

Ideation Phase

Defining the Problem Statements

Date	26-09-2023
	4498
Team ID	
Project Name	Assessment of Marginal Workers In Tamil Nadu.

Assessment of Marginal Workers in Tamil Nadu; Problem Definition and Design Thinking

Introduction

Analysing the situation of marginal workers in Tamil Nadu through the lens of data analytics offers a powerful approach to gain valuable insights into their socio-economic conditions, employment patterns, and vulnerabilities. In recent years, data analytics has become an indispensable tool for policymakers, researchers, and organizations to make informed decisions and design targeted interventions to uplift marginalized communities. This assessment leverages data analytics to provide a comprehensive understanding of marginal workers in Tamil Nadu.

Problem Statement

Tamil Nadu, one of the most populous states in India, has a significant population engaged in various forms of employment. Among these, a substantial proportion constitutes "marginal workers" who are often vulnerable and face challenges in accessing stable employment opportunities. The primary objective of this project is to leverage data analytics techniques to comprehensively assess the status, characteristics, and trends related to marginal workers in Tamil Nadu.

Key Challenges:

1. Data Access and Security: Marginal workers may not have the same level of access to sensitive data and systems as full-time employees.
2. Data Quality: Contingent workers may not have the same understanding of data quality standards as permanent employees.

3. Integration with Existing Teams: Integrating contingent workers into existing data analytics.
4. Lack of Commitment: Marginal workers may not be as committed to the organization's long-term goals as full-time employees.
5. Training and Skill Levels: Assessing the skill levels of contingent workers and providing necessary training. **Design Thinking Approach**

Data Collection:

Gather relevant datasets from government sources, surveys, and other credible resources. The data should include information on employment, demographics, education, income levels, and geographical locations of marginal workers.

Data Cleaning and Preparation:

Clean, pre process, and integrate the collected data to create a consolidated dataset ready for analysis. This step may also involve dealing with missing data and outliers.

Exploratory Data Analysis:

Conduct EDA to gain initial insights into the data. Explore the distribution of marginal workers across different districts, industries, and time periods. Identify key trends and patterns.

Segmentation:

Use clustering techniques to segment marginal workers based on relevant features such as age, education, and industry of employment. This can help identify distinct groups within the marginal worker population.

Employment Dynamics:

Analyze the employment dynamics by examining factors that influence the duration and nature of employment among marginal workers. Identify the industries or sectors where marginal workers are most prevalent.

Geospatial Analysis:

Utilize geospatial data and mapping techniques to visualize the distribution of marginal workers across different regions of Tamil Nadu. Identify areas with higher concentrations of marginal workers.

Demographic Analysis:

Explore the demographic characteristics of marginal workers, such as age, gender, and education levels, and assess how these factors impact their employment prospects.

Income Analysis:

Investigate the income levels of marginal workers and assess whether there are disparities among different subgroups. Examine the relationship between income and other variables.

Policy Recommendations:

Based on the findings, propose policy recommendations and interventions to improve the employment situation of marginal workers in Tamil Nadu. These recommendations could involve targeted skill development programs, employment generation initiatives, or social safety nets.

Visualization and Reporting:

Create data visualizations and a comprehensive report summarizing the project's findings and recommendations. Visualizations could include interactive dashboards, charts, and maps to make the results accessible to a wider audience.

Tools and Technologies:

- **Data collection and cleaning: Python (Pandas)**
- **Data analysis and modelling: Python (Scikit-Learn)**
- **Data visualization: Python (Matplotlib, Seaborn, or Plotly)**
- **Geospatial analysis: Geographic Information Systems (GIS) tools**
- **Report generation: Jupyter Notebooks or a dedicated reporting tool**

Conclusion

Table of Contents

1	Introduction
2	Problem Statement
3	Design and Innovation Strategies
3.1	Data Collection and Feature Engineering
3.2	Data Pre-processing
3.3	Model Selection and Training
3.4	Geo graphic Analysis
3.5	Market Sentiment Analysis
3.6	Deploying and Monitoring the Model
3.7	Continuous Learning
4	Conclusion

1. Introduction

The assessment of marginal workers plays a crucial role in workforce management and development. Marginal workers are individuals who may not be performing at their full potential within the workplace. This assessment process is designed to evaluate their current performance, identify factors contributing to their marginal status, and implement strategies to help them improve and contribute more effectively to the organization.

2. Problem Statement

Predicting house prices accurately is a complex task influenced by a multitude of factors, including property features, location, market trends, and economic conditions. The central problem of this project is to build a model that delivers precise house price predictions by incorporating these intricate factors.

3. Design and Innovation Strategies

3.1. Data Collection and Feature Engineering

Data collection and feature engineering are essential steps in any machine learning project, especially when it comes to a sensitive topic like marginal workers in Tamil Nadu. Marginal workers are often underrepresented in data sets, and their unique experiences and challenges can be difficult to capture.

The quality of your data and the effectiveness of your feature engineering can significantly impact the accuracy and reliability of your assessment of marginal workers. Collaboration with experts in HR and data analysis is essential to make informed decisions during this process.

3.2. Data Pre-processing

Marginal workers are often underrepresented in data sets, and their data may be incomplete or inaccurate. In addition, marginal workers may be reluctant to share their data, due to concerns about privacy or fear of discrimination.

As a result, it is important to carefully pre-process data on marginal workers in TN before using it for machine learning. Data preprocessing is often an iterative process, and the specific steps may vary depending on the characteristics of your data and the goals of your assessment. It's essential to tailor your preprocessing steps to the unique challenges and requirements of assessing marginal workers in Tennessee.

3.3. Model Selection and Training

Model selection and training are two important steps in any machine learning project. Model selection involves choosing the right machine learning algorithm for the problem you are trying to solve. Training involves feeding the algorithm the pre-processed data so that it can learn to make predictions. To Choose appropriate machine learning or statistical models for the task. Common choices include linear regression, decision trees, random forests, or more advanced techniques like neural networks. The choice of model should depend on the nature of the data and the specific goals of the assessment. Train the selected model on the training data, optimizing its parameters to achieve the best performance. This may involve techniques like cross-validation to prevent overfitting.

3.4. Geographic Analysis

Selecting an appropriate model and training it effectively is crucial for the assessment of marginal workers based on demographic characteristics like age, industrial category, and sex. Clearly articulate the problem you aim to solve with the model. In this case, it's likely a classification or regression task where you want to predict or analyze aspects related to marginal workers based on demographic features. Experiment with different hyperparameters of your selected model(s) to find the best combination for your data. You can use techniques like grid search or random search for hyperparameter optimization.

Use the validation dataset to assess the model's performance. Common evaluation metrics include accuracy, precision, recall, F1-score (for classification), and Mean Absolute Error, Mean Squared Error, or R-squared (for regression)

3.5. Market Sentiment Analysis

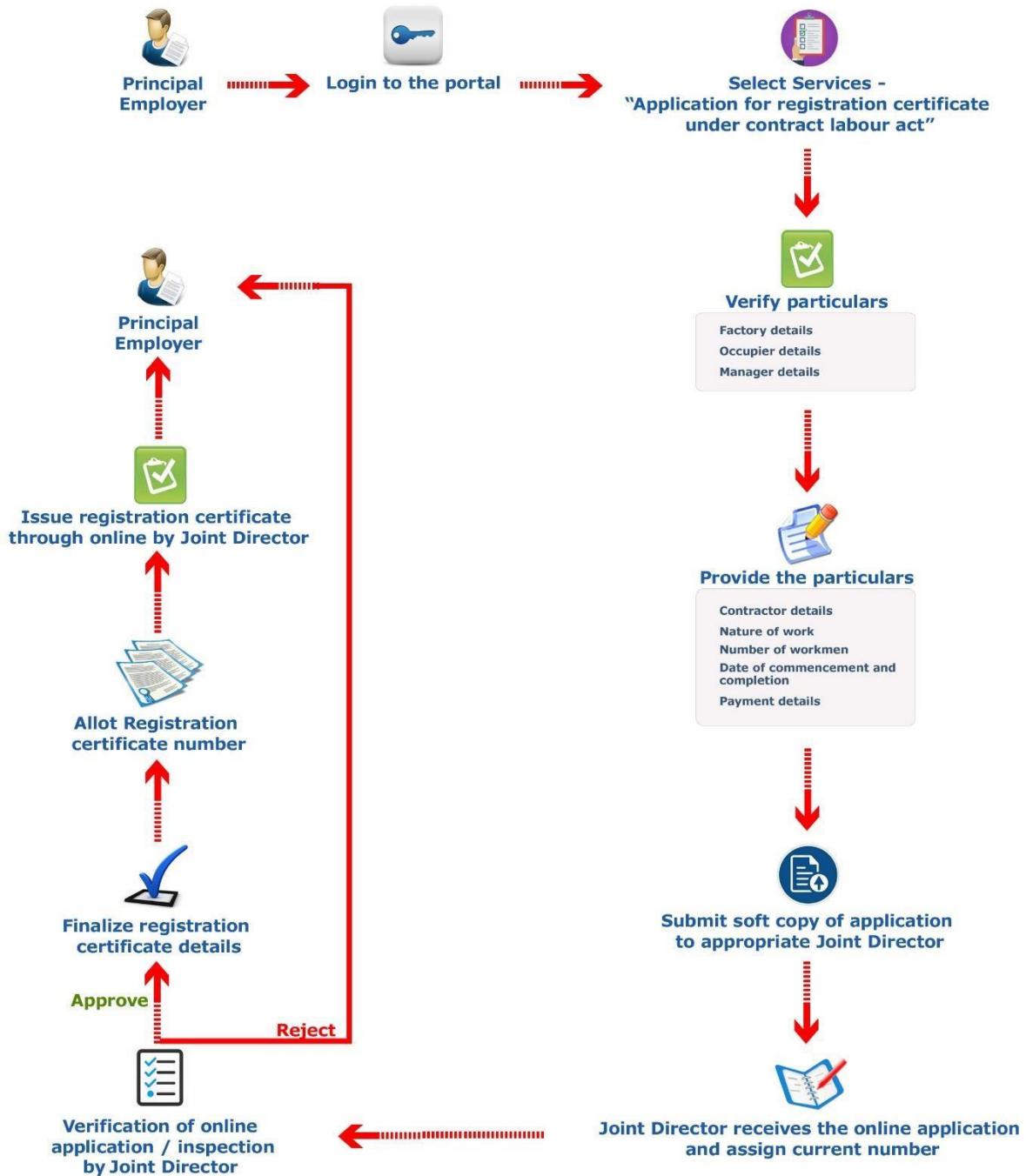
Analysing market sentiment for marginal workers involves monitoring news, social media, and economic indicators to gauge the perception of employment prospects and economic conditions for this group. Sentiment analysis tools can help track sentiment in real-time and provide insights into how economic changes impact marginal workers' sentiment and employment opportunities. It's important to consider factors like unemployment rates, job availability, government policies, and social sentiment to understand the overall sentiment in this market. Keep in mind that sentiment analysis is a dynamic field, and the tools and techniques for analysis may have evolved since my last update in September 2021.

