PHASE – 5

PROJECT DOCUMENTATION AND SUBMISSION

Date	27-10-2023
Team ID	4570
Project Name	Assessment of Marginal Workers in Tamil Nadu

Table of Content:

1. Introduction
2. Problem Statement
3. Objective
4. Analysis Approach
    - 4.1 Data Collection
    - 4.2 Data Cleaning And Preparation
    - 4.3 Demographic Segmentation
    - 4.4 Descriptive Statistics
    - 4.5 Visualization
        - 4.5.1 Age Distribution
        - 4.5.2 Industrial Category Distribution
        - 4.5.3 Sex Distribution
        - 4.5.4 Cross-Category Analysis
5. Loading And Pre – Processing
6. Data Visualization
7. Data Visualization Using Cognos
8. Conclusion

1. INTRODUCTION

This project aims to conduct a comprehensive socioeconomic analysis focusing on marginal workers. The study will delve into the demographic attributes of this vital workforce, including age, industrial categorization, and gender distribution. To facilitate a clear understanding, visualizations in the form of bar charts, pie charts, and heat maps will be employed to depict the distribution across these distinct categories.

In this project, we will outline the problem statement, the steps involved in solving it, and the design thinking approach that will guide our project.

2. PROBLEM STATEMENT

Analyze the demographic characteristics of marginal workers based on age, industrial category, and sex. Create visualizations such as bar charts, pie charts, or heat maps to represent the distribution across different categories.

3. OBJECTIVE

The objective of this project is to conduct a comprehensive analysis of marginal workers based on their demographic characteristics, specifically focusing on age, industrial category, and sex. Marginal workers play a crucial role in the labor force, often facing unique challenges and opportunities. Understanding their demographic profile is essential for policy-making and resource allocation.

4. ANALYSIS APPROACH

4.1 Data Collection:

- Identify authoritative sources for demographic data on marginal workers. These may include government reports, labor market surveys, or specialized research studies.
- Obtain data on age, industrial category, and sex of marginal workers. Ensure the data is comprehensive and representative.

4.2 Data Cleaning and Preparation:

- Perform data cleaning to address any missing or erroneous values. This may involve imputation, removal, or correction of outliers.
- Standardize data formats and units to ensure consistency across variables.

4.3 Demographic Segmentation:

- Categorize marginal workers based on age groups (e.g., 18-25, 26-35, 36-45, 46-60, 60+), industrial categories (e.g., agriculture, manufacturing, construction, services), and sex (male/female).

4.4 Descriptive Statistics:

- Calculate summary statistics for each demographic category, including measures like mean, median, mode, and standard deviation. This provides an initial overview of the distribution.

4.5 Visualization:

Create visual representations using appropriate charts:

4.5.1 Age Distribution:

- Bar chart or histogram to display the distribution of marginal workers across different age groups.

4.5.2 Industrial Category Distribution:

- Pie chart or stacked bar chart to show the proportion of workers in each industrial category.

4.5.3 Sex Distribution:

- Pie chart or bar chart to represent the distribution of male, female, and other genders among marginal workers.

4.5.4 Cross-Category Analysis:

- Heat map to visualize the intersection of age groups, industrial categories, and sex.

This analysis approach provides a structured framework to systematically examine the demographic characteristics of marginal workers and generate meaningful insights.

5. LOADING AND PRE – PROCESSING

```
from google.colab import files
uploaded = files.upload()
```

```
<IPython.core.display.HTML object>
```

```
Saving Dataset.csv to Dataset.csv
```

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score, mean_absolute_error,mean_squared_error
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Lasso
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
```

```
#To know what are the column in the dataset
dataset.columns
```

```
Index(['Table Code', 'State Code', 'District Code', 'Area Name',
       'Total/ Rural/ Urban', 'Age group',
       'Worked for 3 months or more but less than 6 months -  Persons',
       'Worked for 3 months or more but less than 6 months - Males',
       'Worked for 3 months or more but less than 6 months - Females',
       'Worked for less than 3 months - Persons',
       'Worked for less than 3 months - Males',
       'Worked for less than 3 months - Females',
       'Industrial Category - A - Cultivators - Persons',
       'Industrial Category - A - Cultivators - Males',
       'Industrial Category - A - Cultivators - Females',
       'Industrial Category - A - Agricultural labourers - Persons',
       'Industrial Category - A - Agricultural labourers - Males',
       'Industrial Category - A - Agricultural labourers - Females',
       'Industrial Category - A - Plantation, Livestock, Forestry, Fishing,
  Hunting and allied activities - Persons',
       'Industrial Category - A - Plantation, Livestock, Forestry, Fishing,
  Hunting and allied activities - Males',
       'Industrial Category - A - Plantation, Livestock, Forestry, Fishing,
  Hunting and allied activities - Females',
       'Industrial Category - B - Persons', 'Industrial Category - B - Males',
       'Industrial Category - B - Females',
       'Industrial Category - C - HHI - Persons',
```

