**ASSESSMENT OF MARGINAL WORKERS IN TAMIL NADU – A SOCIOECONOMIC ANALYSIS**

**INTRODUCTION:**

The inquiry into the subject entails looking at the sexual orientation, age, and industrial classification-based population demographics of workers who are marginalized in Tamil Nadu. The overarching objective is to accomplish an economic and social investigation and foster illustrations that show how marginal workers are allocated among different sections of the population. In this project, the objectives have been laid out, the data analysis strategy is devised, beneficial visualization kinds have been determined, and the analysis happens to be carried out utilizing Python and data visualization libraries.

**OBJECTIVES:**

**Project Objectives:**

1. Marginal Worker Demographics Analysis:

- Identify and analyze key characteristics of marginal workers.

- Explore factors such as education, income, and employment status.

2. Age and Gender Distribution:

- Break down the dataset by age groups and genders to understand distribution patterns.

- Investigate any correlations between age, gender, and employment.

3. Exploring Industrial Categories:

- Categorize workers based on industries.

- Analyze the distribution of marginal workers across different industrial sectors.

**Analysis Approach:**

1. Data Extraction:

- Identify and source the dataset containing relevant information.

- Ensure the dataset includes variables like demographics, employment details, and industrial categories.

2. Data Cleaning:

- Handle missing or erroneous data points.

- Standardize data formats for consistency.

3. Data Analysis:

- Utilize statistical methods to analyze demographic trends.

- Apply clustering algorithms if necessary to identify patterns in industrial categories.

4. Insights Derivation:

- Derive actionable insights from the analyzed data.

- Look for correlations and trends that can inform decision-making.

**Visualization Selection:**

1. Demographic Distributions:

- Bar charts: For comparing the distribution of different demographic factors.

- Pie charts: To show the proportion of various categories within a demographic variable.

2. Age and Gender Distribution:

- Histograms: Display age distribution.

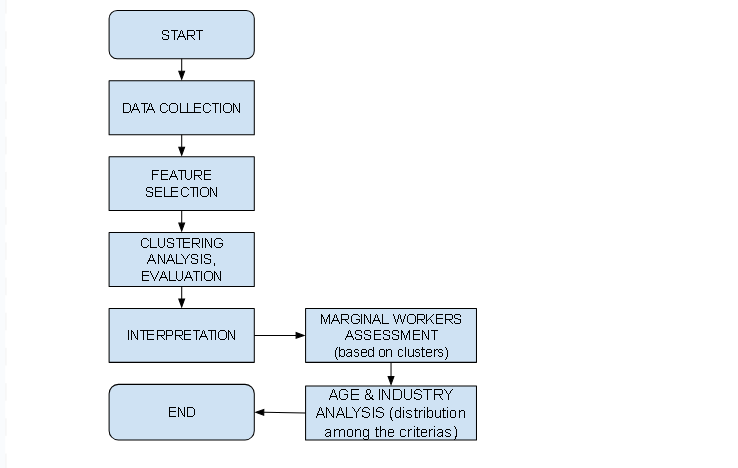
- Stacked bar charts: Showcase gender distribution within each age group.

3. Industrial Categories:

- Heatmaps: Visualize the concentration of workers across different industries.

- Bar charts: Compare the distribution of workers in each industrial category.

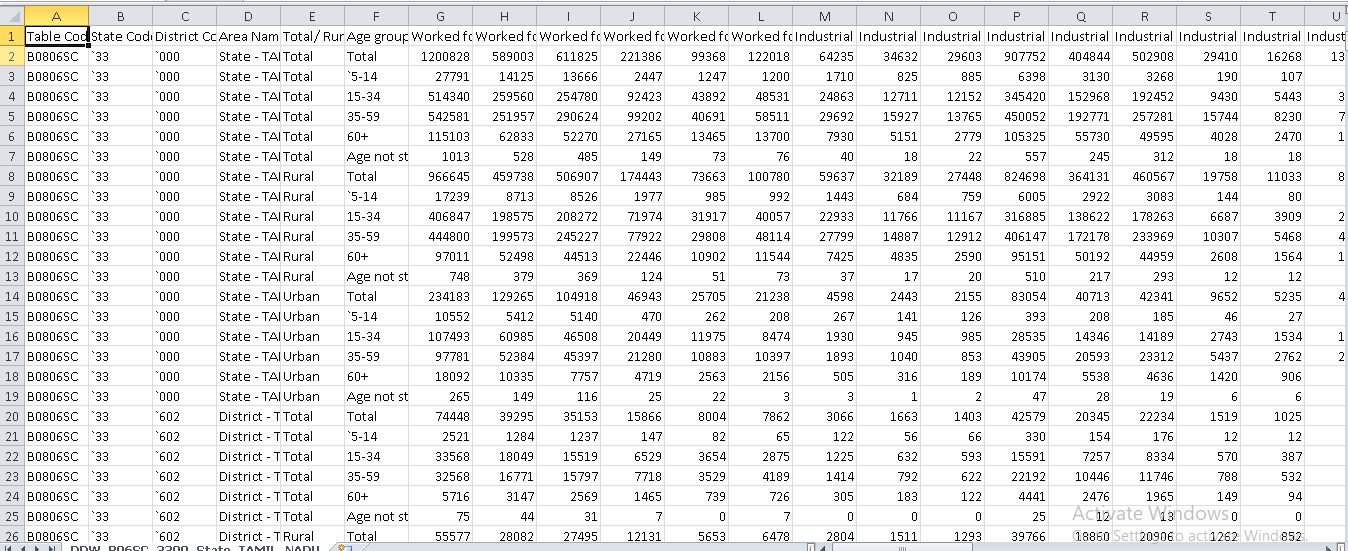
**FLOWCHART:**



**DATA COLLECTION:**

Collect data from the resource provided (<https://tn.data.gov.in/resource/marginal-workers-classified-age-industrial-category-and-sex-scheduled-caste-2011-tamil>)

Here is a snippet of the provided dataset,



**FEATURE SELECTION:**

Identify relevant features for clustering, such as age, industrial category, and sex.

**CLUSTERING ANALYSIS AND EVALUATION:**

Use clustering algorithms (e.g., k-means) to identify patterns among different industrial categories and age groups. Evaluate the quality of clusters to ensure meaningful segmentation. Even though there are many ML algorithms for analyzing we use the k means clustering algorithm to assess the clusters. A dataset can be divided into K unique, non-overlapping subsets (clusters) using the clustering algorithm K-means. Assigning data points to clusters based on how similar they are minimizing the sum of squared distances between each point and the centroid of the cluster to which it is assigned. The approach aims to produce compact and well-separated clusters in the feature space by updating cluster centroids until convergence. In data analysis, pattern identification, and machine learning, K-means is frequently used. The modules and libraries used to implement the analysis in python are pandas, NumPy, matplotlib, sklearn and seaborn.

**INTERPRETATION:**

Interpret the clusters and assign labels to each cluster based on common characteristics.

**MARGINAL WORKERS ASSESSMENT:**

Define criteria for assessing marginal workers based on the cluster labels.

**AGE AND INDUSTRY ANALYSIS:**

Analyze the distribution of marginal workers across different age groups and industrial categories.