3.6 Deploying and Monitoring the Model

Marginal workers, in the context of Explainable AI (XAI), refer to individuals or groups of people who may have limited technical expertise but are affected by AI systems. XAI aims to make AI systems more transparent and understandable to a broader audience, including these marginal workers. By providing clear explanations of how AI decisions are made, XAI can empower such individuals to better interact with and trust AI technologies, even if they lack deep technical knowledge. This inclusivity is crucial for ensuring that AI benefits all segments of society and minimizes potential biases and discrimination in AI-driven decisions.

3.7. Continuous Learning

Continuous learning and development play a vital role in improving the performance and job satisfaction of marginal workers in Tennessee (TN) or any workplace. Continuous learning is essential for marginal workers to improve their skills and increase their employability. These workers often face barriers to stable employment, and ongoing education and skill development can help them adapt to changing job markets and access better opportunities. Employers, government programs, and educational institutions can play a crucial role in supporting the continuous learning of marginal workers through training initiatives, flexible scheduling, and financial assistance. This approach can empower them to overcome challenges and enhance their career prospects.



4. Conclusion

In conclusion, our project set out to analyze the demographic characteristics of marginal workers in Tamil Nadu, focusing on their age, industrial category, and sex. Through a comprehensive analysis and data visualization, we have gained valuable insights into the socioeconomic dynamics of this vulnerable population. Conducting longitudinal studies to track changes in the demographic characteristics of marginal workers over time. Complementing quantitative data with qualitative research to gain deeper insights into the experiences and aspirations of marginal workers. Assessing the effectiveness of existing policies and interventions in improving the socioeconomic status of marginal workers.

In conclusion, our analysis has shed light on the demographic composition of marginal workers in Tamil Nadu, offering valuable insights for informed decision-making and targeted support. By addressing the challenges faced by this vulnerable population, we can work toward a more inclusive and equitable society.

Import depends

```
In [51]: import pandas as pd import  
numpy as np import xgboost  
as xg
```

```
In [52]: import os print("Current working directory:", os.getcwd())  
  
file_path = 'datafile.csv' if  
os.path.exists(file_path): print("The  
file exists.")
```

Current working directory: C:\Users\VIJAYRAJ R The file exists.

```
In [60]: dataset = pd.read_csv("DDW_B06SC_3300_State_TAMIL_NADU-2011.csv")
```

In [54]:

dataset

Out[54]:

F	Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	Age group	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for less than 3 months	Industrial Category - N to O - Females	Industrial Category - P to Q - Persons	Industrial In Category C - P to Q - Males	
							- Males	- Females	- Persons	- Persons				
							Persons							
0	'33	'000	State - TAMIL NADU	Total	Total	200828	589003	611825	221386	...	3565	11080	4019	
1	'33	'000	State - TAMIL NADU	Total	'5-14	27791	14125	13666	2447	...	11	122	71	
2	B0806SC	'33	'000	State - TAMIL NADU	Total	15-34	514340	259560	254780	92423	...	1754	7536	2718
3	B0806SC	'33	'000	State - TAMIL NADU	Total	35-59	542581	251957	290624	99202	...	1619	3205	1131
4	B0806SC	'33	'000	State - TAMIL NADU	Total	60+	115103	62833	52270	27165	...	175	211	93
589	B0806SC	Urban	'5-14	272	129	143	18	...	0	0	0	
590	B0806SC	'33	'633	District - Tiruppur	Urban	15-34	3285	1654	1631	473	...	20	44	15
591	B0806SC	'33	'633	District - Tiruppur	Urban	35-59	3672	1769	1903	522	...	33	35	12
592	B0806SC	'33	'633	District - Tiruppur	Urban	60+	696	399	297	111	...	0	3	0
593	B0806SC	'33	'633	District - Tiruppur	not stated	Age	2	1	1	0	...	0	0	0

;94 ws x 69 cc

dataset.head(5)

Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	Age group	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for less than 3 months	Industrial Category - N to O - Females	Industrial Category - P to Q - Persons	Industrial Category - P to Q - Males	Indus Categ - P to Fem
						- Males	- Females	- Persons	- Persons				
						Persons							
5													

In [55]:						Total	Total	200828	589003	611825	221386	...	3565	11080	4019	7	
Out[55]:	0	B0806SC	'33	'000	TAMIL NADU												
	1	B0806SC	'33	'000	TAMIL NADU	State	Total	'5-14	27791	14125	13666	2447	...	11	122	71	
	2	B0806SC	'33	'000	TAMIL NADU	State	Total	15-34	514340	259560	254780	92423	...	1754	7536	2718 1131 93	4
	3	B0806SC	'33	'000	TAMIL NADU	State	Total	35-59	542581	251957	290624	99202	...	1619	3205		2
	4	B0806SC	'33	'000	TAMIL NADU		Total	60+	115103	62833	52270	27165	...	175	211		

rows × 69 cc

In [43]: dataset.columns

Out[43]:

```
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')
```

In [56]: `dataset.info()`

```
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
```

In [57]: `dataset.describe()`

Out[57]:

	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 - Cat - C
count	5.940000e+02	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000	59
mean	1.617277e+04	7932.700337	8240.067340	2981.629630	1338.289562	1643.340067	865.117845	466.424242	39
std	7.607172e+04	36864.822704	39259.545337	13909.621137	6127.047670	7808.832522	4274.458077	2298.072295	197
min	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	2.872500e+02	147.250000	144.000000	27.000000	14.250000	13.000000	9.000000	5.000000	
50%	2.225500e+03	1147.000000	1076.000000	430.000000	198.500000	213.000000	69.500000	35.500000	3
75%	9.628500e+03	4770.500000	4887.500000	1775.250000	774.250000	946.500000	466.000000	244.250000	20
max	1.200828e+06	589003.000000	611825.000000	221386.000000	99368.000000	122018.000000	64235.000000	34632.000000	2960

8 rows × 63 columns

In [63]: `dataset = dataset.select_dtypes(include=[np.number])`
`correlation_matrix = dataset.corr()`

```
dataset.corr(numeric_only=True)
```

```
In [64]: Industrial 0.987514    0.991627    0.982327 0.993430 0.981375
          Category 0.988538
          - R to U -
          HHI -
          Males
                           0.991781 0.976818
Industrial 0.983226    0.987635    0.977768
          Category 0.985253
          - R to U -
          HHI -
          Females
                           0.913979 0.864574
          Industrial 0.883359    0.898087    0.868345
          Category 0.887969
          - R to U -
          Non HHI
          -
          Persons
                           0.918710 0.871332
          0.890343 0.904637 0.875727
          Industrial
          Category
          - R to U -
          Non
          HHI
          - Males 0.893847 0.874011 0.889046
                           0.8066 0.8021 0.8110
                           78   19    38
          Industrial
          Category
          - R to U -
          Non HHI 0.858721 0.879533
          -
          Females
```

€3 3 columns

```
cleandata=dataset.isnull().sum()
```

In [58]:

In [38]: `print(dataset.isna())`

```
1          False
2          False
3          False
4          False
...
589         ...
590         False
591         False
592         False
593         False
```

Worked for 3 months or more but less than 6 months - Females \

```
0          False
1          False
2          False
3          False
4          False
...
589         ...
590         False
```

In [37]:

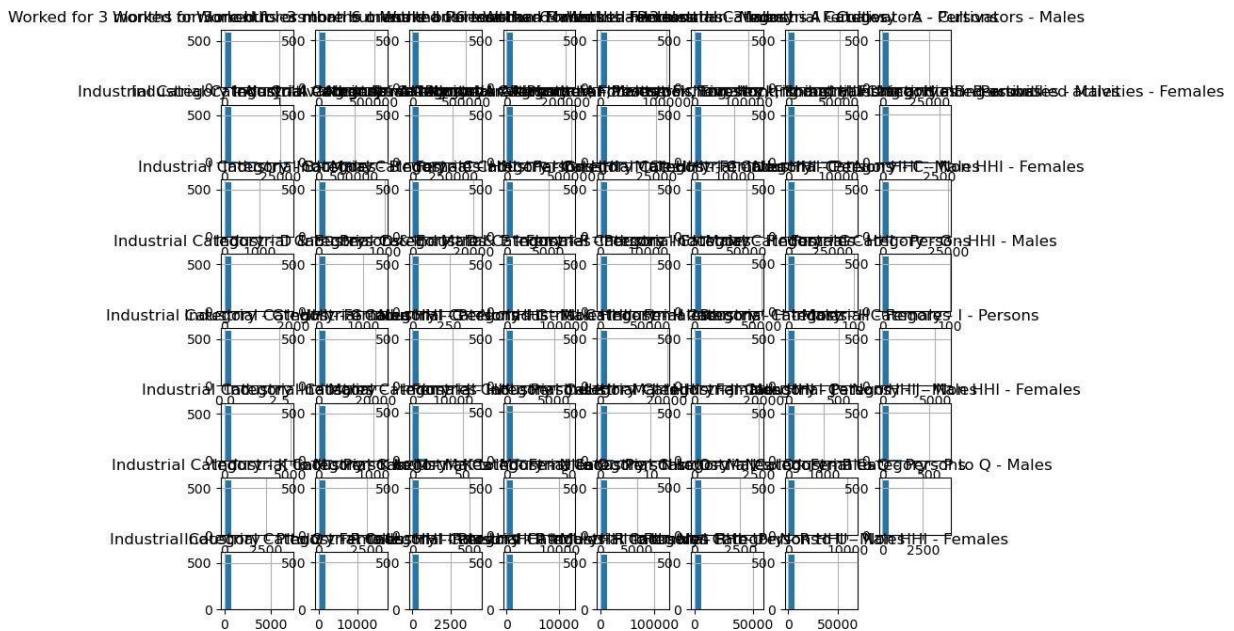
`cleandata`

Out[37]: 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

In [41]: `dataset.hist(figsize=(10,8))`

Out[41]:

```
array([[<Axes: title={'center': 'Worked for 3 months or more but less than 6 months - Persons'}>,
       <Axes: title={'center': 'Worked for 3 months or more but less than 6 months - Males'}>,
       <Axes: title={'center': 'Worked for 3 months or more but less than 6 months - Females'}>,
       <Axes: title={'center': 'Worked for less than 3 months - Persons'}>,
       <Axes: title={'center': 'Worked for less than 3 months - Males'}>,
       <Axes: title={'center': 'Worked for less than 3 months - Females'}>,
       <Axes: title={'center': 'Industrial Category - A - Cultivators - Persons'}>,
       <Axes: title={'center': 'Industrial Category - A - Cultivators - Males'}>,
       [<Axes: title={'center': 'Industrial Category - A - Cultivators - Females'}>,
        <Axes: title={'center': 'Industrial Category - A - Agricultural labourers - Persons'}>,
        <Axes: title={'center': 'Industrial Category - A - Agricultural labourers - Males'}>,
        <Axes: title={'center': 'Industrial Category - A - Agricultural labourers - Females'}>,
        <Axes: title={'center': 'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons'}>,
        <Axes: title={'center': 'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Males'}>,
        <Axes: title={'center': 'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Females'}>,
        <Axes: title={'center': 'Industrial Category - B - Persons'}>],
       [<Axes: title={'center': 'Industrial Category - B - Males'}>,
        <Axes: title={'center': 'Industrial Category - B - Females'}>,
        <Axes: title={'center': 'Industrial Category - C - HHI - Persons'}>,
        <Axes: title={'center': 'Industrial Category - C - HHI - Males'}>,
        <Axes: title={'center': 'Industrial Category - C - HHI - Females'}>,
        <Axes: title={'center': 'Industrial Category - C - Non HHI - Persons'}>,
        <Axes: title={'center': 'Industrial Category - C - Non HHI - Males'}>,
        <Axes: title={'center': 'Industrial Category - C - Non HHI - Females'}>,
        [<Axes: title={'center': 'Industrial Category - D & E - Persons'}>,
         <Axes: title={'center': 'Industrial Category - D & E - Males'}>,
         <Axes: title={'center': 'Industrial Category - D & E - Females'}>,
         <Axes: title={'center': 'Industrial Category - F - Persons'}>,
         <Axes: title={'center': 'Industrial Category - F - Males'}>,
         <Axes: title={'center': 'Industrial Category - F - Females'}>,
         <Axes: title={'center': 'Industrial Category - G - HHI - Persons'}>,
         <Axes: title={'center': 'Industrial Category - G - HHI - Males'}>,
         [<Axes: title={'center': 'Industrial Category - G - HHI - Females'}>,
          <Axes: title={'center': 'Industrial Category - G - Non HHI - Persons'}>,
          <Axes: title={'center': 'Industrial Category - G - Non HHI - Males'}>,
          <Axes: title={'center': 'Industrial Category - G - Non HHI - Females'}>,
          <Axes: title={'center': 'Industrial Category - H - Persons'}>,
          <Axes: title={'center': 'Industrial Category - H - Males'}>,
          <Axes: title={'center': 'Industrial Category - H - Females'}>,
          <Axes: title={'center': 'Industrial Category - I - Persons'}>,
          [<Axes: title={'center': 'Industrial Category - I - Males'}>,
           <Axes: title={'center': 'Industrial Category - I - Females'}>,
           <Axes: title={'center': 'Industrial Category - J - HHI - Persons'}>,
           <Axes: title={'center': 'Industrial Category - J - HHI - Males'}>,
           <Axes: title={'center': 'Industrial Category - J - HHI - Females'}>,
           <Axes: title={'center': 'Industrial Category - J - Non HHI - Persons'}>,
           <Axes: title={'center': 'Industrial Category - J - Non HHI - Males'}>,
           <Axes: title={'center': 'Industrial Category - J - Non HHI - Females'}>,
           [<Axes: title={'center': 'Industrial Category - K to M - Persons'}>,
            <Axes: title={'center': 'Industrial Category - K to M - Males'}>,
            <Axes: title={'center': 'Industrial Category - K to M - Females'}>, <Axes: title={'center': 'Industrial Category - N to O - Persons'}>, <Axes: title={'center': 'Industrial Category - N to O - Males'}>,
            <Axes: title={'center': 'Industrial Category - N to O - Females'}>, <Axes: title={'center': 'Industrial Category - P to Q - Persons'}>,
            <Axes: title={'center': 'Industrial Category - P to Q - Males'}>,
            [<Axes: title={'center': 'Industrial Category - P to Q - Females'}>,
             <Axes: title={'center': 'Industrial Category - R to U - HHI - Persons'}>,
             <Axes: title={'center': 'Industrial Category - R to U - HHI - Males'}>,
             <Axes: title={'center': 'Industrial Category - R to U - HHI - Females'}>,
             <Axes: title={'center': 'Industrial Category - R to U - Non HHI - Persons'}>,
             <Axes: title={'center': 'Industrial Category - R to U - Non HHI - Males'}>,
             <Axes: title={'center': 'Industrial Category - R to U - Non HHI - Females'}>,
             <Axes: title={'center': 'Industrial Category - R to U - Non HHI - Persons'}>, <Axes: title={'center': 'Industrial Category - R to U - Non HHI - Males'}>,
             <Axes: title={'center': 'Industrial Category - R to U - Non HHI - Females'}>]], dtype=object)
```



In [42]: `dataset.dtypes`

Out[42]:

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

In []:

In [48]:

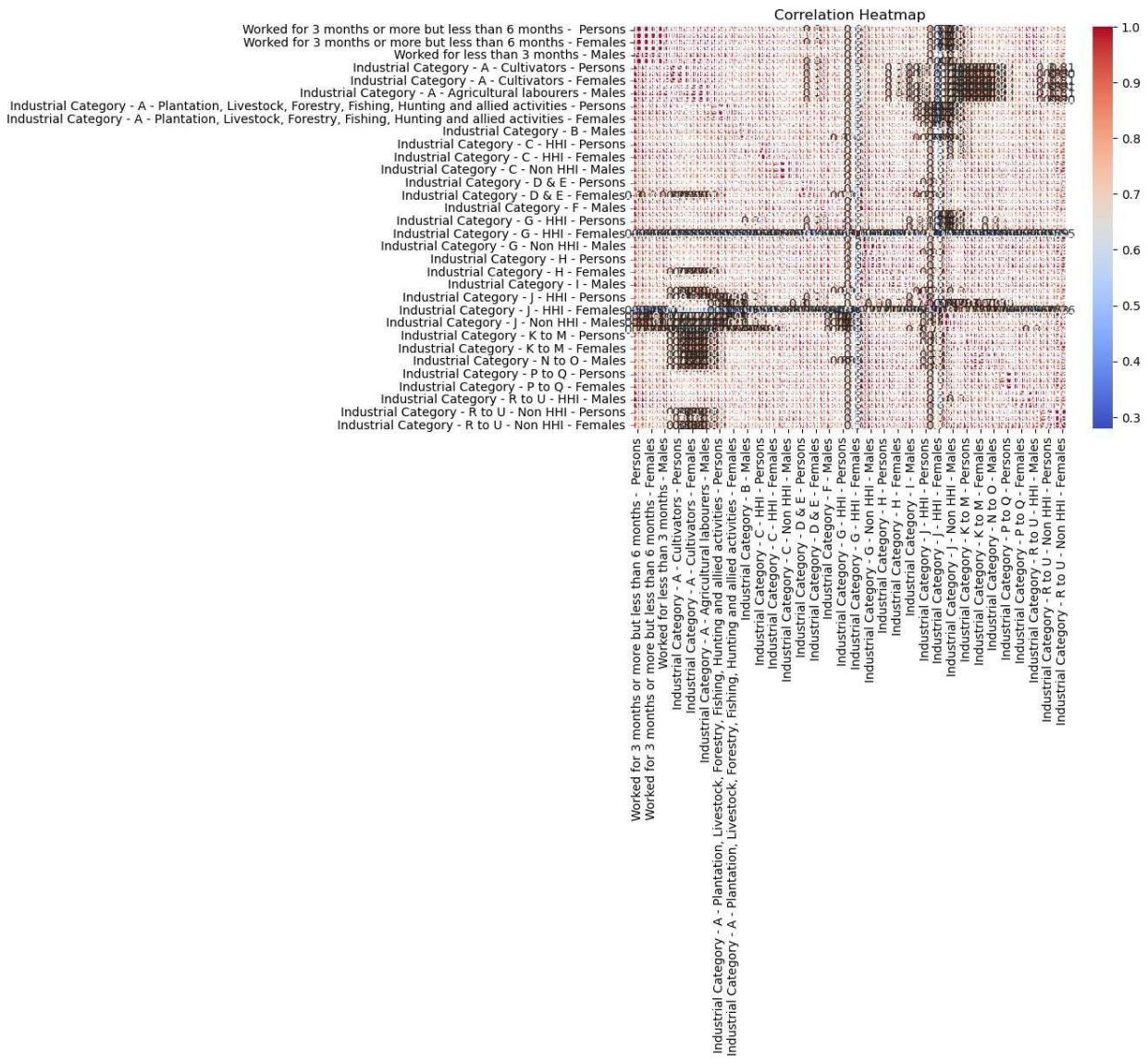
```
import pandas as pd import seaborn
as sns import matplotlib.pyplot as plt
```

```
def visualize_correlation(dataset):

    numeric_columns = dataset.select_dtypes(include='number') correlation_matrix =
    numeric_columns.corr()

    plt.figure(figsize=(8, 6))
    sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
    plt.title('Correlation Heatmap') plt.show()

your_dataset = pd.DataFrame(dataset) visualize_correlation(your_dataset)
```



In []: dataset.shape

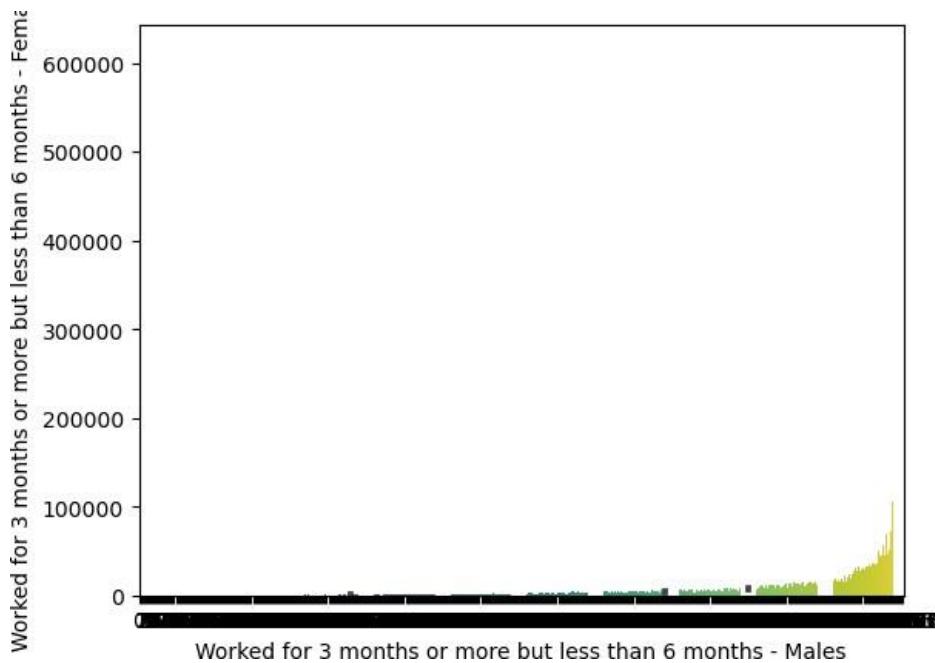
In [50]:

```
import pandas as pd import seaborn
as sns import matplotlib.pyplot as plt

dataset = pd.read_csv("DDW_B06SC_3300_State_TAMIL_NADU-2011.csv")

X = dataset["Worked for 3 months or more but less than 6 months - Males"]
Y = dataset["Worked for 3 months or more but less than 6 months - Females"]

sns.barplot(x=X, y=Y, data=dataset, palette='viridis') plt.show()
```



```
In [*]: plt.figure(figsize=(12,8)) sns.pairplot(dataset)
```

```
In [*]:
```

```
import seaborn as sns
```

```
In [ ]: sns.histplot(dataset, x='Worked for 3 months or more but less than 6 months - People', bins=50, color='r')
```

Import depends

```
In [51]: import pandas as pd  
import numpy as np  
import xgboost as xg
```

```
In [52]: import os  
print("Current working directory:", os.getcwd())  
  
file_path = 'datafile.csv'  
if os.path.exists(file_path):  
    print("The file exists.")
```

Current working directory: C:\Users\VIJAYRAJ R
The file exists.

```
In [60]: dataset = pd.read_csv("DDW_B06SC_3300_State_TAMIL_NADU-2011.csv")
```

10/31/23, 12:11 PM

PHASE-3(4498) - Jupyter Notebook

```
In [54]: dataset
```

Out[54]:

Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	Age group	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for less than 3 months	Industrial Category - N to O - Females	Industrial Category - P to Q - Persons	Industrial Category - P to Q - Males	In C - F	
						- Persons	- Males	- Females	- Persons					
0	B0806SC	'33	'000	State - TAMIL NADU	Total	Total	1200828	589003	611825	221386	...	3565	11080	4019
1	B0806SC	'33	'000	State - TAMIL NADU	Total	5-14	27791	14125	13666	2447	...	11	122	71
2	B0806SC	'33	'000	State - TAMIL NADU	Total	15-34	514340	259560	254780	92423	...	1754	7536	2718
3	B0806SC	'33	'000	State - TAMIL NADU	Total	35-59	542581	251957	290624	99202	...	1619	3205	1131
4	B0806SC	'33	'000	State - TAMIL NADU	Total	60+	115103	62833	52270	27165	...	175	211	93
...
589	B0806SC	'33	'633	District - Tiruppur	Urban	5-14	272	129	143	18	...	0	0	0
590	B0806SC	'33	'633	District - Tiruppur	Urban	15-34	3285	1654	1631	473	...	20	44	15
591	B0806SC	'33	'633	District - Tiruppur	Urban	35-59	3672	1769	1903	522	...	33	35	12
592	B0806SC	'33	'633	District - Tiruppur	Urban	60+	696	399	297	111	...	0	3	0
593	B0806SC	'33	'633	District - Tiruppur	Urban	Age not stated	2	1	1	0	...	0	0	0

In [55]: `dataset.head(5)`

Out[55]:

Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	Age group	Persons	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for less than 3 months	...	Industrial Category - N to O - Females	Industrial Category - P to Q - Persons	Industrial Category - P to Q - Males	Indus Categ - P to Fema
							- Persons	- Females	- Males	Persons	...				
State															
0	B0806SC	'33	'000	TAMIL NADU	Total	Total	1200828	589003	611825	221386	...	3565	11080	4019	7
1	B0806SC	'33	'000	TAMIL NADU	Total	'5-14	27791	14125	13666	2447	...	11	122	71	
2	B0806SC	'33	'000	TAMIL NADU	Total	15-34	514340	259560	254780	92423	...	1754	7536	2718	4
3	B0806SC	'33	'000	TAMIL NADU	Total	35-59	542581	251957	290624	99202	...	1619	3205	1131	2
4	B0806SC	'33	'000	TAMIL NADU	Total	60+	115103	62833	52270	27165	...	175	211	93	

5 rows × 69 columns



```
In [43]: dataset.columns
```

```
Out[43]: 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')
```

In [56]: `dataset.info()`

```
00 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
```

In [57]: `dataset.describe()`

Out[57]:

	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	Cat - C
count	5.940000e+02	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000
mean	1.617277e+04	7932.700337	8240.067340	2981.629630	1338.289562	1643.340067	865.117845	466.424242	39
std	7.607172e+04	36864.822704	39259.545337	13909.621137	6127.047670	7808.832522	4274.458077	2298.072295	197
min	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	2.872500e+02	147.250000	144.000000	27.000000	14.250000	13.000000	9.000000	5.000000	3
50%	2.225500e+03	1147.000000	1076.000000	430.000000	198.500000	213.000000	69.500000	35.500000	3
75%	9.628500e+03	4770.500000	4887.500000	1775.250000	774.250000	946.500000	466.000000	244.250000	20
max	1.200828e+06	589003.000000	611825.000000	221386.000000	99368.000000	122018.000000	64235.000000	34632.000000	2960

8 rows × 63 columns

Preprocessing the data

In [63]: `dataset = dataset.select_dtypes(include=[np.number])`
`correlation_matrix = dataset.corr()`

In [64]: `dataset.corr(numeric_only=True)`

Out[64]:

	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for less than 3 months	Worked for less than 3 months	Worked for less than 3 months	Industrial Category - A - Cultivators - Persons	Industrial Category - A - Cultivators - Males	Industrial Category - A - Cultivators - Females	Industrial Category - A - Agricultural labourers - Persons	Industrial Category - A - ... Industr Categc - N to C Femal	
Persons	- Males	Females	Persons	- Males	Females	Persons	Cultivators - Persons	Cultivators - Males	Cultivators - Females	Agricultural labourers - Persons	... Industr Categc - N to C Femal	
Worked for 3 months or more but less than 6 months - Persons	1.000000	0.999263	0.999351	0.998865	0.994883	0.998631	0.987308	0.985738	0.987987	0.991821	...	0.8862
Worked for 3 months or more but less than 6 months - Males	0.999263	1.000000	0.997232	0.999020	0.997622	0.996757	0.982657	0.981185	0.983225	0.987097	...	0.8979
Worked for 3 months or more but less than 6 months - Females	0.999351	0.997232	1.000000	0.997381	0.990976	0.999052	0.990352	0.988691	0.991132	0.994927	...	0.8740
Worked for less than 3 months - Persons	0.998865	0.999020	0.997381	1.000000	0.997598	0.998522	0.984360	0.983491	0.984228	0.988923	...	0.8941
Worked for less than 3 months - Males	0.994883	0.997622	0.990976	0.997598	1.000000	0.992359	0.973140	0.972236	0.973061	0.977145	...	0.9156
...
Industrial Category - R to U - HHI - Males	0.987514	0.991627	0.982327	0.988538	0.993430	0.981375	0.963460	0.961010	0.965188	0.965973	...	0.9101
Industrial Category - R to U - HHI - Females	0.983226	0.987635	0.977768	0.985253	0.991781	0.976818	0.952677	0.949565	0.955187	0.955424	...	0.9402
Industrial Category - R to U - Non HHI - Persons	0.883359	0.898087	0.868345	0.887969	0.913979	0.864574	0.820028	0.815232	0.824646	0.819429	...	0.9620
Industrial Category - R to U - Non HHI - Males	0.890343	0.904637	0.875727	0.893847	0.918710	0.871332	0.832068	0.827014	0.836972	0.829783	...	0.9454
Industrial Category - R to U - Non HHI - Females	0.874011	0.889046	0.858721	0.879533	0.906405	0.855491	0.806678	0.802119	0.811038	0.807500	...	0.9722

63 rows × 63 columns



In [58]: `cleandata=dataset.isnull().sum()`

```
In [38]: print(dataset.isna())
```

```
1                         False  
2                         False  
3                         False  
4                         False  
..                         ...  
589                         False  
590                         False  
591                         False  
592                         False  
593                         False  
  
Worked for 3 months or more but less than 6 months - Females \\\n0                         False  
1                         False  
2                         False  
3                         False  
4                         False  
..                         ...  
589                         False  
590                         False
```

```
In [37]: cleandata
```

```
Out[37]: 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
```

```
In [41]: dataset.hist(figsize=(10,8))
```

```
Out[41]: array([[<Axes: title={'center': 'Worked for 3 months or more but less than 6 months - Persons'}>,
   <Axes: title={'center': 'Worked for 3 months or more but less than 6 months - Males'}>,
   <Axes: title={'center': 'Worked for 3 months or more but less than 6 months - Females'}>,
   <Axes: title={'center': 'Worked for less than 3 months - Persons'}>,
   <Axes: title={'center': 'Worked for less than 3 months - Males'}>,
   <Axes: title={'center': 'Worked for less than 3 months - Females'}>,
   <Axes: title={'center': 'Industrial Category - A - Cultivators - Persons'}>,
   <Axes: title={'center': 'Industrial Category - A - Cultivators - Males'}>],
  [<Axes: title={'center': 'Industrial Category - A - Cultivators - Females'}>,
   <Axes: title={'center': 'Industrial Category - A - Agricultural labourers - Persons'}>,
   <Axes: title={'center': 'Industrial Category - A - Agricultural labourers - Males'}>,
   <Axes: title={'center': 'Industrial Category - A - Agricultural labourers - Females'}>,
   <Axes: title={'center': 'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons'}>,
   <Axes: title={'center': 'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Males'}>,
   <Axes: title={'center': 'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Females'}>,
   <Axes: title={'center': 'Industrial Category - B - Persons'}>],
  [<Axes: title={'center': 'Industrial Category - B - Males'}>,
   <Axes: title={'center': 'Industrial Category - B - Females'}>,
   <Axes: title={'center': 'Industrial Category - C - HHI - Persons'}>,
   <Axes: title={'center': 'Industrial Category - C - HHI - Males'}>,
   <Axes: title={'center': 'Industrial Category - C - HHI - Females'}>,
   <Axes: title={'center': 'Industrial Category - C - Non HHI - Persons'}>,
   <Axes: title={'center': 'Industrial Category - C - Non HHI - Males'}>,
   <Axes: title={'center': 'Industrial Category - C - Non HHI - Females'}>],
  [<Axes: title={'center': 'Industrial Category - D & E - Persons'}>,
   <Axes: title={'center': 'Industrial Category - D & E - Males'}>,
   <Axes: title={'center': 'Industrial Category - D & E - Females'}>,
   <Axes: title={'center': 'Industrial Category - F - Persons'}>,
   <Axes: title={'center': 'Industrial Category - F - Males'}>,
   <Axes: title={'center': 'Industrial Category - F - Females'}>,
   <Axes: title={'center': 'Industrial Category - G - HHI - Persons'}>,
   <Axes: title={'center': 'Industrial Category - G - HHI - Males'}>],
  [<Axes: title={'center': 'Industrial Category - G - HHI - Females'}>,
   <Axes: title={'center': 'Industrial Category - G - Non HHI - Persons'}>,
   <Axes: title={'center': 'Industrial Category - G - Non HHI - Males'}>,
   <Axes: title={'center': 'Industrial Category - G - Non HHI - Females'}>,
   <Axes: title={'center': 'Industrial Category - H - Persons'}>,
   <Axes: title={'center': 'Industrial Category - H - Males'}>,
   <Axes: title={'center': 'Industrial Category - H - Females'}>,
   <Axes: title={'center': 'Industrial Category - I - Persons'}>],
  [<Axes: title={'center': 'Industrial Category - I - Males'}>,
   <Axes: title={'center': 'Industrial Category - I - Females'}>,
   <Axes: title={'center': 'Industrial Category - J - HHI - Persons'}>,
   <Axes: title={'center': 'Industrial Category - J - HHI - Males'}>,
   <Axes: title={'center': 'Industrial Category - J - HHI - Females'}>,
   <Axes: title={'center': 'Industrial Category - J - Non HHI - Persons'}>,
   <Axes: title={'center': 'Industrial Category - J - Non HHI - Males'}>,
   <Axes: title={'center': 'Industrial Category - J - Non HHI - Females'}>],
  [<Axes: title={'center': 'Industrial Category - K to M - Persons'}>,
   <Axes: title={'center': 'Industrial Category - K to M - Males'}>,
   <Axes: title={'center': 'Industrial Category - K to M - Females'}>,
   <Axes: title={'center': 'Industrial Category - N to O - Persons'}>,
   <Axes: title={'center': 'Industrial Category - N to O - Males'}>,
   <Axes: title={'center': 'Industrial Category - N to O - Females'}>,
   <Axes: title={'center': 'Industrial Category - P to Q - Persons'}>,
   <Axes: title={'center': 'Industrial Category - P to Q - Males'}>],
  [<Axes: title={'center': 'Industrial Category - P to Q - Females'}>,
   <Axes: title={'center': 'Industrial Category - R to U - HHI - Persons'}>,
   <Axes: title={'center': 'Industrial Category - R to U - HHI - Males'}>,
   <Axes: title={'center': 'Industrial Category - R to U - HHI - Females'}>,
   <Axes: title={'center': 'Industrial Category - R to U - Non HHI - Persons'}>,
   <Axes: title={'center': 'Industrial Category - R to U - Non HHI - Males'}>,
   <Axes: title={'center': 'Industrial Category - R to U - Non HHI - Females'}>,
   <Axes: >]], dtype=object)
```

```
In [42]: dataset.dtypes
```

```
Out[42]: 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
```

