**TN MARGINAL WORKERS ASSESSMENT**

**Project's Objectives**

The project aims to conduct a comprehensive socioeconomic analysis of marginal workers in Tamil Nadu, focusing on their age, industrial category, and gender. Utilizing Python and data visualization libraries, the project's objectives include defining clear research goals, designing a robust analysis methodology, selecting suitable visualization techniques, and ultimately creating visual representations. Through this, the project seeks to gain nuanced insights into the demographic composition of marginal workers, enabling informed decision-making for targeted policy interventions and social welfare programs.

**Analysis Approach**

The analysis approach for Tamil Nadu marginal workers data involves several steps.

* These include understanding the dataset, cleaning it, performing exploratory data analysis, enhancing feature engineering, and using visualization techniques.
* The data is then analysed to understand distributions, correlations, and trends, and to compute summary statistics.
* Data visualization is used to represent the data visually, and in-depth analysis is conducted to identify predominant age groups, gender disparities, and employment preferences. Statistical analysis is conducted to determine significant differences in workforce distribution among age groups, categories, and genders.
* Visual representations are provided to communicate the findings effectively.
* This approach provides a comprehensive understanding of the demographic characteristics of marginal workers in Tamil Nadu, leading to valuable insights and data-driven recommendations for policy-making and intervention strategies.

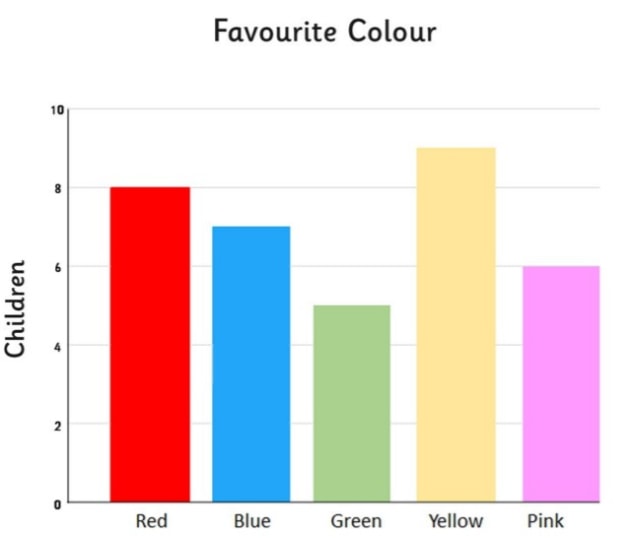
**Visualization Types**

Bar charts are a valuable tool for visualizing marginal labour statistics in Tamil Nadu. They are ideal for comparing disaggregated data, such as technology groups, age groups, or gender, which are important for understanding the demographic composition of an area Bar charts provide clear visual comparisons visibility across teams, enabling stakeholders to quickly identify patterns, anomalies and trends .

Multiple variables can be represented simultaneously, such as age and technology groups, and are easily understood by policymakers, researchers, and the general public they facilitate comparisons within and across groups, and provides complete insight. Bar charts also support trend analysis, allowing analysts to see changes in population distributions over time.

They can also highlight socioeconomic inequalities, such as gender imbalances in specific industries or age-related labour force participation. In conclusion, bar charts are a means of visualizing the profile of marginalized workers in Tamil Nadu, while facilitating complex demographic segmentation, and providing participation are able to quickly grasp key concepts and make informed decisions.

Example of bar chart



**Code Implementation.**

In Python, several modules and libraries are commonly used for data cleansing, data pre-processing, and data analysis tasks. Here are some essential modules and their explanations:

**1. Pandas:**

Pandas is a powerful library for data manipulation and analysis, offering functions for data cleaning, pre-processing, and analysis. It provides data structures like Data Frames and Series, making it easy to handle structured data.

**2. Numpy:**

NumPy is a Python package for data pre-processing and analysis, supporting large, multi-dimensional arrays and matrices. It provides high-level mathematical functions for operations and transformations on data, making it essential for numerical computing.

**3. Matplotlib:**

Matplotlib is a Python-based data visualization library that enables the creation of static, interactive, and animated plots and charts, including line charts, bar charts, histograms, and scatter plots, making it highly customizable.

**4. Seaborn**

Seaborn is a data analysis tool built on Matplotlib, designed for creating informative and visually appealing statistical graphics. It simplifies complex visualizations like heatmaps, violin plots, and pair plots, and integrates well with Pandas Data Frames.

Some of the steps that have to be followed to do the TN marginal worker assessment analysis using the **marginal worker** dataset it consist of

Table Code', 'State Code', 'District Code', 'Area Name’, ‘Total/ Rural/ Urban', 'Age group', 'Worked for 3 months or more but less than 6 months - Persons', 'Worked for 3 months or more but less than 6 months - Males’, ‘Worked for 3 months or more but less than 6 months - Females’, ‘Worked for less than 3 months - Persons’, ‘Worked for less than 3 months - Males’, ‘Worked for less than 3 months - Females’, ‘Industrial Category - A - Cultivators - Persons’, ‘Industrial Category - A - Cultivators - Males’, ‘Industrial Category - A - Cultivators - Females’, ‘Industrial Category - A - Agricultural labourers - Persons’, ‘Industrial Category - A - Agricultural labourers - Males’, ‘Industrial Category - A - Agricultural labourers - Females’, ‘Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons’, ‘Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Males’, ‘Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Females’, ‘Industrial Category - B - Persons', 'Industrial Category - B - Males’, ‘Industrial Category - B - Females’, ‘Industrial Category - C - HHI - Persons’, Industrial Category - C - HHI - Males’, ‘Industrial Category - C - HHI - Females’, ‘Industrial Category - C - Non HHI - Persons', 'Industrial Category - C - Non HHI - Males’, ‘Industrial Category - C - Non HHI - Females', 'Industrial Category - D & E - Persons', 'Industrial Category - D & E - Males’, ‘Industrial Category - D & E - Females’, ‘Industrial Category - F - Persons', 'Industrial Category - F - Males’, ‘Industrial Category -Females’, ‘Industrial Category - G - HHI - Persons', 'Industrial Category - G - HHI - Males', 'Industrial Category - G - HHI - Females’, ‘Industrial Category - G - Non HHI - Persons’, ‘Industrial Category - G - Non HHI - Males’, ‘Industrial Category - G - Non HHI - females’, ‘Industrial Category - H - Persons', 'Industrial Category - H - Males', 'Industrial Category - H - females’, ‘Industrial Category - I - Persons', 'Industrial Category - I - Males’, ‘Industrial Category - I - females’, ‘Industrial Category - J - HHI - Persons’, ‘Industrial Category - J - HHI - Males’, ‘Industrial Category - J - HHI - Females’, ‘Industrial Category - J - Non HHI - Persons’, ‘Industrial Category - J - Non HHI - Males', 'Industrial Category - J - Non HHI - Females’, ‘Industrial Category - K to M - Persons’, ‘Industrial Category - K to M - Males’, ‘Industrial Category - K to M - Females', 'Industrial Category - N to O - Persons’, ‘Industrial Category - N to O - Males’, ‘Industrial Category - N to O - Females’, ‘Industrial Category - P to Q - Persons’, ‘Industrial Category - P to Q - Males’, ‘Industrial Category - P to Q - Females’, ‘Industrial Category - R to U - HHI - Persons’, ‘Industrial Category - R to U - HHI - Males’, ‘Industrial Category - R to U - HHI - Females’, ‘Industrial Category - R to U - Non HHI - Persons’, ‘Industrial Category - R to U - Non HHI - Males’, ‘Industrial Category - R to U - Non HHI - Females'

\*Note:

(For analysis, we have to install the Jupyter notebook. In this environment, we can execute the code to do analysis. Why did I choose a Jupyter notebook? Because Jupyter notebooks support code, text, and images, they provide a flexible and ideal way to manage the iterative exploration process common to data analytics and machine learning.

