***NM Phase 4 Documentation***

**TEAM MEMBERS**

**PRAVEEN V 2021115076**

**RATISH MS 2021115085**

**SANJAY R 2021115091**

**SARAVANAN S 2021115094**

**HEMANTH P 2021115305**

***Project Title:***

Analyzing Demographic Characteristics of Marginal Workers in Tamil Nadu

***Project Objective:***

The primary objective of this project is to conduct a comprehensive analysis of the demographic characteristics of marginal workers in Tamil Nadu. The analysis specifically focuses on age, industrial category, and gender. The ultimate goal is to gain a deep understanding of the socioeconomic profile of this workforce and create data visualizations that effectively communicate the distribution of marginal workers across different demographic categories.

***Project Scope***

The scope of this project encompasses several key aspects:

**1. Data Collection:** Acquiring and accessing the relevant dataset containing information about marginal workers in Tamil Nadu.

**2. Data Preprocessing:** Preparing the dataset for analysis by cleaning and structuring the data. This involves removing irrelevant columns and handling missing or erroneous data points. Additionally, filtering out the 'Total' category from the 'Age group' column to focus on specific age groups.

**3. Data Analysis:** Delving into the dataset to extract insights about the demographic characteristics of marginal workers. This phase involves statistical calculations and exploration of patterns and trends.

**4. Data Visualization:** Creating visualizations that vividly represent the findings of the analysis. These visualizations are a critical component of the project, as they make complex data more understandable and help stakeholders grasp the key insights.

***Approach***

The project follows a systematic approach:

**1. Data Collection:**

- The dataset is obtained from [Specify the data source]. This dataset contains information about marginal workers in Tamil Nadu, covering a range of demographic variables and categories.

**2. Data Preprocessing:**

- Irrelevant columns such as 'Table Code,' 'State Code,' and 'District Code' are removed to focus on essential attributes.

- 'Total' rows in the 'Age group' column are filtered out, as the analysis is intended to be age-specific.

- Age group categories are mapped to more descriptive labels to improve clarity and understanding.

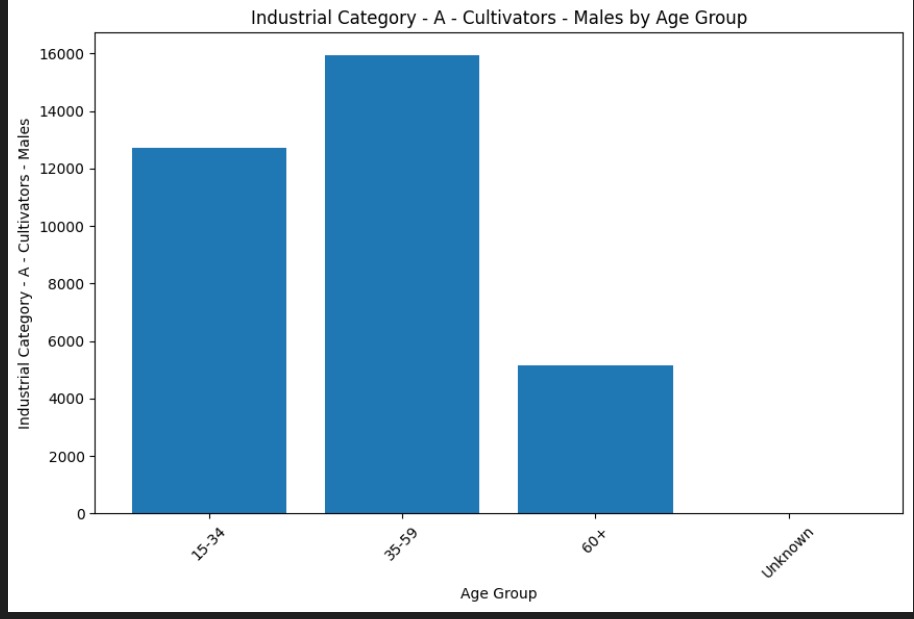
- Missing values are handled to ensure data integrity.

