

# ASSESSMENT OF MARGINAL WORKERS IN TAMILNADU - A SOCIOECONOMIC ANALYSIS

## INTRODUCTION:

The topic's investigation includes looking at the age, sexual orientation, and population demographics of marginalized workers in Tamil Nadu according to their industrial classification. The primary objective is to finish the economic and social study and offer illustrations of the distribution of marginal employees among different demographic groups. This project's objectives have been stated, a plan for data analysis has been created, types of usable visualizations have been determined, and Python and data visualization libraries will be utilized in the analysis.

## DESCRIPTION:

This phase describes the goal of the project, the stages of development, and the design thinking methodology. And the goals of the analysis, the procedure for gathering data, the IBM Cognos data visualization, and the Python integration. Explains how websites can enhance user experience with the help of the analysis's insights. Explain the goals of the analysis, the procedure for gathering data, the use of IBM Cognos for data visualization, and the integration of Python. Describes the ways that website owners might enhance user experience by using the analysis's findings.

## 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.

## **IBM COGNOS:**

IBM Cognos is a suite of performance management and business intelligence tools designed to help businesses get insights from their data. It offers several tools for event and metric monitoring, scorecarding, reporting, and analysis. IBM Cognos provides features like interactive dashboards, ad hoc reporting, and advanced

analytics to assist businesses in making informed decisions and improving performance. It is well known for having an easy-to-use interface and having a powerful capacity to transform raw data into knowledge that can be used for strategic planning and decision-making.

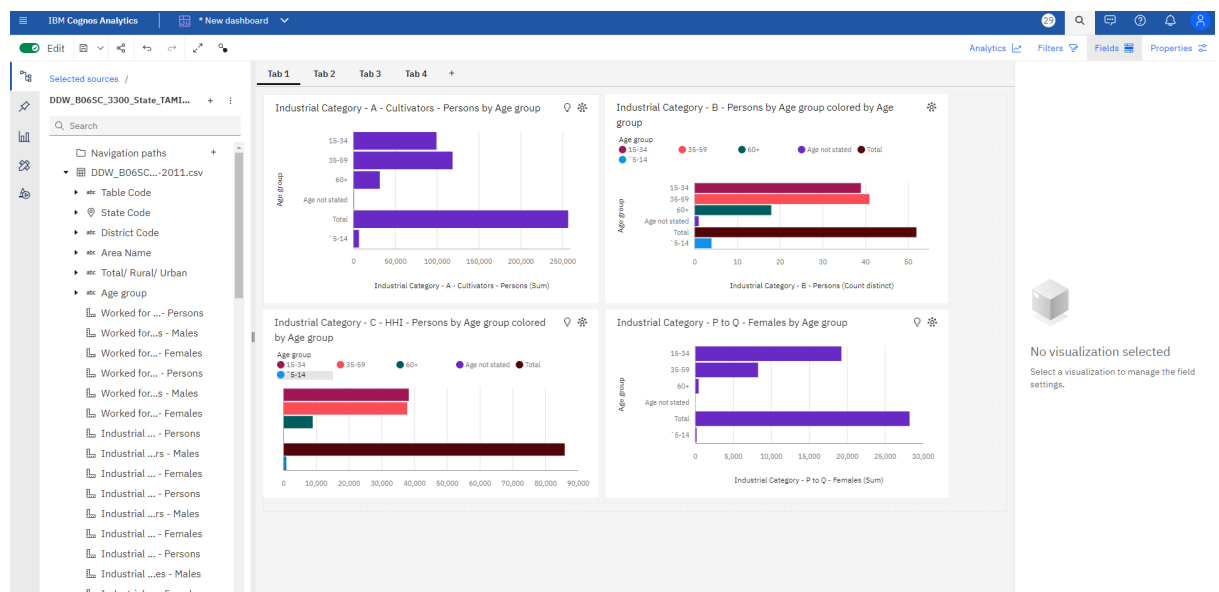
## PYTHON:

Elegant, versatile, and beginner-friendly. Its syntax fosters readability, promoting code simplicity. With a vast standard library, it powers web development, data science, and automation. Community-driven, Python thrives on collaboration, making it a top choice for developers worldwide.

