

TN Marginal Workers Assessment

Team Member

Name : A R HRUDAYABHIRAM

Register Number : 211521243019

Applied Data Science Phase-1 document



Problem Statement:

A Socioeconomic Analysis: Analyze the demographic characteristics of marginal workers based on age, industrial category, and sex. Create visualizations such as bar charts, pie charts, or heatmaps to represent the distribution across different categories.

Phase 1 : Problem Definition and Design Thinking

Problem Definition:

The project involves analyzing the demographic characteristics of marginal workers in Tamil Nadu based on their age, industrial category, and sex. The objective is to perform a socioeconomic analysis and create visualizations to represent the distribution of marginal workers across different categories. This project includes defining objectives, designing the analysis approach, selecting appropriate visualization types, and performing the analysis using Python and data visualization libraries.

Design Thinking:

Project Objectives:

Marginal Worker Demographic Analysis:

Age group: This column provides information about different age groups.

Total/ Rural/ Urban: This column likely indicates the type of area the individuals are from (total, rural, urban).

Work Duration: Various columns with labels like Worked for X months or more but less than Y months likely provide information about the duration of employment for different individuals.

Industrial Categories:

The columns starting with Industrial Category - represent various industrial categories. These categories are further divided based on different attributes like type of work, gender, and HHI (Household Income) status. For example:

Category - A: This category might relate to specific types of agricultural work.

Industrial Category - B: Represents another category of industries.

Similarly, Industrial Category - C to Industrial Category - R to U represent different industrial categories.

1. Age Distribution:

The different categories of workers age groups are:

5-9 , 10-14 ,15-19 , 20-24, 25-29 , 30-34, 35-39, 40-49,
50-59 ,60-69, 70-79 , 80+ , 'Age not stated'

The total number of workers are categorized in the all age groups (including male and female) :

<u>Category</u>	<u>Total</u>
Worked for 3 months or more but less than 6 months	4218884
Worked for less than 3 months	723891
Industrial Category - A – Cultivators	393082
'Industrial Category - A - Agricultural labourers - Persons	2372446
'Industrial Category - A - Plantation, Livestock	12509
Forestry, Fishing, Hunting and allied activities - Persons'	
Industrial Category - B – Persons	14979

Industrial Category - C - HHI - Persons',	154133
'Industrial Category - C - Non HHI - Persons',	306528
'Industrial Category - D & E - Persons',	7137
'Industrial Category - F – Persons	390275
'Industrial Category - G - HHI – Persons	510
'Industrial Category - G - Non HHI – Persons	171440
'Industrial Category - H - Persons',	84686
'Industrial Category - I - Persons'	42321
'Industrial Category - J - HHI - Persons',	463
'Industrial Category - J - Non HHI - Persons',	23293
'Industrial Category - K to M - Persons',	26047
'Industrial Category - N to O - Persons',	56495
'Industrial Category - P to Q - Persons',	58788
'Industrial Category - R to U - HHI – Persons	89703
'Industrial Category - R to U - Non HHI – Persons	625350

2. Understanding Gender Distribution:

Description: Gain insights into the gender distribution of marginal workers and identify any gender-based disparities.

Tasks - **Analyze the distribution of marginal workers by gender.**

The Marginal workers are classified into two Genders Males and Females based on different industrial categories such as :

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 - Males

Worked for less than 3 months – Females

And so on.,

Calculate proportions or percentages for different gender categories.

For example I calculated for ‘workers who worked less than 3 months -**Males** and workers who worked for less than 3 months -**Females**

Proportion of Males: 0.47

Proportion of Females: 0.53

Percentage of Males: 46.59%

Percentage of Females: 53.41%

4. Exploring Industrial Categories:

Description: Investigate the distribution of marginal workers across different industrial categories to understand their employment patterns.

