

3.2 Methodology

A one-way ANOVA test was conducted to compare unemployment rates across different states. The null hypothesis (H_0) and alternative hypothesis (H_1) for this test are as follows:

- H_0: The mean unemployment rate is the same across all states.
- H_1: There is a significant difference in the mean unemployment rate among the states.

3.3 Data Analysis

The following R code was used to conduct the one-way ANOVA and summarize the results.

3.3.1 R Code

```
# Load necessary libraries
library(dplyr)
library(stats)

# Load the dataset
data <- read.csv("C:/Users/UTTATI/Downloads/Unemployment␣in␣India.csv")

# Run one-way ANOVA test
anova_result <- aov(EstimatedUnemploymentRate ~ Region, data = data)

# Display the ANOVA summary
summary(anova_result)
```

3.4 Results

The results of the one-way ANOVA test are summarized below:

- **F-Statistic**: `13.35`
- **p-Value**: `< 2e-16`
- **Degrees of Freedom**: `Region:27, Residuals:712`

Based on the p-value, we **reject** the null hypothesis, concluding that there **is** a significant difference in unemployment rates across different regions.

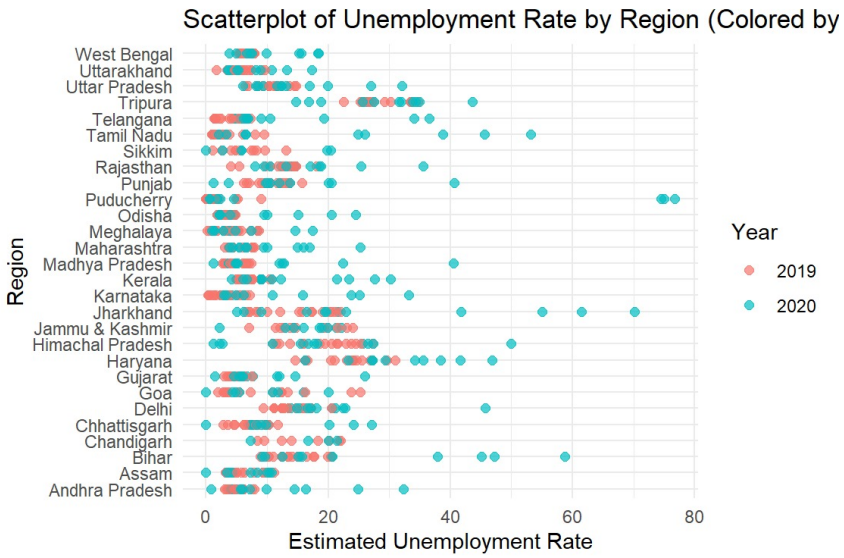

Figure 2: Scatterplot of Unemployment Rates by Region

In 2019, unemployment rate was less as compared to 2020. Also in Puducherry, Jharkhand, Bihar the unemployment rate increased in 2020 due to pandemic.

3.5 Conclusion

This analysis provides insights into regional variations in unemployment rates. The findings can be valuable for policy formulation aimed at addressing regional employment disparities.

4 Unemployment Rates Across Regions and areas in India.

4.1 Performing Two-way ANOVA

This report investigates whether there are significant differences in unemployment rates across regions (states) and areas (rural or urban) in India. To examine these relationships, we employ a two-way ANOVA test, which also allows us to assess any interaction effect between regions and area types.

4.2 Methodology

The two-way ANOVA test includes two factors: region (representing different states) and area (rural or urban). The null hypotheses for the test are:

- H_{01}: There is no significant difference in unemployment rates across different regions.
- H_{02}: There is no significant difference in unemployment rates between rural and urban areas.
- H_{03}: There is no interaction effect between region and area on unemployment rates.

4.3 Data Analysis

The following R code was used to perform the two-way ANOVA and summarize the results.