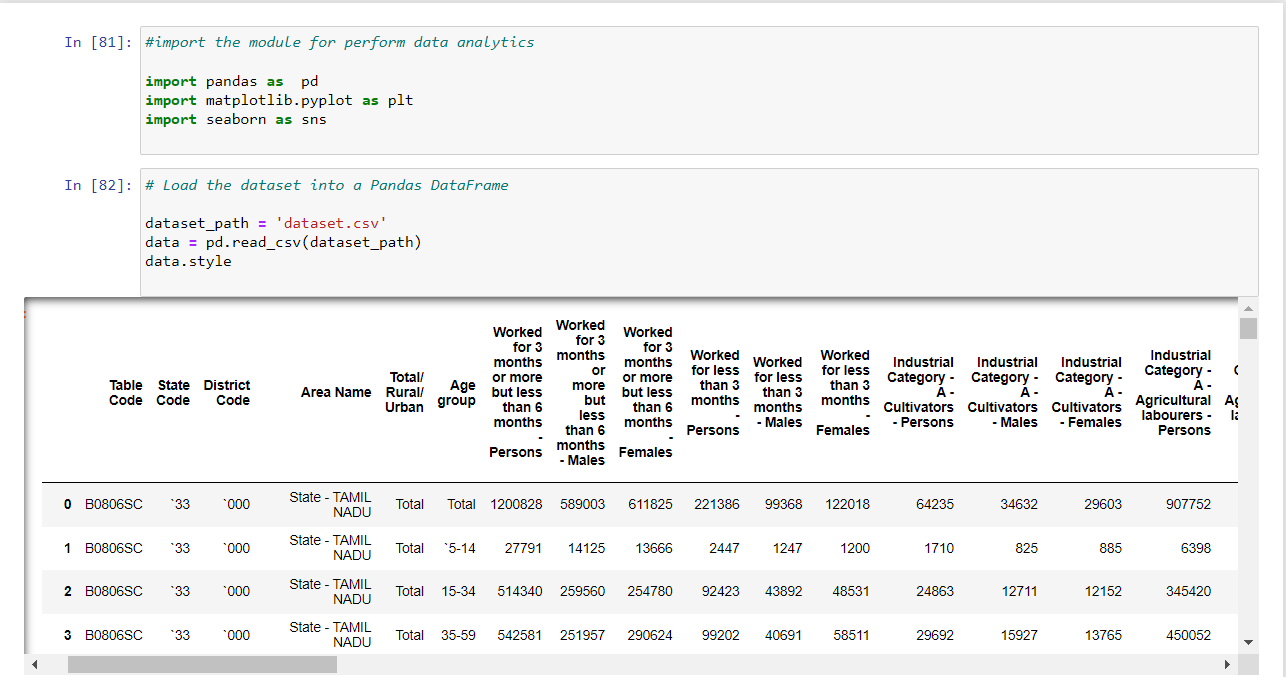
**PROCEDURE:**

**PART 1:**

We have to do analysis first we have to clean and pre-processing the data so we use the pandas for manipulate the data.

**STEP 1:**

We have to load the dataset into the data frame for access the data

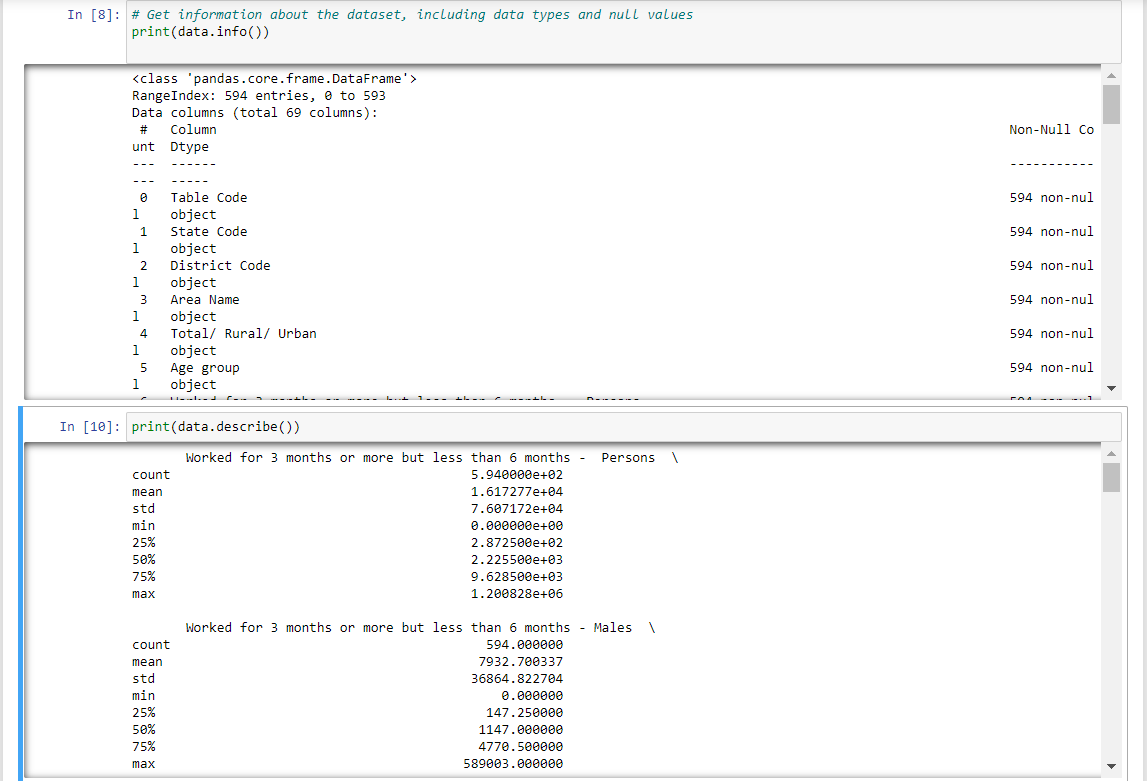


The line “read\_csv” is used to load the dataset that we are going to perform the analysis why we use the csv means CSV files are widely used for storing and exchanging tabular data, making them an excellent fit for importing and analysing data in Jupyter Notebook.

The .style is used to display all the data in the dataset in the table manner

**STEP 2:**

After that we have to the information if it has the null statement are not and describe of the data

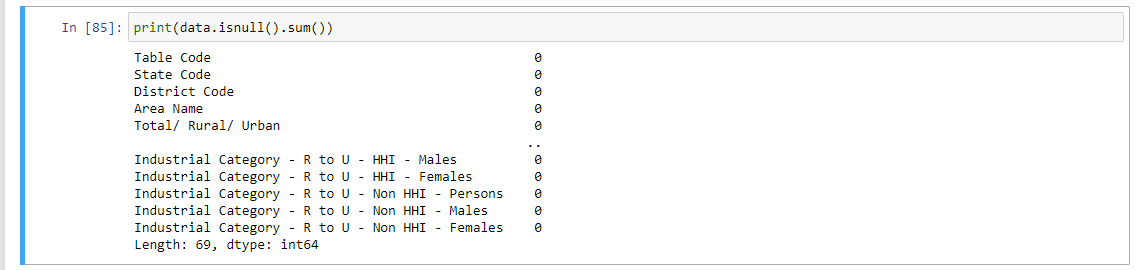


The info() method prints information about the DataFrame. The information contains the number of columns, column labels, column data types, memory usage, range index, and the number of cells in each column (non-null values).

The describe() method returns description of the data in the Data Frame. If the Data Frame contains numerical data, the description contains these information for each column: count - The number of not-empty values. Mean - The average (mean) value.

**STEP 3:**

We have to check the given data is null or not and sum up the null value.



**STEP 4:**

We have to check the missing value that are present in the dataset and remove the duplicate data.

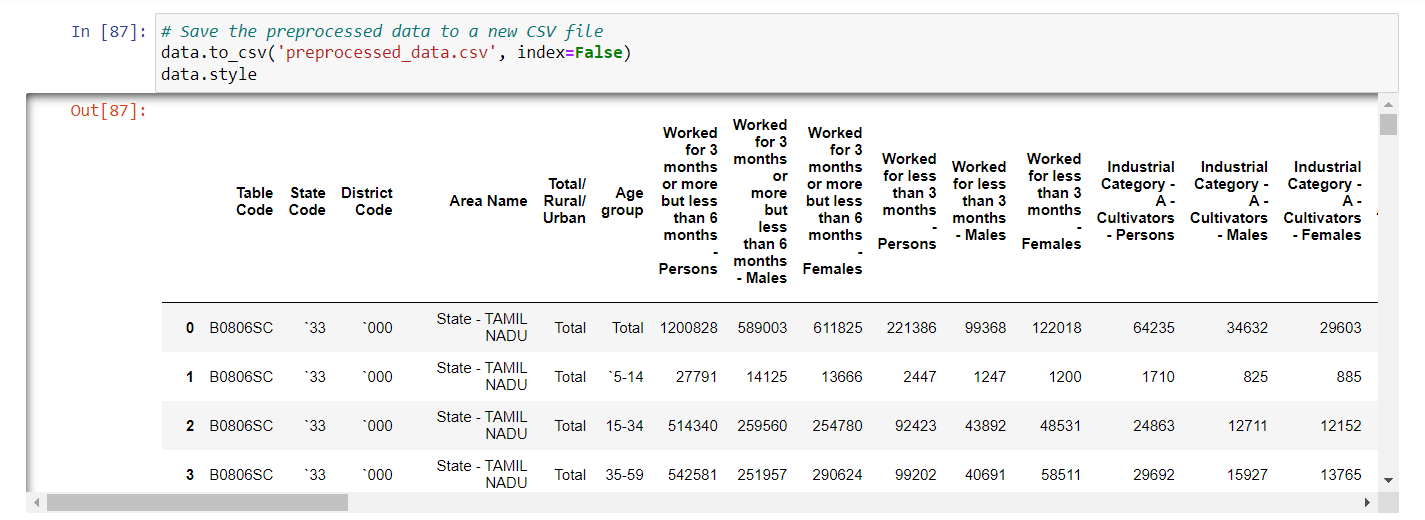


The dropna() method removes the rows that contains NULL values. The dropna() method returns a new DataFrame object unless the inplace parameter is set to True , in that case the dropna() method does the removing in the original DataFrame instead.

