Data Analysis Report: COVID-19 Employment and Unemployment Trends (2019–2020)

Author: Muhammad Abdurrehman
FAST-NUCES

1. Data Sources and Preparation

The analysis was conducted using structured employment datasets from 2019 and 2020. These datasets contained region-wise records for estimated employment, unemployment rates, and labour participation rates. The following key steps were carried out to prepare the data:

- Data from multiple sources were loaded into two primary subsets, df1_subset and df2_subset.
- Both datasets were concatenated into a unified DataFrame, combined_df, using pd.concat().
- Missing values in important fields such as Area, Region.1, longitude, and latitude were filled using the most frequent (mode) value per region or mapping from non-null reference rows.
- The Date column was converted to datetime format, and a new period-based field COVID_Period was introduced to categorize rows into "Pre-COVID" and "COVID Period" based on date ranges.
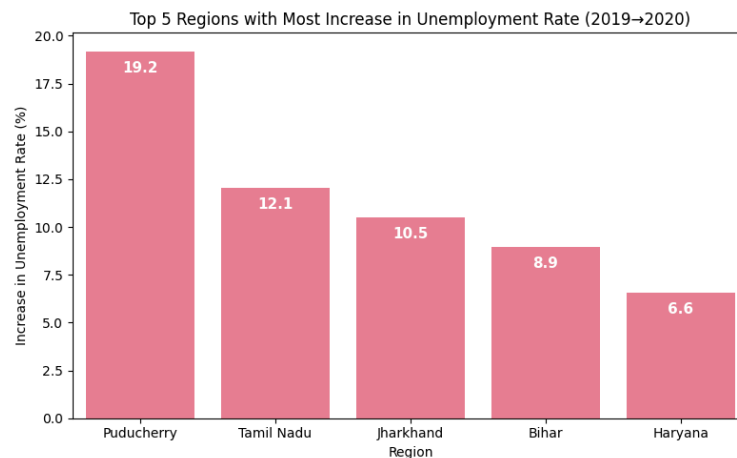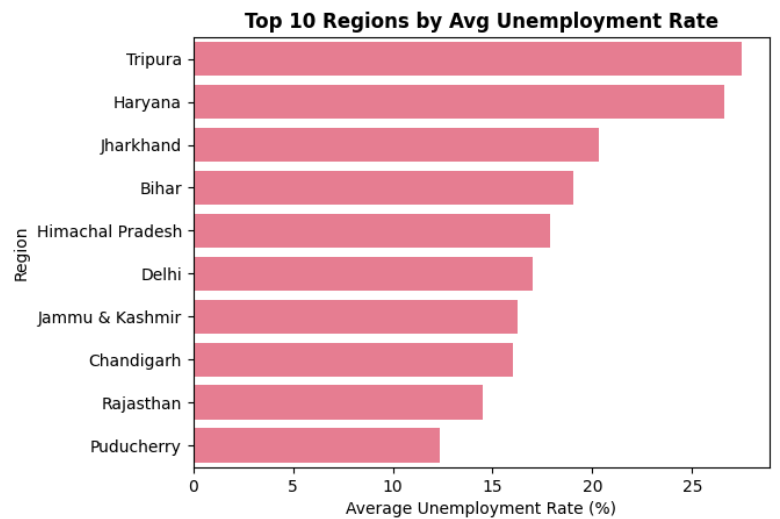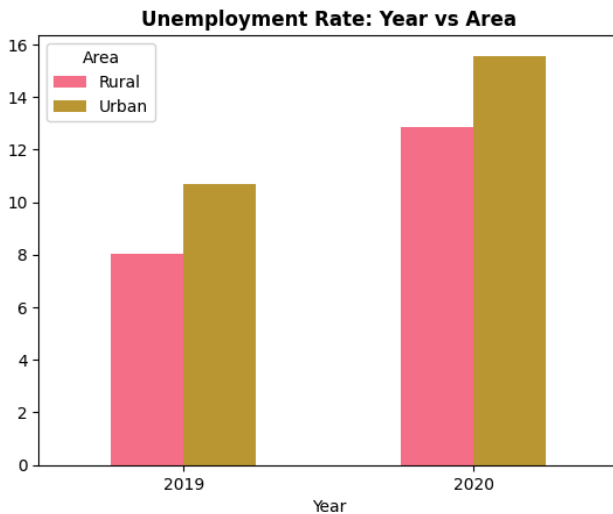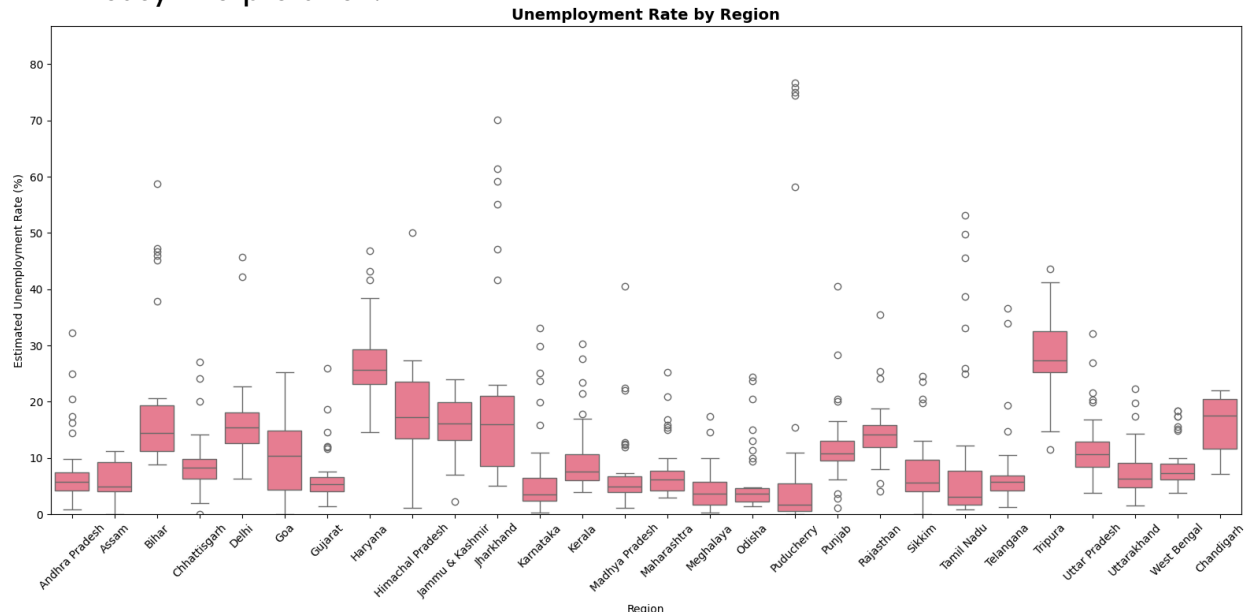
2. Regional Employment and Labour Insights

