**ASSESSMENT OF MARGINAL WORKERS IN TAMIL NADU-A SOCIOECONOMIC ANALYSIS**

**Project’s objectives:**

The socio-economic status of marginal workers in Tamil Nadu is characterized by a complex web of challenges, encompassing low-income levels, limited access to education and healthcare, precarious employment opportunities, and enduring social disparities. This project seeks to comprehensively assess the multifaceted issues faced by marginal workers in the region, with the overarching aim of identifying barriers to their socio-economic advancement and formulating evidence-based recommendations to enhance their well-being and prospects for social and economic inclusion.

**Design Thinking:**

Design thinking will be integral to this assessment of marginal workers in Tamil Nadu, as it offers a human-centered approach to understand, empathize with, and address their unique challenges. By adopting this methodology, we will engage with marginal workers and the community to co-create solutions that cater to their specific needs, fostering innovation and promoting actionable, people-centric recommendations. This approach will ensure that the socio-economic analysis is not only data-driven but also deeply resonates with the lived experiences of marginal workers, leading to more effective, sustainable, and inclusive solutions for their empowerment and betterment.

**Development Phase:**

*Overview of the process:*

Data Collection: "Assessment of Marginal Workers in Tamil Nadu - A Socio-Economic Analysis" involved a multi-faceted approach. It included gathering demographic data such as age, gender, and caste/ethnicity, as well as detailed information on employment status, income levels, access to education and healthcare, living conditions, and social mobility among marginal workers. Both primary data collection methods, such as surveys and interviews, and secondary data sources, like government reports and existing socio-economic studies, were leveraged to ensure a comprehensive dataset. This hybrid approach allowed for a holistic analysis of the socio-economic challenges faced by marginal workers in Tamil Nadu.

Data Cleaning: Before loading the data into Cognos, you need to clean it. This involves handling missing values, correcting errors, and ensuring data consistency.

Data Loading: You can load the cleaned data into IBM Cognos by connecting to a database or importing data from a file. IBM Cognos typically supports various data sources and formats.

Data Preprocessing: Data preprocessing for the "Assessment of Marginal Workers in Tamil Nadu - A Socio-Economic Analysis" is a critical step to ensure the data is clean, structured, and ready for analysis. The process involved the following key tasks:

**Data Cleaning:** Raw data collected from various sources were reviewed for

inconsistencies, errors, and missing values. Incomplete or erroneous data points were either

corrected or imputed, and duplicates were removed to ensure data accuracy.

**Data Integration:** Data from diverse sources, such as surveys, government reports, and

secondary studies, were integrated into a unified dataset to facilitate a comprehensive

analysis.

**Data Transformation:** Categorical variables, such as caste/ethnicity, were one-hot

encoded to convert them into a numerical format for analysis. Additionally, data may have

been normalized or scaled to ensure that different variables were on a similar scale.

**Feature Engineering:** New features may have been created to enhance the analysis.

For example, an indicator variable for access to education or healthcare might have been

generated based on the available data.

**Handling Missing Data:** Missing data were addressed using appropriate techniques

like imputation, removal, or specialized handling, depending on the nature and extent of the

missing values.

**Outlier Detection:** Outliers, if any, were identified and either treated or noted for their

potential impact on the analysis.

**Data Validation:** Data integrity was confirmed by checking for data inconsistencies or

errors that could affect the analysis.

**Data Reduction (if necessary):** In cases where the dataset was extensive,

dimensionality reduction techniques, like principal component analysis (PCA), may have

been applied to simplify the analysis.

**Data Splitting:** The dataset may have been divided into training and testing subsets to

facilitate model training and validation.

Data preprocessing in this analysis ensured that the data was ready for the subsequent stages of the study, including exploratory data analysis, statistical modeling, and deriving actionable insights from the socio-economic data of marginal workers in Tamil Nadu.