```
       'Industrial Category - C - HHI - Males',
       'Industrial Category - C - HHI - Females',
       'Industrial Category - C - Non HHI - Persons',
       'Industrial Category - C - Non HHI - Males',
       'Industrial Category - C - Non HHI - Females',
       'Industrial Category - D & E - Persons',
       'Industrial Category - D & E - Males',
       'Industrial Category - D & E - Females',
       'Industrial Category - F - Persons', 'Industrial Category - F - Males',
       'Industrial Category - F - Females',
       'Industrial Category - G - HHI - Persons',
       'Industrial Category - G - HHI - Males',
       'Industrial Category - G - HHI - Females',
       'Industrial Category - G - Non HHI - Persons',
       'Industrial Category - G - Non HHI - Males',
       'Industrial Category - G - Non HHI - Females',
       'Industrial Category - H - Persons', 'Industrial Category - H - Males',
       'Industrial Category - H - Females',
       'Industrial Category - I - Persons', 'Industrial Category - I - Males',
       'Industrial Category - I - Females',
       'Industrial Category - J - HHI - Persons',
       'Industrial Category - J - HHI - Males',
       'Industrial Category - J - HHI - Females',
       'Industrial Category - J - Non HHI - Persons',
       'Industrial Category - J - Non HHI - Males',
       'Industrial Category - J - Non HHI - Females',
       'Industrial Category - K to M - Persons',
       'Industrial Category - K to M - Males',
       'Industrial Category - K to M - Females',
       'Industrial Category - N to O - Persons',
       'Industrial Category - N to O - Males',
       'Industrial Category - N to O - Females',
       'Industrial Category - P to Q - Persons',
       'Industrial Category - P to Q - Males',
       'Industrial Category - P to Q - Females',
       'Industrial Category - R to U - HHI - Persons',
       'Industrial Category - R to U - HHI - Males',
       'Industrial Category - R to U - HHI - Females',
       'Industrial Category - R to U - Non HHI - Persons',
       'Industrial Category - R to U - Non HHI - Males',
       'Industrial Category - R to U - Non HHI - Females'],
      dtype='object')
```

12.
```
# TO view TOP Five rows in dataset
dataset.head()
```

13.
```
  Table Code State Code District Code          Area Name Total/ Rural/ Urban  \
0    B0806SC        `33          `000  State - TAMIL NADU               Total
1    B0806SC        `33          `000  State - TAMIL NADU               Total
2    B0806SC        `33          `000  State - TAMIL NADU               Total
3    B0806SC        `33          `000  State - TAMIL NADU               Total
4    B0806SC        `33          `000  State - TAMIL NADU               Total
```

	Age group	Worked for 3 months or more but less than 6 months - Persons	\
0	Total	1200828	
1	`5-14	27791	
2	15-34	514340	
3	35-59	542581	
4	60+	115103	

	Worked for 3 months or more but less than 6 months - Males	\
0	589003	
1	14125	
2	259560	
3	251957	
4	62833	

	Worked for 3 months or more but less than 6 months - Females	\
0	611825	
1	13666	
2	254780	
3	290624	
4	52270	

	Worked for less than 3 months - Persons	...	\
0	221386	...	
1	2447	...	
2	92423	...	
3	99202	...	
4	27165	...	

	Industrial Category - N to O - Females	\
0	3565	
1	11	
2	1754	
3	1619	
4	175	

	Industrial Category - P to Q - Persons	\
0	11080	
1	122	
2	7536	
3	3205	
4	211	

	Industrial Category - P to Q - Males	\
0	4019	
1	71	
2	2718	
3	1131	
4	93	

	Industrial Category - P to Q - Females	\
0	7061	
1	51	

```
2                                    4818
3                                    2074
4                                     118

   Industrial Category - R to U - HHI - Persons  \
0                                          16833
1                                            427
2                                           8346
3                                           6591
4                                           1457

   Industrial Category - R to U - HHI - Males  \
0                                         4266
1                                          169
2                                         2127
3                                         1487
4                                          483

   Industrial Category - R to U - HHI - Females  \
0                                         12567
1                                           258
2                                          6219
3                                          5104
4                                           974

   Industrial Category - R to U - Non HHI - Persons  \
0                                           122088
1                                            19305
2                                            68929
3                                            26498
4                                             7065

   Industrial Category - R to U - Non HHI - Males  \
0                                           55801
1                                            9774
2                                           32803
3                                            9675
4                                            3394

   Industrial Category - R to U - Non HHI - Females
0                                           66287
1                                            9531
2                                           36126
3                                           16823
4                                            3671

