**TN Marginal Workers Assessment**

**PHASE 2:INNOVATION**

**Introduction:**

Clustering analysis can help in industrial categories and age groups within a dataset.

The approaches for this analysis are

* Data collection
* Data preprocessing
* Choosing a cluster algorithm
* Clustering
* Validation
* Labeling

**Data collection:**

**Dataset Link:**[**https://tn.data.gov.in/resource/marginal-workers-classified-age-industrial-category-and-sex-scheduled-caste-2011-tamil**](https://tn.data.gov.in/resource/marginal-workers-classified-age-industrial-category-and-sex-scheduled-caste-2011-tamil)

**Data preprocessing:**

To clean and preprocess the data. This involves handling missing values

**Choosing a cluster algorithm:**

Some clustering algorithms are K-means, Hierarchical clustering, DBSCAN

**Clustering:**

Creating cluster based on both age and industrial category.

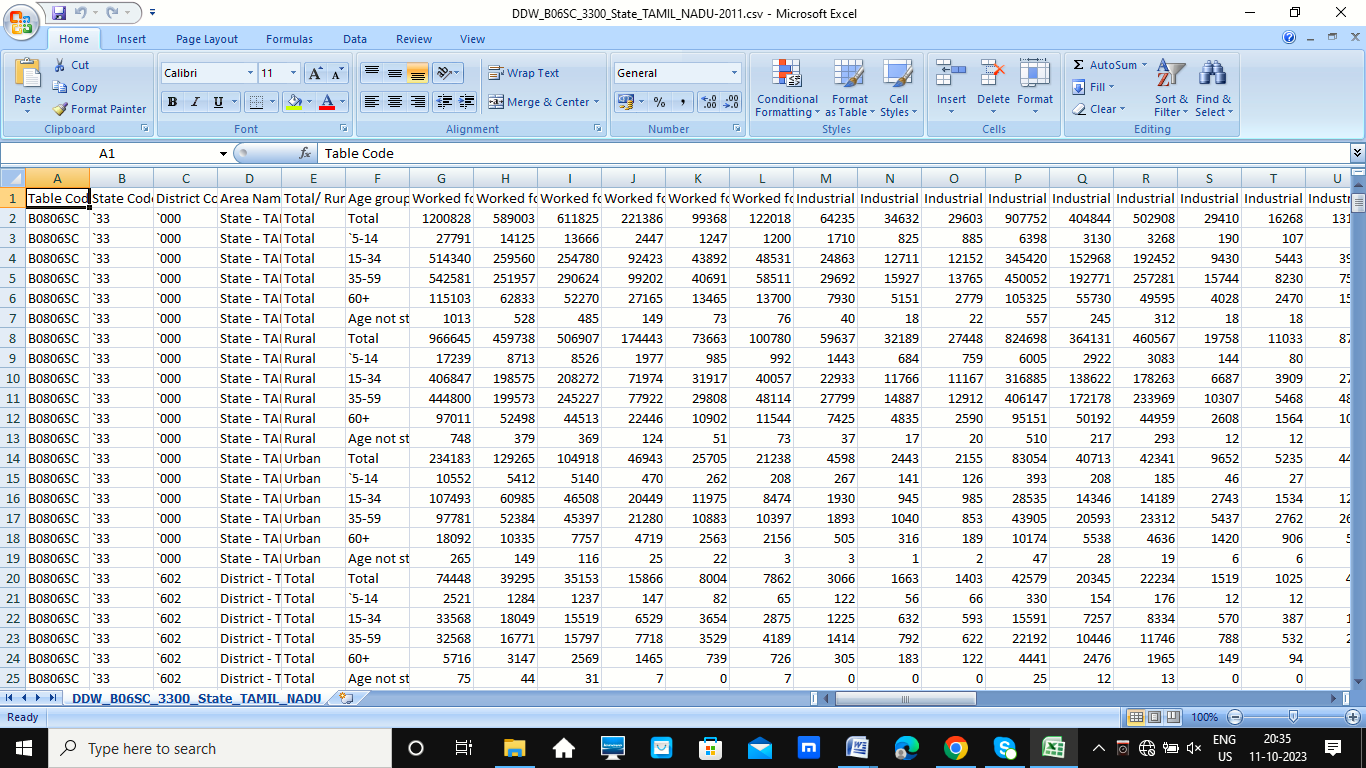
**Validation**:

* Assess the quality of the cluster.
* For clustering with mixed data types (e.g., numerical age and categorical industrial categories), we need to use appropriate validation metrics.

**Labeling:**

Assign labels to the clusters, representing different age groups and industrial categories.

The given data set



**Program:**

import pandas as pd

import matplotlib.pyplot as plt

# Load your dataset

data = pd.read\_csv("your\_data.csv") # Replace with the path to your dataset

# Assuming you have columns 'industry\_category' and 'age' in your dataset

industry\_category = data['industry\_category']

age = data['age']

# Create a bar chart to visualize the patterns

plt.figure(figsize=(10, 6))

plt.bar(industry\_category, age)

plt.xlabel('Industry Category')

plt.ylabel('Age')

plt.title('Age Distribution by Industry Category')

plt.xticks(rotation=90) # Rotate x-axis labels for readability

plt.show()

**Conclusion:**

**Pattern Recognition:**

* Clustering analysis helps identify natural patterns and groupings within the data.
* It can reveal how individuals or entities with certain age groups tend to belong to specific industrial categories.
* Clustering analysis is a valuable tool for identifying relationships between age groups and industrial categories.