DATA TRANSFORMATION: **It involves the systematic restructuring and preparation of the**

**dataset collected from various sources. This process may include cleaning and**

**standardizing data to ensure consistency, converting categorical variables into numerical**

**formats for modeling, scaling or normalizing features, and engineering new variables to**

**capture relevant socio-economic indicators. Additionally, data transformation may involve**

**aggregating data at appropriate geographical or temporal levels to facilitate meaningful**

**analysis. The goal is to create a well-organized, uniform dataset that can be effectively**

**used for socio-economic analysis to gain insights into the status and conditions of**

**marginal workers in Tamil Nadu.**

EXPLORATORY DATA ANALYSIS: **During EDA, various statistical and visualization techniques**

**are applied to identify trends, patterns, and outliers within the data. EDA helps in**

**uncovering key variables, relationships, and potential areas for in-depth analysis, enabling**

**researchers to formulate hypotheses and refine their research questions for a more**

**focused and informed socio-economic analysis of this specific demographic group.**

FEATURE SECTION: **By selecting the most influential socio-economic indicators and**

**characteristics of marginal workers, feature selection enhances the precision and focus of**

**the analysis, allowing for a more targeted exploration of the socio-economic factors that**

**affect this specific demographic group in Tamil Nadu.**

MODEL BUILDING AND ANALYSIS: **Use Cognos to build models for water quality prediction,**

**classification, or clustering, depending on your specific objectives. You can choose from a**

**variety of data analysis techniques, including regression, decision trees, clustering**

**algorithms, and more.**

MODEL EVALUATION: **It involves the rigorous assessment of the predictive models**

**developed to understand and analyze the socio-economic dynamics of marginal workers.**

**This process encompasses various techniques such as cross-validation, metrics like F1**

**score or accuracy, and the assessment of model performance against a set of predefined**

**criteria. By evaluating these models, researchers can determine the validity and reliability**

**of their findings, ensuring that the analysis accurately captures the socio-economic factors**

**impacting marginal workers in Tamil Nadu. The evaluation helps in selecting the most**

**appropriate model, making data-driven inferences, and drawing meaningful conclusions to**

**inform policy recommendations and socio-economic interventions for this vulnerable**

**demographic group.**

## 

## DEPLOYMENT AND AUTOMATION: If the analysis needs to be automated for real-time monitoring, you can deploy the models and dashboards within IBM Cognos to ensure ongoing data analysis.

## MAINTENANCE: Regularly update and maintain your data analysis process, including refreshing data, retraining models, and adapting to changes in data quality or data sources.

DOCUMENTATION: **Maintain comprehensive documentation of your work, including data**

**sources, preprocessing steps, model details, and results.**

**PROGRAMMING:**

import pandas as pd

import matplotlib.pyplot as plt

df= pd.read\_csv(‘Marginal Workers.csv’)

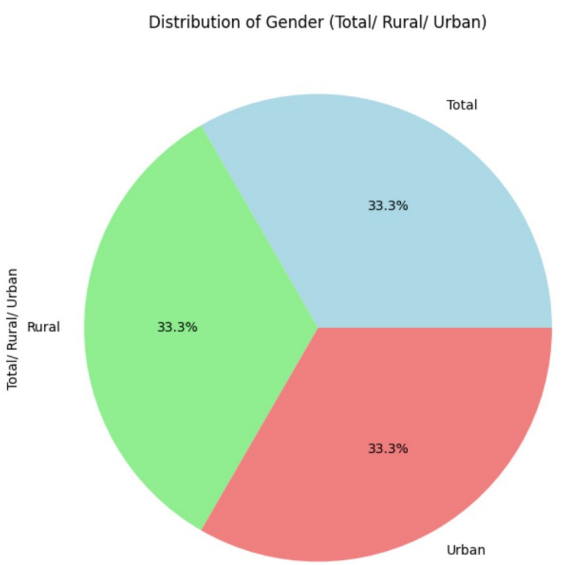
**# Demographic Analysis: Gender**

gender-counts=df[‘Total/Rural/Urban’].value\_counts()

gender\_counts.plot(kind=’pie’, autopct=’%1.11%%’,colors=[‘lightblue’, ‘lightgreen’, ‘lightcoral’],figsize=(8,8))

plt.title(‘Distribution of Gender (Total/Rural/Urban)’)

plt.show()

******

***#*Demographic Analysis: Age Groups**

age\_group\_counts = df [‘Age group’]. value\_counts()

age\_group\_counts.plot(kind=’bar’,figsize=(8,6),color=’skyblue’)

