

Data Analytics with Cognos

Group 2

Project : Marginal workers in Tamilnadu -A Socioeconomic Analysis

A.project objectives:

1. Analyzing Marginal Worker

Demographics:

This objective involves examining the demographic characteristics (e.g., age, gender, education, location) of marginal workers to gain insights into their profiles.

2. Understanding Age and Gender

Distribution:

This objective focuses specifically on age and gender, aiming to understand how these factors are distributed among marginal workers.

3. Exploring Industrial Categories:

This objective involves exploring the different industrial sectors or categories that marginal workers are employed in, potentially identifying patterns or trends.

B.Analysis Approach:

1. Data Collection:

Identify and obtain the dataset that contains relevant information about marginal workers, including demographics and industrial categories.

2. Data Cleaning:

Inspect the dataset for missing values, outliers, and inconsistencies. Impute or remove missing data as needed. Address outliers appropriately, considering their impact on analysis. Standardize or normalize data if necessary.

3. Data Exploration:

Generate summary statistics to get an initial understanding of the dataset. Create visualizations (e.g., histograms, box plots) to explore data distributions. Identify any patterns, trends, or anomalies in the data.

4. Feature Engineering (if required):

Create new features or transform existing ones that might be relevant to your analysis. For example, you could calculate age groups or categorize industries into broader categories.

5. Hypothesis Formulation:

Based on your project objectives, formulate specific hypotheses or research questions that you aim to address through analysis.

6. Statistical Analysis:

Perform statistical tests or analyses to test your hypotheses. For example, you might use chi-square tests to analyze gender distribution across industries.

7. Data Visualization:

Select appropriate visualization types (e.g., bar charts, pie charts, heatmaps) to represent demographic distributions, age and gender insights, and industrial category exploration effectively. Create visualizations that support your findings and make them easy to understand for your audience.

8. Interpretation and Insights:

Interpret the results of your analysis in the context of your project objectives. Identify key findings and insights, including any significant correlations, trends, or differences in demographics and industrial categories.

9. Report and Presentation:

Document your analysis process and findings in a clear and organized report or presentation. Communicate your insights effectively to stakeholders or your target audience.

10. Validation and Sensitivity Analysis (if necessary):

Validate your findings using additional techniques or sensitivity analyses to ensure the robustness of your conclusions.

11. Conclusion and Recommendations:

Summarize your main findings and provide recommendations or implications based on your analysis.

12. Documentation and Code:

Maintain well-documented code and data sources for reproducibility and future reference. Selecting suitable visualizations is crucial for effectively representing demographic distributions. Here are some visualization types you can consider based on your objectives:

C.Visualization :

1. Bar Charts:

Use bar charts to display counts or percentages of demographic groups (e.g., age groups, gender) within specific categories (e.g., industries). Stacked bar charts can show the composition of each category.

2. Pie Charts:

Pie charts are useful for displaying the distribution of a whole (e.g., gender distribution among marginal workers).

They work well when you want to show the proportion of each category relative to the total.

3. Histograms:

Histograms are great for visualizing the distribution of a single continuous variable (e.g., age). They help understand the distribution's shape, central tendency, and spread.