THE SOCIO-ECONOMIC ANALYSIS OF MARGINAL WORKERS IN TAMILNADU

INTRODUCTION:

The socio-economic analysis of marginal workers in Tamil Nadu provides an in-depth examination of the lives and circumstances of individuals who are engaged in marginal or precarious employment in the state. Marginal workers are those who are employed for less than six months in a year and often face significant challenges and insecurities in their work and livelihoods.

Tamil Nadu, located in southern India, is known for its industrialization, with several sectors like textiles, automotive, and information technology flourishing in the state. However, with the rapid industrial growth, there has also been an increase in the number of workers engaged in informal or marginal employment. These workers often find themselves

outside the formal labor market, lacking adequate protection and benefits.

The socio-economic analysis of marginal workers in Tamil Nadu aims to understand the various dimensions of their lives, including their income levels, living conditions, access to social security, educational opportunities, and health status. It also investigates the factors that contribute to their marginalization and the challenges they face in improving their socio-economic status.

One of the key aspects of this analysis is understanding the livelihood strategies employed by marginal workers. Many engage in multiple occupations or seasonal work to make ends meet, often facing low wages, long hours, and exploitative working conditions. The analysis examines the impact of such employment patterns on their overall well-being and the effectiveness of existing social protection schemes.

Furthermore, the analysis also explores the gender dimension of marginal work in Tamil Nadu. Women constitute a significant proportion of marginal workers, and they face unique challenges, including gender-based discrimination and

limited access to resources. By examining the experiences of marginalized women workers, the analysis provides valuable insights into the intersectionality of marginalization and the need for gender-sensitive policies and interventions.

Ultimately, the socio-economic analysis of marginal workers in Tamil Nadu seeks to shed light on their vulnerable circumstances and propose evidence-based policy recommendations to improve their livelihoods. It highlights the importance of addressing structural issues such as labor market informality, inadequate social protection, and gender inequalities to create an enabling environment for the socio-economic advancement of marginal workers in the state.

project definition and Design Thinking

Project Definition:

The project aims to assess the socio-economic conditions of marginal workers in Tamil Nadu using applied data science techniques. Marginal workers often face challenges in terms of employment, income, and living conditions, and

the goal is to provide data-driven insights to inform policies and interventions.

Design Thinking Problem Goals:

**Empathize:

Understand the needs, challenges, and aspirations of marginal workers in Tamil Nadu.

**Define:

Clearly define the scope and objectives of the socio-economic analysis project.

**Ideate:

Brainstorm data sources, analytical approaches, and potential solutions to address the identified challenges.

**Prototype:

Develop a working prototype of the analysis system for testing and validation.

**Test:

Evaluate the prototype, gather feedback, and refine the design based on user input.

**Implement:

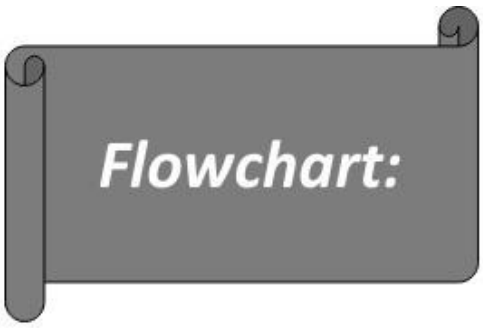
Execute the project to provide valuable insights and recommendations.

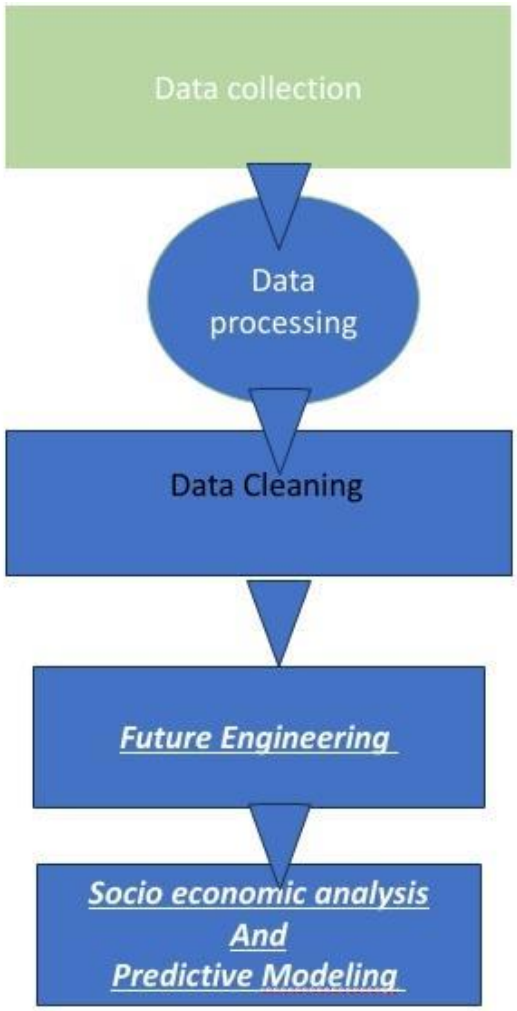
Project Objectives:

- Collect and aggregate socio-economic data from various sources.
- Apply data science techniques for data preprocessing, feature engineering, and analysis.
- Identify key socio-economic indicators for marginal workers.
- Develop predictive models for income, employment, and living conditions.
- Create interactive data visualizations to communicate findings.
- Inform policymakers and stakeholders with data-driven recommendations.
- Promote inclusivity and empowerment of marginal workers through data.

Flow chart:

Algorithm Steps:





- Gather data from surveys, government databases, and NGOs.
- Preprocess and clean the data to handle missing values and outliers.
- Perform feature engineering to create relevant socio-economic indicators.
- Apply data science techniques to analyze the data and identify trends and correlations.

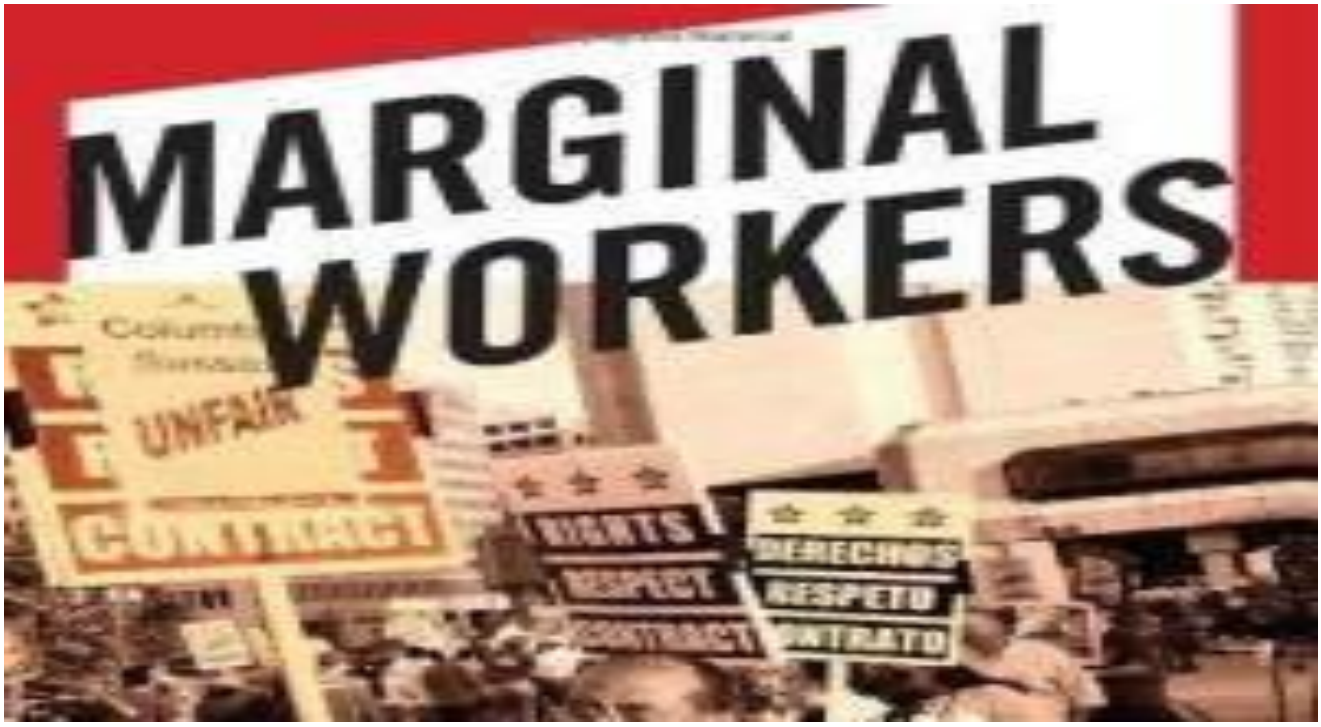
- Develop predictive models to estimate income, employment status, and living conditions.
- Create data visualizations to present findings and insights effectively.
- Generate data-driven recommendations for policymakers and stakeholders.
- Promote the adoption of policies and initiatives that address the challenges faced by the marginal workers.