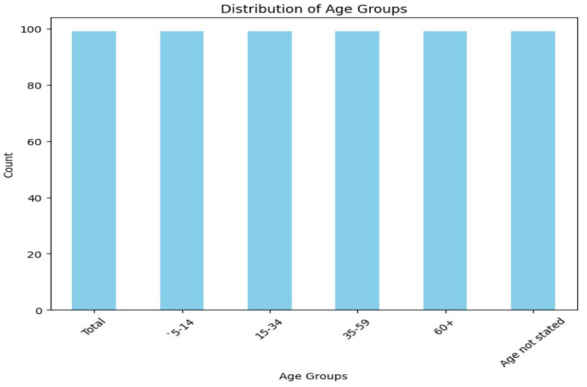
plt.title(‘Distribution of Age Groups’)

plt.xlabel(‘Age Groups’)

plt.ylabel(‘Count’)

plt.xticks(rotation=45)

plt.show()

******

**#Demographic Analysis: Total/Rural/Urban Distribution**

urban\_rural\_counts=df[‘Total/Rural/Urban’].value\_counts()

urban\_rural\_counts.plot(kind=’bar’, color=[‘lightgreen’, ‘lightblue’],figsize=(6,6))

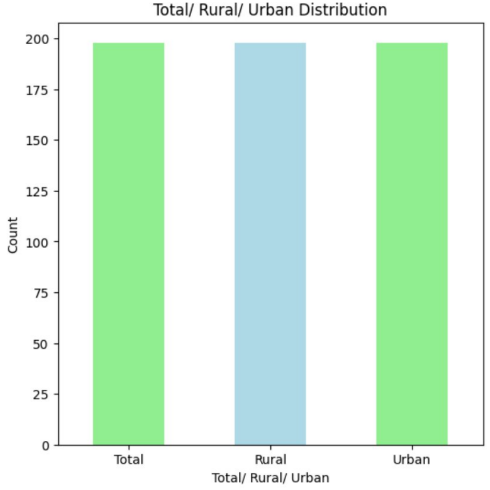
plt.title(“Total/Rural/Urban Distribution’)

plt.xlabel(“Total/Rural/Urban’)

plt.ylaabel(‘Count’)

plt.xsticks(rotation=0)

plt.show()

**

***DATA AGGREGATION***

Import pandas as pd

Df = pd.read\_csv(‘Marginal Workers.csv’)

# Group by age group , industrial categories – a - cultivators – persons ,and worked for 3 months or more but less than 6 months – persons then aggregate the counts

marginal\_workers\_distribution = df.groupby([‘Age group’, ‘industrial Category – A -Cultivators – persons’ , ‘Worked for 3 months or more but less than 6 months – Persons’])[‘Industrial Category - A – Cultivators – Males’].sum().reset\_index()

# Print the resulting DataFrame (optional)

Print(marginal\_workers\_distribution)

#To save the aggregated data to a new CSV file (optional)

#marginal\_workers\_distribution.to\_csv(‘marginal\_workers\_distribution.csv’. index=false)

Age group … Industrial Category - A – Cultivators – Males

0 15-34 … 0

1 15-34 … 3

2 15-34 … 6

3 15-34 … 4

4 15-35 … 8 .. … … …

540 ‘5-14 … 133

541 ‘5-14 … 137

542 ‘5-14 … 141

543 ‘5-14 … 664

544 ‘5-14 … 825

[545 rows x 4 columns]

***Data Visualization***

***Countplot:***

**#1**

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

df = pd.read\_csv("Marginal Workers.csv")

sns.set(style="whitegrid")

plt.figure(figsize=(10,6))

sns.countplot(data=df, x="Area Name")

plt.title("Count of Marginal Workers by Area Name")

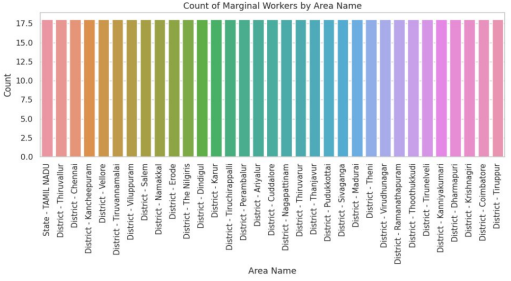
plt.xlabel("Area Name")

plt.ylabel("Count")

plt.xticks(rotation=90)

plt.tight\_layout()

plt.show()