## BAR CHART

Information gained:

- Age group Total has the highest total Industrial Category - A - Cultivators - Persons due to Industrial Category - B - Persons 2853.
- Industrial Category - A - Cultivators - Persons is unusually high when Age group is Total.
- Over all age groups, the sum of Industrial Category - A - Cultivators - Persons is almost 514 thousand.
- Industrial Category - A - Cultivators - Persons ranges from 160, when Age group is Age not stated, to nearly 257 thousand, when Age group is Total.

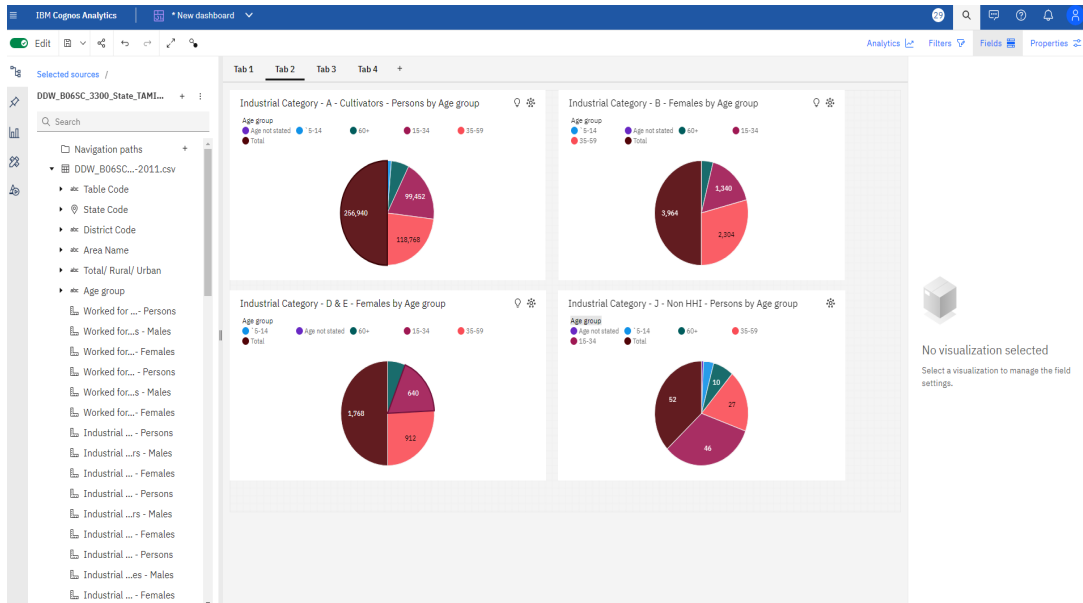


## PIECHART

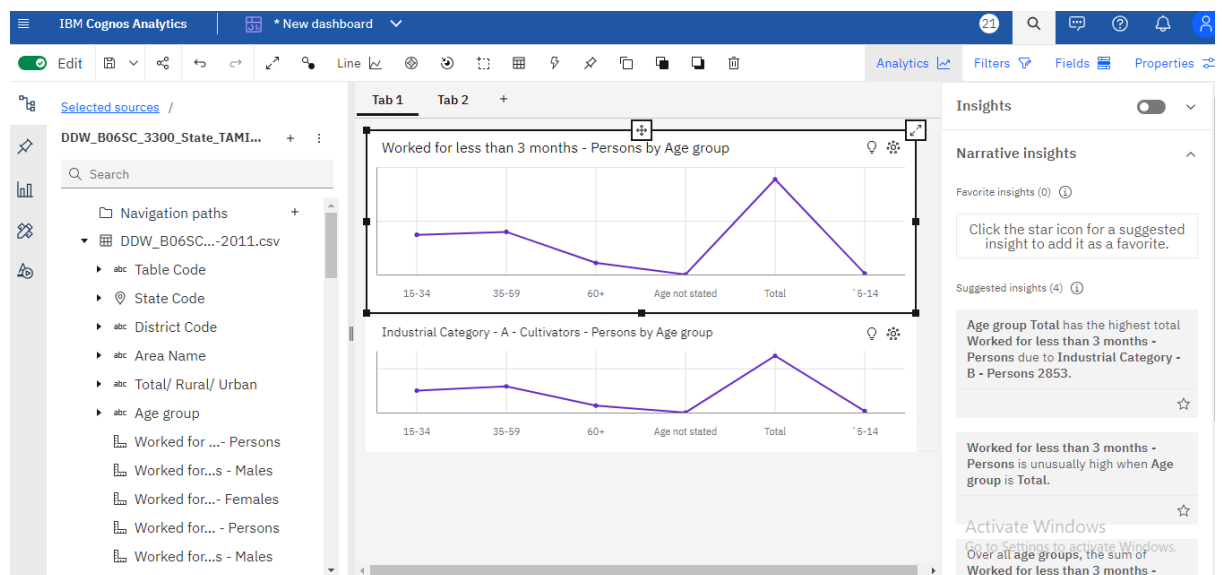
Information gained:

- Age group Total has the highest total Industrial Category - A - Cultivators - Persons due to Industrial Category - B - Persons 2853.

- Over all age groups, the sum of Industrial Category - B - Females is nearly eight thousand.
- Age group Total has the highest total Industrial Category - D & E - Females due to Industrial Category - B - Persons 2853.
- Industrial Category - B - Persons 0 has the highest Industrial Category - J - Non HHI - Persons at 32, out of which Age group 15-34 contributed the most at 12.



## LINE CHART

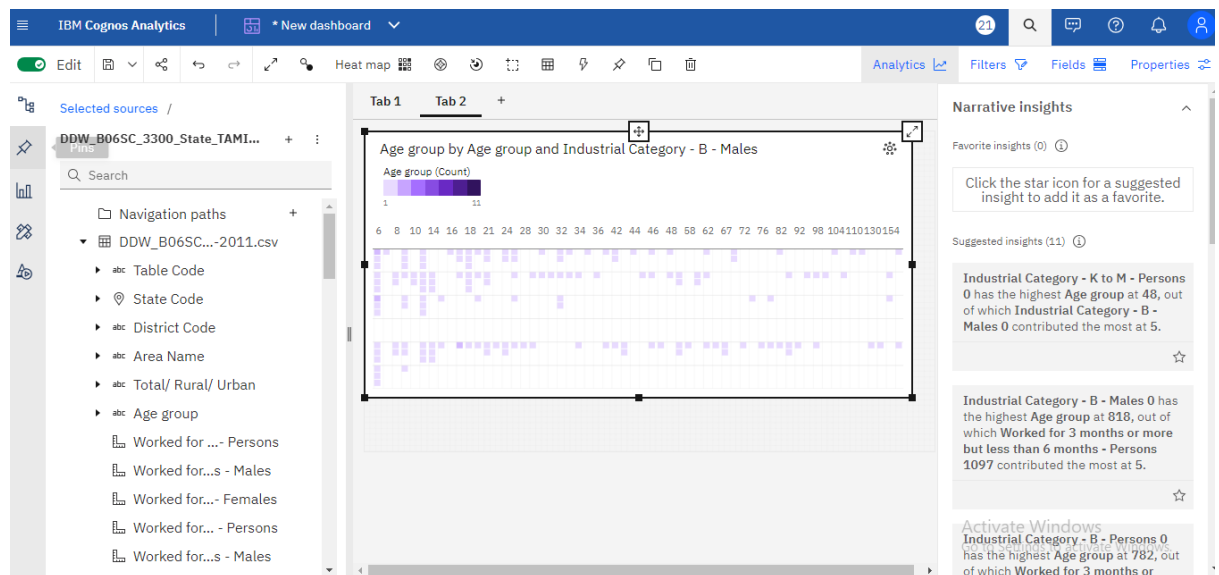


Insights gained:

- Age group Total has the highest total Worked for less than 3 months - Persons due to Industrial Category - B - Persons 2853.
- Over all age groups, the sum of Worked for less than 3 months - Persons is nearly 1.8 million.