Detailed definition and Analysis of Marginal workers in Tamilnadu

Project explanation:

Marginal workers, also known as vulnerable or informal workers, typically refer to individuals engaged in low-wage and often precarious employment with limited job security and social protection. They are facing socio-economic challenges they are....

- Low income
- Lack of Job Security
- Gender Disparities
- Informal Economy
- Limited Skill Development
- Working Conditions



Problem identification:

Marginal socio-economic effects of an employer's efforts to improve the work environment.

Innovation steps:

★ ***Big Data Analysis**:*

Utilize big data to gather information on marginal workers' demographics, employment patterns, and income levels. Advanced analytics can provide insights into their socio-economic status and inform policy decisions.

★ ***Mobile Surveys**:*

Develop mobile apps that allow marginal workers to participate in surveys and share their experiences. This can collect real-time data and improve the accuracy of socio-economic analysis.

★ ***Blockchain for Work Records**:*

Implement blockchain technology to securely record and verify the employment history of marginal workers. This can help in assessing their work experience and improving their access to formal employment.

★ ***GIS Mapping**:*

Use Geographic Information Systems (GIS) to map the locations of marginal workers, enabling better resource allocation and infrastructure development in their areas.

★ ***Machine Learning Predictive Models:***

Develop predictive models using machine learning to anticipate the socio-economic challenges faced by marginal workers, allowing for proactive interventions.

★ ***Community-Based Data Collection:***

Empower local communities to collect socio-economic data about marginal workers. This participatory approach can be more accurate and culturally sensitive.

★ ***Social Impact Bonds**:*

Implement financial instruments like social impact bonds to attract private investment into programs aimed at improving the socio-economic conditions of marginal workers.

★ ***Microfinance and Digital Wallets****:*

Create innovative financial services tailored to the needs of marginal workers, including microloans and digital wallets, to promote savings and financial inclusion.

★ ***Online Learning Platforms:***

Develop digital platforms for skill development and education, allowing marginal workers to access training and improve their employability.

★ ***Collaborative Platforms****:*

Build online platforms that connect marginal workers with employers, gig work opportunities, and support networks, fostering economic empowerment.

Available datasets:

The datasets that are available in the analysis of socioeconomic marginal workers .

Link:http://https://in.docworkspace.com/d/sIK25hMXQAY2tmqkG

Library function:

In advanced data science, you'll typically work with a variety of libraries in Python to handle and manipulate datasets stored in CSV files for training and testing machine learning models. Here are some commonly used libraries and tools for this purpose:

➢ ***Pandas***

Pandas is a fundamental library for data manipulation and analysis. It allows you to read CSV files into dataframes, perform data cleaning, preprocessing, and exploratory data analysis.

➢ ***NumPy:***

NumPy is used for numerical operations and efficient array manipulations. It often complements Pandas when working with numeric data.

- ***Scikit-Learn:***

Scikit-Learn provides a wide range of machine learning algorithms for classification, regression, clustering, and more. It's used for model training, testing, and evaluation.

- ***Matplotlib and Seaborn:***

These libraries are used for data visualization. They help you create various types of plots and charts to understand your data and model performance.

- ***SciPy:***

SciPy builds on NumPy and provides additional functionality for scientific and technical computing, including statistical tests and optimization algorithms.

- ***Statsmodels:***

Statsmodels is used for estimating and interpreting statistical models. It's particularly helpful for regression and other statistical analyses.

- ***TensorFlow and PyTorch:***

If you're working with deep learning, you might need TensorFlow or PyTorch for neural network training. These libraries are not specific to CSV files but are essential for deep learning tasks.