**

**#2**

import seaborn as sns import

matplotlib.pyplot as plt

import pandas as pd

#Assuming ’Age group’ is the column representing age groups in your dataset

df = nd pd.read\_csv("Marginal Workers.csv")

sns.set(style="whitegrid")

plt.figure(figsize = (10, 6)

sns.countplot x=^ prime Aqe groupdata=dfpalette='viridis)

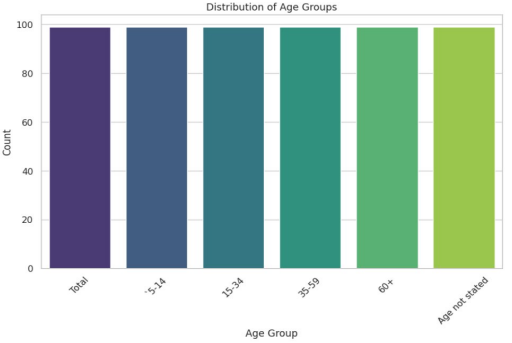
plt.title('Distribution of Age Groups')

plt.xlabel('Age Group')

plt.ylabel('Count)

plt.xticks (rotation = 45)

plt.show()

**

***Line Graph:***

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read\_csv("Marginal Workers.csv")

x\_values =df['Age group']

y\_values =df['Area Name']

plt.figure(figsize (10, 6))

plt.plot(x\_values, y\_values, marker='o', linestyle='-')

plt.title("Line Graph for Marginal Workers Dataset")

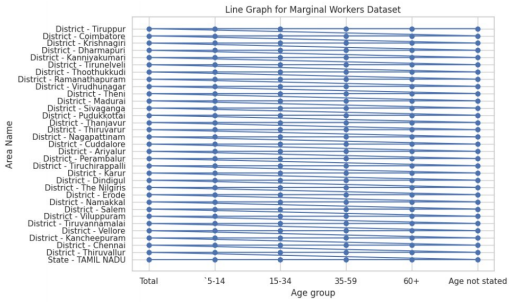
plt.xlabel("Age group")

plt.ylabel("Area Name")

plt.grid(True)

plt.tight\_layout()

plt.show()



***Histogram:***

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read\_csv("Marginal Workers.csv")

data=df['Age group']

plt.figure(figsize-(10, 6))

plt.hist(data, bins=20, color='skyblue', edgecolor='black')

plt.title("Histogram of Age in Marginal Workers Dataset")

plt.xlabel("Age group")

plt.ylabel("Frequency")

plt.tight\_layout()

plt.show()



***Scatter Plot:***

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read\_csv("Marginal Workers.csv")

x\_values =df['Worked for 3 months or more but less than 6 months - Persons']

y\_values =df['Worked for 3 months or more but less than 6 months - Males']

plt.figure(figsize-(10, 6))

plt.scatter(x\_values, y\_values, c='blue', alpha=0.5, edgecolors='k')

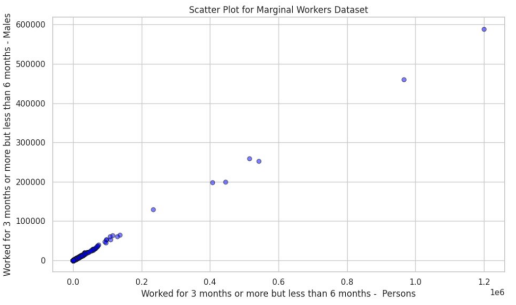
plt.title("Scatter Plot for Marginal Workers Dataset")

plt.xlabel("Worked for 3 months or more but less than 6 months - Persons")

plt.ylabel("Worked for 3 months or more but less than 6 months - Males")

pit tight\_layout()

plt.show()