[5 rows x 69 columns]
```

14. ```
    # To view BOTTOM five rows in dataset
    dataset.tail()
    ```
    ```

15.

	Table Code	State Code	District Code	Area Name
589	B0806SC	`33	`633	District - Tiruppur
590	B0806SC	`33	`633	District - Tiruppur
591	B0806SC	`33	`633	District - Tiruppur
592	B0806SC	`33	`633	District - Tiruppur
593	B0806SC	`33	`633	District - Tiruppur

	Total/ Rural/ Urban	Age group
589	Urban	`5-14
590	Urban	15-34
591	Urban	35-59
592	Urban	60+
593	Urban	Age not stated

	Worked for 3 months or more but less than 6 months - Persons
589	272
590	3285
591	3672
592	696
593	2

	Worked for 3 months or more but less than 6 months - Males
589	129
590	1654
591	1769
592	399
593	1

	Worked for 3 months or more but less than 6 months - Females
589	143
590	1631
591	1903
592	297
593	1

	Worked for less than 3 months - Persons	...
589	18	...
590	473	...
591	522	...
592	111	...
593	0	...

	Industrial Category - N to O - Females
589	0
590	20
591	33
592	0
593	0

	Industrial Category - P to Q - Persons
589	0
590	44

591		35
592		3
593		0

	Industrial Category - P to Q - Males \
589	0
590	15
591	12
592	0
593	0

	Industrial Category - P to Q - Females \
589	0
590	29
591	23
592	3
593	0

	Industrial Category - R to U - HHI - Persons \
589	0
590	62
591	36
592	10
593	0

	Industrial Category - R to U - HHI - Males \
589	0
590	6
591	9
592	3
593	0

	Industrial Category - R to U - HHI - Females \
589	0
590	56
591	27
592	7
593	0

	Industrial Category - R to U - Non HHI - Persons \
589	228
590	675
591	279
592	81
593	0

	Industrial Category - R to U - Non HHI - Males \
589	104
590	247
591	103
592	35
593	0

```
     Industrial Category - R to U - Non HHI - Females
589                                                124
590                                                428
591                                                176
592                                                 46
593                                                  0

[5 rows x 69 columns]
```

16. `dataset.describe()`

17.
```
       Worked for 3 months or more but less than 6 months -  Persons  \
count                                       5.940000e+02
mean                                        1.617277e+04
std                                         7.607172e+04
min                                         0.000000e+00
25%                                         2.872500e+02
50%                                         2.225500e+03
75%                                         9.628500e+03
max                                         1.200828e+06

       Worked for 3 months or more but less than 6 months - Males  \
count                                         594.000000
mean                                         7932.700337
std                                         36864.822704
min                                             0.000000
25%                                           147.250000
50%                                          1147.000000
75%                                          4770.500000
max                                        589003.000000

       Worked for 3 months or more but less than 6 months - Females  \
count                                         594.000000
mean                                         8240.067340
std                                         39259.545337
min                                             0.000000
25%                                           144.000000
50%                                          1076.000000
75%                                          4887.500000
max                                        611825.000000

       Worked for less than 3 months - Persons  \
count                                 594.000000
mean                                 2981.629630
std                                 13909.621137
min                                     0.000000
25%                                    27.000000
50%                                   430.000000
75%                                  1775.250000
max                                221386.000000
```

```
       Worked for less than 3 months - Males  \
count                             594.000000
mean                             1338.289562
std                              6127.047670
min                                 0.000000
25%                                14.250000
50%                               198.500000
75%                               774.250000
max                             99368.000000

       Worked for less than 3 months - Females  \
count                               594.000000
mean                               1643.340067
std                                7808.832522
min                                   0.000000
25%                                  13.000000
50%                                 213.000000
75%                                 946.500000
max                              122018.000000

       Industrial Category - A - Cultivators - Persons  \
count                                        594.000000
mean                                         865.117845
std                                         4274.458077
min                                            0.000000
25%                                            9.000000
50%                                           69.500000
75%                                          466.000000
max                                        64235.000000

       Industrial Category - A - Cultivators - Males  \
count                                      594.000000
mean                                       466.424242
std                                       2298.072295
min                                          0.000000
25%                                          5.000000
50%                                         35.500000
75%                                        244.250000
max                                      34632.000000

       Industrial Category - A - Cultivators - Females  \
count                                        594.000000
mean                                         398.693603
std                                         1978.682322
min                                            0.000000
25%                                            4.000000
50%                                           32.000000
75%                                          204.750000
max                                        29603.000000

       Industrial Category - A - Agricultural labourers - Persons  ...  \
count                                          594.000000           ...
```

```
mean                                        12225.616162          ...
std                                         60458.382586          ...
min                                             0.000000          ...
25%                                            79.250000          ...
50%                                          1094.000000          ...
75%                                          6279.750000          ...
max                                        907752.000000          ...