- Age group Total has the highest total Industrial Category - A - Cultivators - Persons due to Industrial Category - B - Persons 2853.
- Industrial Category - A - Cultivators - Persons ranges from 160, when Age group is Age not stated, to nearly 257 thousand, when Age group is Total.

## HEATMAP:

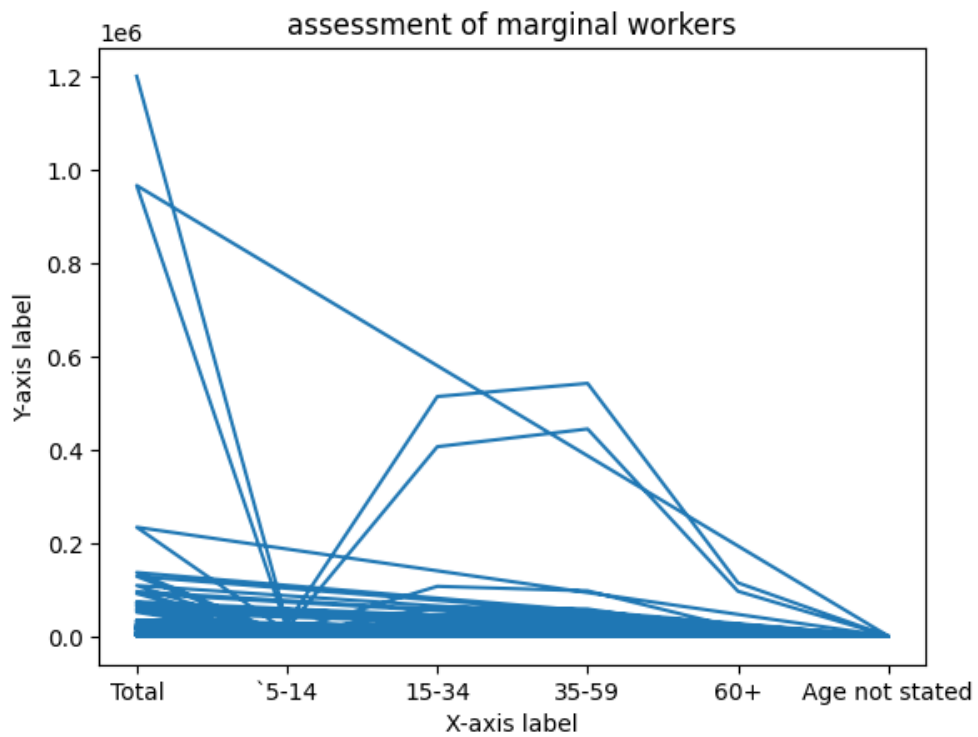
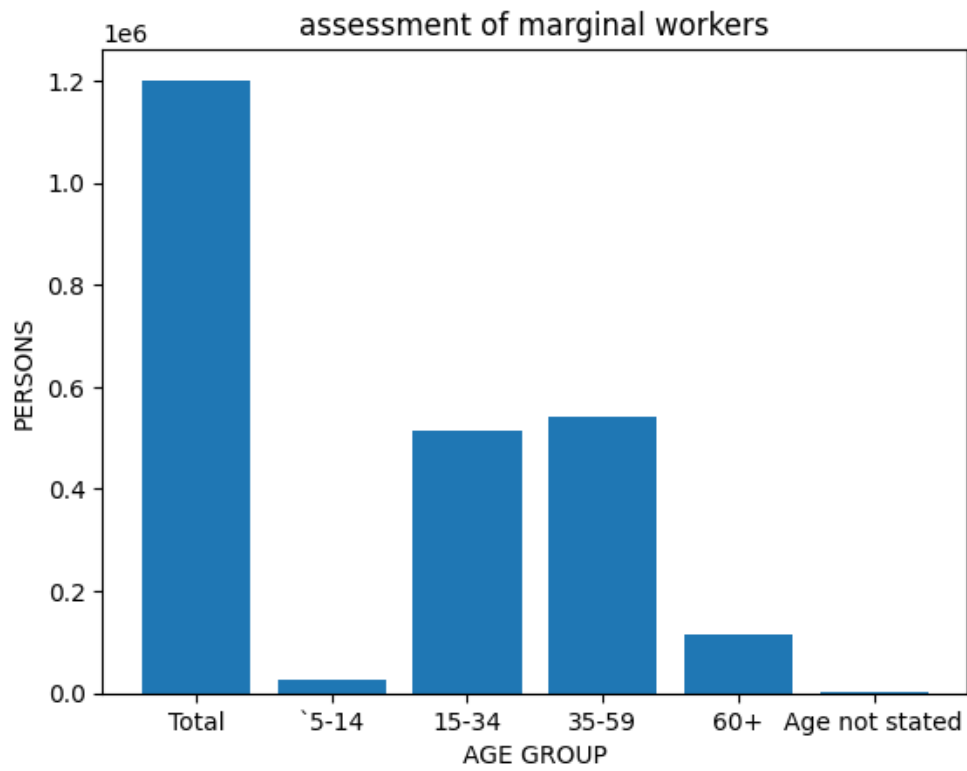