4.3.1 R Code

```r
# Load necessary libraries
library(dplyr)
library(stats)

# Load the dataset
data <- read.csv("C:/Users/UTTATI/Downloads/Unemployment in India.csv")

# Run two-way ANOVA test
two_way_anova <- aov(EstimatedUnemploymentRate ~ Region * Area, data = data)

# Display the ANOVA summary
summary(two_way_anova)
```

4.4 Results

The results of the two-way ANOVA test are summarized below:

- **F-Statistic for Region**: `13.513`
- **p-Value for Region**: `< 2e-16`
- **F-Statistic for Area**: `19.385`
- **p-Value for Area**: `1.24e-05`
- **F-Statistic for Interaction (Region * Area)**: `0.636`
- **p-Value for Interaction (Region * Area)**: `0.92`

Based on the p-values, we **reject** the null hypotheses. Both Region and Area have significant individual effects on unemployment rates, meaning that: **Unemployment rates vary considerably from region to region**. There is a consistent difference between urban and rural unemployment rates across regions. However, **there is no significant interaction between Region and Area**, suggesting that the difference in unemployment rates between urban and rural areas is relatively uniform across regions.

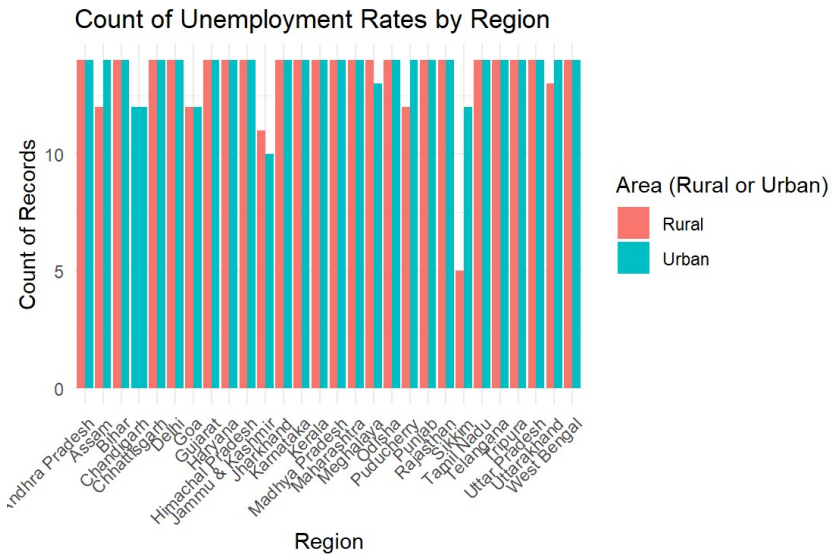

Figure 3: countplot of Unemployment Rates by Region

Majority of the states have an equal share of unemployment rate in Rural as well as Urban areas. Sikkim has the lowest unemployment rate in Rural areas.

4.5 Conclusion

This analysis provides insight into how unemployment rates vary across regions and between rural and urban areas. The findings may inform targeted policies to address specific regional and area-based unemployment issues.

5 Analysis of Unemployment Rate of Different year (2019 and 2020) in India

5.1 Performing t test

In this report, we analyze the unemployment rate in different year (2019 and 2020) of India. The purpose is to determine whether there is a statistically significant difference in unemployment rates between 2019 and 2020. To achieve this, we employ a two-sample t-test.

5.2 Methodology

The data for this analysis includes unemployment rates, categorized by 2019 and 2020 data. We conducted a two-sample t-test to examine the differences between these groups. The null hypothesis (H_0) and alternative hypothesis (H_1) are defined as follows:

- H_0: There is no difference in unemployment rates between year 2019 and 2020.
- H_1: There is a difference in unemployment rates between year 2019 and 2020.

5.3 Data Analysis

The R code below was used to load the data, perform the t-test, and obtain summary statistics.