       Industrial Category - N to O - Females  \
count                                594.000000
mean                                  48.013468
std                                  222.553500
min                                    0.000000
25%                                    0.000000
50%                                    2.000000
75%                                   18.000000
max                                 3565.000000

       Industrial Category - P to Q - Persons  \
count                                594.000000
mean                                 149.225589
std                                  696.553730
min                                    0.000000
25%                                    0.000000
50%                                   14.500000
75%                                   99.750000
max                                11080.000000

       Industrial Category - P to Q - Males  \
count                              594.000000
mean                                54.127946
std                                253.067862
min                                  0.000000
25%                                  0.000000
50%                                  6.000000
75%                                 35.750000
max                               4019.000000

       Industrial Category - P to Q - Females  \
count                                594.000000
mean                                  95.097643
std                                  444.011425
min                                    0.000000
25%                                    0.000000
50%                                    6.500000
75%                                   64.000000
max                                 7061.000000

       Industrial Category - R to U - HHI - Persons  \
count                                     594.000000
mean                                      226.707071
std                                      1039.953069
```

```
min                                          0.000000
25%                                          0.000000
50%                                         27.000000
75%                                        126.750000
max                                      16833.000000

       Industrial Category - R to U - HHI - Males  \
count                                    594.000000
mean                                      57.454545
std                                      265.230865
min                                        0.000000
25%                                        0.000000
50%                                        7.500000
75%                                       32.000000
max                                     4266.000000

       Industrial Category - R to U - HHI - Females  \
count                                      594.000000
mean                                       169.252525
std                                        776.206806
min                                          0.000000
25%                                          0.000000
50%                                         20.000000
75%                                         97.500000
max                                      12567.000000

       Industrial Category - R to U - Non HHI - Persons  \
count                                         594.000000
mean                                         1644.282828
std                                          7325.241597
min                                             0.000000
25%                                            64.500000
50%                                           263.500000
75%                                           994.000000
max                                        122088.000000

       Industrial Category - R to U - Non HHI - Males  \
count                                       594.000000
mean                                        751.528620
std                                        3352.811737
min                                           0.000000
25%                                          34.000000
50%                                         123.000000
75%                                         447.750000
max                                       55801.000000

       Industrial Category - R to U - Non HHI - Females
count                                         594.000000
mean                                          892.754209
std                                          3988.125301
min                                             0.000000
25%                                            30.500000
```

```
50%                                              135.000000
75%                                              500.000000
max                                            66287.000000

[8 rows x 63 columns]
```

```
Warning: Total number of columns (63) exceeds max_columns (20) limiting to first
(20) columns.
```

```
#To find how many null values
print(dataset.isnull().sum())
```

```
Table Code                                             0
State Code                                             0
District Code                                          0
Area Name                                              0
Total/ Rural/ Urban                                    0
                                                      ..
Industrial Category - R to U - HHI - Males             0
Industrial Category - R to U - HHI - Females           0
Industrial Category - R to U - Non HHI - Persons       0
Industrial Category - R to U - Non HHI - Males         0
Industrial Category - R to U - Non HHI - Females       0
Length: 69, dtype: int64
```

```
#How does the data looks like
print(dataset.shape)          #Total ROWS and COLUMN
print("\n")
print(dataset.dtypes)         #to know the data type each column
```

```
(594, 69)


Table Code                                             object
State Code                                             object
District Code                                          object
Area Name                                              object
Total/ Rural/ Urban                                    object
                                                        ...
Industrial Category - R to U - HHI - Males              int64
Industrial Category - R to U - HHI - Females            int64
Industrial Category - R to U - Non HHI - Persons        int64
Industrial Category - R to U - Non HHI - Males          int64
Industrial Category - R to U - Non HHI - Females        int64
Length: 69, dtype: object
```

```
dataset.info()    #This will provide
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 594 entries, 0 to 593
Data columns (total 69 columns):
 #   Column
Non-Null Count  Dtype
```