***Heatmap:***

import seaborn as sns

import pandas as pd

import matplotlib.pyplot as plt

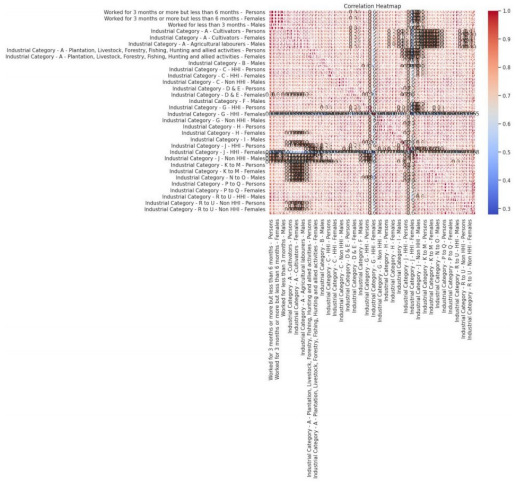
df = pd.read\_csv('Marginal Workers.csv')

plt.figure(figsize=(10,8))

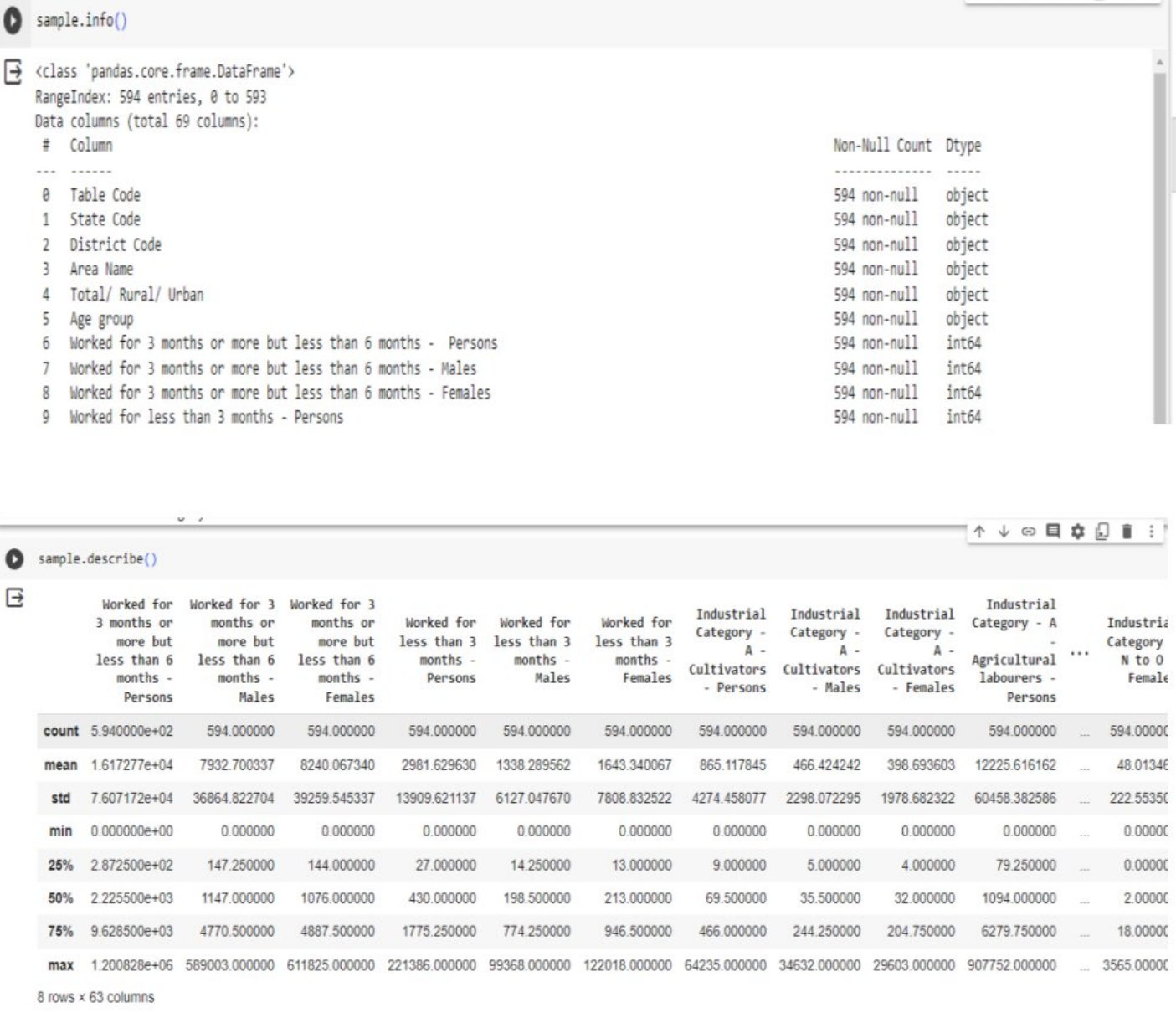
sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt=".2f") plt.title("Correlation Heatmap')

plt.show()