5.3.1 R Code

```r
# Load necessary libraries
library(dplyr)

# Load the dataset
data <-read.csv("C:\Users\UTTATI\Downloads\Unemployment␣in␣India.csv")

#Create a year column
install.packages("lubridate")
library(lubridate)
data$year <- year(dmy(data$Date))

# Convert year to numeric if it's not already
data$year <- as.numeric(as.character(data$year))

# Perform the t-test
t_test_result <- t.test(EstimatedUnemploymentRate ~ year, data = data)

# Display the result
print(t_test_result)
```

5.4 Results

The results of the t-test are summarized below:

- **t-Statistic**: `-6.7513`
- **p-Value**: `4.7e-11`
- **Degrees of Freedom**: `435.25`
- **Mean Unemployment Rate (2019)**: `9.399047`
- **Mean Unemployment Rate (2020)**: `15.101581`

- **95% confidence Interval** : `-7.362655 to -4.042414`

Based on the p-value obtained, we **reject** the null hypothesis, concluding that there **is** a significant difference in unemployment rates between year 2019 and 2020. The 95% confidence interval provides a range of values for the difference in means. Since both bounds are negative, it indicates that, on average, year 2019 have a lower unemployment rate than year 2020.

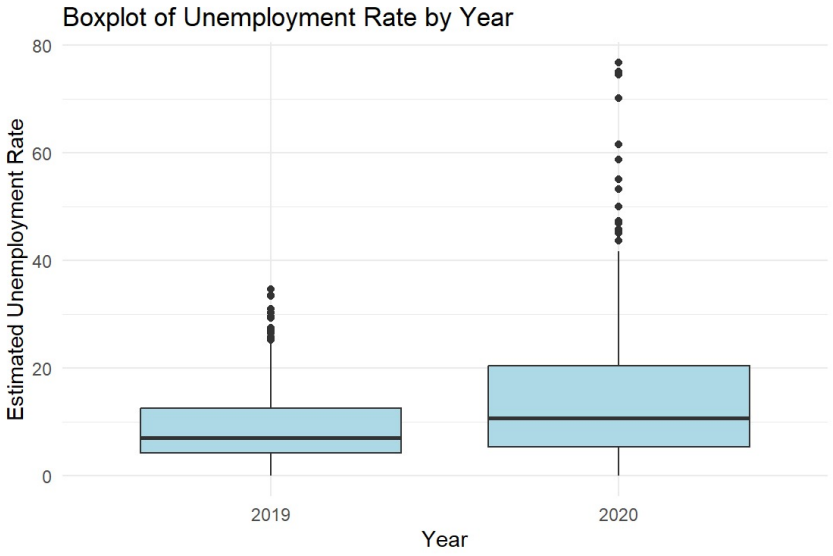

Figure 4: Boxplot of Unemployment Rates by year

There are no outliers. Due to Covid-19 pandemic, there was sudden increase in the Unemployment Rate in 2020 due to Lockdown

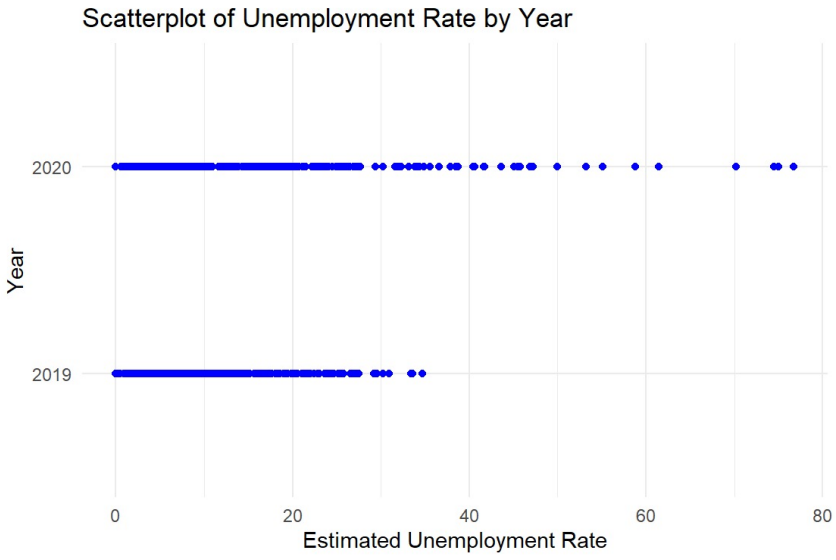

Figure 5: Scatterplot of Unemployment Rates by year

We can clearly see that in 2020 there was an increase in the unemployment rate after the rise of Covid-19.