Pandas drop\_duplicates() function removes duplicate rows from the DataFrame. Its syntax is: drop\_duplicates(self, subset=None, keep="first", inplace=False) subset: column label or sequence of labels to consider for identifying duplicate rows. By default, all the columns are used to find the duplicate rows.

**STEP 5:**

So that we have clean and pre-processing the data and we create a new dataset that contain the preprocesssing\_data.csv as saved in our system.



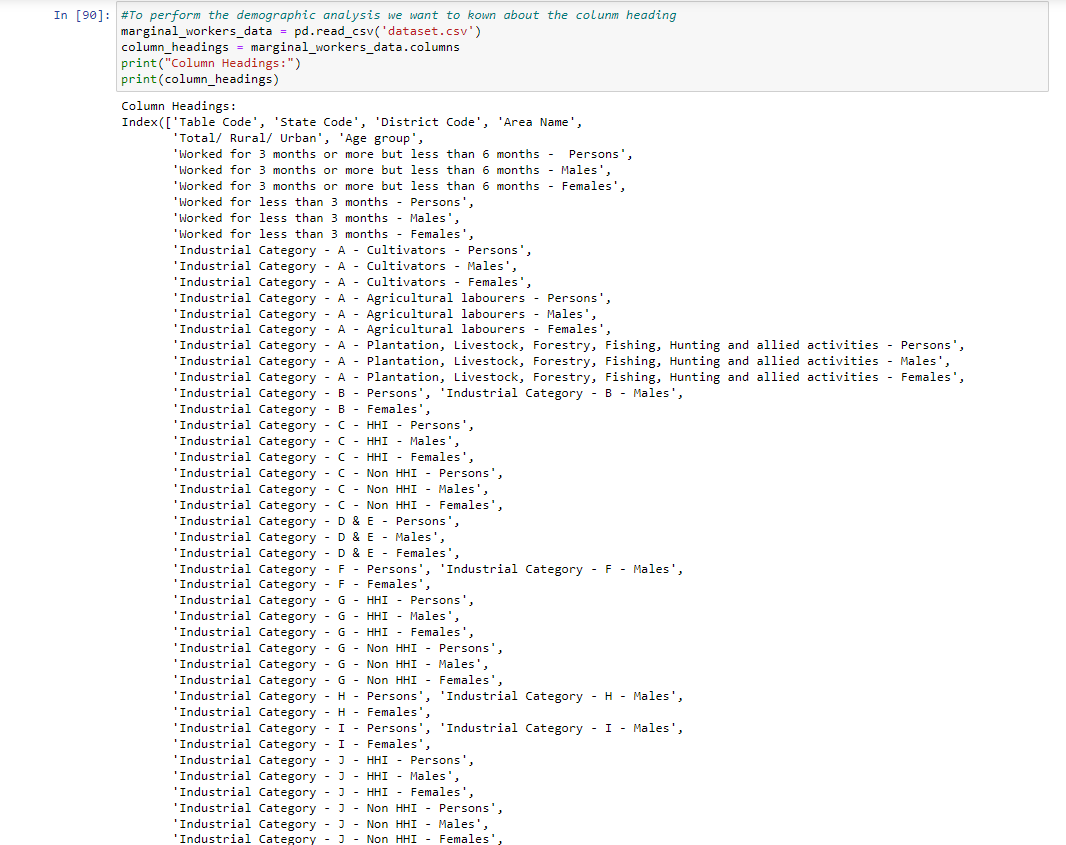
In the procedure of part 1 we have successfully clean the dataset and pre-process the data by using the dataset we are going to perform the demographic visualization of marginal workers.

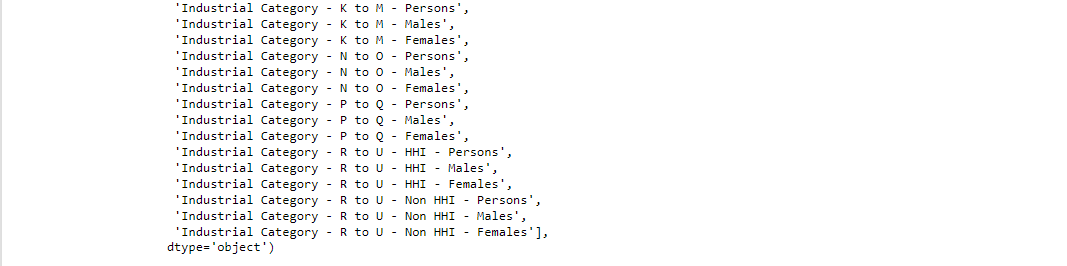
**PROCEDURE:**

**PART 2:**

**STEP 1:**

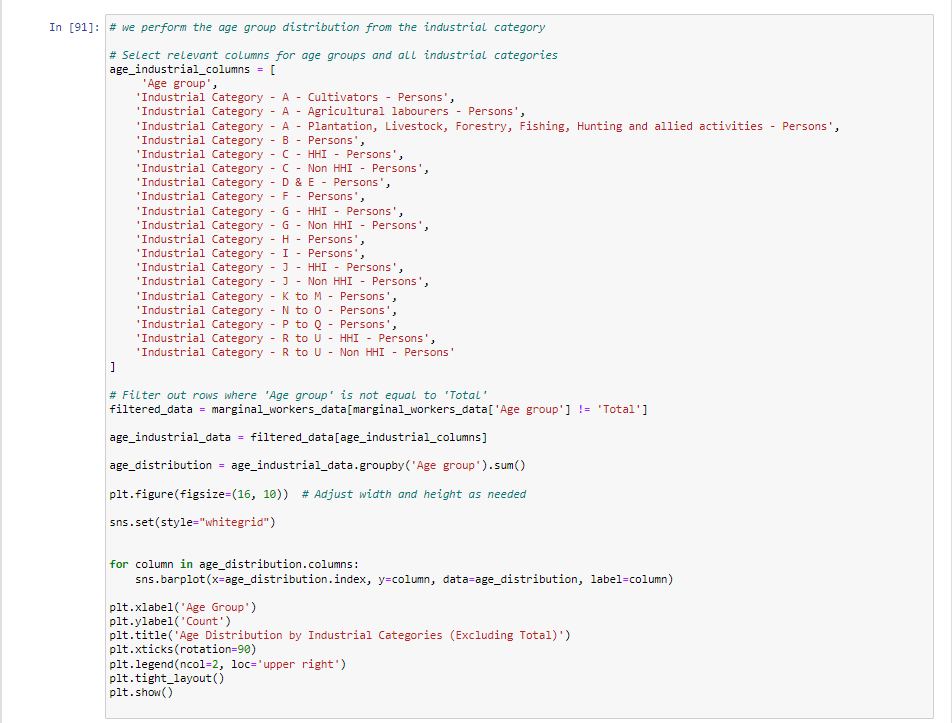
To perform the demographic analysis we want to know the column headings of the dataset

