**Phase 2: Innovation**

**Problem:Assessment of marginal workers in Tamilnadu - A socio economic analysis**

**INTRODUCTION:**

Marginal Workers are the significant workforce of our state and face challenges in terms of wages, employment opportunities and working conditions. The project aims at conducting a comprehensive study of the marginal workers of Tamilnadu focusing on the factors such as age, gender, industry category and demography.

**NEED FOR ANALYSIS OF DATA :**

* ***Policy formulation*** : Analysis of marginal workers yields results that aid the policy makers to frame policies to improve the life of marginal workers.
* ***Decision Making:***Analysis of marginal workers data drives decision making in the state level, district level as well as in the local level.
* ***Labor Market Planning and Development:*** The project provides essential insights into the labor market and can guide in labor market planning industrial growth strategies and workforce development strategies.
* ***Gender Equality****:* Analyzing the gender distribution of marginal workers gives light on the status of participation of both male and female employees in the workforce. It ensures equal opportunities for both male and female workers.
* The analysis is needed to provide a more equitable, inclusive and informed approach to address the challenges faced by the marginal workers of our society.

**DATA ANALYSIS:**

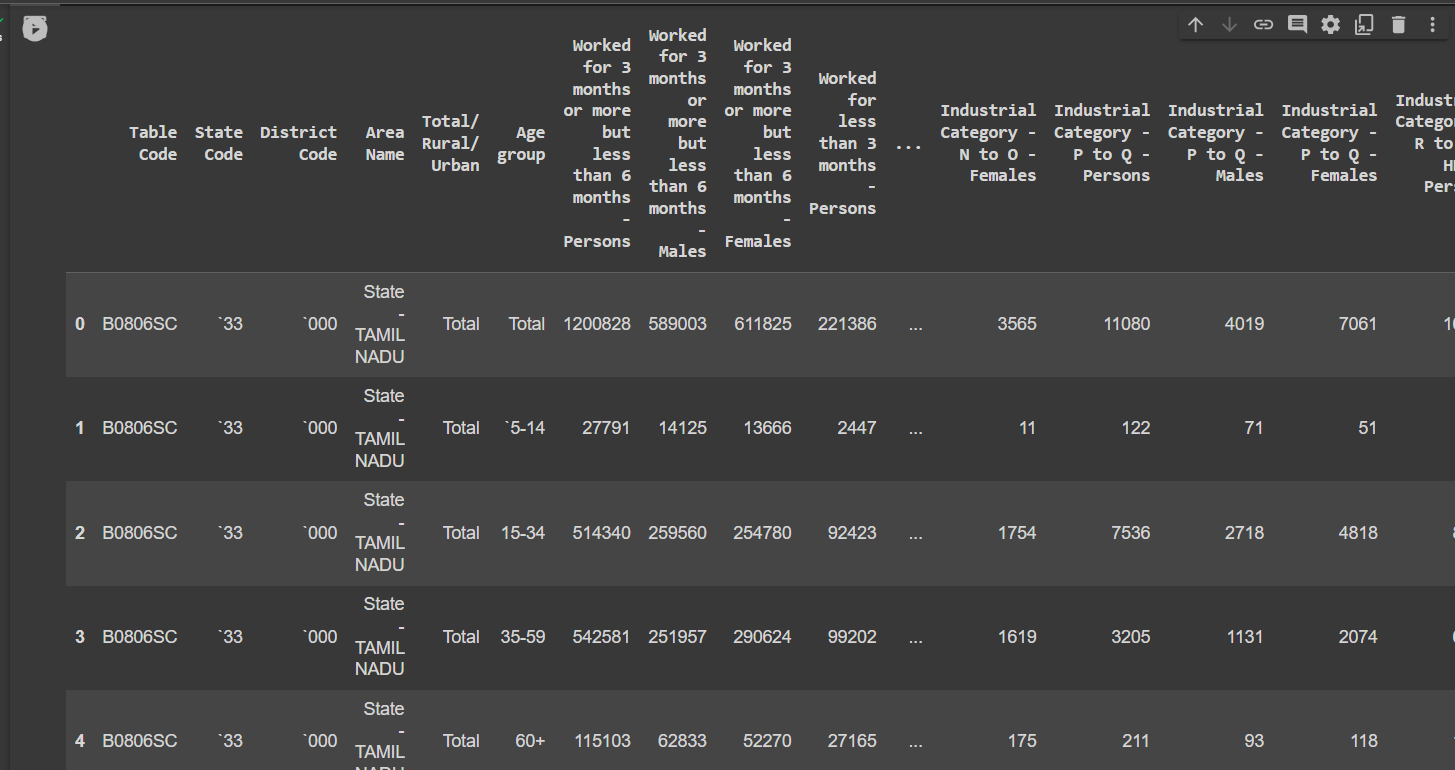
* From the given dataset which is provided, we can able to identify what kind of information is available.
* Also, we can able to identify the key variables like District location, Area name, Age group, Worked period, Industrial Category etc.

