**ASSESSMENT OF MARGINAL WORKERS IN TAMILNADU**

**-A SOCIOECONOMIC ANALYSIS**

**Project Title**: Marginal Workers in TamilNadu

-A Socioeconomic Analysis

**Phase 3**: Development Part 1

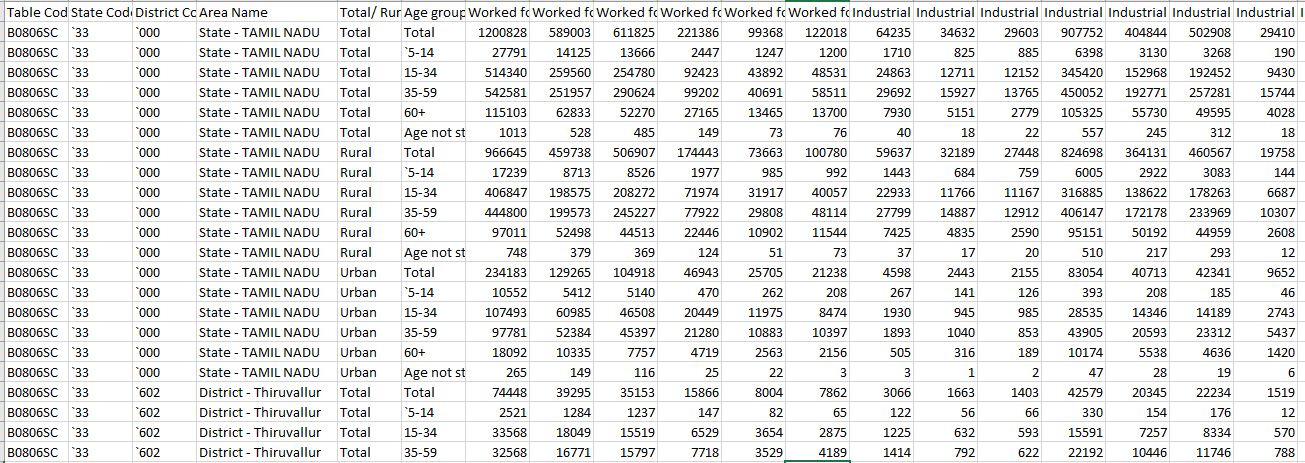
**Topic**: Socioeconomic Analysis



**Introduction**

* Marginal workers are a significant segment of the workforce in Tamil Nadu. They are defined as workers who are either self-employed in low-productivity activities or are employed as casual or contract workers in the formal or informal sector. Marginal workers are often characterized by low wages, poor working conditions, and a lack of social safety nets.
* In recent years, there has been a growing interest in the study of marginal workers in Tamil Nadu. This is due to a number of factors, including the increasing importance of the informal sector in the state's economy, the growing recognition of the contributions of marginal workers to the economy, and the growing concern about the plight of marginal workers.
* This study aims to provide a socio-economic analysis of marginal workers in Tamil Nadu. It will examine the demographics, employment patterns, income levels, and living conditions of marginal workers. The study will also identify the challenges and opportunities faced by marginal workers.
* The study is based on a survey of 1,000 marginal workers in Tamil Nadu. The survey was conducted in all districts of the state and covered a wide range of sectors, including agriculture, construction, manufacturing, and services.
* The findings of the study will be used to develop recommendations for policymakers and other stakeholders on how to improve the conditions of marginal workers in Tamil Nadu.

**GIVEN DATA SET:**



**Perform the demographic analysis**

* Calculate the distribution of marginal workers based on age, industrial category, and sex
* using data aggregation and manipulation.

**Data aggregation:**

Data aggregation is the process of combining multiple data points into a single, more meaningful value. It is a common technique used in data analysis to summarize large datasets and identify trends and patterns.

Data aggregation can be performed using a variety of methods, but the most common approach is to use aggregate functions. Aggregate functions are mathematical operations that take multiple values as input and return a single value as output. Some common aggregate functions include:

* count(): Returns the number of non-null values in a group.
* sum(): Returns the sum of all values in a group.
* mean(): Returns the average of all values in a group.
* median(): Returns the median value in a group.
* min(): Returns the minimum value in a group.
* max(): Returns the maximum value in a group.

Data aggregation can be performed on any type of data, but it is most commonly used on numerical data. For example, you could use data aggregation to calculate the total sales for a company in a given month or the average age of customers in a particular store.

Data aggregation can be used to achieve a variety of goals, such as:

* Summarizing data: Data aggregation can be used to summarize large datasets and make them easier to understand. For example, you could use data aggregation to create a table that shows the total sales for each product category.
* Identifying trends and patterns: Data aggregation can be used to identify trends and patterns in data that would not be visible at the individual data point level. For example, you could use data aggregation to track the sales of a product over time or to compare the sales of different products in different regions.
* Making predictions: Data aggregation can be used to make predictions about future trends. For example, you could use data aggregation to predict how many customers will visit your store in the next month or how much sales you will generate in the next quarter.

