

TN Marginal Workers

Data Collection and Preparation:

- Collect data related to marginal workers in Tamil Nadu, focusing on variables such as industrial categories, age, occupation, education level, income, etc.
- Clean and preprocess the data to handle missing values, outliers, and inconsistencies. Transform the data into a suitable format for clustering analysis.

Feature Selection:

- Identify the relevant features (variables) that are crucial for clustering analysis, such as industrial category and age.

Clustering Algorithm Selection:

- Choose a suitable clustering algorithm for this analysis. Commonly used algorithms include K-means clustering, hierarchical clustering, and DBSCAN (Density-Based Spatial Clustering of Applications with Noise).

Determine the Number of Clusters:

- Use techniques like the Elbow Method or Silhouette Score to determine the optimal number of clusters for the data.

Clustering Analysis:

- Apply the chosen clustering algorithm to the selected features.
- Assign each data point (marginal worker) to a specific cluster based on the algorithm's output.

Interpretation and Visualization:

- Analyze the clusters to identify patterns among different industrial categories and age groups.
- Create visualizations such as scatter plots or heatmaps to illustrate the clusters and their characteristics.

Pattern Identification:

- Examine the clusters to identify commonalities and differences among various industrial categories and age groups.
- Interpret the patterns and insights obtained from the clustering analysis.

Insights and Recommendations:

- Provide insights and recommendations based on the patterns identified, aiming to support decision-making for policies or interventions related to marginal workers in different industrial categories and age groups.

Documentation and Reporting:

- Document the entire process, including data preparation, clustering analysis steps, results, and interpretations.
- Prepare a comprehensive report summarizing the findings, insights, and recommendations for stakeholders and further research.