

ASSESSMENT OF MARGINAL WORKERS IN TN

PHASE - 2

Clustering analysis is a powerful technique to identify patterns within data. In your case, you can use clustering to uncover patterns among different industrial categories and age groups among marginal workers.

Here's how you can proceed:

1. Data Preparation :

Ensure your dataset is ready, with features such as industrial category and age group properly labeled and prepared for analysis.

2. Feature Selection :

Decide which features (variables) you want to include in the clustering analysis. In your case, you would use industrial category and age group.

3. Data Scaling :

It's essential to scale or normalize the data if the features have different scales or units to give each feature equal importance in the clustering process.

4. Choosing a Clustering Algorithm :

Select an appropriate clustering algorithm based on your data and objectives. Common algorithms include K-Means, Hierarchical Clustering, and DBSCAN. The choice depends on the nature of your data and the number of clusters you want to identify.

5. Determine the Number of Clusters :

You may need to decide how many clusters (groups) you want to identify. This can be determined using methods like the elbow method, silhouette score, or domain knowledge.

6. Perform Clustering :

Apply the chosen clustering algorithm to your data. The algorithm will group similar data points (industrial categories and age groups) into clusters.

7. Visualize the Clusters :

To visualize the clusters and patterns, you can use techniques such as scatter plots, heatmaps, or parallel coordinates plots. These visualizations will help you understand the relationships between industrial categories and age groups within each cluster.

8. Interpret the Clusters :

Analyze the results to interpret what each cluster represents. For example, you might find clusters that indicate certain age groups are more prevalent in specific industrial categories, revealing patterns in employment.

9. Validate and Refine :

It's essential to validate the results and potentially refine the analysis. You can use internal validation metrics (e.g., silhouette score) or domain knowledge to assess the quality of the clusters.

10. Report Findings :

Present your findings and insights from the clustering analysis to your audience. Use visualizations and clear explanations to communicate the patterns you've identified.