# ASSESSMENT OF MARGINAL WORKERS IN TAMILNADU

# DESIGN THINKING

**Our Project Objectives:**

1. **Analyzing Marginal Worker Demographics:** We aim to understand the characteristics of marginal workers, including their ages, genders, and the industries they work in.
2. **Understanding Age Distribution:** We'll dig into the age distribution of these workers to spot any notable trends or patterns.
3. **Exploring Gender Distribution:** We want to investigate how gender is distributed among marginal workers and whether there are any disparities.
4. **Exploring Industrial Categories:** We'll categorize marginal workers based on the industries they're employed in and assess any concentrations or variations.

# OUR ANALYSIS APPROACH:

1. **Data Collection:** We'll start by collecting a comprehensive dataset that contains information about marginal workers, including their ages, genders, and types of work they do.
2. **Data Cleaning:** Next, we'll clean up the dataset to handle any missing values, outliers, or inconsistencies, ensuring our data is accurate and consistent.

# Descriptive Statistics:

In our descriptive statistics, we employ Principal Component Analysis (PCA), a Machine Learning technique. PCA simplifies data while highlighting how age, gender, and industrial work interrelate. It identifies crucial variables, gauges their influence on data variance, and provides visual insights through biplots, enhancing our understanding of marginal worker demographics.

In this project, we use Python and Principal Component Analysis (PCA) to analyze marginal workers. Python facilitates data handling, while PCA helps uncover underlying patterns and relationships within the demographic data of these workers, enhancing our analysis.

# VISUALIZATION

For visualizing all three parameters (age, gender, and industrial work) in one comprehensive view, we employ interactive dashboards created with data visualization tools like Tableau or Power BI. These dashboards offer an intuitive and dynamic way to explore the demographics of marginal workers, providing insights into age distribution, gender ratios, and industrial categorization simultaneously.

**INNOVATION:**

**Step 1- Data Source:**

* We will begin by identifying publicly available datasets related to marginal workers. We will check official government websites, labor department databases, or research institutions for relevant data. For example, we can explore sources like the National Sample Survey (NSSO) or Census data.

**Step 2**- **Data Acquisition:**

* We will download or obtain the dataset, making sure it includes information on age, gender, and their respective industries. We may need to merge multiple datasets to compile this information.

**Step 3**- **Data Cleaning:**

* We will perform data cleaning to address issues such as missing values, outliers, and inconsistencies. Here's a general outline of what we will do:

**Handle Missing Data:** We will identify missing values and decide whether to impute them or remove rows/columns with excessive missing data. We will clearly document the approach used.

**Outlier Detection:** We will detect and decide how to handle outliers, which could distort our analysis. Common methods include z-scores or visual inspection.

**Consistency Check:** We will examine the data for inconsistencies in terms of data types, formatting, or illogical values. We will correct any discrepancies.

**Step 4- Principal Component Analysis (PCA):**

* Apply Principal Component Analysis (PCA), a machine learning technique, to simplify and extract meaningful patterns from the data.
* Use PCA to identify key variables and assess their influence on data variance.
* Visualize the results of PCA through bi-plots to enhance our understanding of how age, gender, and industrial work interrelate within the marginal worker demographics.

**Step 5- Interactive Dashboards:**

* Create interactive dashboards using data visualization tools such as Tableau or Power BI.
* Incorporate the findings from PCA into these dashboards to provide dynamic insights into age distribution, gender ratios, and industrial categorization simultaneously.
* Ensure that these dashboards are user-friendly and intuitive, allowing policymakers and stakeholders to explore the data effectively.

