

Final Data Science Project: Unemployment Analysis (1976-2018)

Congratulations, you've been hired by a national economic research institution to conduct a comprehensive analysis of the **"Unemployment by State 1976 - 2018"** dataset. The organization is keen to gain valuable insights into the historical patterns of unemployment across different US states. The economic research institution is particularly interested in understanding the factors influencing unemployment rates, variations across states, and potential indicators for economic downturns or recoveries. Your analysis will be presented to policymakers, economists, and researchers, impacting strategic decisions related to labor markets and economic development.

Tasks:

1. Data Loading and Overview:

- Load the dataset into a DataFrame.
- Display the first few rows to understand the structure of the data.
- To maintain consistency, create a copy of the DataFrame which will be used to perform any changes to the data.

2. Data Cleaning:

- Check for missing or duplicate values and handle them appropriately.
- Handle any inconsistencies or outliers in the data.
- Ensure consistent data types for each column.
- Consider converting date columns to a datetime format.
- For easier coding, the names of the columns in the dataset should not contain any spaces.

3. Descriptive Statistics:

- Generate summary statistics for key numerical columns.
- Explore the distribution of unemployment rates across states and years.

4. Time Series Analysis:

- Visualize the overall trend in unemployment rates over the years.
- Identify any significant fluctuations or patterns.
- Explore seasonality in unemployment rates.

5. State-wise Comparison:

- Analyze and compare unemployment rates among different states.
- Identify states with consistently high or low unemployment.
- Create a table showing the average unemployment rate per state, ordered from higher to lower and print the first 10 rows of the

6. Labor Force Participation:

- Investigate the correlation between the labor force and unemployment rates.
- Explore the percentage of the population actively participating in the labor force.

7. Data Exploration:

- Conduct exploratory data analysis to identify any interesting patterns or anomalies.
- Perform univariate and bivariate analysis to understand relationships between variables.
- Create a new feature to represent the change in unemployment rates from the previous month or year.
- Engineer features that represent the percentage of the labor force employed and unemployed.

8. Visualization:

- Utilize Matplotlib, Seaborn, and Plotly to create visualizations that effectively communicate your findings.
- Include time series plots, state-wise choropleth maps, and other relevant visualizations.

9. Documentation:

- Document your entire analysis in a well-structured Jupyter Notebook.
- Clearly explain each step, including the rationale behind your decisions.
- Use Markdown cells to provide context, insights, and any recommendations.

10. Predictive Modeling (Optional):

- If you feel confident, attempt to build a predictive model to forecast future unemployment rates.