```
---  ------
--------------  -----
 0   Table Code
594 non-null    object
 1   State Code
594 non-null    object
 2   District Code
594 non-null    object
 3   Area Name
594 non-null    object
 4   Total/ Rural/ Urban
594 non-null    object
 5   Age group
594 non-null    object
 6   Worked for 3 months or more but less than 6 months -  Persons
594 non-null    int64
 7   Worked for 3 months or more but less than 6 months - Males
594 non-null    int64
 8   Worked for 3 months or more but less than 6 months - Females
594 non-null    int64
 9   Worked for less than 3 months - Persons
594 non-null    int64
 10  Worked for less than 3 months - Males
594 non-null    int64
 11  Worked for less than 3 months - Females
594 non-null    int64
 12  Industrial Category - A - Cultivators - Persons
594 non-null    int64
 13  Industrial Category - A - Cultivators - Males
594 non-null    int64
 14  Industrial Category - A - Cultivators - Females
594 non-null    int64
 15  Industrial Category - A - Agricultural labourers - Persons
594 non-null    int64
 16  Industrial Category - A - Agricultural labourers - Males
594 non-null    int64
 17  Industrial Category - A - Agricultural labourers - Females
594 non-null    int64
 18  Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting
and allied activities - Persons  594 non-null    int64
 19  Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting
and allied activities - Males    594 non-null    int64
 20  Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting
and allied activities - Females  594 non-null    int64
 21  Industrial Category - B - Persons
594 non-null    int64
 22  Industrial Category - B - Males
594 non-null    int64
 23  Industrial Category - B - Females
594 non-null    int64
 24  Industrial Category - C - HHI - Persons
594 non-null    int64
 25  Industrial Category - C - HHI - Males
```

```
594 non-null    int64
 26  Industrial Category - C - HHI - Females
594 non-null    int64
 27  Industrial Category - C - Non HHI - Persons
594 non-null    int64
 28  Industrial Category - C - Non HHI - Males
594 non-null    int64
 29  Industrial Category - C - Non HHI - Females
594 non-null    int64
 30  Industrial Category - D & E - Persons
594 non-null    int64
 31  Industrial Category - D & E - Males
594 non-null    int64
 32  Industrial Category - D & E - Females
594 non-null    int64
 33  Industrial Category - F - Persons
594 non-null    int64
 34  Industrial Category - F - Males
594 non-null    int64
 35  Industrial Category - F - Females
594 non-null    int64
 36  Industrial Category - G - HHI - Persons
594 non-null    int64
 37  Industrial Category - G - HHI - Males
594 non-null    int64
 38  Industrial Category - G - HHI - Females
594 non-null    int64
 39  Industrial Category - G - Non HHI - Persons
594 non-null    int64
 40  Industrial Category - G - Non HHI - Males
594 non-null    int64
 41  Industrial Category - G - Non HHI - Females
594 non-null    int64
 42  Industrial Category - H - Persons
594 non-null    int64
 43  Industrial Category - H - Males
594 non-null    int64
 44  Industrial Category - H - Females
594 non-null    int64
 45  Industrial Category - I - Persons
594 non-null    int64
 46  Industrial Category - I - Males
594 non-null    int64
 47  Industrial Category - I - Females
594 non-null    int64
 48  Industrial Category - J - HHI - Persons
594 non-null    int64
 49  Industrial Category - J - HHI - Males
594 non-null    int64
 50  Industrial Category - J - HHI - Females
594 non-null    int64
 51  Industrial Category - J - Non HHI - Persons
594 non-null    int64
```