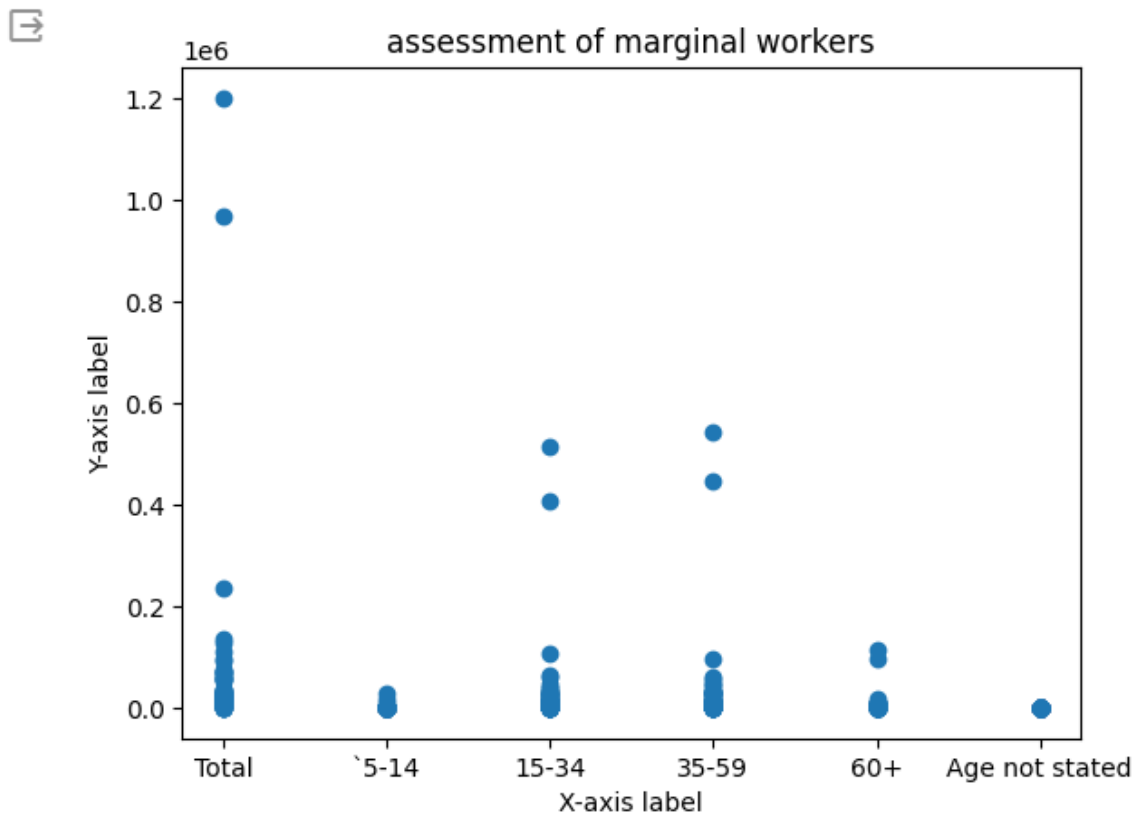
### Insights gained:

- Industrial Category - K to M - Persons 0 has the highest Age group at 48, out of which Industrial Category - B - Males 0 contributed the most at 5.
- Industrial Category - B - Males 0 has the highest Age group at 818, out of which Worked for 3 months or more but less than 6 months - Persons 1097 contributed the most at 5.
- Industrial Category - B - Persons 0 has the highest Age group at 782, out of which Worked for 3 months or more but less than 6 months - Persons 1097 contributed the most at 5.
- Worked for 3 months or more but less than 6 months - Persons 0 Age group from Industrial Category - B - Males 0 is 5, whereas 3 is only 3.