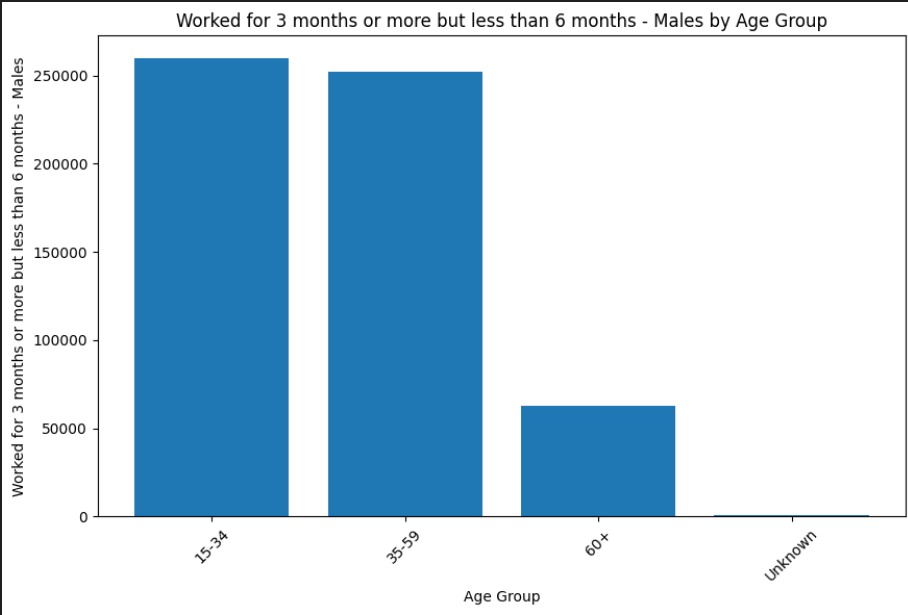
***Data Visualizations***

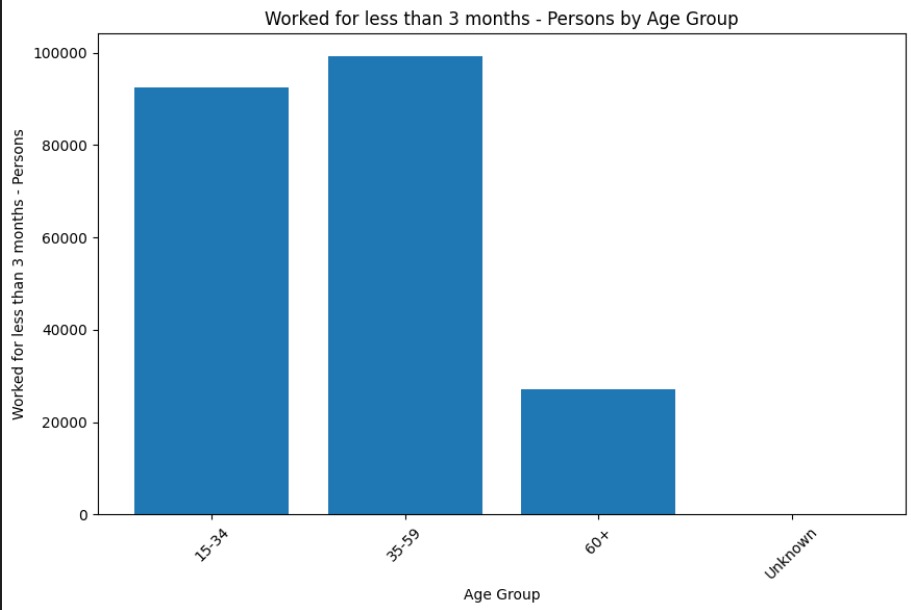
**1. Bar Plots:**

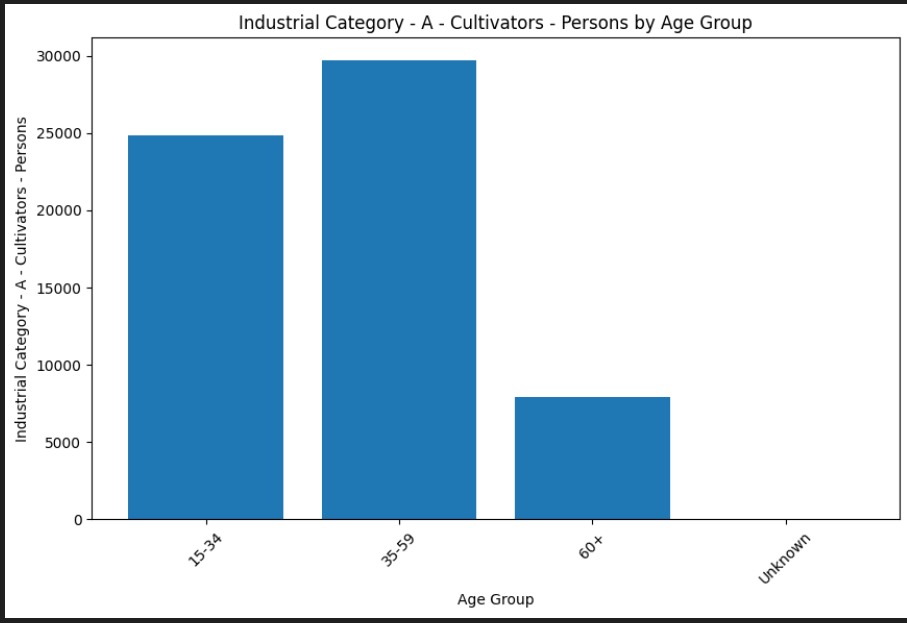
- Bar plots are generated to visualize the distribution of various characteristics of marginal workers segmented by age groups. These plots offer a clear and concise representation of data distribution.