Tasks

- **Categorize marginal workers based on the industrial sector they are employed in.**

The workers are categorized on:

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 - F - 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'

Plan the steps to extract, clean, and analyze the dataset to derive insights:

Step 1: Data Extraction

Load the Dataset:

Use a suitable programming language (e.g., Python with Pandas) to load the dataset into a DataFrame.

```
import pandas as pd
```

```
df = pd.read_csv('marginal_workers_dataset.csv')
```

Inspect the Data:

Display the first few rows to get an overview of the dataset's structure.

```
print(df.head())
```

Step 2: Data Cleaning and Preparation

Handle Missing Values:

Identify and handle any missing or NaN values in the dataset. Options include filling with averages or dropping rows/columns.

```
print(df.isnull().sum())
```

```
df['Column_Name'].fillna(df['Column_Name'].mean(),  
inplace=True)
```

Data Type Conversion:

Ensure that the data types of columns are appropriate for analysis. Convert categorical variables if needed.

```
df['Categorical_Column'] =  
df['Categorical_Column'].astype('category')
```


Step 3: Exploratory Data Analysis (EDA)

Descriptive Statistics:

Calculate summary statistics to understand the central tendency and variability of the data.

```
print(df.describe())
```

Step 4: Derive Insights

- **Demographic Analysis:**
Analyze demographic characteristics like age, gender, and location
- **Gender Distribution:**
Analyze and visualize the distribution of workers by gender.
- **Industrial Sector Analysis:**
Categorize and analyze workers based on the industrial sector they are employed in.

Some Potential Insights:

1. Age Distribution of Marginal Workers:

- Identify the age groups that have a higher representation among marginal workers.
- Determine if there are any age groups that stand out in terms of workforce participation.

2. Gender Distribution of Marginal Workers:

- Analyze the proportion of male and female workers among marginal workers.
- Identify any gender-based disparities in workforce participation.

3. Duration of Employment:

- Understand how many workers have been employed for different durations (e.g., 3 months to 6 months, less than 3 months).
- Explore any trends or patterns in employment duration.

4. Industrial Categories:

- Categorize workers based on the industrial sector they are employed in.
- Identify which industries have a higher representation of marginal workers.

5. Demographic Details:

Gain insights into demographic characteristics like age, gender, and location.

6. Comparative Analysis:

Compare different aspects of marginal workers, such as employment duration, industrial categories, or age groups, to identify any disparities or patterns.

7. Potential Disparities:

Identify potential disparities in workforce participation based on demographic factors (age, gender) or employment characteristics (duration, industry).

8. Recommendations:

Based on the insights gained, consider providing recommendations for interventions or policies to address any disparities or improve the conditions of marginal workers.

Visualization Selection: Determine suitable visualization types (e.g., bar charts, pie charts, heatmaps) to represent demographic distributions effectively.

1. Age Distribution:

- Bar Chart: Display the count of workers in each age group.
- Histogram: Provide a visual representation of the distribution of ages.

2. Gender Distribution:

- Pie Chart: Show the proportion of male and female workers.

3. Urban/Rural Distribution:

- Bar Chart or Pie Chart: Display the distribution of workers in urban, rural, and total areas.

4. Duration of Employment:

- Stacked Bar Chart: Show the distribution of workers based on different durations of employment.

5. Industrial Categories:

- Stacked Bar Chart or Grouped Bar Chart: Represent the distribution of workers across different industrial categories.

6. Comparison between Categories:

- Grouped Bar Chart: Compare the distribution of workers between different categories (e.g., age groups, gender) for a specific attribute (e.g., employment duration, industrial category).

7. Location-based Distribution:

*Choropleth Map: Display the distribution of workers across different regions or districts.

8. Comparative Analysis:

- Scatter Plot or Line Plot: Show relationships or trends between different demographic factors (e.g., age vs. employment duration).

9. Heatmap (if applicable):

- If there are multiple attributes to compare (e.g., industrial categories by gender), a heatmap can provide a comprehensive view of the distribution.

Conclusion:

In this analysis, we delved into the demographic characteristics of marginal workers in Tamil Nadu, focusing on their age, industrial category, and gender distribution. We employed a systematic approach, beginning with data extraction and cleaning, followed by insightful visualizations to represent the distribution patterns effectively.

While this initial examination has provided a foundation, it's important to note that further observations and analyses may yield deeper insights. As we continue to explore this dataset, we anticipate uncovering additional trends and patterns that will inform more comprehensive socio-economic assessments.

We look forward to the continued exploration of this dataset, which promises to offer valuable insights into the dynamics of marginal employment in Tamil Nadu.
