Unemployment Prediction: Data Science Summary Report
---

1. Objective:
Analyze and predict the unemployment rate in India using historical data.

2. Data Source:
- Dataset: Unemployee.csv
- Features: Region, Area, Date, Estimated Employed, Labour Participation Rate

3. Exploratory Data Analysis:
- Time-based trends show increased unemployment post-COVID-19 (March 2020).
- The analysis suggests that the COVID-19 pandemic had a profound impact on
- unemployment rates, particularly in rural areas, highlighting the need for targeted policy interventions.
- The histogram shows that unemployment rate is right-skewed, with most values concentrated between 10% and 15%.
- Urban areas typically exhibit higher unemployment rates than rural.
- Regional differences observed; some regions consistently show higher unemployment.

4. Outlier Detection:
- Outliers in 'Estimated Unemployment Rate' were detected using IQR and removed.

5. Analysis of Data Distribution:
- The bar plot of record counts by month demonstrates consistent data availability across the observed period, supporting reliable time-series analysis.
- The count plot by Region reveals an uneven distribution of records across states, suggesting some regions are overrepresented.
- The Area distribution shows a relatively balanced number of records between Urban and Rural areas, enabling fair comparisons.
- The line plot of unemployment rate over time shows greater volatility in Rural areas, suggesting sensitivity to external factors.
- The horizontal bar chart reveals significant regional disparities in average unemployment rate, indicating structural or policy-based differences.
- The histogram shows that unemployment rate is right-skewed, with most values concentrated between 10% and 15%.
- The KDE curve indicates that this range is the most frequent, with a long right tail suggesting fewer areas have very high unemployment.
- The boxplot shows that rural areas tend to have a higher and more variable unemployment rate than urban areas.

6. Findings from Correlation and Variable Relationships
- The heatmap reveals a strong negative correlation between Estimated Employed and Estimated Unemployment Rate (%), implying that higher employment levels are associated with lower unemployment.
- Estimated Labour Participation Rate (%) also shows a positive correlation with employment reinforcing the link between workforce engagement and job availability.
- The scatterplot highlights how this relationship varies across urban and rural areas, with rural regions generally exhibiting greater variability.
- The analysis suggests that while urban areas have a more stable employment -unemployment relationship, rural areas are more susceptible to fluctuations.

5. Feature Engineering:
- COVID period flag introduced.