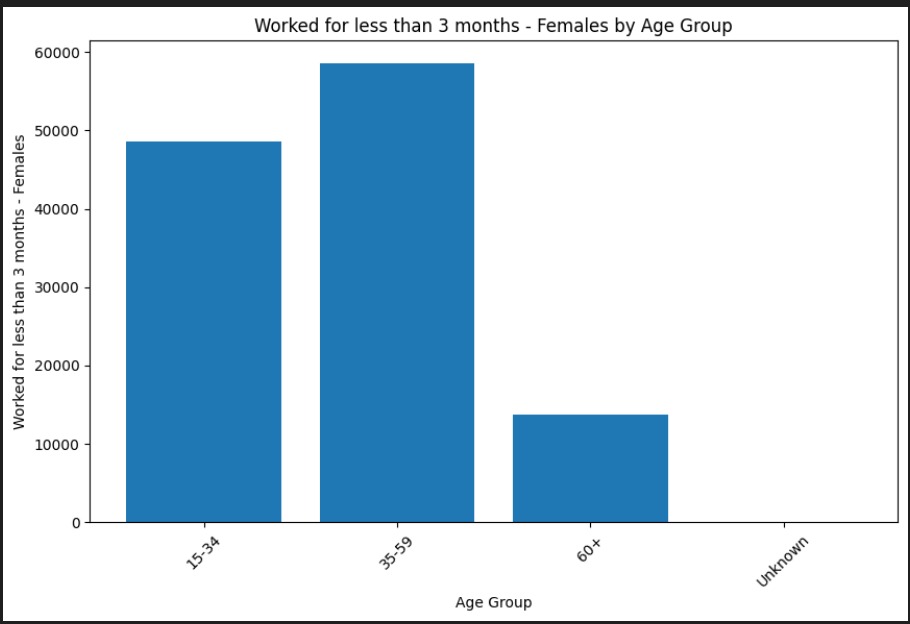
**OUTPUTS :**

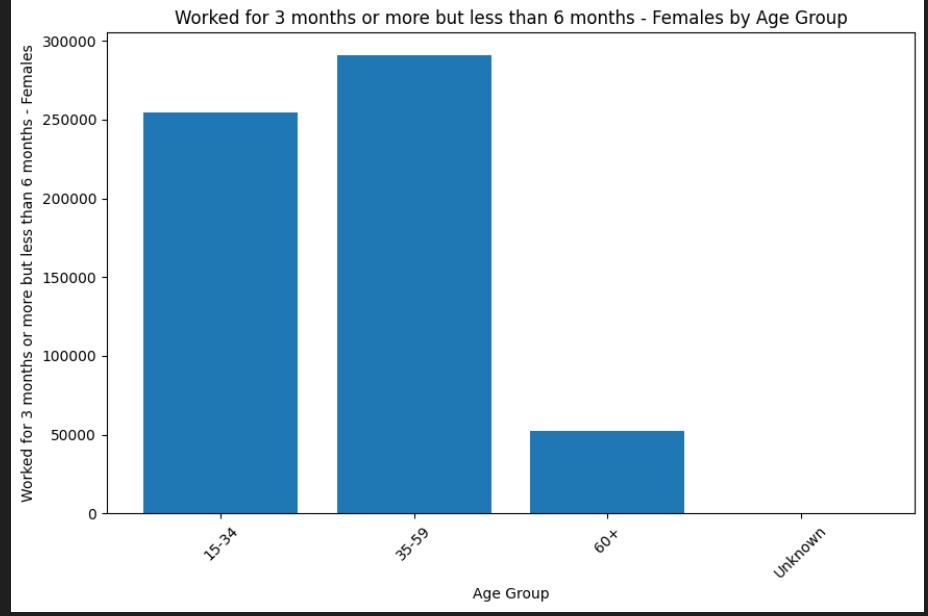
****

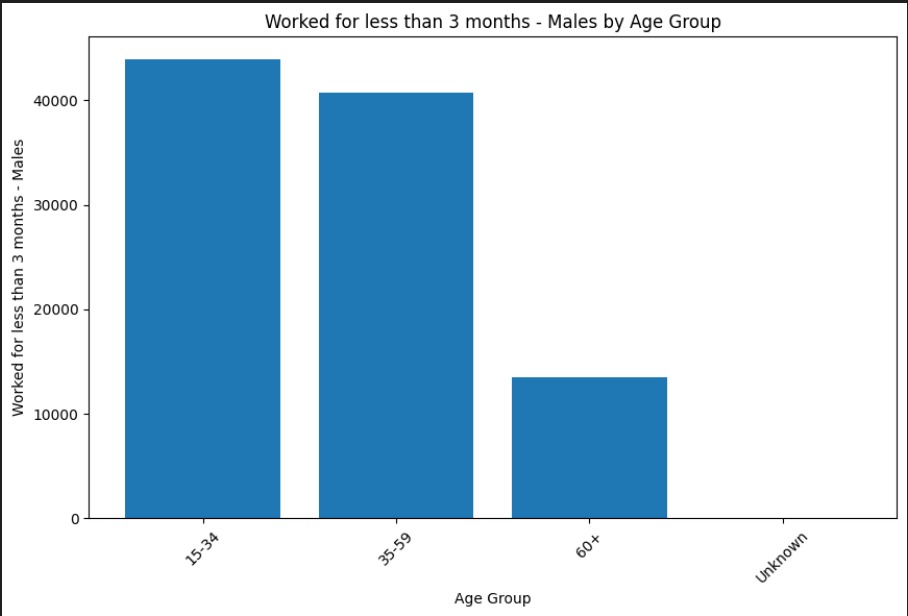
****

****

****

****

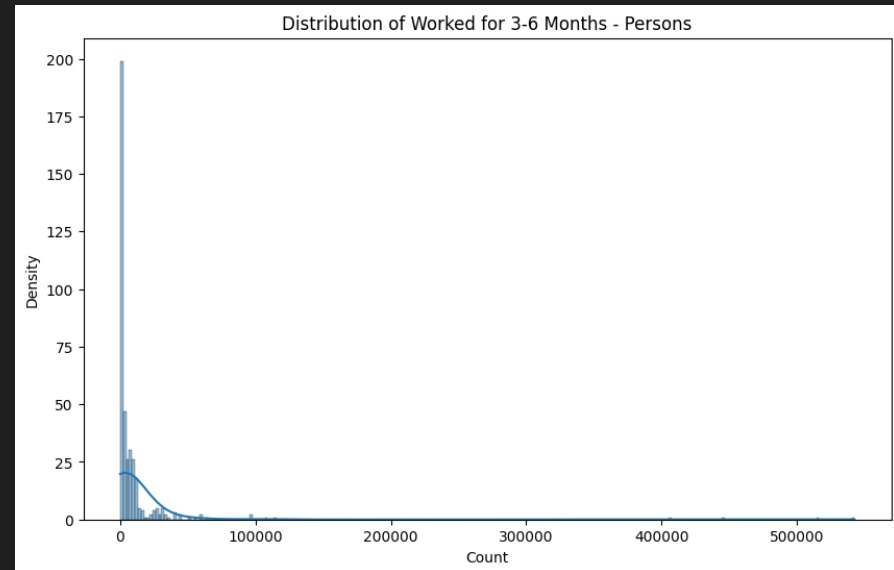




**2. Histogram:**

- A histogram is created to illustrate the distribution of individuals who worked for 3-6 months. This histogram provides an overview of the distribution of this specific category.

