

Analyzing and Forecasting Unemployment Trends in India

Objective

Utilize state-level unemployment data in India (urban vs. rural) to analyze patterns over time, demographic differences, and the economic impact of events like COVID-19. Build dashboards and predictive models to forecast future unemployment rates and inform policy.

Dataset

- **Source:** Kaggle — *Unemployment in India* dataset (2016–2021)
- **Link:** [Unemployment in India dataset on Kaggle]([kaggle.com](https://www.kaggle.com/datasets/rajatsharma/unemployment-in-india))

Use Cases

- Track unemployment trends across states and between rural/urban areas
- Compare participation rates across regions
- Forecast future unemployment based on time-series patterns
- Visualize impact of lockdown or economic shocks
- Provide policy recommendations for high-risk regions

SECTION A: Python & Data Cleaning

- Load the dataset and inspect structure, data types, and nulls.
- Convert Date to datetime and extract Year, Month.
- Handle any missing values or outliers in employment or participation fields.
- Add derived features:
 - **Unemployment Rate** (% if not present)
 - **Gender Gap** if gender data is available, or rural–urban gap
 - **YoY Change** in unemployment rate per region

SECTION B: SQL Operations

- Import cleaned data into SQLite or pandasql environment.
- Example SQL queries:
 - Top 5 regions with highest average unemployment rate during 2020–21
 - Compare participation rates between rural and urban regions
 - Query regional unemployment trends over years
 - Identify regions with greatest improvement post-2020
 - Count months/years where unemployment rate exceeded national average

SECTION C: EDA & Descriptive Statistics

Exploratory Data Analysis

- Line charts of unemployment rate over time for select states (e.g., Delhi, Maharashtra)
- Bar charts comparing rural vs urban unemployment rates
- Heatmap: region × year showing unemployment intensity
- Box plots: participation rate distribution across regions
- Time series of employment and participation rates

Descriptive Statistics

- Mean, median, standard deviation for unemployment and participation
- Compute coefficient of variation by region
- Rank top 3 states with highest variability in unemployment
- Identify months/years with extreme peaks or dips aka outliers (IQR method)
- Calculate correlation between unemployment rate and participation rate

SECTION D: Tableau Dashboard

- Interactive map of India showing unemployment rates by state per year
- Filters: Region type (rural/urban), Year, Month
- Line chart of unemployment trends per region
- Bar chart comparison of participation vs unemployment by region
- KPI cards: highest unemployment rate, average participation
- Heatmap: region vs year unemployment values
- Trend absorber: highlight policy impact years like 2020

SECTION E: Machine Learning

Regression Task

- Forecast future unemployment rate (e.g., for 2022–2023) using features such as:
 - Current & previous rates, participation rate, region type
- Models: Linear Regression, Random Forest Regressor

Classification Task

- Categorize unemployment rate into buckets: Low (<5%), Medium (5–10%), High (>10%)
- Train models: Logistic Regression, Decision Tree, Random Forest
- Evaluate using confusion matrix, accuracy, precision/recall
- Feature importance to identify drivers of high unemployment