Analysis Criteria:

* The analysis of marginal workers data may be based on district, industry category or age group.

Preprocessing the data:

* Data preprocessing is a crucial step in the data analysis and machine learning pipeline. It involves cleaning, transforming, and organizing raw data into a format suitable for analysis or model training.
* Data preprocessing can be carried out using python modules like pandas, sci-kit learn and numpy.



* The above image shows the first 5 rows of the data set before preprocessing. The column Table Code is not needed for our analysis. So we drop it using pandas drop command.

df.drop(columns=[‘Table Code’],inplace=True)

Similarly the state code column is also not needed for our analysis thus we remove it using

df.drop(columns=[‘State Code’],inplace=True)

The District Code column is a string and thus we convert it to integer for better understanding

df [‘District Code’]=df[‘District Code’].astype(int)

We label the age group into categories from A to F to simplify understanding

Total – A  
5-14 – B

15-34 – C

35-59 – D

60+ - E

Age not stated – F

df1['Age group']=df1['Age group'].replace('Total', 'A')

df1['Age group']=df1['Age group'].replace('`5-14', 'B')

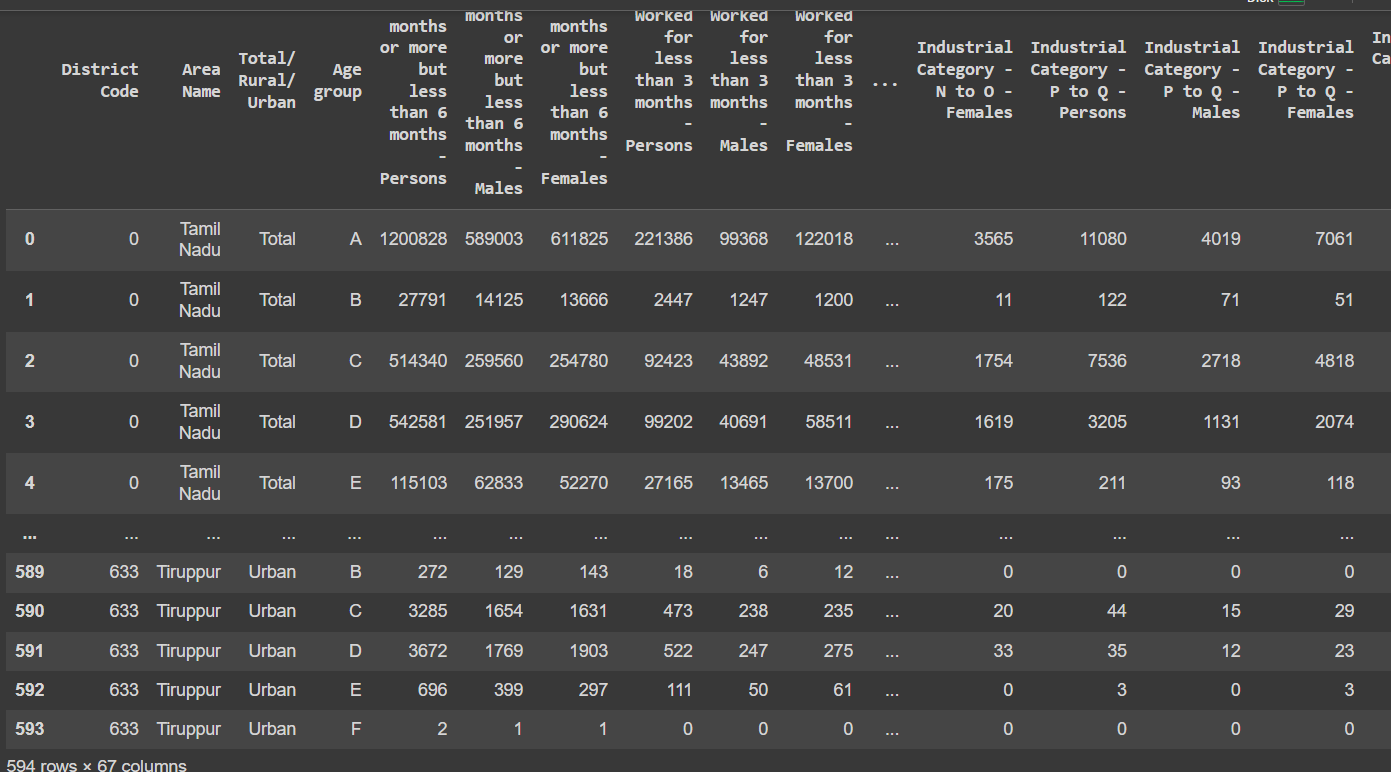
df1['Age group']=df1['Age group'].replace('`15-34', 'C')