****

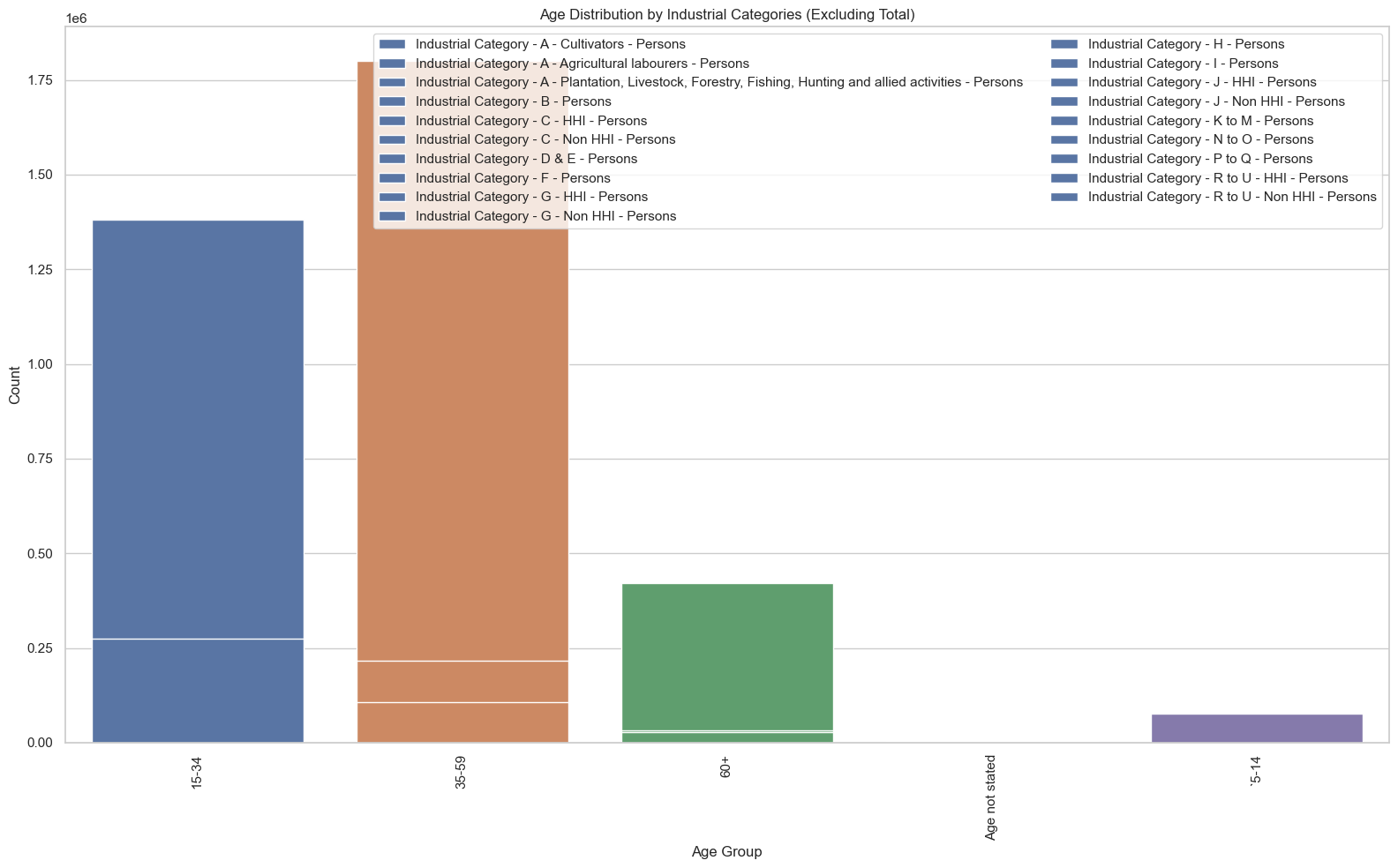
****

**STEP 2:**

To perform the demographic analysis for the age group under the industrial category using the bar chart we visualize the distribution



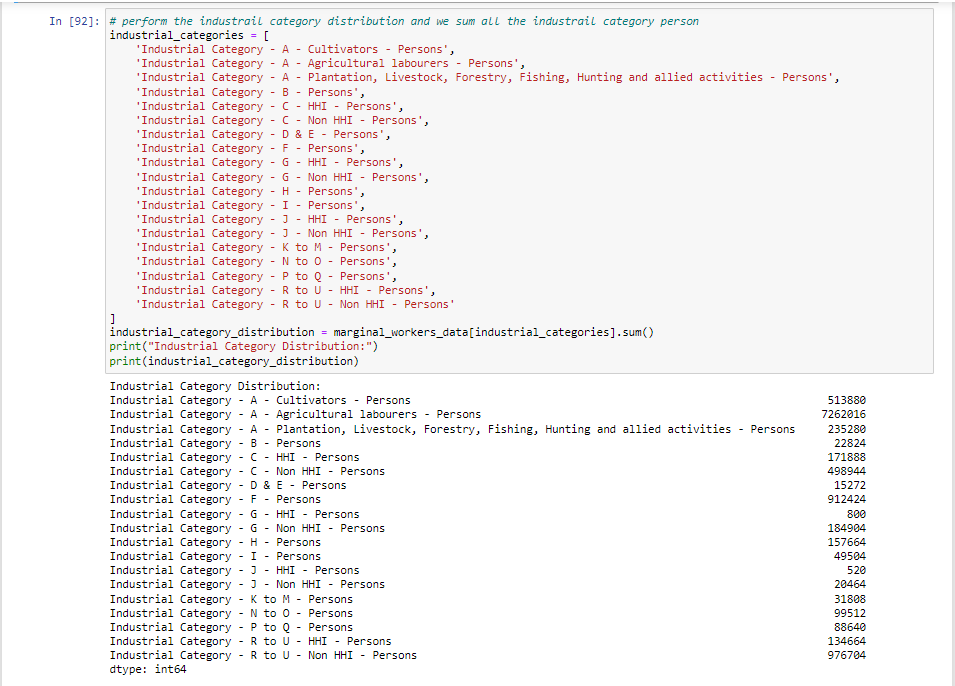
**OUTPUT:**

****

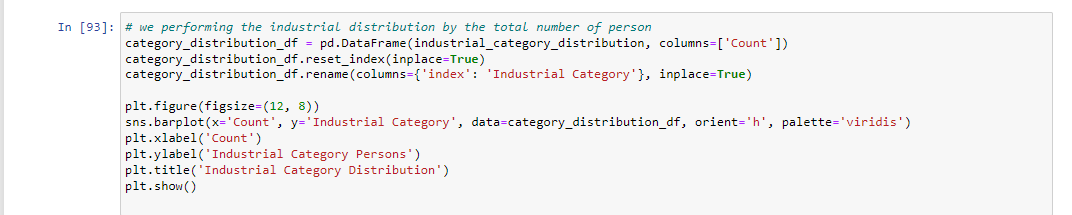
The code constructs a bar plot to visualize the age distribution across industrial sectors in Tamil Nadu, with a focus on marginal workforce. These analyses can help address issues such as insights into workforce demographics, successful age groups, comparing across groups, identifying marginalized workers, understanding workforce development, and providing informed decisions about social and of financial management.

**STEP 3:**

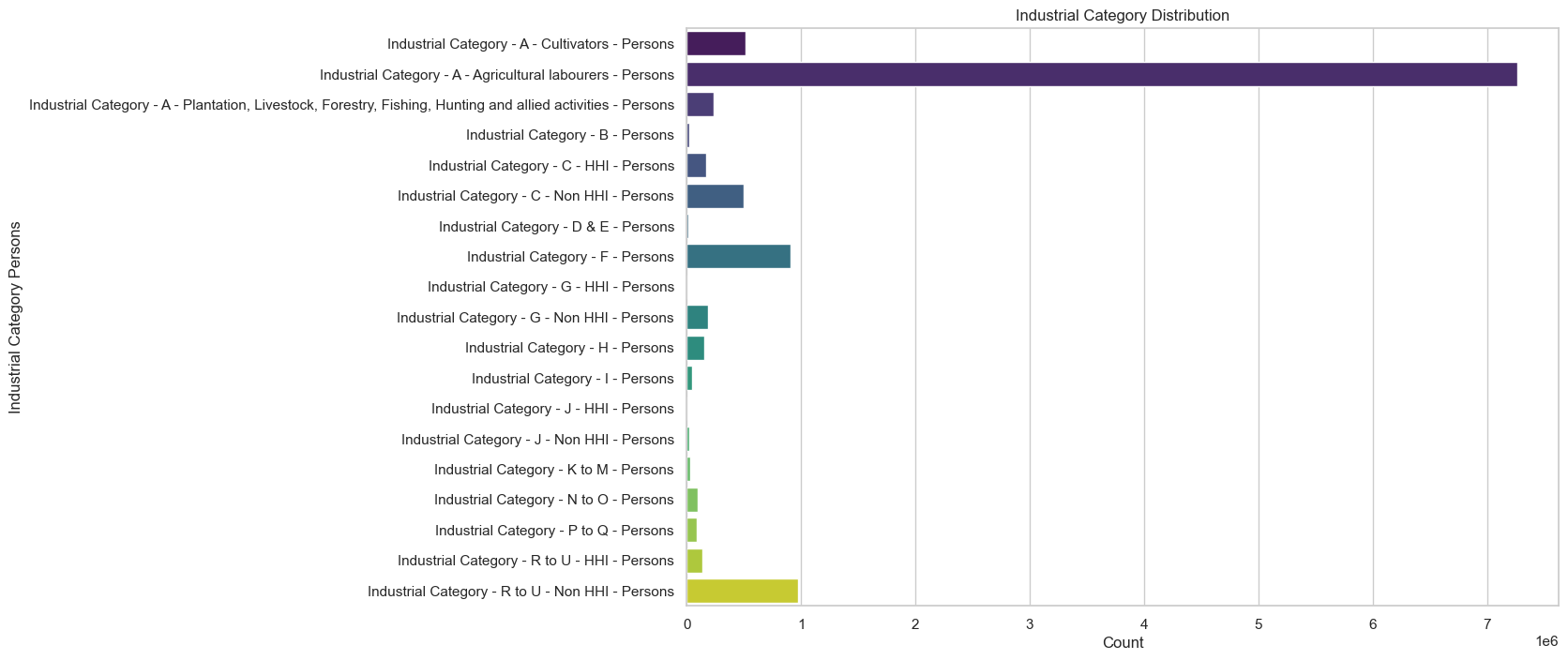
Performing the demographic analysis for the industrial category for persons and summing number of people are contain in each industrial category.