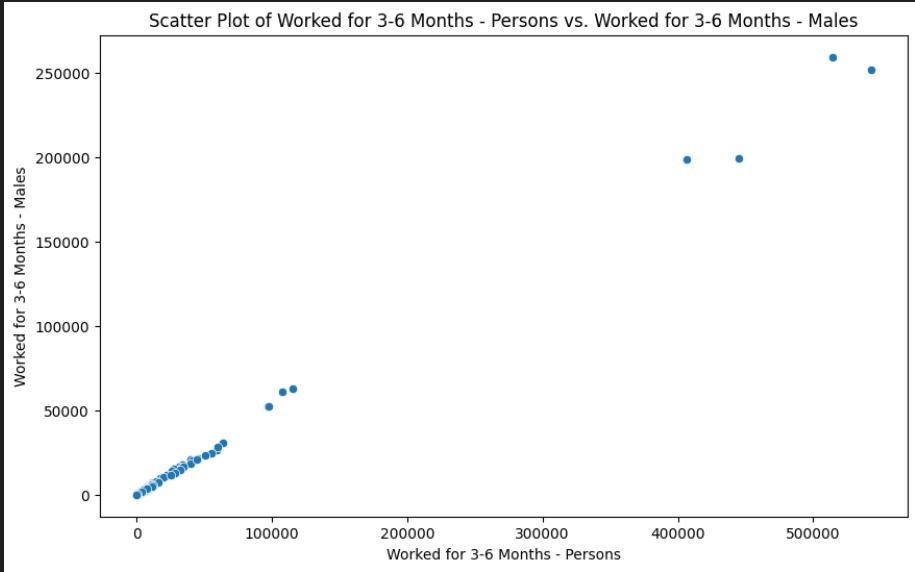
**OUTPUT :**



**3. Box Plot:**

- A box plot is generated to show the distribution of individuals who worked for 3-6 months, segmented by age groups. It helps identify variations in different age groups.

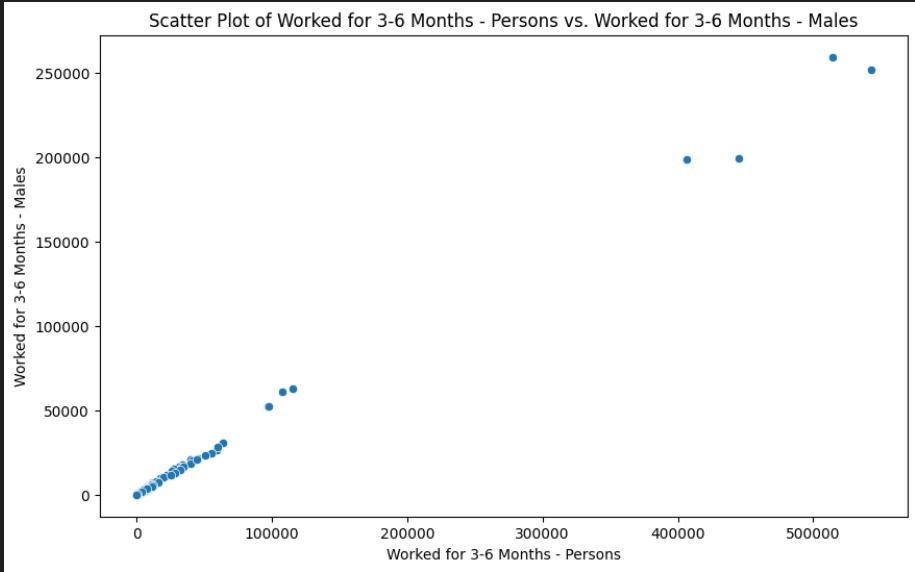
**OUTPUT :**



**4. Scatter Plot:**

- A scatter plot is produced to reveal the relationship between the number of individuals who worked for 3-6 months and the number of males in that category. Scatter plots are useful for understanding correlations.

