# Development part-2

**Project name: Assessment of marginal workers in tamilnadu -A socioeconomic analysis.**

## Perform the demographic analysis

Data Collection: Gather the most recent data on marginal workers in Tamil Nadu. This data may come from the Census, Labor Force Surveys, or other official government sources.

Define Marginal Workers: Clearly define what constitutes a marginal worker based on the criteria used in the data source. Marginal workers are often those who are employed for less than six months in a year.

Demographic Variables: Collect demographic information such as age, gender, education level, and location (urban/rural) for the marginal workers in Tamil Nadu.

Data Analysis: Use statistical and data analysis tools to examine the demographic characteristics of marginal workers. This may involve creating tables, charts, and conducting statistical tests to identify patterns and trends.

Interpretation: Analyze the results to draw meaningful conclusions about the demographic composition of marginal workers in Tamil Nadu. Are there any significant variations by age, gender, education, or location?

Report Findings: Summarize your analysis in a report or presentation, highlighting key findings, trends, and any policy implications.

**Exploratory analysis**

To calculate the distribution of marginal workers based on age, industrial category, and sex, we can follow these steps:

Load the necessary data: You need to load data on the labor force, such as age, sex, and occupation, as well as data on the unemployment rate, if available.

Preprocess the data: Clean and preprocess the data to handle missing values, outliers, and format inconsistencies.

Identify marginal workers: Determine the criteria for a marginal worker, such as those with the lowest education levels or lowest employment status. Alternatively, if you have unemployment rate data, you can use it to identify marginal workers.

Calculate the distribution: Once you have identified the marginal workers, you can calculate the distribution based on age, industrial category, and sex using various aggregation and manipulation techniques. For example, you can use the pandas library in Python to calculate the percentage of marginal workers in each age group, industry category, and sex.import pandas as pd

**CODE:**

# Assuming ‘data’ is a DataFrame containing the preprocessed data

# and ‘marginal\_workers’ is a boolean Series indicating whether a worker is marginal

# Calculate the distribution based on age

Data[‘age\_group’] = pd.cut(data[‘age’], bins=[0, 15, 25, 35, 45, 55, 65, 75, 100])

Marginal\_age\_distribution = data.groupby(‘age\_group’)[‘marginal\_workers’].mean()

# Calculate the distribution based on industrial category

Marginal\_industry\_distribution = data.groupby(‘industry’)[‘marginal\_workers’].mean()

# Calculate the distribution based on sex

Marginal\_sex\_distribution = data.groupby(‘sex’)[‘marginal\_workers’].mean()

After running the code, you will have three pandas Series (marginal\_age\_distribution, marginal\_industry\_distribution, and marginal\_sex\_distribution) containing the distribution of marginal workers based on age, industrial category, and sex, respectively.

## Create visualizations using data visualization libraries

1.Start by importing the necessary libraries:

Import matplotlib.pyplot as plt

Import seaborn as sns

2.Assuming you have a pandas DataFrame df containing your data, you can visualize the data using different libraries and functions. For example, to create a histogram of the age column, you can use the following code:

# Create a histogram using Matplotlib

Plt.hist(df[‘age’])

Plt.xlabel(‘Age’)

Plt.ylabel(‘Frequency’)

Plt.title(‘Histogram of Age’)

Plt.show()

3.For creating the bar chart, you can use the following code:

# Create a bar chart using Seaborn

Sns.barplot(x=’sex’, y=’age’, data=df)

Plt.xlabel(‘Sex’)

Plt.ylabel(‘Age’)

Plt.title(‘Bar Chart of Age by Sex’)

Plt.show()

4.To create a scatter plot, you can use the following code:

# Create a scatter plot using Matplotlib

Plt.scatter(df[‘age’], df[‘income’])

Plt.xlabel(‘Age’)

Plt.ylabel(‘Income’)

Plt.title(‘Scatter Plot of Age vs Income’)

Plt.show()

5.To create a box plot, you can use the following code:

# Create a box plot using Seaborn

Sns.boxplot(x=’sex’, y=’income’, data=df)

Plt.xlabel(‘Sex’)

Plt.ylabel(‘Income’)

Plt.title(‘Box Plot of Income by Sex’)

Plt.show()