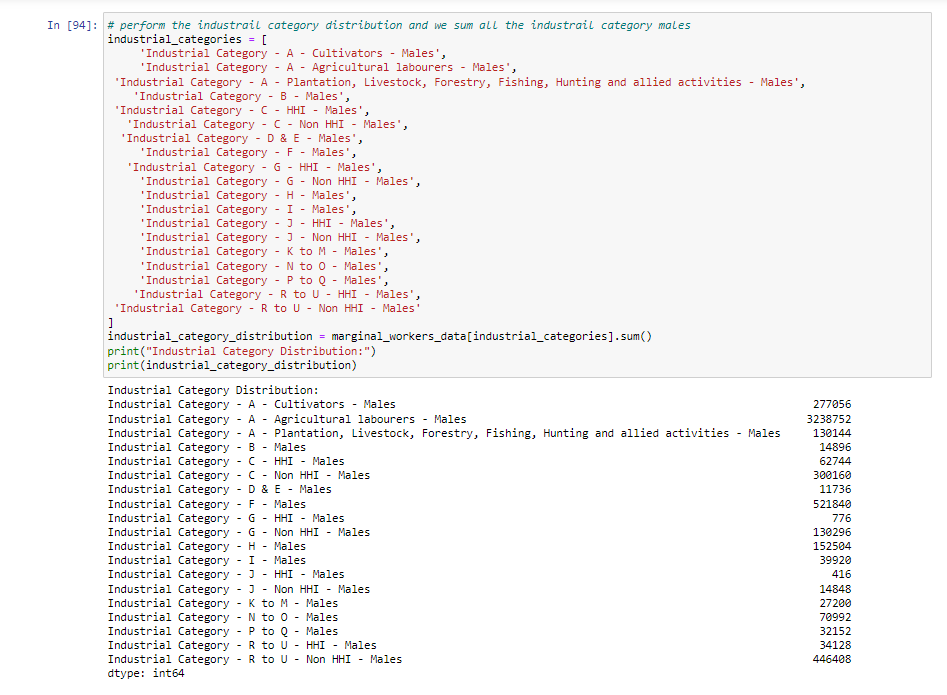
After that we have to visualize the distribution for industrial category for persons.



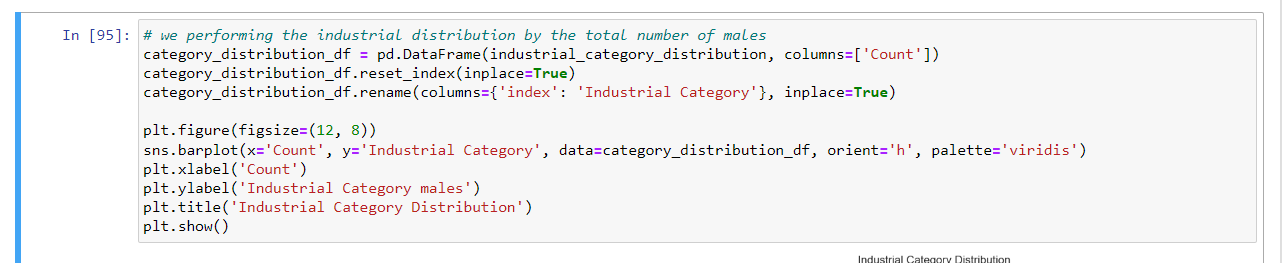
**OUTPUT:**

****

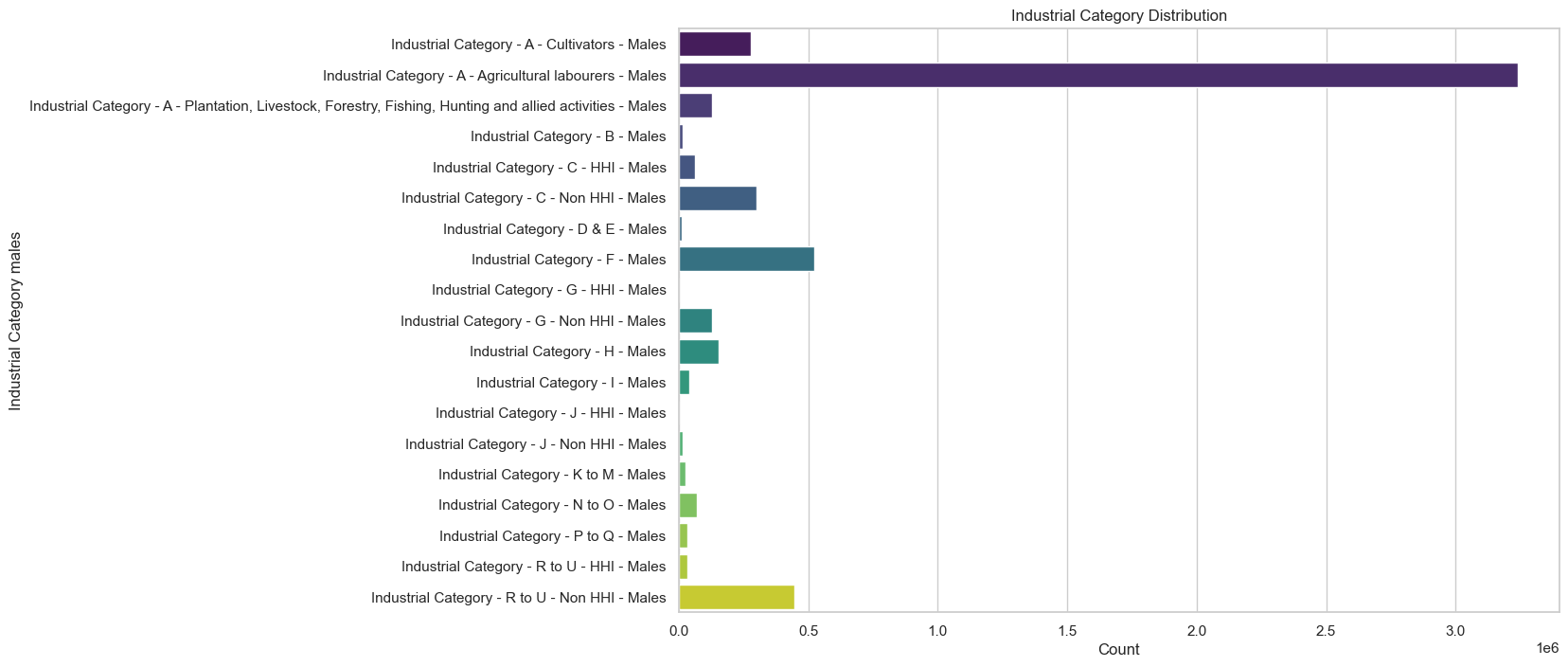
Performing the demographic analysis for the industrial category for Males and summing number of people are contain in each industrial category.