In []:

```
In [48]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

def visualize_correlation(dataset):

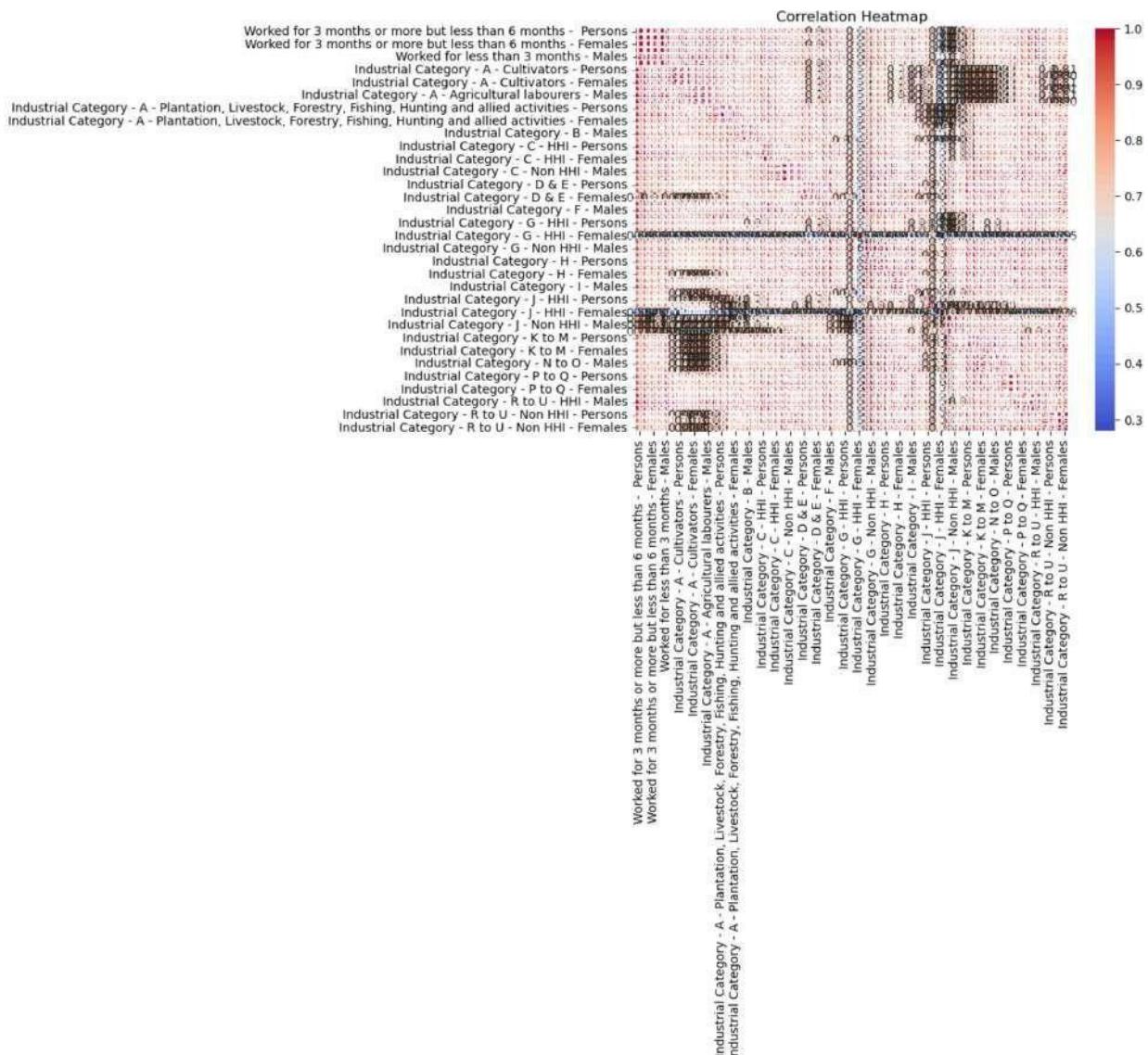
    numeric_columns = dataset.select_dtypes(include='number')

    correlation_matrix = numeric_columns.corr()

    plt.figure(figsize=(8, 6))
    sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
    plt.title('Correlation Heatmap')
    plt.show()

your_dataset = pd.DataFrame(dataset)

visualize_correlation(your_dataset)
```



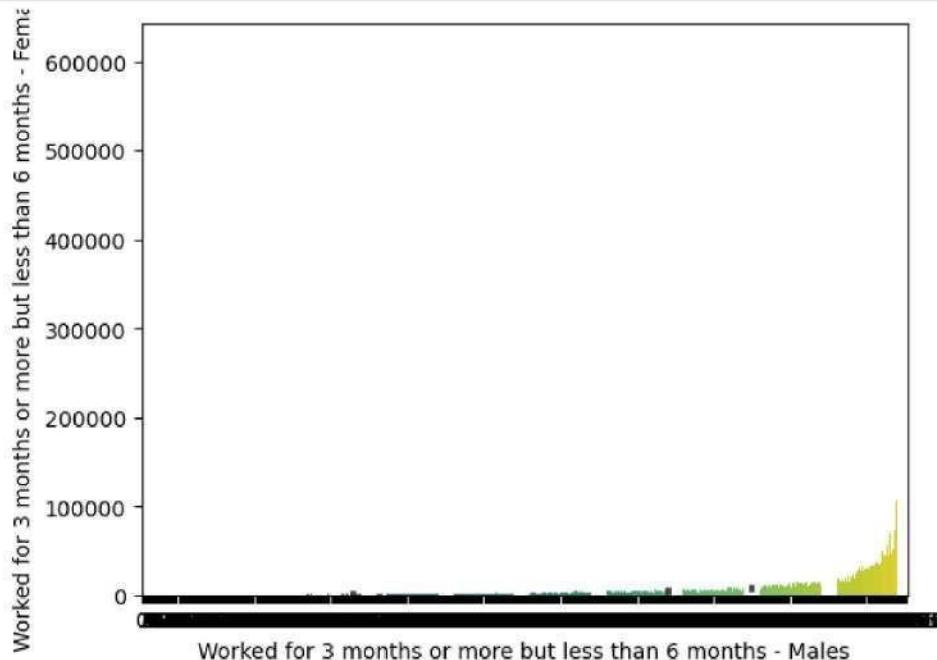
```
In [ ]: dataset.shape
```

```
In [50]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

dataset = pd.read_csv("DDW_B06SC_3300_State_TAMIL_NADU-2011.csv")

X = dataset['Worked for 3 months or more but less than 6 months - Males']
Y = dataset['Worked for 3 months or more but less than 6 months - Females']

sns.barplot(x=X, y=Y, data=dataset, palette='viridis')
plt.show()
```



```
In [*]: plt.figure(figsize=(12,8))
sns.pairplot(dataset)
```

```
In [*]: import seaborn as sns
sns.histplot(dataset, x='Worked for 3 months or more but less than 6 months - People', bins=50, color='r')
```

```
In [ ]:
```

```
<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
```

```
[1]: import pandas as pd

# Create a pandas DataFrame from your dataset
data = {
    'Industrial Category - A - Cultivators - Persons': [64235, 190],
    'Industrial Category - A - Cultivators - Males': [34632, 107],
    'Industrial Category - A - Cultivators - Females': [29603, 83],
    'Industrial Category - A - Agricultural labourers - Persons': [907752, 2853],
    'Industrial Category - A - Agricultural labourers - Males': [404844, 1862],
    'Industrial Category - A - Agricultural labourers - Females': [502908, 991],
    'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons': [29410, 3],
    'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Males': [16268, 3],
    'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Females': [13142, 0],
}

df = pd.DataFrame(data)
df
```



```
[1]: Industrial Category - A - Cultivators - Persons \
0 64235
1 190

Industrial Category - A - Cultivators - Males \
0 34632
1 107

Industrial Category - A - Cultivators - Females \
0 29603
1 83

Industrial Category - A - Agricultural labourers - Persons \
0 907752
1 2853
```