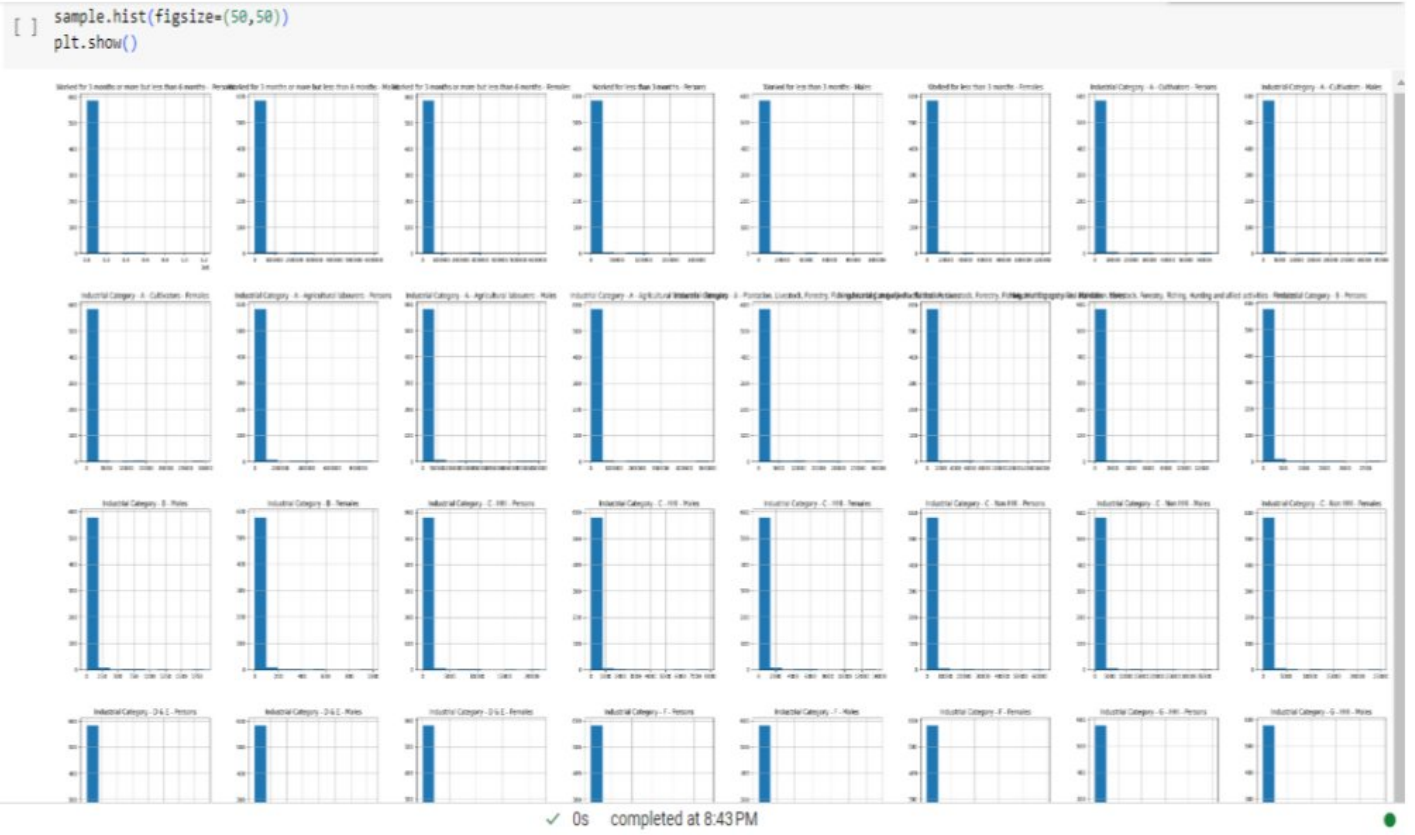




**OPERATION:**

****

**After replacement:**

****

**C0NCLUSION:** 