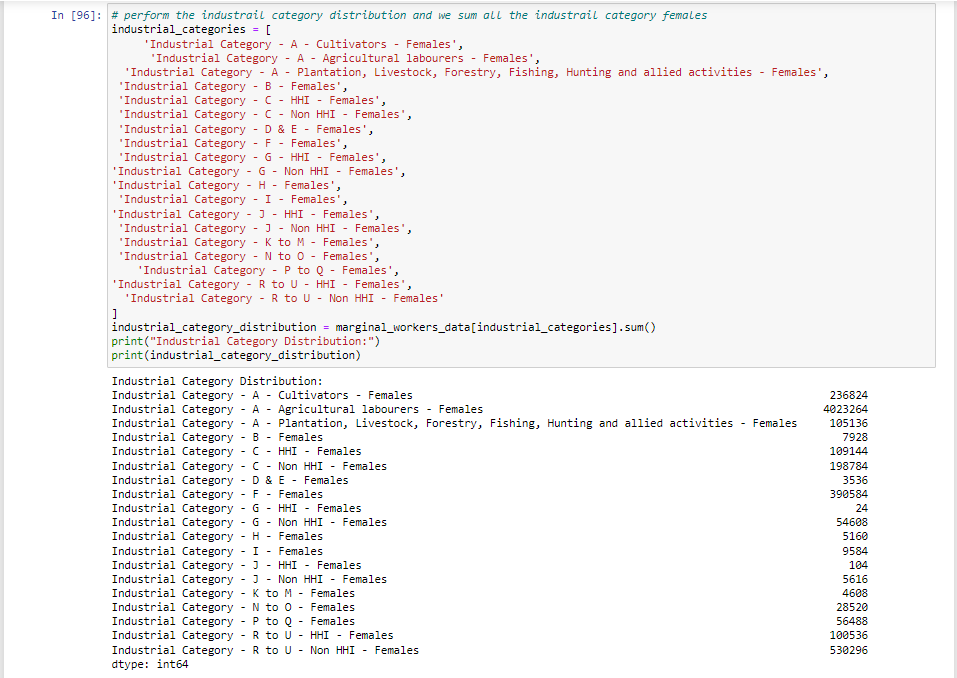
After that we have to visualize the distribution for industrial category for males.

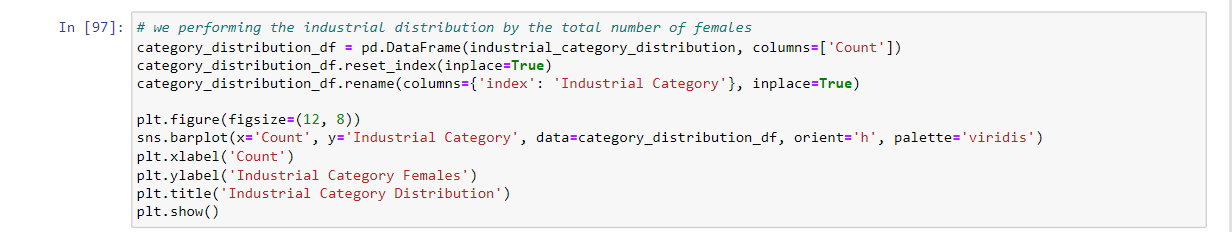


**OUTPUT:**

****

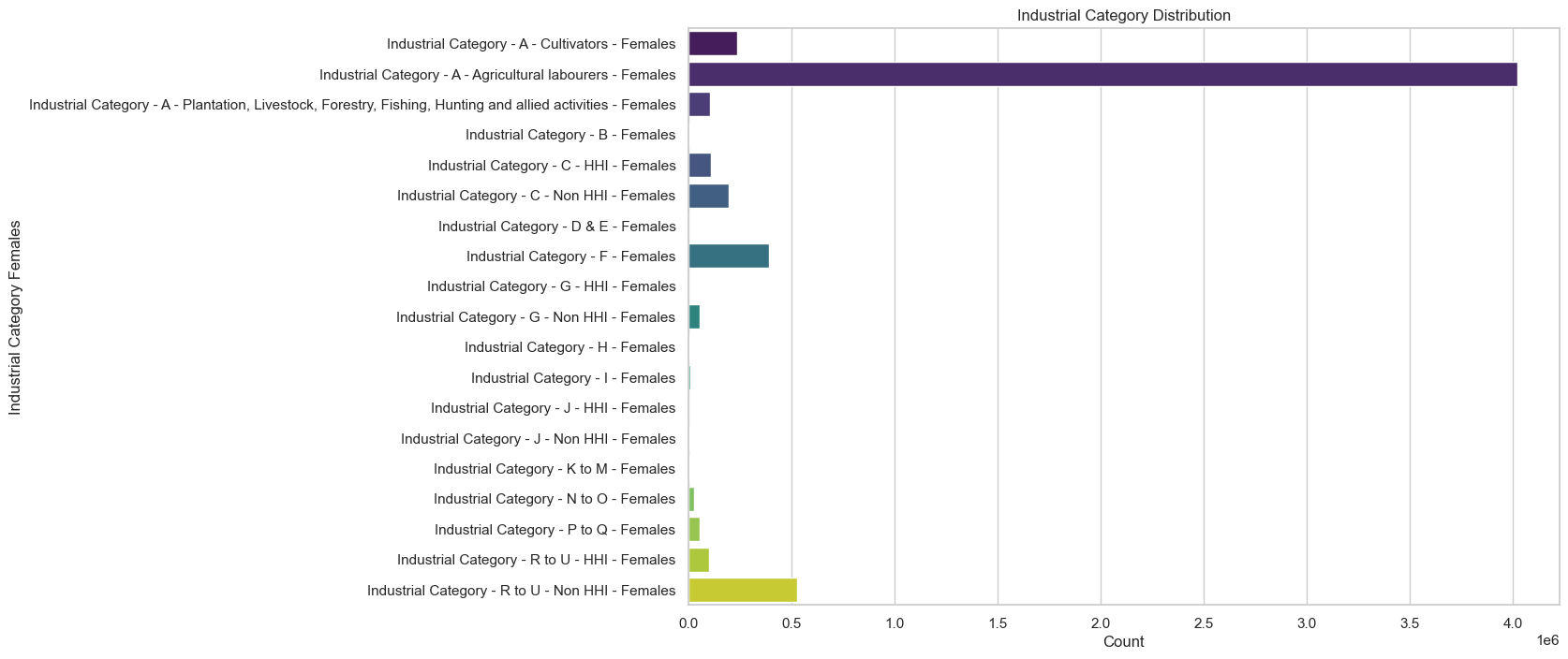
Performing the demographic analysis for the industrial category for Female and summing number of people are contain in each industrial category.





After that we have to visualize the distribution for industrial category for Females.

**OUTPUT:**



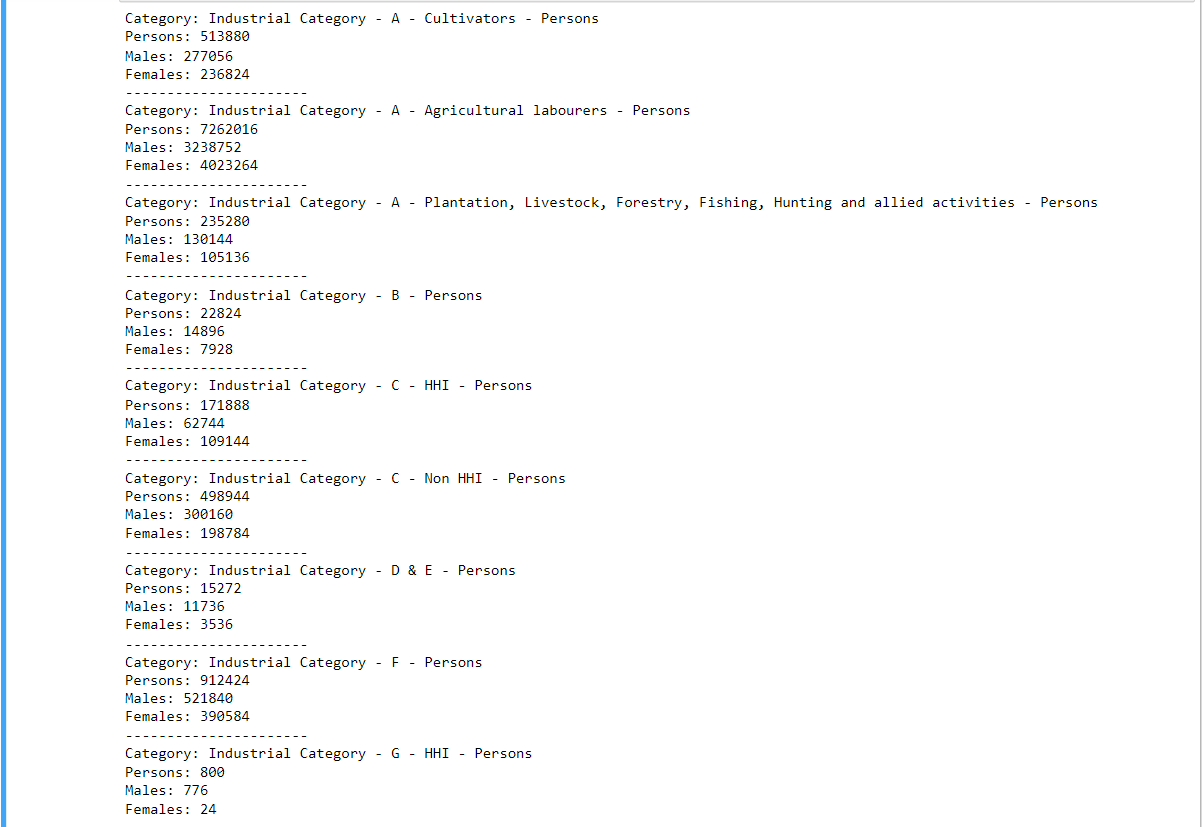
Analysis of the marginalized labor force in Tamil Nadu provides valuable insights into the labor force, classification into sectors This information helps policy makers target interventions and understand the social and economic development, thus informing decisions in knowledge, policy making and depth seeking

**STEP 4:**

Performing the sex distribution for the marginal worker using the industrial category finds the sum of males, females and persons.