df1['Age group']=df1['Age group'].replace('`35-59', 'D')

df1['Age group']=df1['Age group'].replace('`60+', 'E')

df1['Age group']=df1['Age group'].replace('Age not stated', 'F')

* The below figure shows the data after preprocessing



Filtering and Extracting the Data:

* By using tools like pandas in Python, we can able to filter and extract data based on analysis criteria.
* For example, data can be filtered based on district or based on industry category or age group.

df2=df[df['Area Name']=='Tiruppur']

The above command is used to extract rows for which the Area Name is Tiruppur.

The output is given as:



Visualization of Data:

* In order to gain a deeper insight into patterns and trends within segmented data, it is necessary to employ visualization. Time series plots, bar charts, and maps are frequently employed to facilitate a better understanding of data set.
* IBM Cognos analytics can be used to plot the different graphs.
* Data drilling can be applied to analyze subsets of data.
* Simple visualization of data can be done as follows:

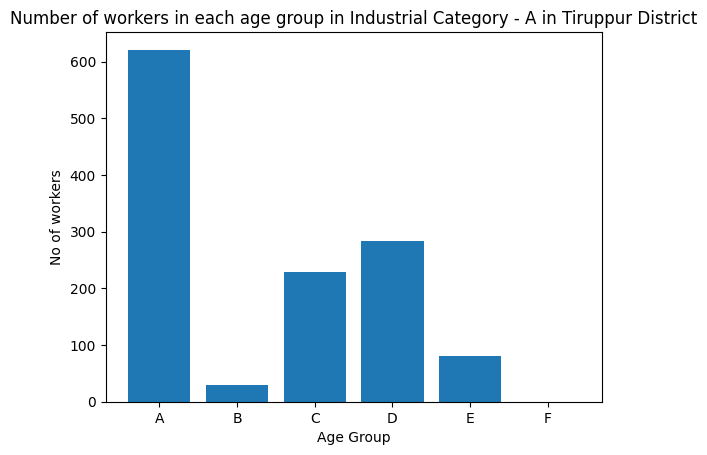
import matplotlib.pyplot as plt

plt.bar(df2[df2['Total/ Rural/ Urban']=='Total']['Age group'],df2[df2['Total/ Rural/ Urban']=='Total']['Industrial Category - A - Cultivators - Persons'])

plt.title("Number of workers in each age group in Industrial Category - A in Tiruppur District")

plt.xlabel("Age Group")

plt.ylabel("No of workers")



**Innovation that can be applied on analysis of marginal workers:**

1. Cluster Analysis for Industrial Patterns – Application of algorithms like K Means to identify patterns within industrial categories.
2. Gender based employment analysis in Industries- Participation of male and female employees among different industry types is analyzed
3. Correlation analysis- It is used to understand the relation between different industrial categories.
4. Map based visualization of industries across state to provide an interactive source of information.

**Conclusion:**

Thus the above mentioned steps can be carried out to analyze the Marginal workers dataset obtained during the 2011 census using IBM Cognos and Python.