```
 52  Industrial Category - J - Non HHI - Males
594 non-null    int64
 53  Industrial Category - J - Non HHI - Females
594 non-null    int64
 54  Industrial Category - K to M - Persons
594 non-null    int64
 55  Industrial Category - K to M - Males
594 non-null    int64
 56  Industrial Category - K to M - Females
594 non-null    int64
 57  Industrial Category - N to O - Persons
594 non-null    int64
 58  Industrial Category - N to O - Males
594 non-null    int64
 59  Industrial Category - N to O - Females
594 non-null    int64
 60  Industrial Category - P to Q - Persons
594 non-null    int64
 61  Industrial Category - P to Q - Males
594 non-null    int64
 62  Industrial Category - P to Q - Females
594 non-null    int64
 63  Industrial Category - R to U - HHI - Persons
594 non-null    int64
 64  Industrial Category - R to U - HHI - Males
594 non-null    int64
 65  Industrial Category - R to U - HHI - Females
594 non-null    int64
 66  Industrial Category - R to U - Non HHI - Persons
594 non-null    int64
 67  Industrial Category - R to U - Non HHI - Males
594 non-null    int64
 68  Industrial Category - R to U - Non HHI - Females
594 non-null    int64
dtypes: int64(63), object(6)
memory usage: 320.3+ KB
```

6. DATA VISUALIZATION:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score, mean_absolute_error,mean_squared_error
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Lasso
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
```

```
from google.colab import files
uploaded = files.upload()
```

```
<IPython.core.display.HTML object>
```

```
Saving Dataset.csv to Dataset.csv
```

```
df=pd.read_csv('Dataset.csv')
```

```
import matplotlib.pyplot as plt

# Assuming you have a DataFrame called df with your data
# You can extract the required columns
age_groups = df["Age group"]
males = df["Worked for 3 months or more but less than 6 months - Males"]
females = df["Worked for 3 months or more but less than 6 months - Females"]

# Create a figure and a set of subplots
fig, ax = plt.subplots()

# Define the width of the bars
width = 0.25

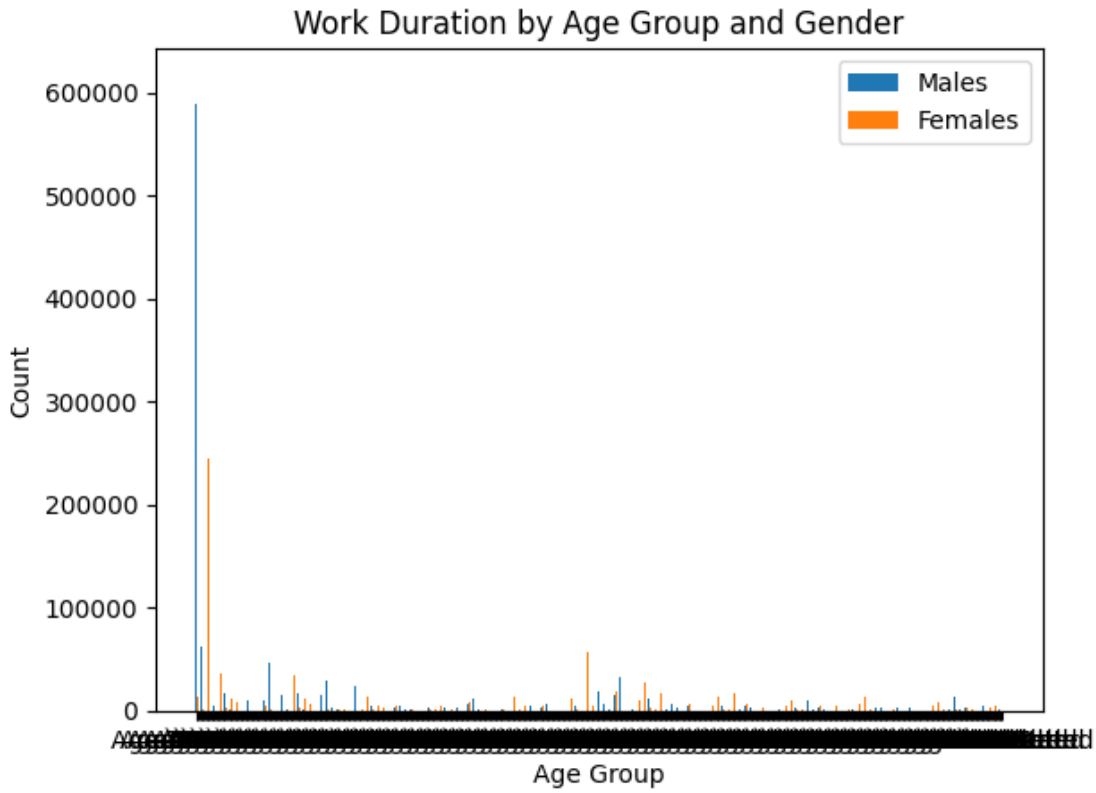
# Set the positions of the bars on the x-axis
x = range(len(age_groups))

# Create the bars
plt.bar([i + width for i in x], males, width, label='Males')
plt.bar([i + width*2 for i in x], females, width, label='Females')

# Set the x-axis ticks and labels
plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Work Duration by Age Group and Gender')
plt.xticks([i + width for i in x], age_groups)

# Add a legend
plt.legend()

# Show the plot
plt.show()
```




```
import matplotlib.pyplot as plt

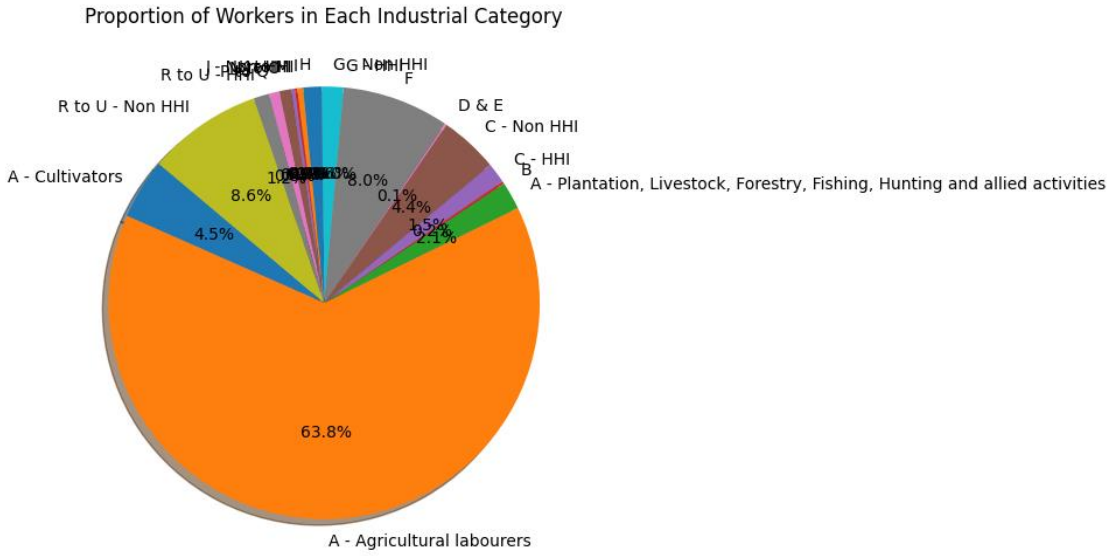
# Define the categories and corresponding column names
categories = ['A - Cultivators', 'A - Agricultural labourers', 'A - Plantation,
Livestock, Forestry, Fishing, Hunting and allied activities',
              'B', 'C - HHI', 'C - Non HHI', 'D & E', 'F', 'G - HHI', 'G - Non
HHI', 'H', 'I', 'J - HHI', 'J - Non HHI', 'K to M',
              'N to O', 'P to Q', 'R to U - HHI', 'R to U - Non HHI']

column_names = ['Industrial Category - A - Cultivators - Persons', 'Industrial
Category - A - Agricultural labourers - Persons', 'Industrial Category - A -
Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons',
              'Industrial Category - B - Persons', 'Industrial Category - C - HHI
- Persons', 'Industrial Category - C - Non HHI - Persons', 'Industrial Category - D
& E - Persons', 'Industrial Category - F - Persons', 'Industrial Category - G - HHI
- Persons', 'Industrial Category - G - Non HHI - Persons', 'Industrial Category - H
- Persons', 'Industrial Category - I - Persons', 'Industrial Category - J - HHI -
Persons', 'Industrial Category - J - Non HHI - Persons', 'Industrial Category - K
to M - Persons',
              'Industrial Category - N to O - Persons', 'Industrial Category - P
to Q - Persons', 'Industrial Category - R to U - HHI - Persons', 'Industrial
Category - R to U - Non HHI - Persons']

# Extract the data for industrial categories
category_counts = [df[column].sum() for column in column_names]
```