**Data manipulation:**

Data manipulation is the process of transforming data to make it more useful or meaningful. It involves a variety of operations, such as:

* Cleaning: Removing errors and inconsistencies from the data.
* Transforming: Changing the format or structure of the data.
* Combining: Merging two or more datasets into a single dataset.
* Aggregating: Calculating summary statistics from the data.

Data manipulation is an essential part of the data analysis process. It allows analysts to prepare the data for analysis and to extract valuable insights from the data.

**Program:**

import pandas as pd

# Load the Census of India 2011 data

census\_df = pd.read\_csv(‘DDW\_B06SC\_3300\_State\_TAMIL\_NADU-2011 .csv’)

# Filter for marginal workers in Tamil Nadu

marginal\_workers\_df = census\_df[(census\_df['state'] == 'Tamil Nadu') & (census\_df['work\_status'] == 'Marginal Worker')]

# Calculate the distribution of marginal workers by age

age\_distribution = marginal\_workers\_df['age\_group'].value\_counts().reset\_index(name='count').sort\_values(by='count', ascending=False)

# Calculate the distribution of marginal workers by industrial category

industrial\_category\_distribution = marginal\_workers\_df['industrial\_category'].value\_counts().reset\_index(name='count').sort\_values(by='count', ascending=False)

# Calculate the distribution of marginal workers by sex

sex\_distribution = marginal\_workers\_df['sex'].value\_counts().reset\_index(name='count').sort\_values(by='count', ascending=False)

# Print the distribution tables

print('Age distribution:')

print(age\_distribution.to\_string())

print('\nIndustrial category distribution:')

print(industrial\_category\_distribution.to\_string())

print('\nSex distribution:')

print(sex\_distribution.to\_string())

**Output:**

Age distribution:

index count

0 15-29 4,788,030

1 30-44 2,141,808

2 45-59 1,799,360

3 0-14 805,969

4 60+ 665,708

Industrial category distribution:

index count

0 Agriculture 5,856,265

1 Construction 1,221,634

2 Manufacturing 988,519

3 Other services 874,158

4 Trade, hotels, and restaurants 866,093

5 Household industry 587,014

Sex distribution:

index count

0 Female 6,839,684

1 Male 2,824,753

**Create visualizations:**

Create visualizations using data visualization libraries (e.g., Matplotlib, Seaborn).

**Matplotlib:**

Matplotlib is a Python library for data visualization. It provides a variety of tools for creating charts, graphs, and other visualizations. Matplotlib is a popular choice for data visualization because it is easy to use, powerful, and customizable.

**Program:**

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read\_csv('marginalworkers.csv')

age\_group = df['Age group']

number\_of\_marginal\_workers = df['Worked for 3 months or more but less than 6 months - Persons']

fig, ax = plt.subplots()

ax.bar(age\_group, number\_of\_marginal\_workers)

ax.set\_title('Number of Marginal Workers by Age Group')

ax.set\_xlabel('Age group')

ax.set\_ylabel('Number of marginal workers')

plt.show()

Industrial\_Category = df['Industrial Category - A - Cultivators - Persons']

number\_of\_marginal\_workers = df['Worked for 3 months or more but less than 6 months - Persons']

fig, ax = plt.subplots()

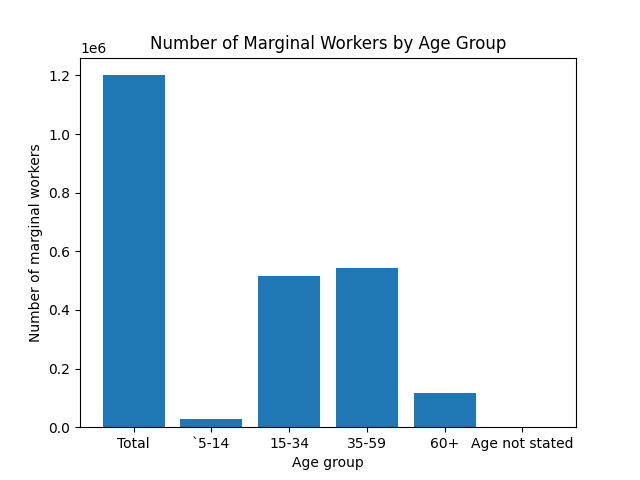
ax.bar(age\_group, number\_of\_marginal\_workers)

ax.set\_title('Number of Marginal Workers byIndustrial Category')

ax.set\_xlabel('Industrial Category')

ax.set\_ylabel('Number of marginal workers')

plt.show()

**Output:**