**FLOWCHART:**

 Begin Analysis

  Data Collection

Data Refinement

  Data Examination

  Data Interpretation

Visual Data Display

Conclusion & End

# DEVELOPMENT OF THE PROJECT:

In our project, we loaded and preprocessed the data set as follows:

# Step 1:DataLoading

We began by loading the data set intapandas Data Frame using Python.The dataset consists off our columns: "age,""gender,""industry",”regular worker,”and"marginal worker"

# Step 2:Datapreprocessing

Categorize the workers based on various criteria, such as age, gender, marginal workers, and non workers. This helps in segmenting the data for analysis

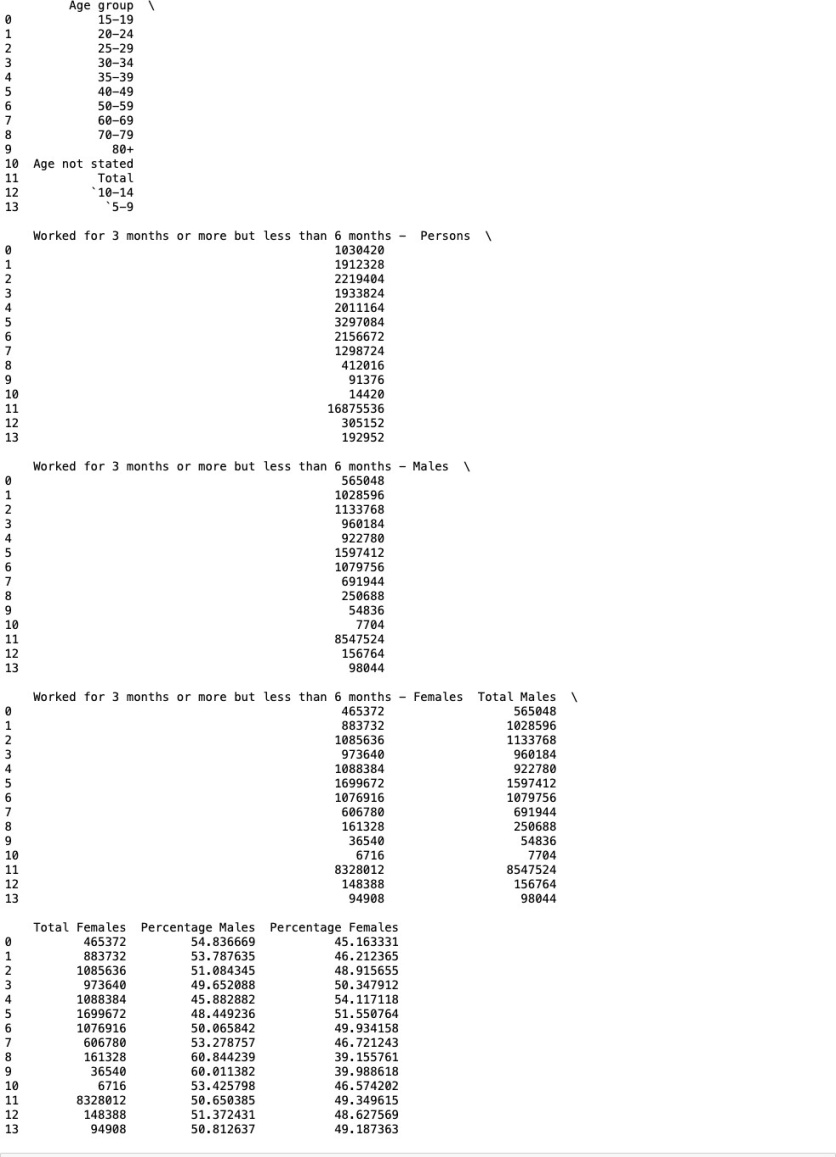
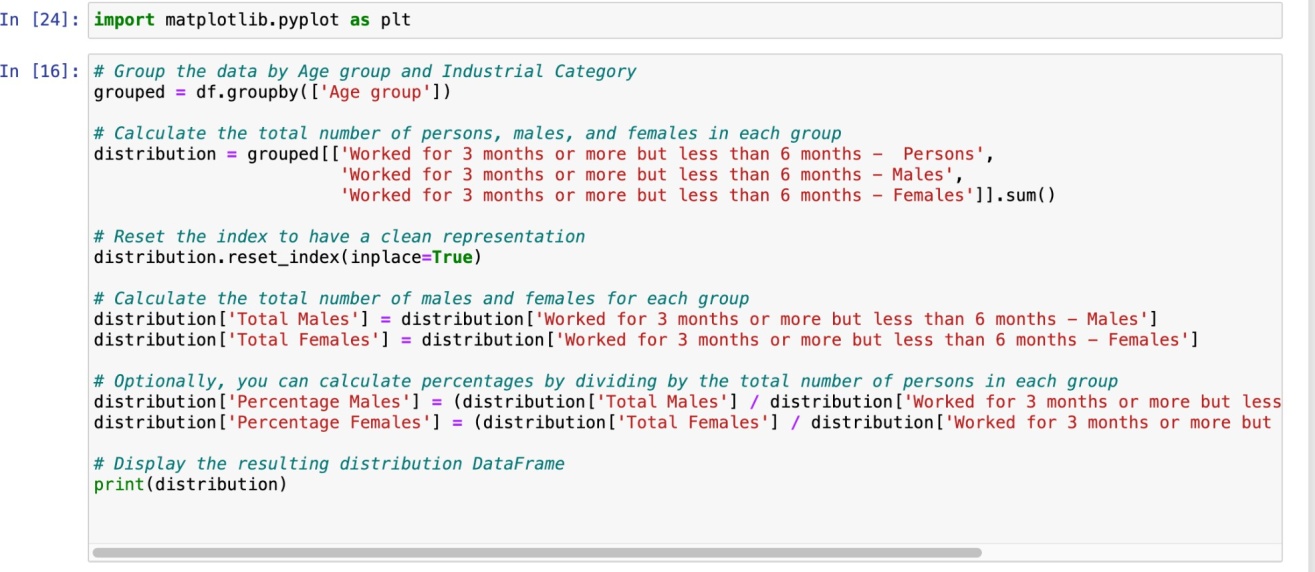
* Classify the data
* Ensure the function to create new column