Industrial Category - A - Agricultural labourers - Males \	
0	404844
1	1862
Industrial Category - A - Agricultural labourers - Females \	
0	502908
1	991
Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons \	
0	29410
1	3
Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Males \	
0	16268
1	3
Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Females	
0	13142
1	0

```
[3]: data = {
    'Age group': ['Total', '5-14'],
    'Worked for 3 months or more but less than 6 months - Persons': [1200828, 27791],
    'Worked for 3 months or more but less than 6 months - Males': [589003, 14125],
    'Worked for 3 months or more but less than 6 months - Females': [611825, 13666],
}

df = pd.DataFrame(data)
df
```

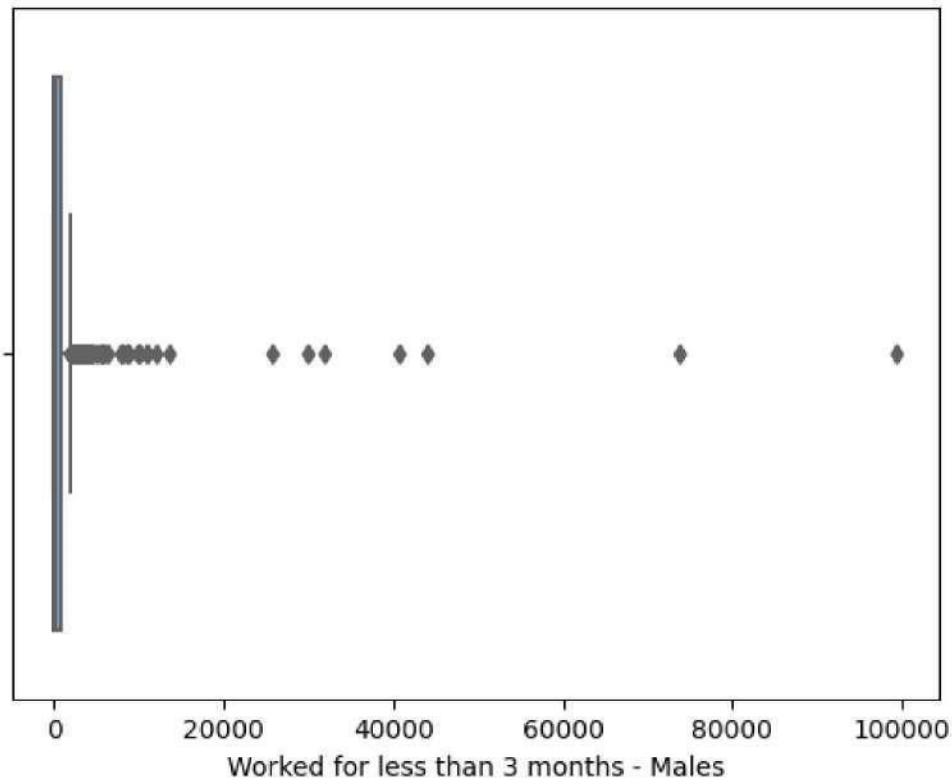
[3]: Age group	Worked for 3 months or more but less than 6 months - Persons \	
0	Total	1200828
1	5-14	27791
	Worked for 3 months or more but less than 6 months - Males \	
0		589003
1		14125
	Worked for 3 months or more but less than 6 months - Females	
0		611825

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]

```
[9]: sns.boxplot(dataset, x='Worked for less than 3 months - Males',
    palette='Blues')
```

```
[9]: <Axes: xlabel='Worked for less than 3 months - Males'>
```



```
[11]: dataset.hist(figsize=(155,100))
```

```
[11]: array([[<Axes: title={'center': 'Worked for 3 months or more but less than 6 months - Persons'}>,
   <Axes: title={'center': 'Worked for 3 months or more but less than 6 months - Males'}>,
   <Axes: title={'center': 'Worked for 3 months or more but less than 6 months - Females'}>,
   <Axes: title={'center': 'Worked for less than 3 months - Persons'}>,
   <Axes: title={'center': 'Worked for less than 3 months - Males'}>,
   <Axes: title={'center': 'Worked for less than 3 months - Females'}>,
   <Axes: title={'center': 'Industrial Category - A - Cultivators - Persons'}>,
   <Axes: title={'center': 'Industrial Category - A - Cultivators - Males'}>,
   [<Axes: title={'center': 'Industrial Category - A - Cultivators - Females'}>,
    <Axes: title={'center': 'Industrial Category - A - Agricultural labourers - Persons'}>,
    <Axes: title={'center': 'Industrial Category - A - Agricultural labourers - Males'}>,
```

IBM COGNOS VISUALIZATION:

10/31/23, 11:27 PM

New dashboard 2

Tab 1

Worked for 3 months or more but less than 6 months - Persons compared to Age group for Worked for 3 months or more but less than 6 months - Males

6 ↓

Age group

9.61M (-100%)

Worked for 3 months or more but less than 6 months - Persons



10/31/23, 11:27 PM

New dashboard 2

Age group by Worked for 3 months or more but less than 6 months - Persons, Worked for 3 months or more but less than 6 months - Persons, Worked for less than 3 months - Persons, Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons, Industrial Category - G - HHI - Persons and Industrial Category - N to O - Persons colored by State Code

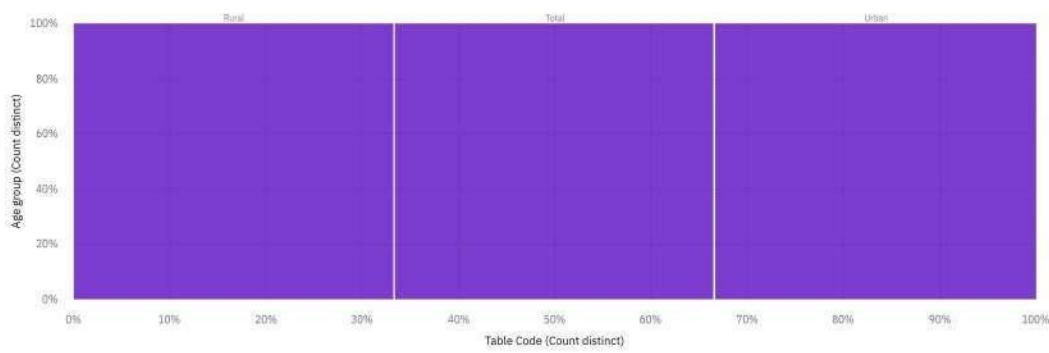
State Code
● : 33



10/31/23, 11:27 PM

New dashboard 2

Age group and Table Code for Total/ Rural/ Urban

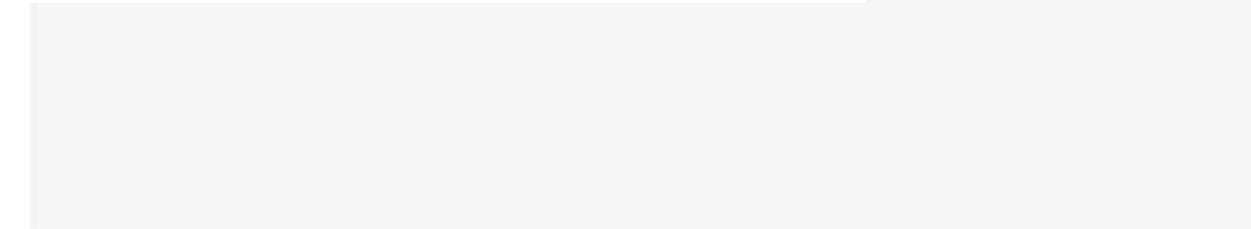
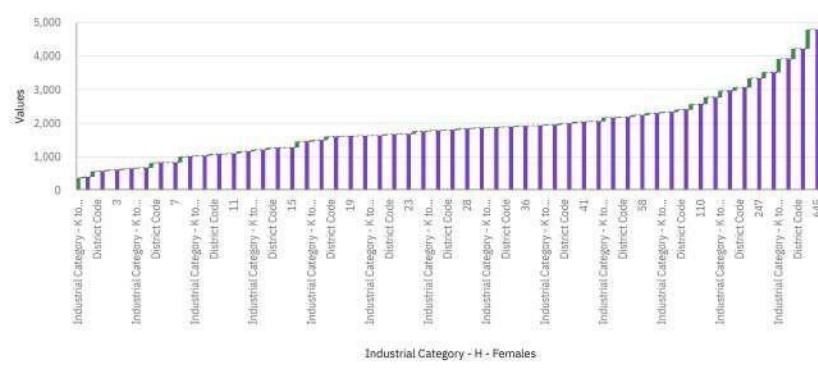


10/31/23, 11:27 PM

New dashboard 2

Industrial Category - K to M - Females and District Code for Industrial Category - H - Females