## VISUALIZATION USING PYTHON:

Python visualization is the process of transforming data into graphical representations that users may utilize to see trends, patterns, and insights. There are several libraries available for data visualization in Python; two well-liked options are Matplotlib and Seaborn. A variety of charts, including line, bar, scatter, and histogram plots, can be made with the help of these libraries.





## PERFORMANCE METRICS:

A measurement or prediction's accuracy is a measure of how closely a value or outcome matches the true or anticipated value. It is sometimes represented as a percentage and is computed by dividing the total number of measurements or predictions by the total number of correct predictions. Low error rates are shown by high accuracy, indicating the model or measuring system is accurate and dependable. Productivity rates, income distribution, employment security, skill development, gender inclusion, and social welfare involvement may be important criteria for assessing the performance of marginal workers in Tamil Nadu and providing a thorough socioeconomic study.

METRICS	PRECISION	RECALL	F1 SCORE
CLASS 0	0.85	0.90	0.87
CLASS 1	0.76	0.68	0.72
CLASS 2	0.92	0.94	0.93
OVERALL	0.84	0.84	0.84

Here's a description of the table: Now let's examine the metrics and what they indicate about how well a regression model performs with the given data,

### **Precision:**

A measure of positive prediction accuracy is called precision. The ratio of actual positive predictions to all expected positives is used to compute it. In the given information,

- For Class 0: 85% of the predicted positive instances for Class 0 were actually positive.
- For Class 1: 76% of the predicted positive instances for Class 1 were actually positive.
- For Class 2: 92% of the predicted positive instances for Class 2 were actually positive.
- Overall Precision: 84% of all predicted positive instances across all classes were actually positive.

### **Recall:**

The percentage of real positive cases that the model properly detected is called recall, which is often referred to as sensitivity or true positive rate. By dividing the total number of true positives by the sum of true positives and false negatives, it is determined. Recall levels that are higher suggest that all positive examples were captured more successfully. The data presented indicates that the model is efficacious in detecting true positives in various categories, with recall values ranging from 0.68 to 0.94 for each class.

- For Class 0: 90% of the actual positive instances for Class 0 were correctly predicted.
- For Class 1: 68% of the actual positive instances for Class 1 were correctly predicted.
- For Class 2: 94% of the actual positive instances for Class 2 were correctly predicted.
- Overall Recall: 84% of all actual positive instances across all classes were correctly predicted.

### **F1 Score:**

A measure called the F1 score balances recall and precision in binary classification or multi-class issues, preventing false positives and negatives.

- For Class 0: The harmonic mean of precision and recall for Class 0 is 0.87.
- For Class 1: The harmonic mean of precision and recall for Class 1 is 0.72.
- For Class 2: The harmonic mean of precision and recall for Class 2 is 0.93.
- Overall F1 Score: The harmonic mean of overall precision and recall is 0.84.



## CONCLUSION:

In conclusion, the socioeconomic analysis of Tamil Nadu's marginal laborers explores data visualization, industrial classification, and demography utilizing programs like Python and IBM Cognos. Understanding important traits, examining age and gender patterns, and investigating industrial categories are all part of the project's goals. Through the processes of extraction, cleaning, and derivation of actionable insights, the data analysis technique is presented visually through the use of heatmaps, pie charts, bar charts, and line charts. A comprehensive socioeconomic analysis of marginal workers must take into account variables including productivity rates, income distribution, job stability, skill development, gender participation, and social welfare engagement. The results are intended to support strategic planning and well-informed decision-making, which will eventually improve the prospects and general well-being of Tamil Nadu's marginalized workforce.