The report primarily focuses on regional variation in employment statistics. The following insights were drawn:
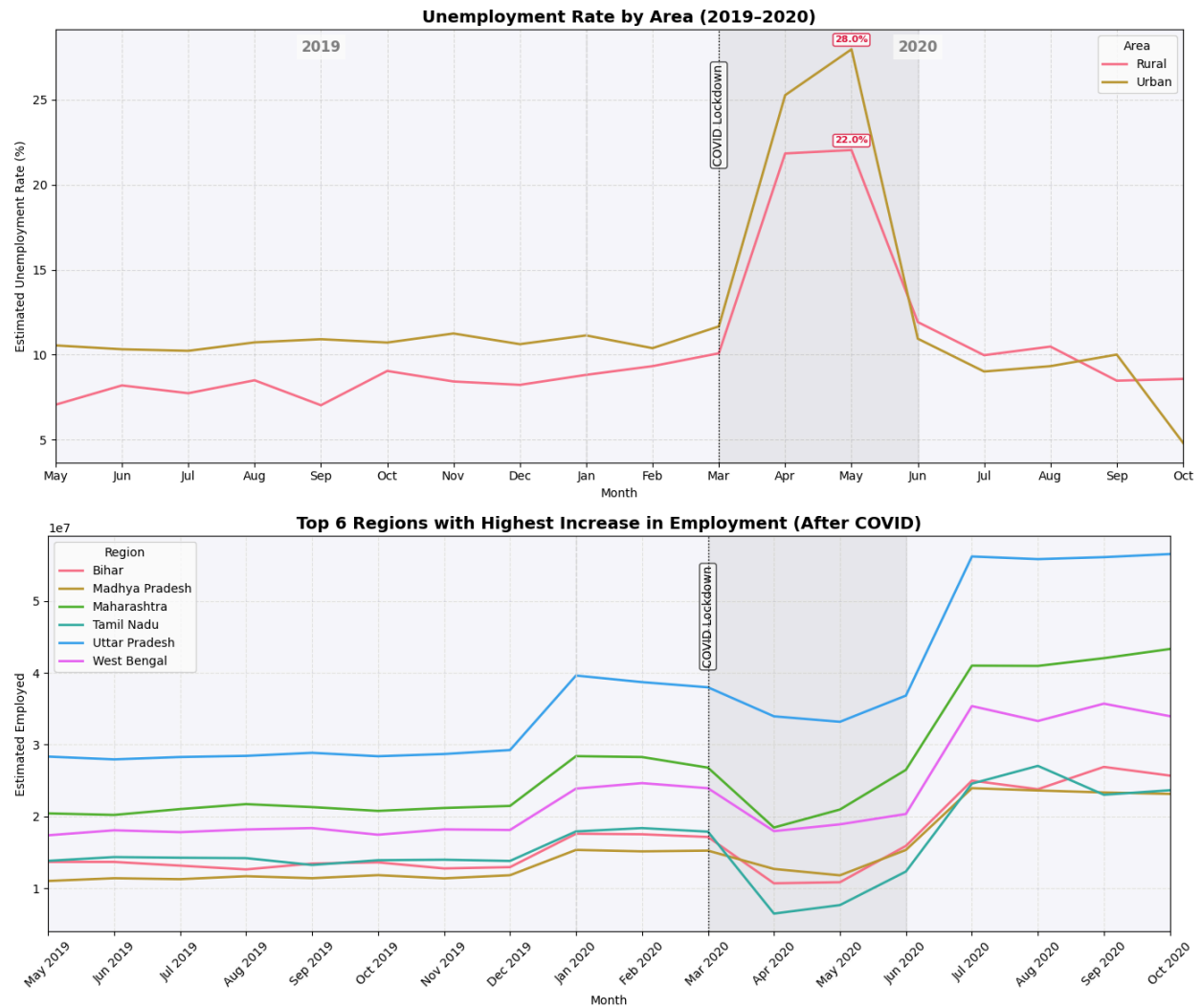
- Labour participation rates were aggregated by region and visualized using bar plots.
- Unemployment rate distributions were plotted using violin plots, revealing differences in variability across regions.
- Area labels (Urban/Rural) were inferred where missing, using the mode value associated with each region.
- Geographic zone data (Region.1) and coordinates (latitude, longitude) were restored where applicable.