5.5 Conclusion

This analysis provides insights into the distribution of unemployment rates in year 2019 and 2020. These findings can inform policy decisions focused on addressing area-specific unemployment issues.

6 Correlation Analysis between various variables

6.1 Correlation Analysis

This report examines the correlation between various variables in the unemployment in India dataset, including estimated unemployment rate, estimated employed, estimated labor participation rate, and year. Understanding these correlations can help identify relationships between different economic indicators.

6.2 Methodology

To calculate the correlation between the variables, we utilized the Pearson correlation coefficient, which measures the linear relationship between two continuous variables. The coefficients range from -1 to 1, where:

- 1 indicates a perfect positive correlation,
- -1 indicates a perfect negative correlation,
- 0 indicates no correlation.

6.3 Data Analysis

The following R code was used to compute the correlation matrix for the relevant columns in the dataset.

6.3.1 R Code

```
# Load necessary libraries
library(dplyr)

# Load the dataset
data <- read.csv("C:/Users/UTTATI/Downloads/Unemployment␣in␣India.csv")

install.packages("corrplot")
library(corrplot)
# Select only numeric columns
numeric_data <- data[sapply(data, is.numeric)]

# Calculate the correlation matrix
correlation_matrix <- cor(numeric_data, use = "complete.obs")

# Plot the correlation matrix
corrplot(correlation_matrix, method = "color", type = "upper",
         tl.col = "black", tl.srt = 45,
         title = "Correlation␣Matrix␣of␣Numeric␣Variables", mar = c(0,0,1,0))
```

6.4 Results

The correlation matrix obtained from the analysis is shown below:

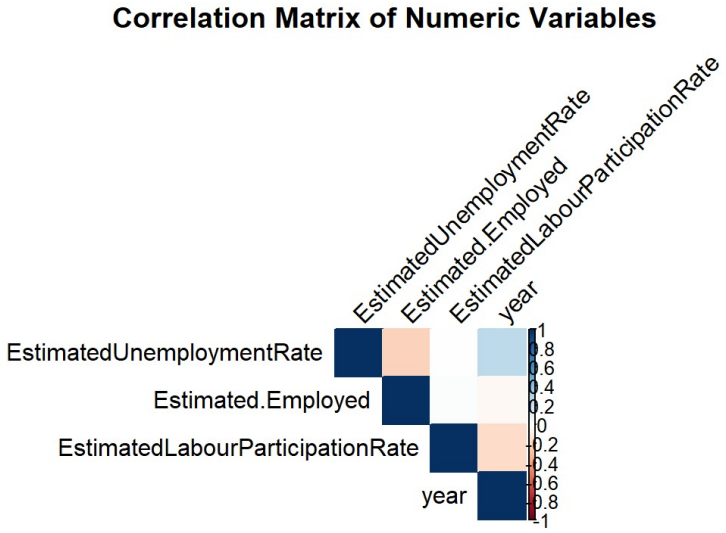

Figure 6: Correlation Matrix

Interpretation of the correlation coefficients:

- There is a positive correlation between the estimated unemployment rate and estimated employed individuals, indicating that as employment increases, unemployment tends to increase.
- A moderate correlation (almost 0) exists between the estimated unemployment rate and estimated labor participation rate.
- The correlation between year and estimated unemployment rate is positive, which may indicate a trend over time.

6.5 Conclusion

The correlation analysis reveals significant relationships between unemployment rates and various economic indicators. These insights can inform policymakers and researchers about trends and relationships that may require further investigation.