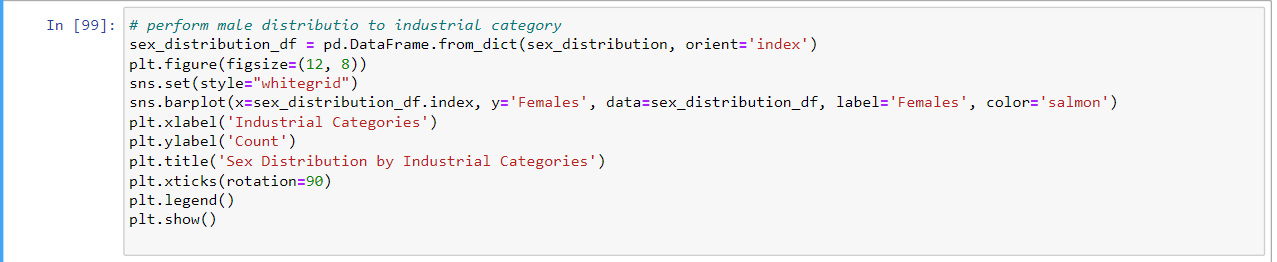


**OUTPUT:**

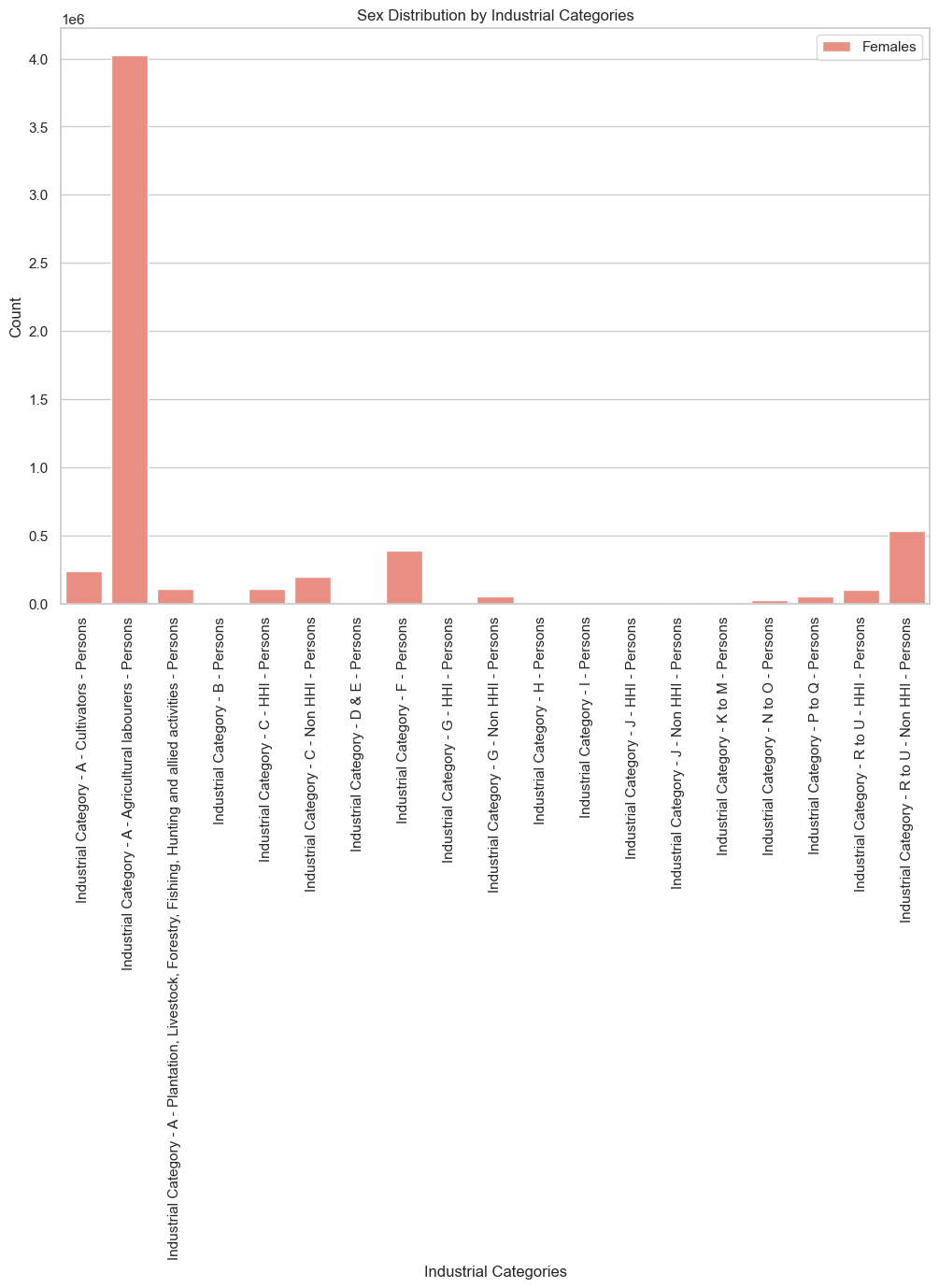
****



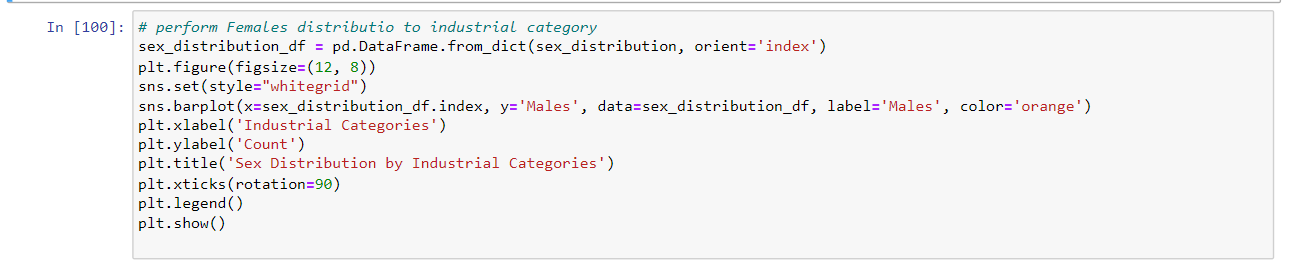
Perform male distribution to industrial category for visualization.