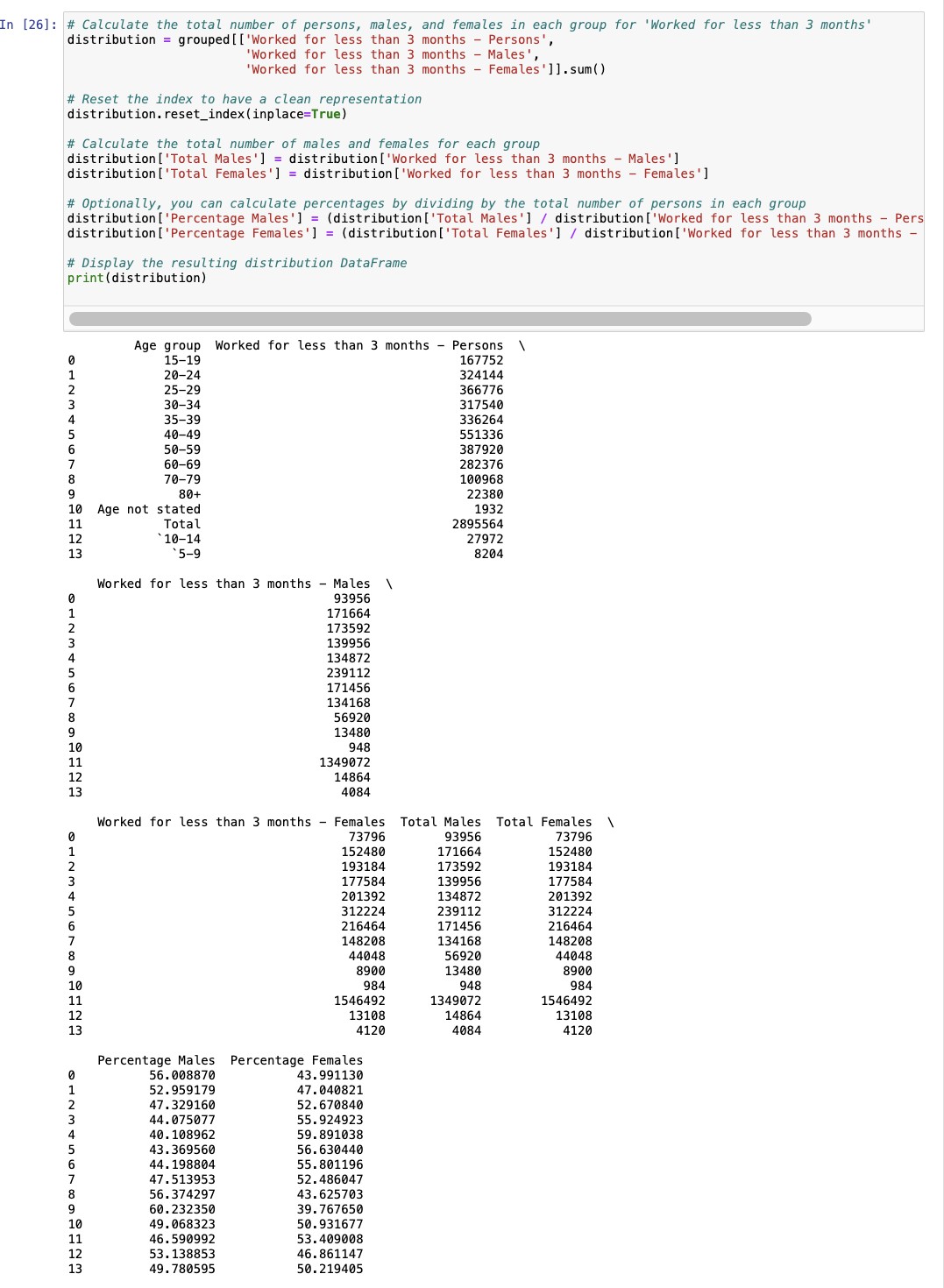
Aggregate the data by age ,gender and industry