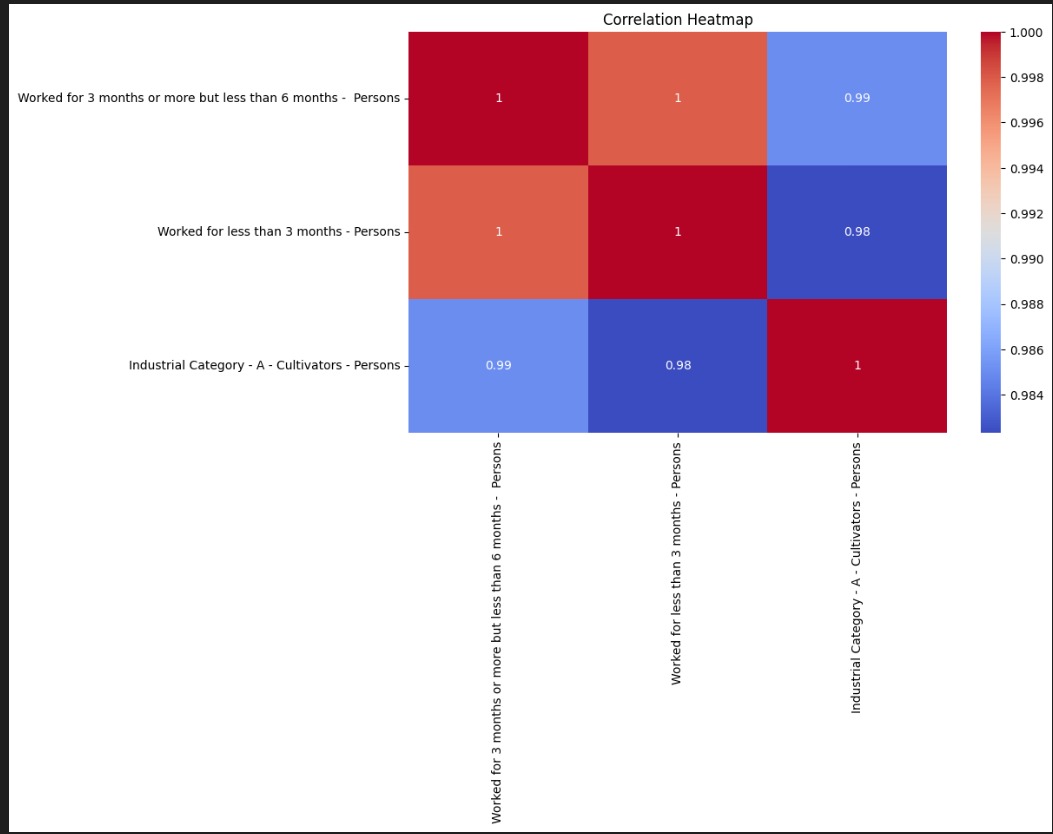
**OUTPUTS :**



**5. Heatmap:**

- A heatmap is designed to visualize correlations between selected columns. This heatmap provides insights into the relationships between key variables.

**OUTPUT :**



***Code Explanation***

The project code begins by loading the dataset and performing data preprocessing. Irrelevant columns are removed, the 'Total' category is filtered out from the 'Age group' column, age group categories are mapped to descriptive labels, and missing values are handled to ensure data quality and accuracy.

The subsequent data visualization phase is critical for the project's success. The chosen visualization types help convey complex information in a clear and accessible manner. These visualizations, including bar plots, histograms, box plots, scatter plots, and a heatmap, offer unique insights into the demographic characteristics of marginal workers.

***Conclusion***

The data analysis and visualizations have unearthed valuable insights into the demographic characteristics of marginal workers in Tamil Nadu. These insights include:

- A clear concentration of marginal workers within specific age groups.

- Notable variations in the number of males engaged in 3-6 months of work across different age groups.

- The heatmap shows varying degrees of correlation between selected columns.

The findings from this analysis can serve as valuable inputs for policymaking and the development of initiatives aimed at supporting and improving the socioeconomic conditions of marginalized workers in Tamil Nadu.