Column values
● Increase ● Decrease ● Total

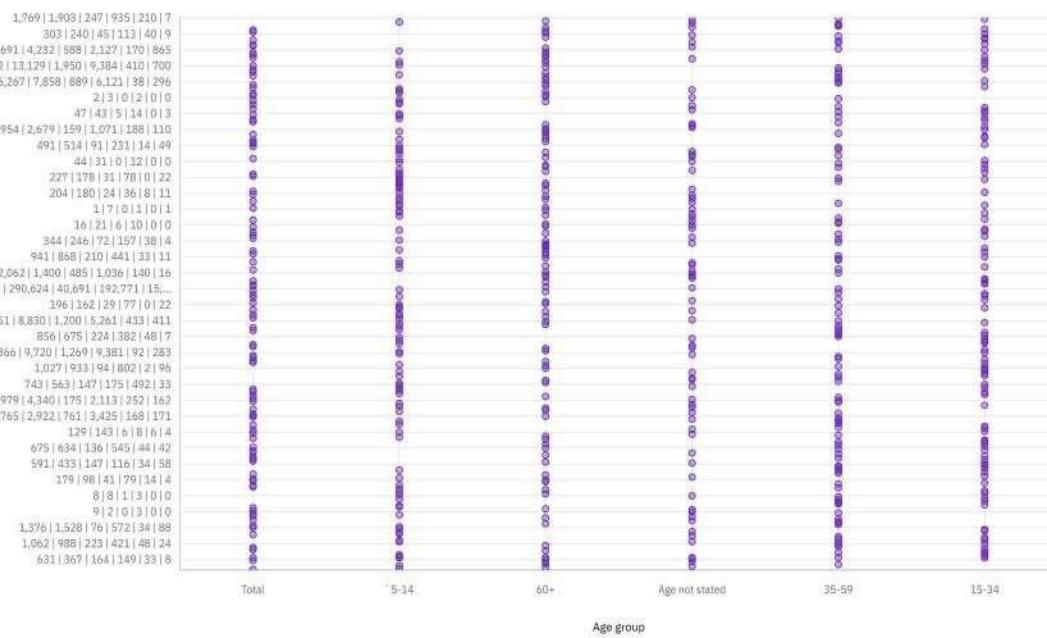


10/31/23, 11:27 PM

New dashboard 2

Click to edit title

Worked for 3 months or more but less than 6 months - Males (Sum) - Worked for 3 months or more but less than 6 months - Females (Sum)



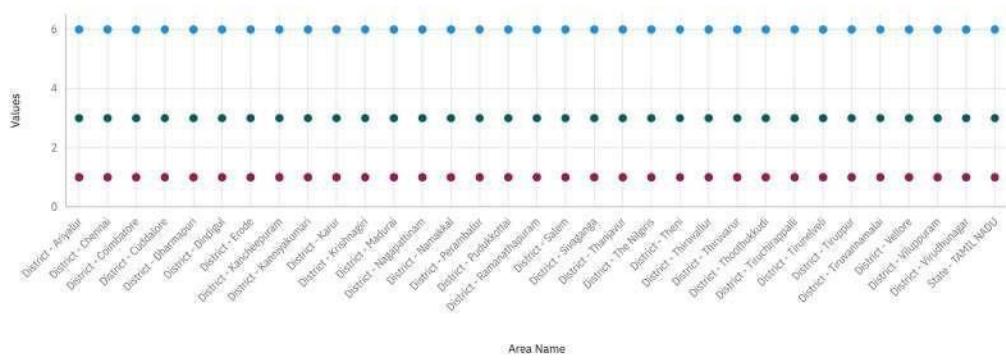
10/31/23, 11:27 PM

New dashboard 2

Age group, Total/ Rural/ Urban and District Code by Area Name

Measures

● Age group ● Total/ Rural/ Urban ● District Code



10/31/23, 11:27 PM

New dashboard 2

District Code, Area Name, Total/ Rural/ Urban and Age group by Industrial Category - C - Non HHI - Males, Industrial Category - F - Males and Industrial Category - I - Persons

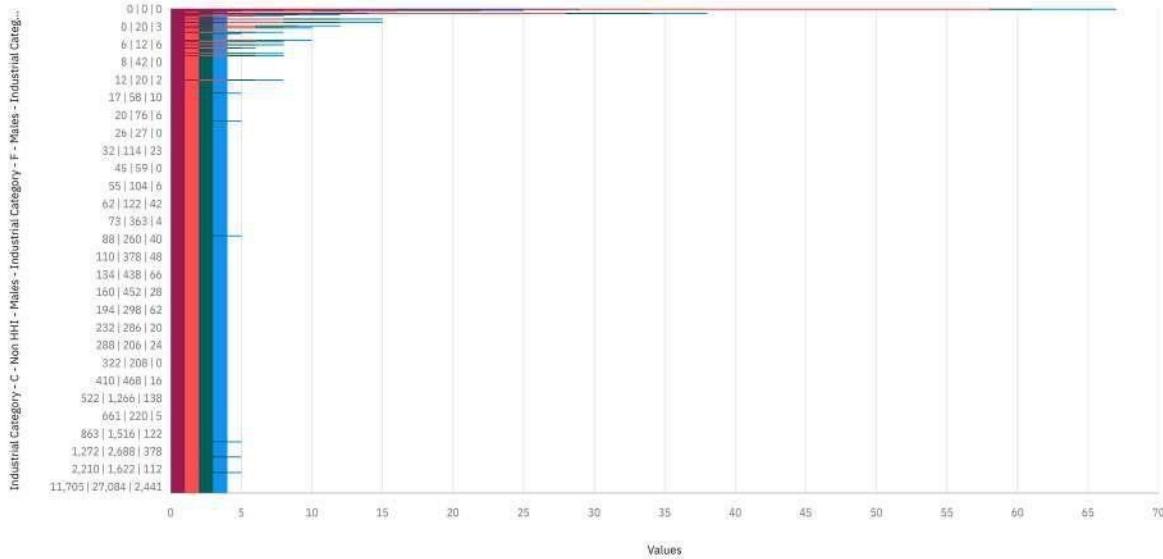
Measures

District Code

Area Name

Total/ Rural/ Urban

Age group

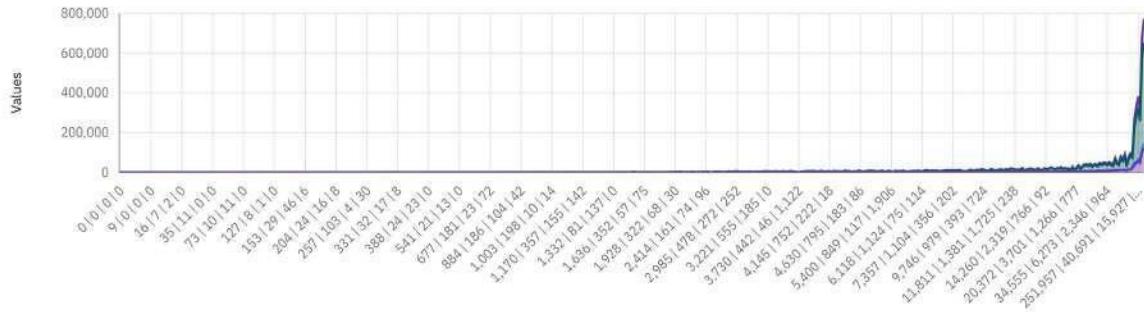


Tab 1

Worked for less than 3 months - Females, Industrial Category - C - Non HHI - Females, Industrial Category - A - Agricultural labourers - Females and Worked for less than 3 months - Females by Worked for 3 months or more but less than 6 months - Males, Worked for less than 3 months - Males, Industrial Category - A - Cultivators - Males, Industrial Category - A - Plantation, Livestock, Forestry, Fishing and Hunting and allied activities - Males

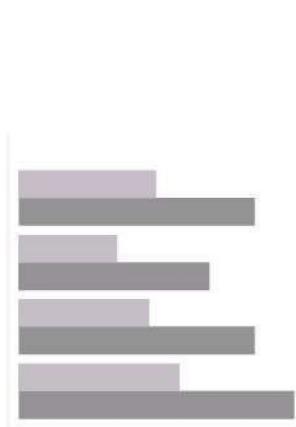
Measures

Worked for less than 3 months - Females, Industrial Category - C - Non HHI - Females, Industrial Category - A - Agricultural labourers - Females and Worked for less than 3 months - Females by Worked for 3 months or more but less than 6 months - Males, Worked for less than 3 months - Males, Industrial Category - A - Cultivators - Males, Industrial Category - A - Plantation, Livestock, Forestry, Fishing and Hunting and allied activities - Males



Worked for 3 months or more but less than 6 months - Males - Worked for less than 3 months - Males - Industrial Category - A - Cultivators - Males - Industrial Categ...

Tab 2



Conclusion:

This project aims to provide a comprehensive analysis of the employment status of marginal workers in Tamil Nadu, shedding light on various demographic and socio-economic factors influencing their employment. In doing so, we aim to offer actionable policy recommendations that can address the challenges faced by this vulnerable group and enhance their livelihoods. The project not only serves as a valuable source of insights into the employment landscape of marginal workers in Tamil Nadu but also contributes to informed decision-making and the formulation of targeted interventions to uplift their lives.