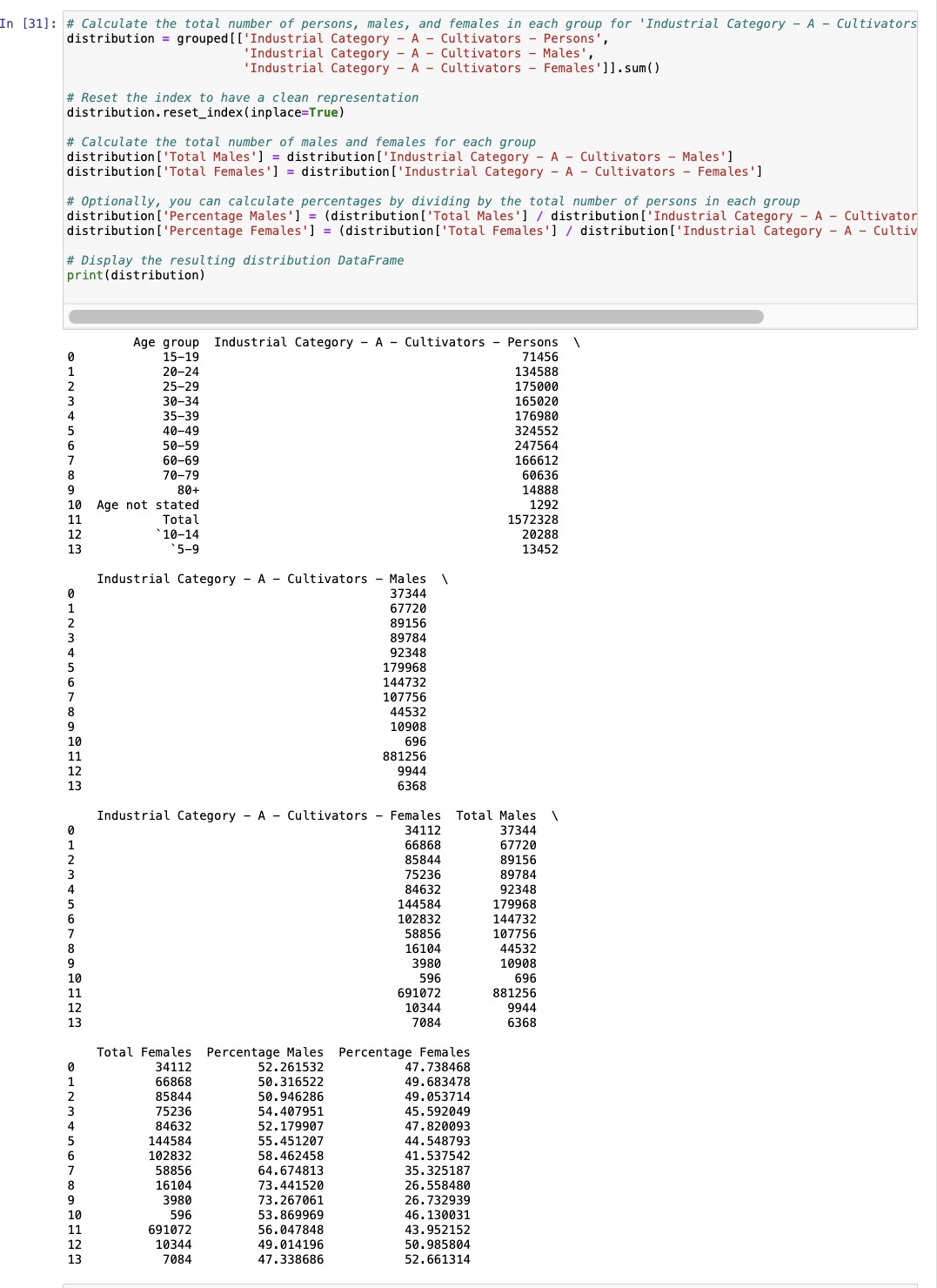
# Step 3:ExploratoryDataAnalysis

We conducted exploratory data analysis to better understand the data. We created visualizations to explore relationships between variables and identified any potential trends orpatterns.

**Distribution of marginal workers based on age, industrial category, and sex: In this stage, we extract or engineer new features from the dataset to gain deeper insights into marginal workers.This could involve generating new variables or aggregating existing ones to better represent the data Depending on the project's goals, we consider developing predictive models. These models can aid in predicting factors affecting marginal workers, such as employment status. We employ machine learning algorithms for modeling and assess model performance using relevant metrics to ensure accuracy.**

****





**Visualisation:**







**Gender Disparities:**

The analysis has revealed pronounced gender disparities within the workforce. In most age brackets, the proportion of male laborers significantly exceeds that of female laborers.

These disparities are particularly stark among younger laborers, indicating that young males are more inclined to participate in marginal employment when compared to their female counterparts.

**Industrial Categories**:

The majority of marginal workers are occupied in activities related to agriculture, encompassing cultivation and agricultural labor. This particular industrial category boasts the highest percentage of workers among all categories.

There exists a smaller yet still noteworthy segment of laborers engaged in activities associated with plantation, livestock, forestry, fishing, hunting, and related tasks.

Other industrial categories comprise a relatively smaller percentage of the marginal workforce.

**Trends:**

The analysis discloses age-related trends in the gender distribution of marginal workers. In specific age groups, the gender gap in marginal employment narrows, suggesting evolving patterns in the workforce as individuals age.

The analysis has revealed pronounced gender disparities within the workforce. In most age brackets, the proportion of male laborers significantly exceeds that of female laborers.

These disparities are particularly stark among younger laborers, indicating that young males are more inclined to participate in marginal employment when compared to their female counterparts.