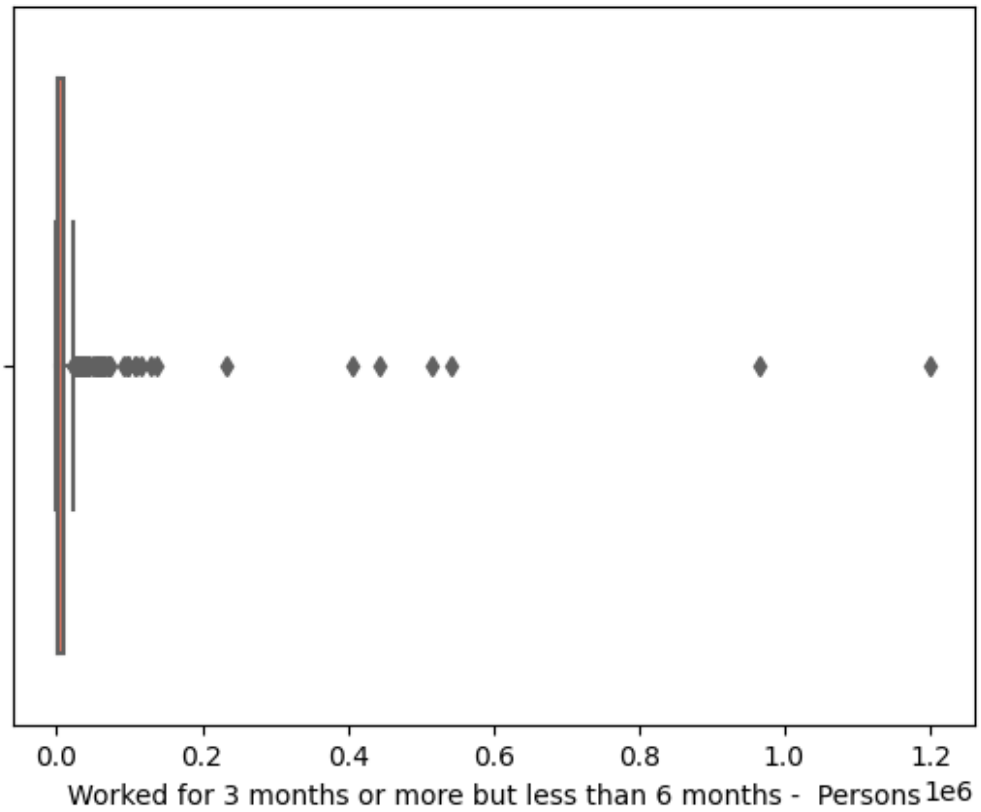
```
# Create a pie chart
plt.figure(figsize=(10, 6))
plt.pie(category_counts, labels=categories, autopct='%1.1f%%', startangle=140,
shadow=True)
plt.title('Proportion of Workers in Each Industrial Category')

# Display the pie chart
plt.show()
```