3. Time Series Analysis (2019–2020)

A comprehensive time series analysis was conducted for unemployment rates across the two-year window. Steps included:

- Aggregation of monthly unemployment rates by area (urban, rural).
- Conversion of date fields to monthly periods for coherent temporal plotting.
- Annotation of each year label at its midpoint, and use of vertical lines and shaded spans to represent year transitions and the COVID period respectively.
- Highlighting of peak unemployment values per area with labeled annotations for easy interpretation.

Unemployment Rate by Region

Unemployment Rate: Year vs Area

Top 10 Regions by Avg Unemployment Rate

Top 5 Regions with Most Increase in Unemployment Rate (2019→2020)

Unemployment Rate by Area (2019–2020)

Top 6 Regions with Highest Increase in Employment (After COVID)
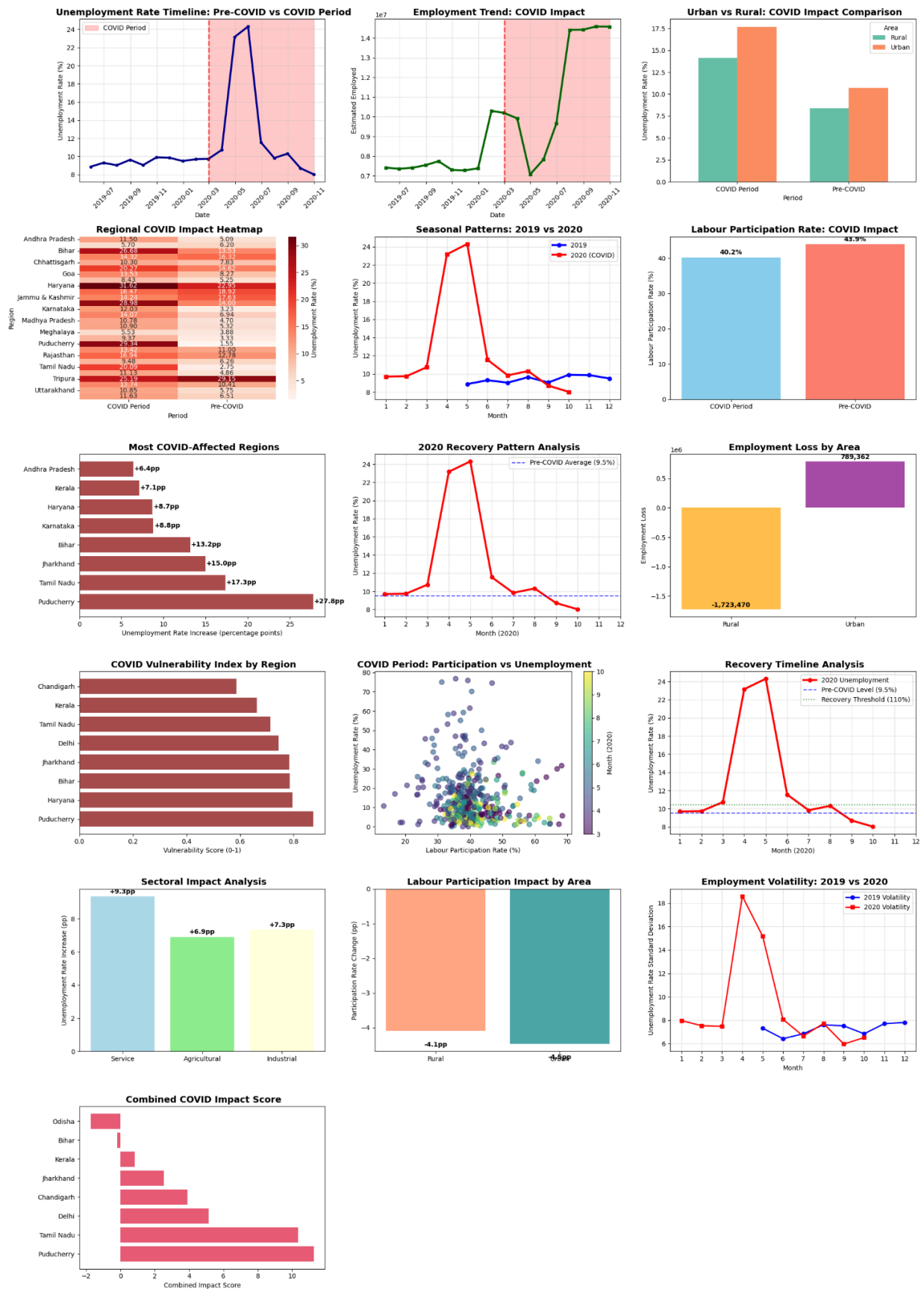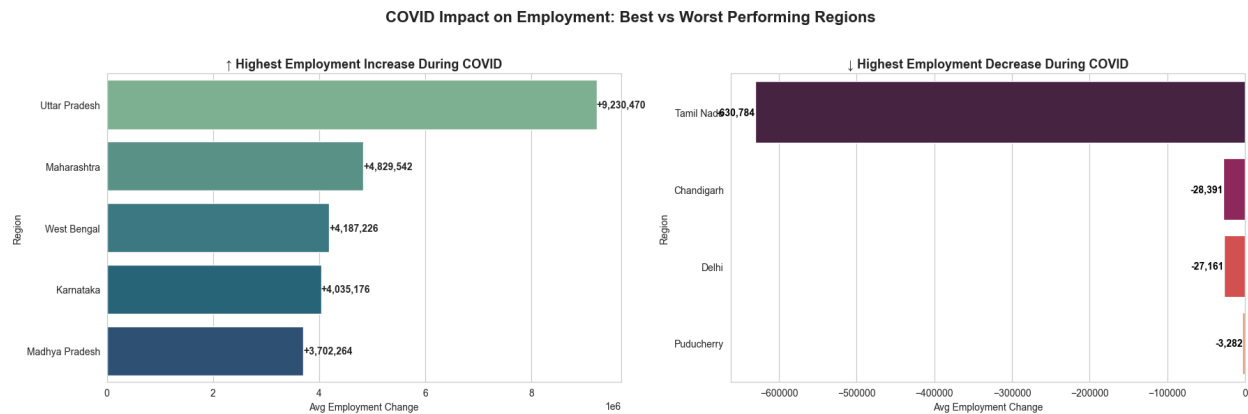
4. COVID-19 Impact Evaluation

