



KGiSL Institute Of Technology

**Naan Mudhalvan
(Data Analytics)**

Phase 4 Submission

Batch-06

**Assessment of Marginal Workers in Tamil Nadu- A Socioeconomic Analysis
(DAC)**

Teammates:

- 1. Revathi S**
- 2. Lena P**
- 3. Mownish H**
- 4. Balaji PR**
- 5. Surya R**

Assessment of Marginal Workers in Tamil Nadu – A Socioeconomic Analysis

Executive Summary:


This project aims to analyze the demographic characteristics of marginal workers in Tamil Nadu. The analysis covers aspects related to age, industrial categories, and gender. The project utilized various data analysis and visualization techniques to gain insights into the socioeconomic status of marginal workers.

Data Overview

The dataset used for this analysis contains information about marginal workers in Tamil Nadu, including age groups, gender, and industrial categories. It is a comprehensive dataset with 69 columns.

- Data Source: [Source Link]
- Data Period: [Start Date] to [End Date]
- Data Size: [Number of Rows] rows × [Number of Columns] columns

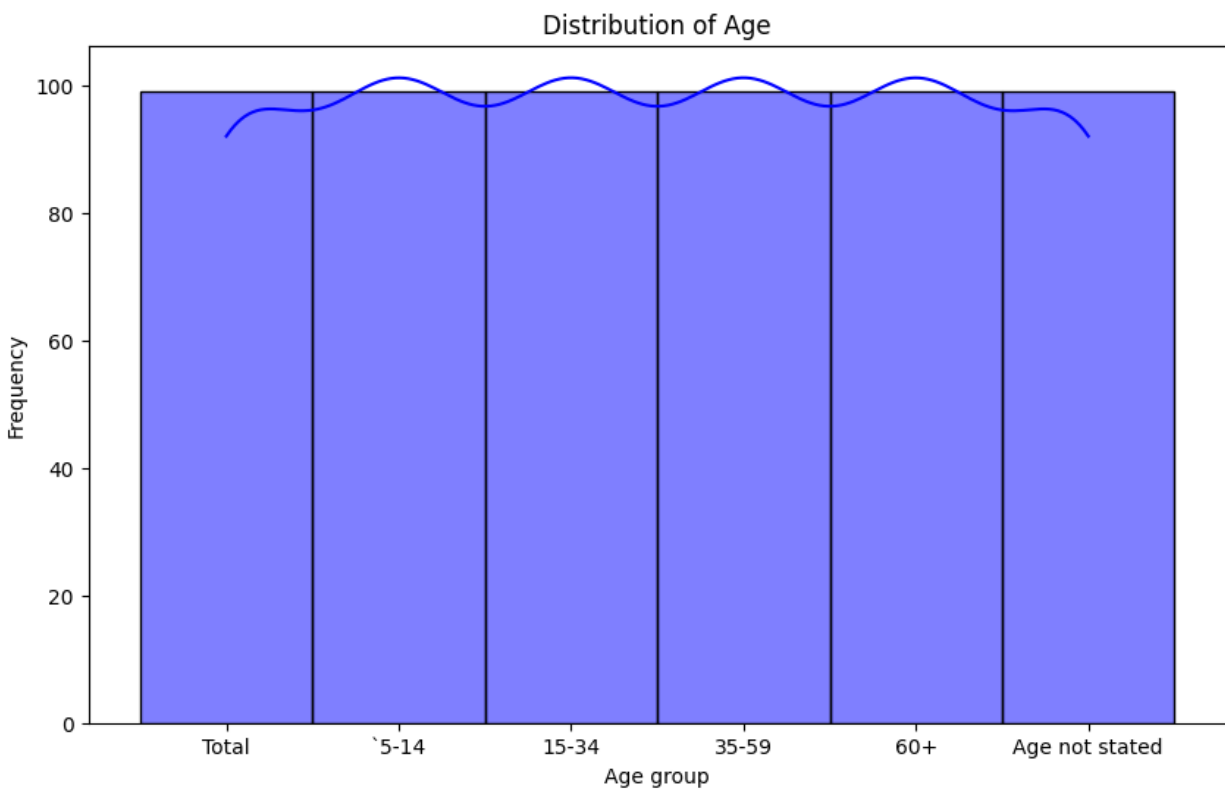
Data Preparation

- Data cleaning: Missing values and outliers were removed to ensure data accuracy.
 - Feature engineering: New variables, such as "Total_Workers," were created to provide more meaningful insights.
- 