7 Policy Recommendations

Based on the findings from the data analysis of unemployment rates and other economic indicators, the following actionable policy recommendations are proposed:

- **Targeted Regional Employment Initiatives**
  - **Finding**: The ANOVA test indicates significant regional differences in unemployment rates.
  - **Recommendation**: Implement region-specific employment programs that address local needs in high-unemployment regions, including:
    - **Skill Development and Vocational Training**: Provide training aligned with regional industry demands (e.g., agriculture in rural areas, technology in urban areas).
    - **Incentives for Businesses**: Offer tax incentives or subsidies to companies establishing operations in high-unemployment regions.
  - **Expected Outcome**: These initiatives aim to reduce regional unemployment disparities and promote balanced economic growth across the country.
- **Urban Employment and Skill Alignment Programs**
  - **Finding**: T-test results show higher unemployment rates in urban areas compared to rural areas.

  - **Recommendation**: Develop programs targeting urban unemployment, focusing on:
    * **Skill Mismatch**: Address skill mismatches by providing urban-specific training in digital skills, technical skills, and soft skills.
    * **Support for Small Businesses and Start-ups**: Facilitate urban entrepreneurship through access to loans, grants, and mentorship, especially in sectors such as technology and retail.
  - **Expected Outcome**: This approach aims to reduce urban unemployment by addressing skill gaps and encouraging entrepreneurship.

- **Enhancing Rural Workforce Participation**
  - **Finding**: Labor participation rate trends suggest room for improvement in rural workforce engagement.
  - **Recommendation**: Focus on policies that facilitate workforce entry and retention in rural areas:
    * **Agricultural Diversification**: Encourage employment in non-agricultural sectors like rural manufacturing and services.
    * **Infrastructure and Connectivity Improvements**: Invest in transportation and digital infrastructure to facilitate workforce participation and entrepreneurship in rural areas.
  - **Expected Outcome**: These strategies are expected to boost rural workforce participation and reduce poverty, creating a more resilient rural economy.

- **Sector-Specific Job Creation and Training Programs**
  - **Finding**: Correlation analysis suggests that unemployment and labor participation rates may be affected by socio-economic factors like industry presence.
  - **Recommendation**: Implement sector-specific training programs tailored to regional strengths:
    * **Agriculture and Allied Activities**: In predominantly rural areas, provide training in modern agriculture, agro-processing, and sustainable farming.
    * **Industrial Skill Development**: Near industrial hubs, collaborate with local industries for skill training in manufacturing, construction, and logistics.
  - **Expected Outcome**: This policy is intended to enhance employability in key sectors, facilitating a smoother transition from unemployment to employment.

- **Continuous Data Monitoring and Policy Adjustment**
  - **Finding**: Unemployment and labor participation rates fluctuate over time, requiring adaptable policies.
  - **Recommendation**: Establish systems for regular data monitoring and policy adjustments:
    * **Real-Time Labor Market Data**: Develop a platform to track unemployment rates, labor participation, and job vacancies by region.
    * **Feedback Mechanism**: Gather feedback from regional governments and industries to tailor policies based on current economic trends.
  - **Expected Outcome**: Continuous monitoring allows for data-driven policy adjustments, ensuring policies remain effective and responsive to emerging trends.

- **Support Skills Development Programs:** Given the changing dynamics of the labor market, investing in skills development and training programs is crucial. Tailored training programs that align with the needs of high-demand sectors can equip job seekers with the necessary skills, thereby reducing unemployment rates.

- **Monitor Economic Trends Over Time:** The positive correlation between the year and estimated unemployment rate suggests the necessity for ongoing monitoring of economic trends. Establishing a robust framework for analyzing labor market data can help policymakers anticipate changes and implement timely interventions to mitigate rising unemployment rates.

- **Foster Economic Diversification:** Encouraging economic diversification can reduce vulnerability to sector-specific downturns. By supporting various industries, particularly those that are labor-intensive, the government can create a more resilient economy with lower unemployment rates.
- **Implement Targeted Regional Policies:** Since unemployment rates can vary significantly by region, targeted policies that address specific regional challenges are essential. Investments in infrastructure, education, and local business development can help alleviate regional disparities in employment rates.

These recommendations aim to provide a framework for addressing unemployment in India effectively, fostering a more inclusive and robust labor market.