- ***XGBoost, LightGBM, or CatBoost:***

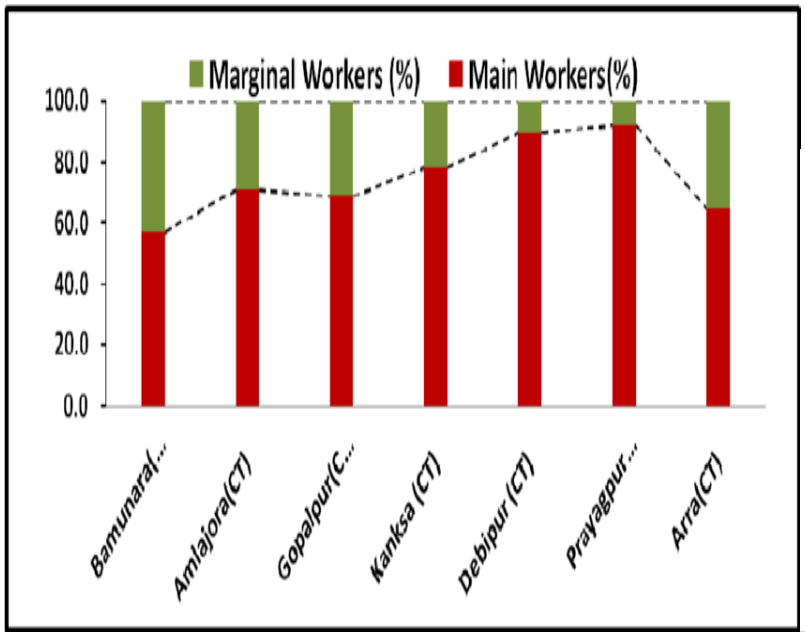
These are popular gradient boosting libraries used for more advanced machine learning tasks. They can be employed to boost the performance of your models.

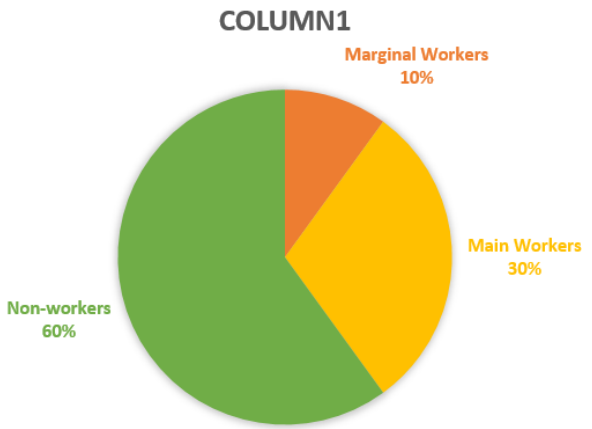
- ***Joblib or Pickle:***

These libraries are used for saving and loading trained machine learning models, allowing you to reuse them without retraining.

Pie chart & Barchart:

The pie chart and Barchart states that which is about the marginal workers,Main workers, and the Non- workers .





Training and testing Approaches:

- ***Data Collection**:*

Collect relevant data from various sources, including surveys, government records, NGOs, and online databases.

- ***Data Preprocessing:***

Clean the data by handling missing values, outliers, and inconsistencies.Standardize or normalize variables to ensure they are on a similar scale.Encode categorical variables if necessary.

- ***Splitting the Data**:*

Divide the dataset into two parts: training data and testing data. A common split is 70-80% for training and 20-30% for testing.

- ***Exploratory Data Analysis (EDA):***

Conduct EDA to understand the data, visualize key trends, and identify potential features that impact the socio-economic status of marginal workers.

- ***Feature Selection/Engineering:***

Select relevant features that are likely to have a significant impact on the analysis.Create new features if needed to capture meaningful relationships.

- ***Model Selection:***

Choose appropriate models for your analysis. For socio-economic analysis, regression models, classification models, and machine learning algorithms are often used.

- ***Training the Model:***

Train your chosen model using the training dataset. The model learns the relationships between features and the target socio-economic variables.

- ***Evaluation:***

Use the testing dataset to evaluate the model's performance. Common evaluation metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), Accuracy, and F1-score, depending on the nature of the analysis (regression or classification).