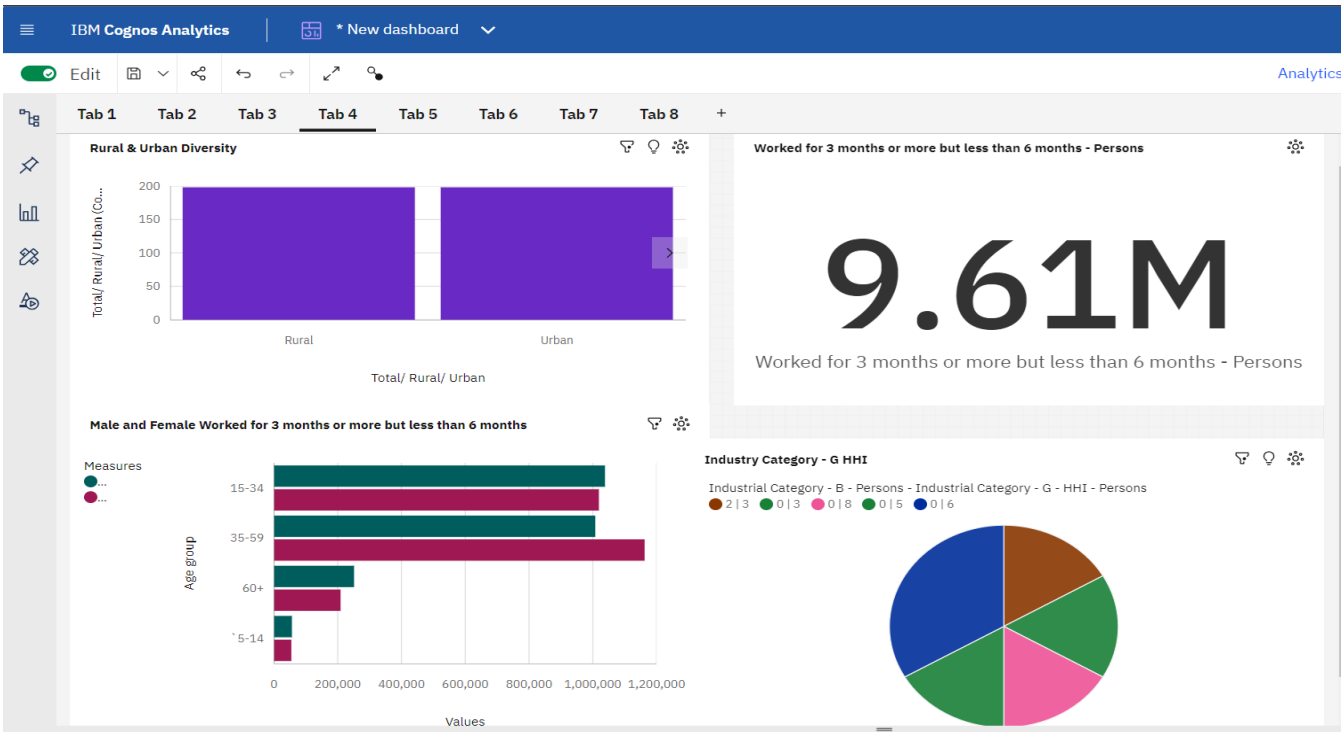



```
sns.boxplot(dataset, x='Worked for 3 months or more but less than 6 months -
Persons', palette='Reds')
```

```
<Axes: xlabel='Worked for 3 months or more but less than 6 months -  Persons'>
```



7. DATA VISUALIZATION USING COGNOS



8. CONCLUSION:

The project concludes by summarizing the data analysis work, emphasizing the use of visualization libraries like Matplotlib and Seaborn, and the application of data-driven techniques for understanding marginal workers In this project, we embarked on a comprehensive journey to understand and optimize marginal workers through data analysis. By employing a structured approach and leveraging powerful data analysis tools, we've unveiled insights and established a foundation for data-driven decision-making within the socio economic sector. Throughout the project, we've emphasized the importance of data preprocessing as a critical step. It is essential for refining and enhancing the quality of the data, which, in turn, paves the way for more accurate predictions and insights. These insights have the potential to support a wide range of marginal workers improvement initiatives, ultimately benefiting commuters and urban development. Through the fusion of cutting-edge technologies and methodologies, our ultimate goal is to provide a comprehensive and insightful solution for evaluating and enhancing marginal workers efficiency.