

Remote Data Analytics Internship

Report on Unemployment in India

By Amna Bibi

Report on My First Task as a Data Analyst Intern

Introduction:

As a Data Analyst Intern, my first task involved analyzing a dataset titled "Unemployment in India." The story behind this dataset is how lock-down affects employment opportunities and how the unemployment rate increases during the Covid-19. This dataset presented an opportunity to apply data analysis techniques and gain practical experience in handling real-world data. I divided the task into several steps, including data exploration, data cleaning, data formatting, and data visualization using Power BI.

Step 1: Data Exploration:

The initial step in the process was to explore the dataset. This involved familiarizing myself with the structure of the data, understanding the various columns and their significance, and identifying potential areas that required attention. I examined the dataset for any anomalies, inconsistencies, or missing values that could affect the quality and reliability of the analysis.

This was the dataset I had to analyze: [Unemployment in India \(kaggle.com\)](https://www.kaggle.com/datasets/india-unemployment)

Step 2: Data Cleaning and Formatting:

The next step was to perform data cleaning and formatting. This phase was crucial to ensure that the data was accurate, consistent, and ready for analysis.

The following tasks were performed:

Removing Duplicates: Identified and removed duplicate entries to prevent skewed results and ensure data integrity.

Handling Errors: Corrected errors in the dataset, such as incorrect data entries or outliers that could distort the analysis.

Handling Missing Values: Addressed blanks or missing data points by either filling them with appropriate values or removing them, depending on the context and significance of the data.

Data Formatting: Reformatted the data to a consistent structure, including standardizing date formats and aligning numerical values for better readability and analysis.

After ensuring that the data was clean and properly formatted, I used pivot tables to draw preliminary insights from the data. Pivot tables allowed me to summarize and analyze the data efficiently, providing a clearer understanding of the unemployment trends in India.

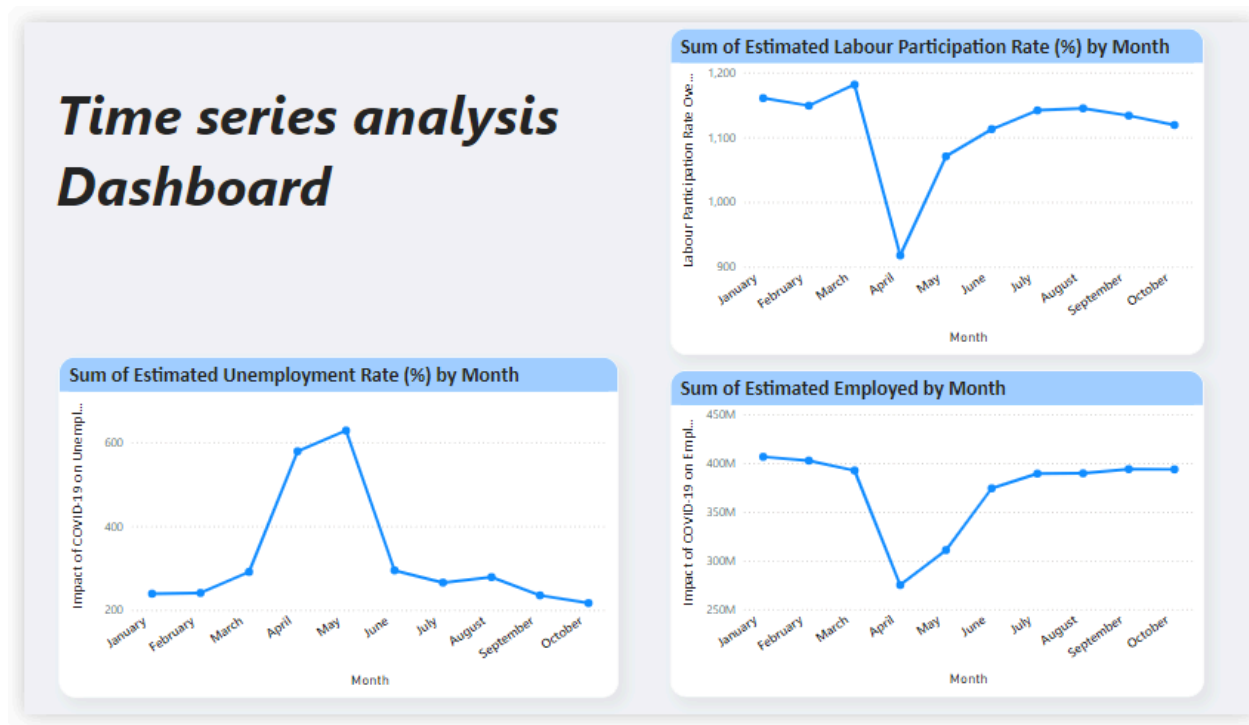
Step 3: Data Visualization:

Once the data was prepared, I moved on to the data visualization phase. I used Power BI, a powerful data visualization tool, to create multiple dashboards. These dashboards provided visual insights into the unemployment trends, highlighting key patterns and relationships within the data.

The Dashboards I created includes:

Time Series Analysis:

Shows the impact of COVID-19 on the unemployment rate, labor participation rate, and estimated employment in India by month.