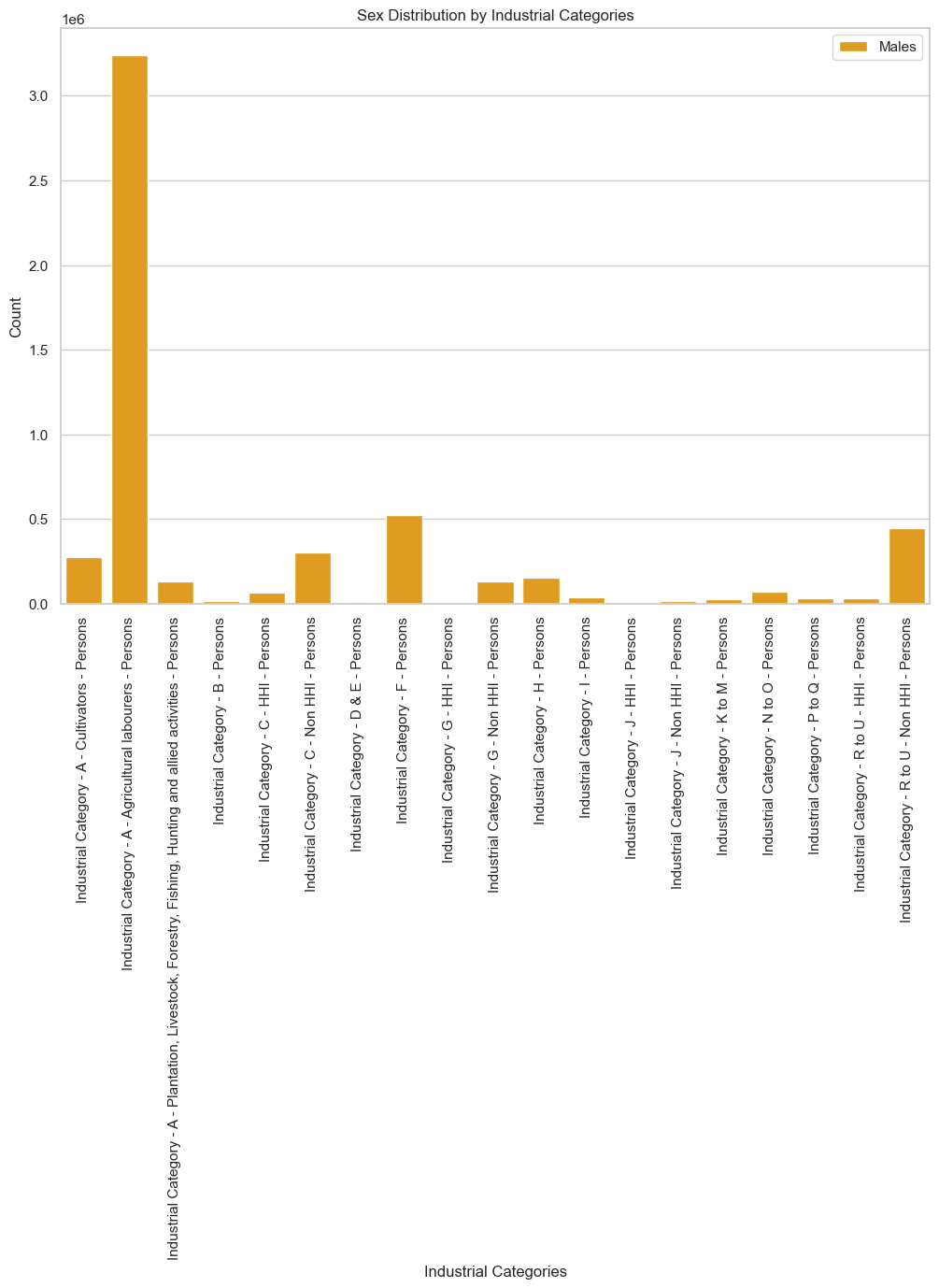
**OUTPUT:**

****

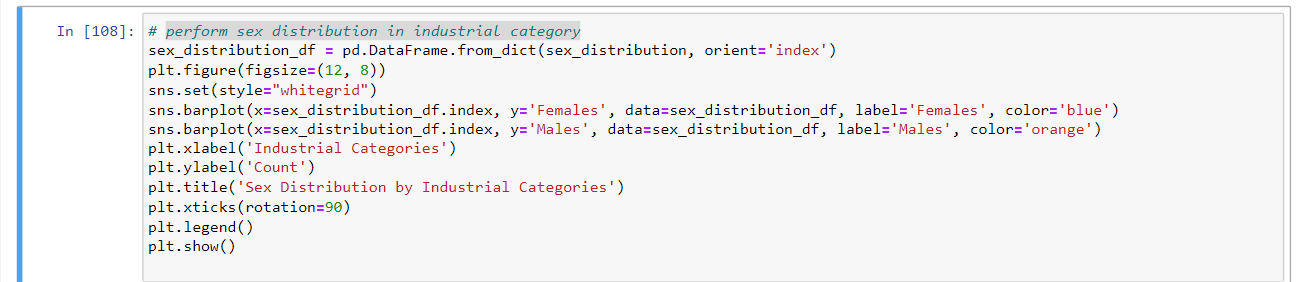
Perform Females distribution to industrial category for visualization.

****

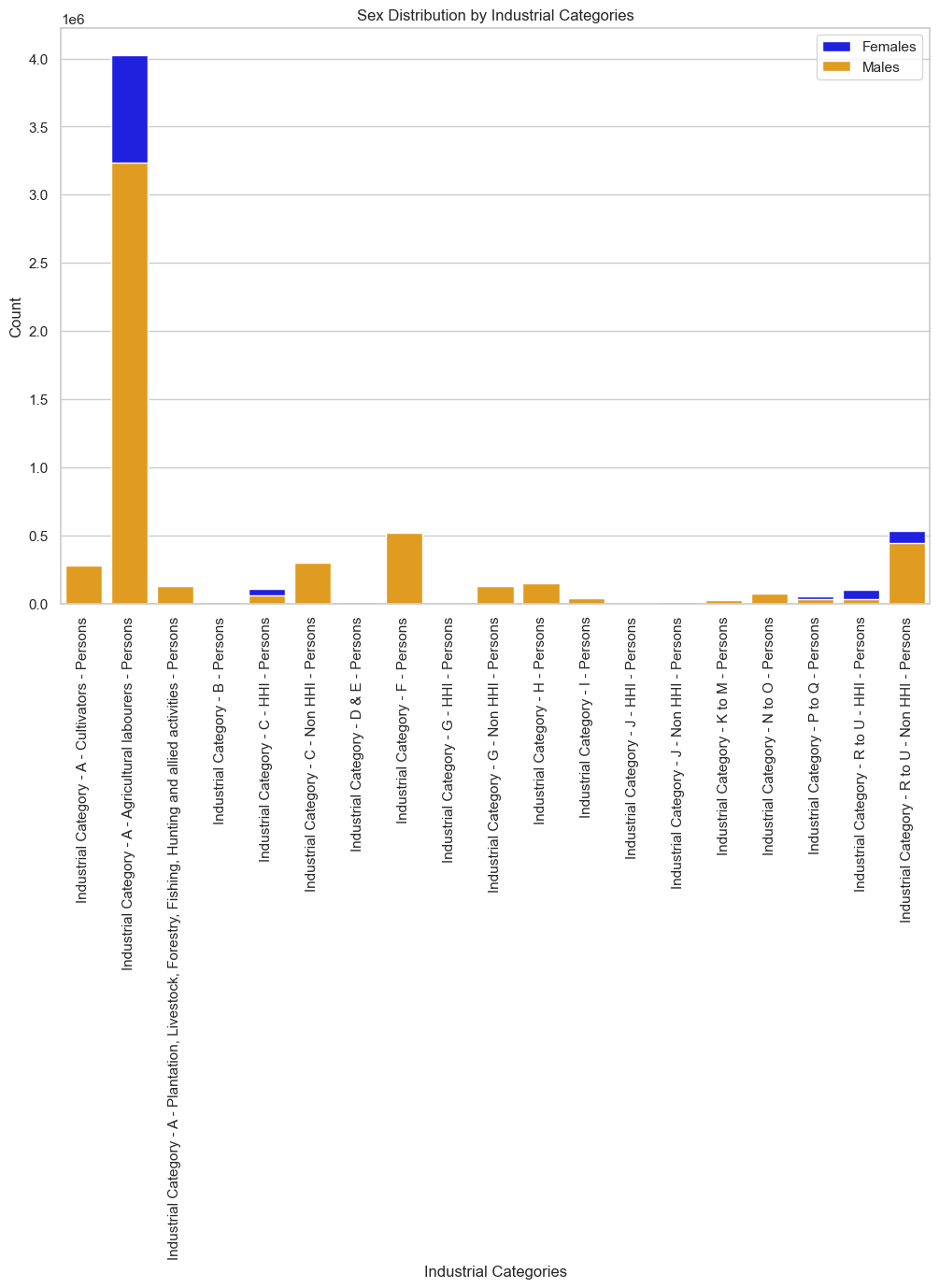
**OUTPUT:**

****

Perform sex distribution in industrial category and comparison the male and female distribution in bar chart.



**OUTPUT:**

****

The study examines the demographic makeup of marginal workers in Tamil Nadu, focusing on specific industrial categories and gender distribution. It provides a nuanced view of employment patterns and gender disparities within specific industries. The findings are valuable for policymakers to address gender imbalances and enhance opportunities for marginalized workers, fostering a more inclusive workforce in Tamil Nadu.

**CONCLUSION:**

Analysing the demographic distribution across age groups, industrial categories, and sexes provides a comprehensive understanding of the labor landscape. It illuminates the nuanced patterns of employment, shedding light on which age groups are predominantly engaged in specific industries and how gender disparities manifest within these sectors. Such insights are instrumental for crafting targeted policies that address the unique needs of different demographic groups. Understanding the age-wise, sector-specific, and gender-based variations in the workforce not only aids policymakers in designing inclusive initiatives but also empowers businesses to create diverse and equitable work environments. Moreover, it fosters a more nuanced understanding of societal employment dynamics, facilitating informed decision-making for both public and private sectors.