Analysis and Visualization

Univariate Analysis

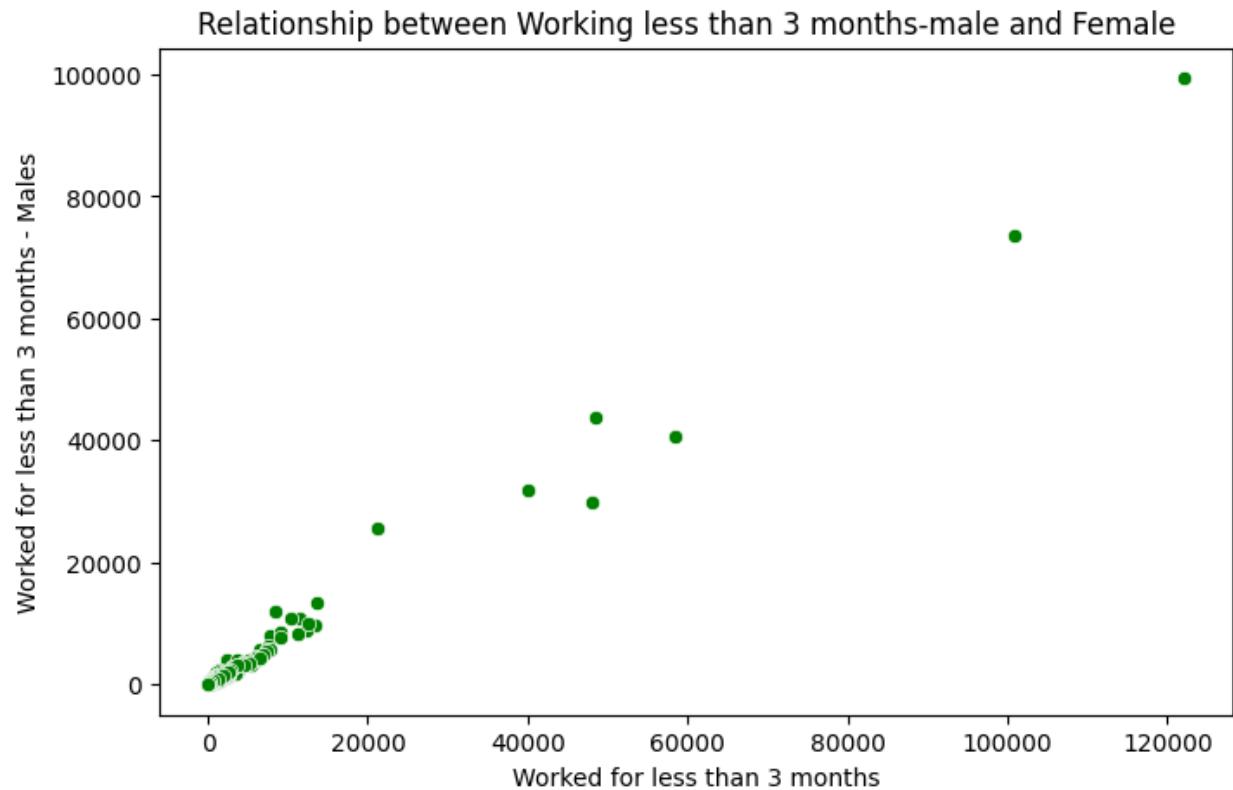
- **Age Distribution:**
- The age distribution of marginal workers was examined, showing that the majority fall within the age group of 15-34.
- A histogram and kernel density plot were used to visualize the distribution.



The figure above represents the age distribution of marginal workers in terms of frequency