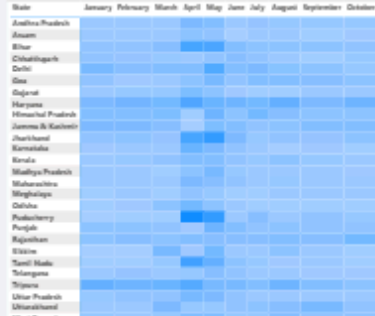


Geographical analysis:

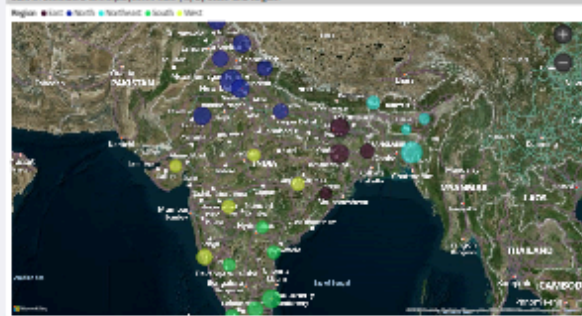
Demonstrates the impact of COVID-19 on the unemployment rate, labor participation rate, and estimated employment in India by state and region.

Geographical analysis Report

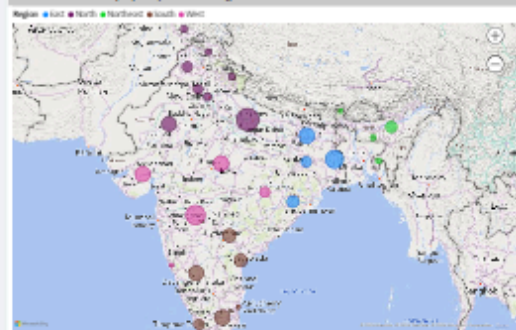
Heatmap of Unemployment Rates by State



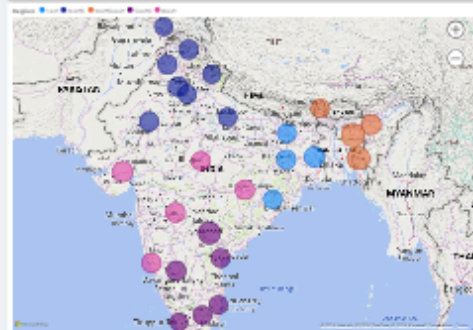
Sum of Estimated Unemployment Rate (%) by State and Region



Sum of Estimated Employed by State and Region



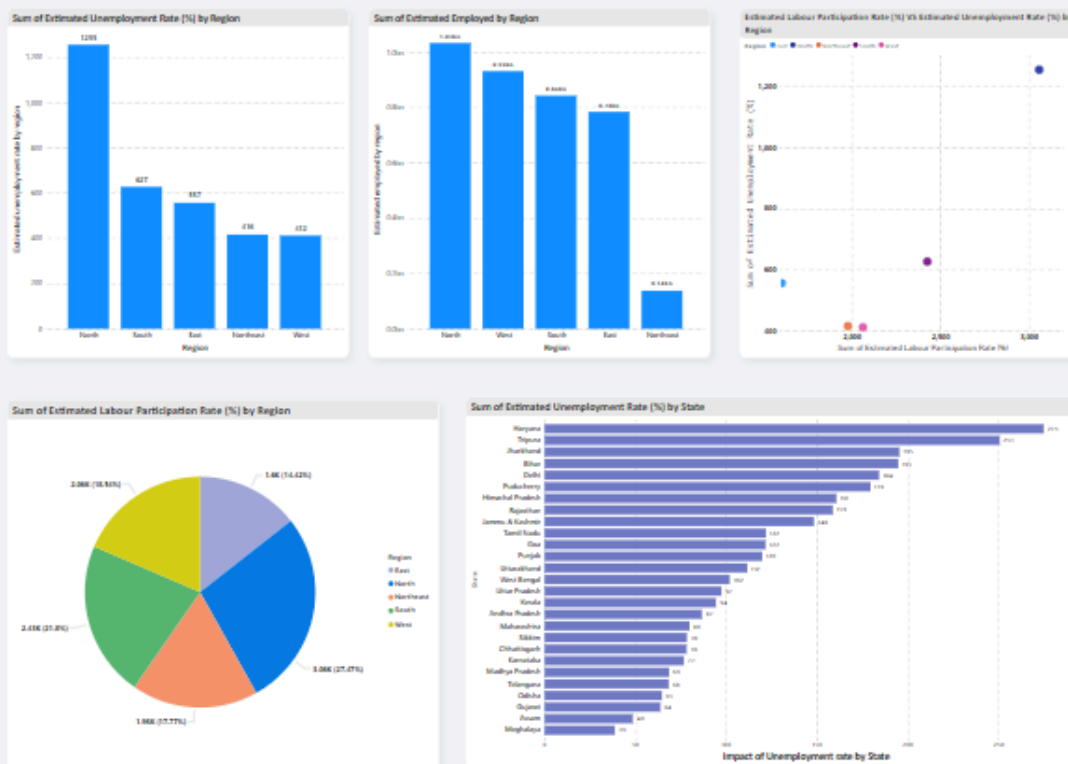
Sum of Estimated Labour Participation Rate (%) by State and Region



Use of different visuals:

I have used different visuals, such as bar charts and bubble maps, to display the same results because each visual type can provide a unique perspective. I also learned about the significance of using different visuals to present the same data.

Unemployment Rate Analysis Report



Learning Outcomes:

This task was a significant learning experience, as it allowed me to apply theoretical knowledge to practical scenarios. I gained hands-on experience in data cleaning, formatting, and visualization, and learned how to use Power BI to create impactful visual representations of data. Moreover, this exercise helped me develop a deeper understanding of data analysis workflows and the importance of maintaining data integrity throughout the process.