**PROBLEM STATEMENT:**

Start the data analysis by loading and preprocessing the dataset.

Load the dataset using Python and data manipulation libraries (e.g., pandas).

**Dataset:** Link: [https://tn.data.gov.in/resource/marginal-workers-classified-age-industrial-category-and-sex-scheduled-caste-2011-tamil](https://tn.data.gov.in/resource/marginal-workers-classified-age-industrial-category-and-sex-scheduled-caste-2011-tamil" \t "/home/surya/Documents\\x/_blank)

**Solution:**

To start the data analysis process by loading and preprocessing a dataset using Python, you'll typically use the “pandas” library. I'll walk you through the steps to do this assuming you have a dataset in a CSV file. Make sure you have `pandas` installed. If not, you can install it using”pip”:

Code to install panadas

pip install pandas

Here's how to load and preprocess a dataset:

**1.** **Import the necessary libraries:**

Program to import library

import pandas as pd

**2. Load the dataset using “pandas”:**

Program to load the data

dataset = pd.read\_csv('tn\_age\_distribution\_dataset.csv')

**3. Explore the dataset:**

* To get a quick overview of the dataset, you can use “head()”to display the first few rows of the data:

Code to get first five rows

print(dataset.head())

* To get a summary of the dataset's statistics, you can use “describe()”:

Program to get summary of the datas:

print(dataset.describe())

* To check for missing values, you can use “isnull()” and “sum()”:

Program to remove the null values

print(dataset.isnull().sum())

**4. Preprocess the dataset:**

* Handle missing values: Depending on the dataset, you might want to fill missing values with appropriate values or drop rows or columns with missing data.
* Data type conversion: Ensure that the data types of columns are appropriate. You can use “astype()”to convert data types.
* Remove duplicates: Use “drop\_duplicates”` to remove duplicate rows, if necessary.
* Feature engineering: Create new columns, modify existing ones, or encode categorical variables as needed.
* Normalize or scale numerical features, if required.

**5. Save the preprocessed dataset (optional):**

If you want to save the preprocessed dataset for future use, you can use “to\_csv()”:

Program to save cleaned data:

# Replace “tn\_age\_ditribution\_preprocessed\_dataset.csv”

your desired file name

dataset.to\_csv(“tn\_age\_ditribution\_preprocessed\_dataset.csv”, index=False)