To assess the economic impact of COVID-19, comparisons were made between the "Pre-COVID" and "COVID Period" intervals. The following metrics and visualizations were used:

- Average estimated employment before and during COVID were calculated and compared region-wise.
- Two bar plots were displayed side-by-side: one for the top five regions with the largest employment increase, and the other for the five with the largest decrease.
- A dumbbell plot was generated to compare pre-COVID and COVID unemployment rates for the ten most affected regions, emphasizing the absolute and percentage point changes.

- A dedicated marker for March 2020 and a shaded region through June 2020 helped visually define the onset and height of the lockdown's economic impact.

COVID Impact on Employment: Best vs Worst Performing Regions

↑ Highest Employment Increase During COVID / ↓ Highest Employment Decrease During COVID

5. Key Findings

- Puducherry experienced the highest surge in unemployment, jumping from 1.55% pre-COVID to 29.34% during COVID, an increase of 27.79 percentage points.
- Other significantly affected regions include Tamil Nadu, Jharkhand, Bihar, and Karnataka.
- Despite the overall negative trend, some regions showed resilience or even an increase in employment during COVID, as reflected in the comparative bar charts.

6. Visualization and Techniques

The project employed a wide variety of visualization techniques:

- Time series line plots to analyze monthly unemployment trends.
- Annotated peak markers for critical data points.
- Violin plots for distribution analysis across states.
- Dual bar plots and dumbbell plots to clearly communicate before-and-after changes.
- Use of matplotlib, seaborn, and pandas plotting extensions to format and style visuals appropriately.

Conclusion and Recommendations

This study successfully identifies regional disparities in employment and unemployment trends caused by the COVID-19 pandemic using public labor data. The time series and comparative visualizations provide strong evidence of the economic impact across various Indian states.

For future work, it is recommended to:

- Extend the dataset to include data from 2021 and beyond to assess post-pandemic recovery.
- Introduce geospatial analysis for visualizing unemployment spatially across the map.
- Integrate markdown documentation within the notebook for improved narrative clarity.

This report can serve as a foundation for policymaking and academic exploration into labor market disruptions caused by global crises.