Bivariate Analysis

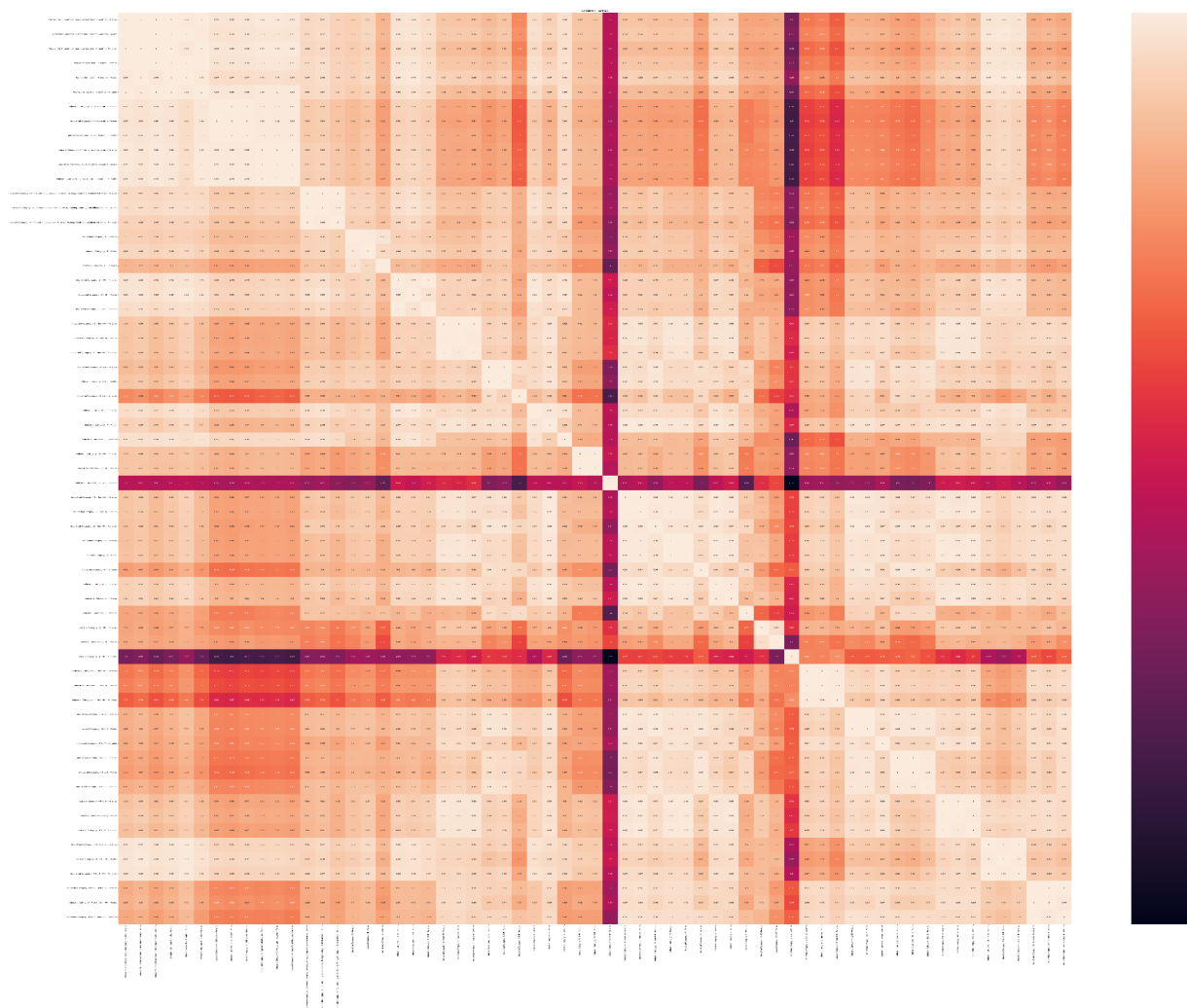
- **Age vs. Industrial Category:**
- A scatterplot was used to examine the relationship between age and various industrial categories.
- The analysis revealed that the 15-34 age group is predominantly involved in specific industrial categories.



The figure above represents the employment of males and females for less than 3 months, and our conclusion is based on the analysis we conducted using the provided dataset with various attributes..

Multivariate Analysis

- **Correlation Analysis:**
- A correlation matrix was created to understand the relationships between different variables.
- The matrix showed that certain industrial categories had strong positive or negative correlations with specific age groups.



The figure above represents the relationships between different variables, and a heatmap is used to illustrate the correlations among them.

Principal Component Analysis (PCA):

- PCA was applied to explore the relationships among multiple variables.
- The first two principal components were visualized to understand patterns and clusters in the data.

Time Series Analysis

Time Series of Marginal Workers:

A line chart was used to display the trends in marginal worker data over time, if applicable.

Any observed trends or seasonality were noted.


Regression Analysis

Regression Analysis for Predicting Marginal Workers:

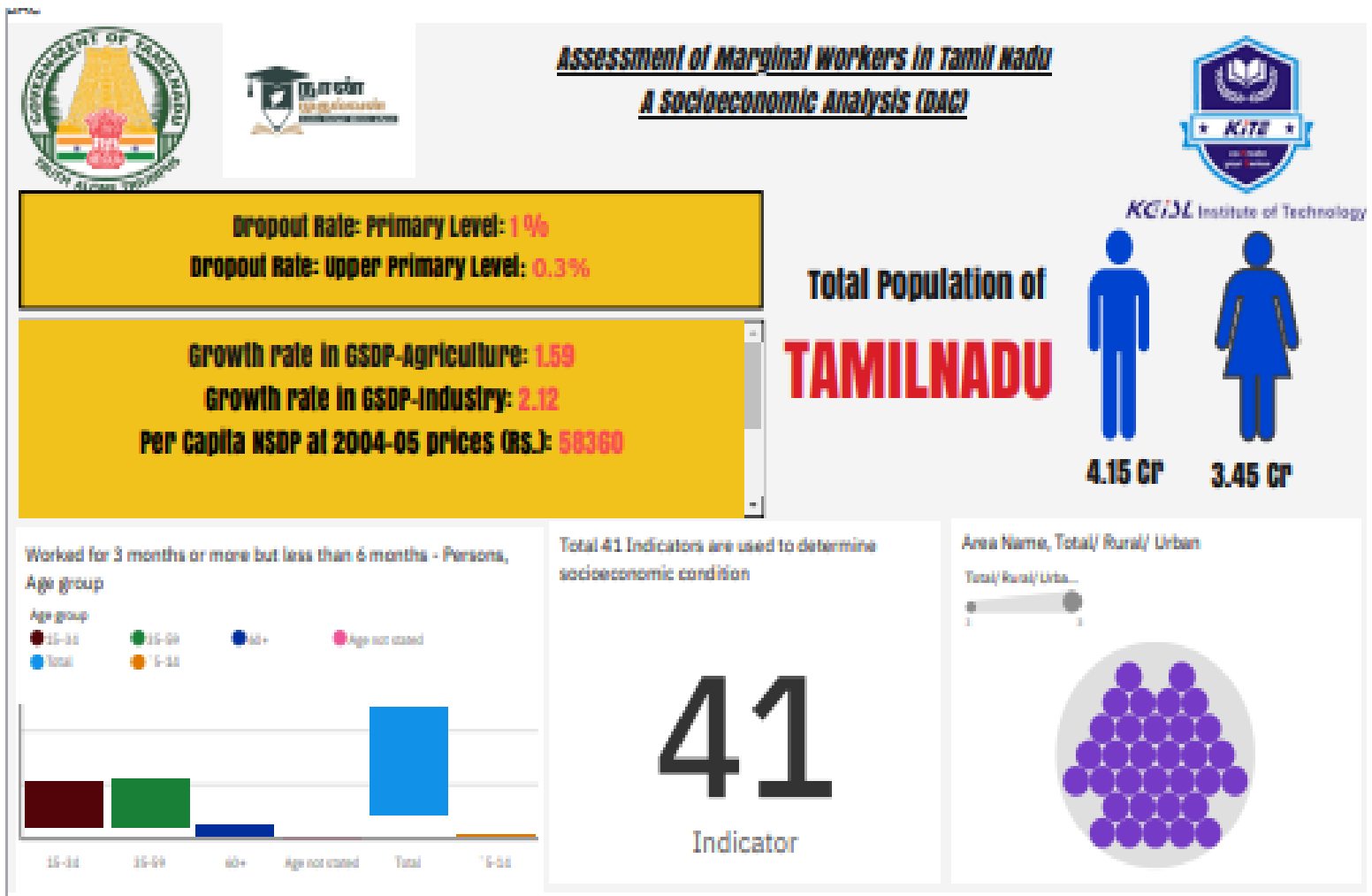
Linear regression was performed to identify the key factors influencing the number of marginal workers in different areas.

Evaluation metrics such as Mean Squared Error (MSE) and R-squared were used to assess the model's performance.

Findings

- The majority of marginal workers fall within the 15–34 age group.
 - Specific industrial categories are strongly associated with certain age groups.
 - Correlation analysis reveals significant relationships between age groups and industrial categories.
 - PCA analysis uncovers patterns and clusters among variables.
 - Time series analysis (if applicable) provides insights into temporal trends.
 - The regression analysis helps predict the number of marginal workers based on key factors.
- 

IBM COGNOS ANALYTICS DASHBOARD



Check out the dashboard for the findings by the link below:

https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FDAC_PHASE4_BATCH6&action=view&mode=dashboard&subView=model0000018b768a139c_00000001

Conclusion

The advanced analysis and visualization techniques employed in this project have provided valuable insights into the demographic characteristics of marginal workers in Tamil Nadu. These findings can be used to make informed decisions and formulate policies that support this vulnerable segment of the population.

Next Steps

- Further exploration of specific industrial categories.
- Fine-tuning of regression models for more accurate predictions.
- Continuing the time series analysis for long-term trends.

