**IBM NAAN MUDHALVAN**

**PHASE 3 - DEVELOPMENT PHASE PART 1**

**DATA LOADING AND PREPROCESSING FOR MARGINAL WORKERS DATASET**

|  |  |
| --- | --- |
| **DOMAIN :** | **DATA ANALYTICS**  **WITH COGNOS** |
| **PROJECT TITLE :** | **TN MARGINAL WORKERS ASSESSMENT**  **A SOCIOECONOMIC ANALYSIS** |
| **TEAM MEMBERS :** | **BHARATH KUMAR B-420421104012**  **SANJAY KUMAR A-420421104068**  **NOOR MOHAMED K-420421104048**  **NITHISH KANNAN G-420421104302**  **MOHAMMED JAVITH M-420421104042** |

**Objective:**

The objective of this project is to load and preprocess the "Marginal Workers Classified by Age, Industrial Category, and Sex for Scheduled Caste (2011)" dataset from the Tamil Nadu government data portal.

**Tools and Libraries:**

* Python (3.x recommended)
* Jupyter Notebook (or any preferred code editor)
* Libraries: Pandas, NumPy, scikit-learn

**Steps:**

**1.Import Libraries:**

* Import necessary python libraries like pandas, sckit-learn



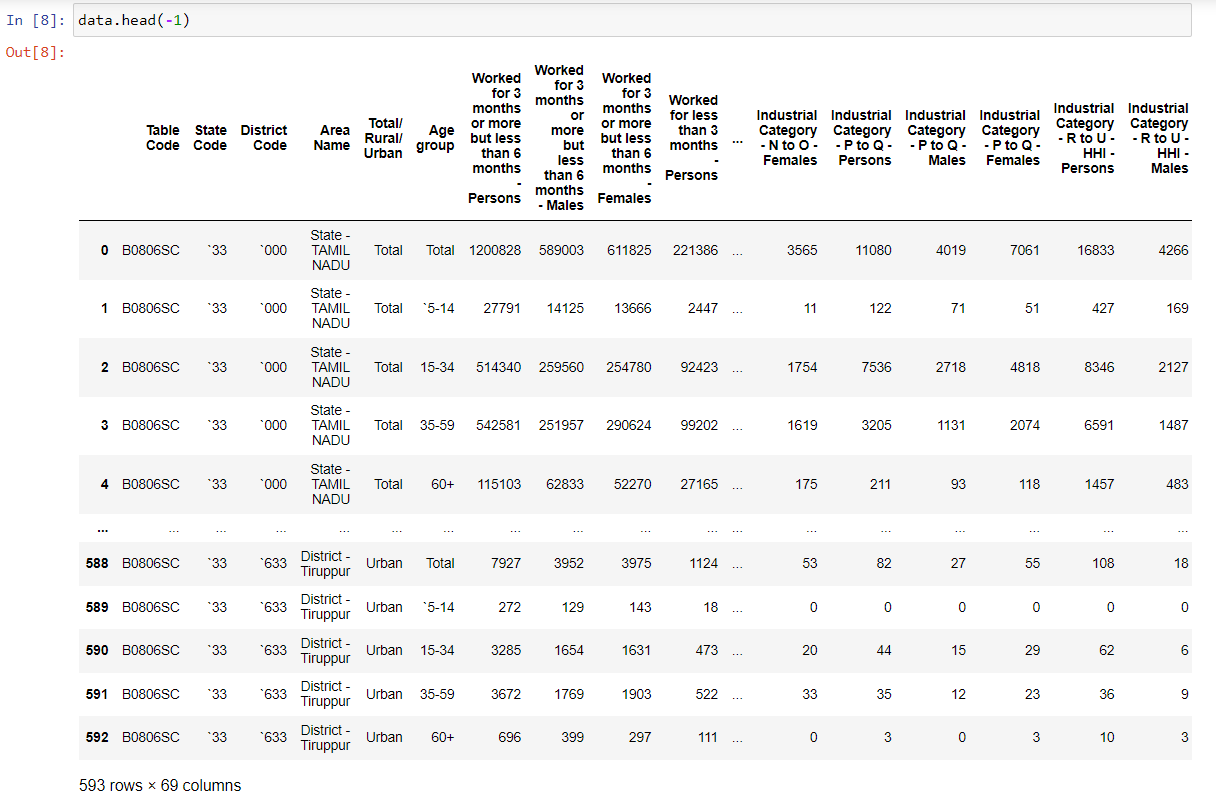
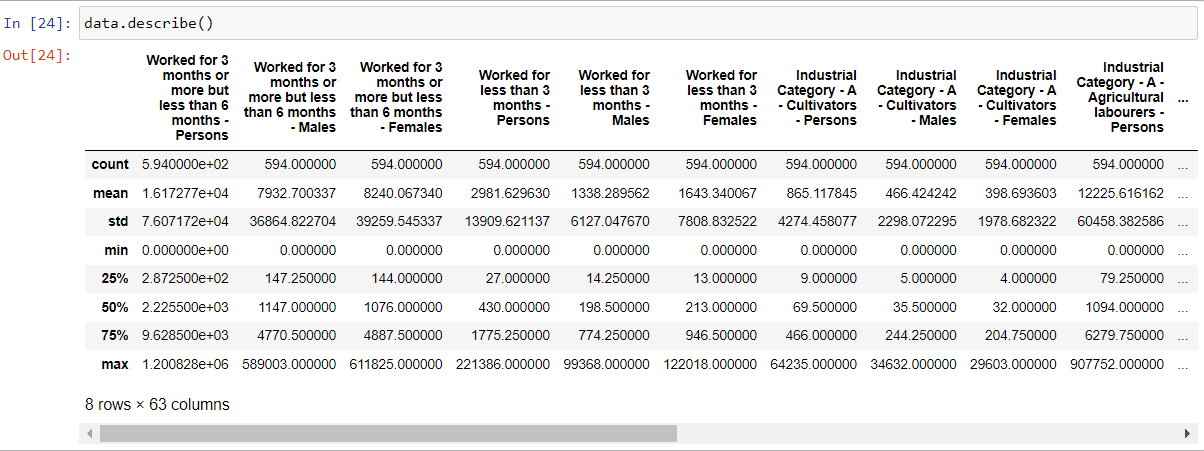
**2. Data Retrieval:**

