Assessment of Marginal Workers in TamilNadu

Step 1: Data Loading

We began by loading the dataset into a pandas DataFrame using Python.

```
In [1]: import pandas as pd

In [3]: data = pd.read_csv('dataset.csv')
        df = pd.DataFrame(data)
```

Step 2: Data Inspection

We then inspected the dataset to gain an understanding of its structure. We examined the first few rows to ensure data integrity and reviewed the data types and any missing values.

```
In [4]: print(data.head())
     Table Code State Code District Code        Area Name Total/ Rural/ Urban  \
  0     B0706        `33          `000  State - TAMIL NADU               Total
  1     B0706        `33          `000  State - TAMIL NADU               Total
  2     B0706        `33          `000  State - TAMIL NADU               Total
  3     B0706        `33          `000  State - TAMIL NADU               Total
  4     B0706        `33          `000  State - TAMIL NADU               Total

    Age group  Worked for 3 months or more but less than 6 months -  Persons  \
  0     Total                                              4218884
  1      `5-9                                                48238
  2    `10-14                                                76288
  3     15-19                                               257605
  4     20-24                                               478082

    Worked for 3 months or more but less than 6 months - Males  \
  0                                              2136881
  1                                                24511
  2                                                39191
  3                                               141262
```

```
In [5]: print(data.info())
  <class 'pandas.core.frame.DataFrame'>
  RangeIndex: 1386 entries, 0 to 1385
  Data columns (total 69 columns):
   #   Column
  Non-Null Count  Dtype
  ---  ------
  --------------  -----
   0   Table Code
  1386 non-null   object
   1   State Code
  1386 non-null   object
   2   District Code
  1386 non-null   object
   3   Area Name
  1386 non-null   object
   4   Total/ Rural/ Urban
  1386 non-null   object
   5   Age group
  1386 non-null   object
```

Step 3: Handling Duplicates

We checked for duplicate record and to remove them

```
In [8]: data = data.drop_duplicates()
```