* Visit the dataset URL: <https://tn.data.gov.in/resource/marginal-workers-classified-age-industrial-category-and-sex-scheduled-caste-2011-tamil>.
* Download the dataset in CSV format or any other suitable format if available.
* Save the dataset to your local machine for further analysis.

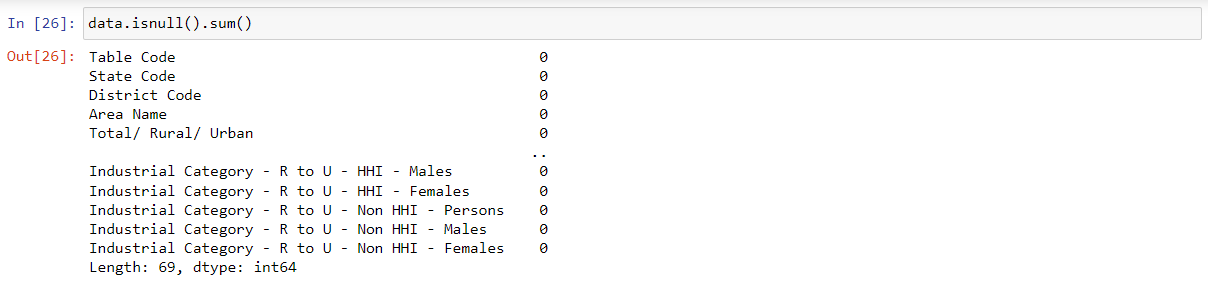
**3. Data Loading:**

* Open your Jupyter Notebook or preferred code editor.
* Create a new Python script.
* ****Use pandas to load a dataset

**4. Data Exploration:**

* Explore the dataset to understand its structure and contents.
* Check the first few rows to get a glimpse of the data.
* Examine basic statistics.****

**5. Data Preprocessing:**

* Handle missing values if any.
* Clean and format the data as needed.
* ****Perform any necessary data transformations.

**6. Save the Preprocessed Data:**

* After preprocessing, save the cleaned dataset to a new CSV file for future use
* Now transform the preprocessed data using Standard Scalar



**Conclusion:**

This document outlines the process of loading and preprocessing the "Marginal Workers Classified by Age, Industrial Category, and Sex for Scheduled Caste (2011)" dataset from the Tamil Nadu government data portal. Following these steps will enable you to prepare the data for further analysis and modeling in your project.