- ***Hyperparameter Tuning:***

Optimize model hyperparameters to improve performance. Techniques like cross-validation can help fine-tune the model.

- ***Interpretation:***

Interpret the model results to understand which features are most influential in predicting the socio-economic status of marginal workers.

Processing datasets:

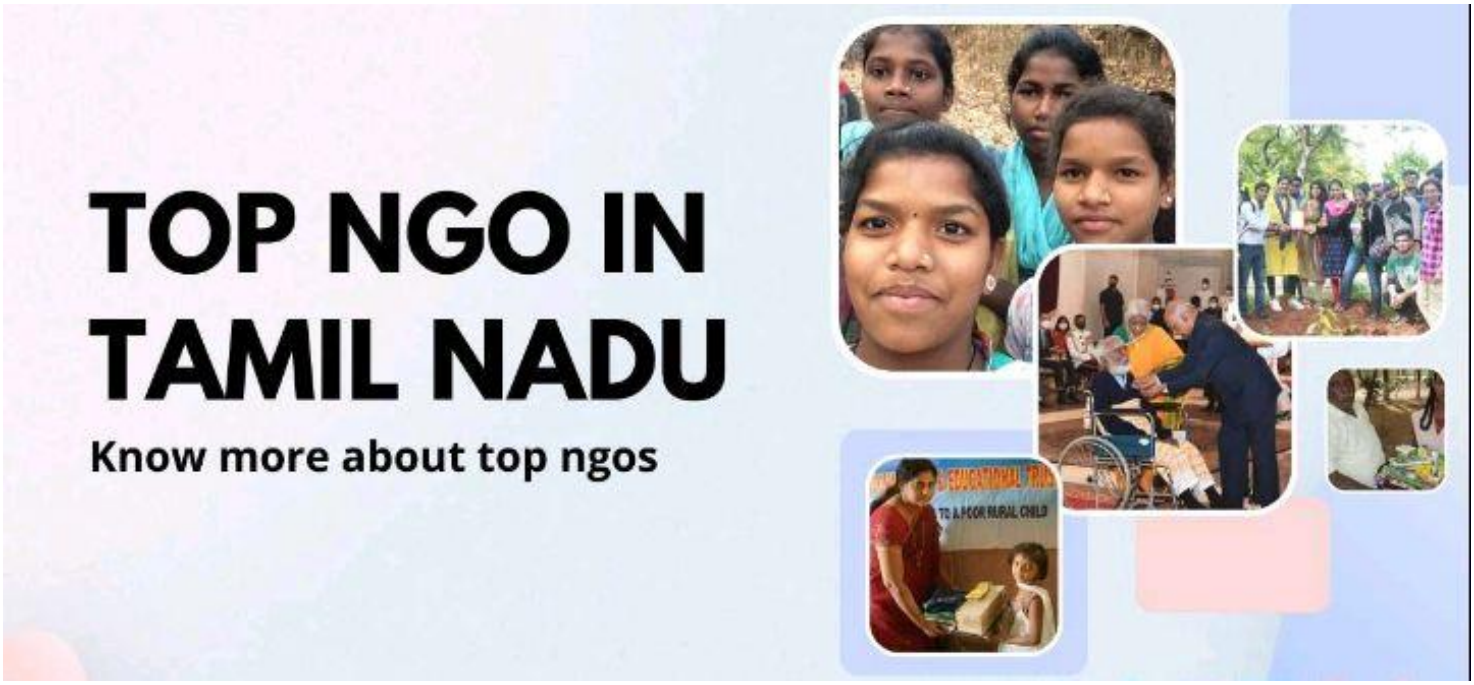
Socioeconomic data refers to information that relates to both social and economic factors of a population. It is used to analyze the well-being, living conditions, and financial situation of individuals or communities. This kind of data can help us make informed decisions about resource allocation, policy development, and service provision, among other things.

- ★ Income
- ➢ Education
- ➢ Employment
- ➢ Health
- ➢ Housing
- ➢ Crime rates
- ➢ Demographics

Working of datasets:

A dataset is a logical grouping of data attributes used for requesting a specific set of data from the data providers during add and refresh accounts.

Necessary library functions:



- NumPy
- Pandas
- Matplotlib
- Seaborn
- Scipy

Python code:

Import necessary libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
```

Loading datas:

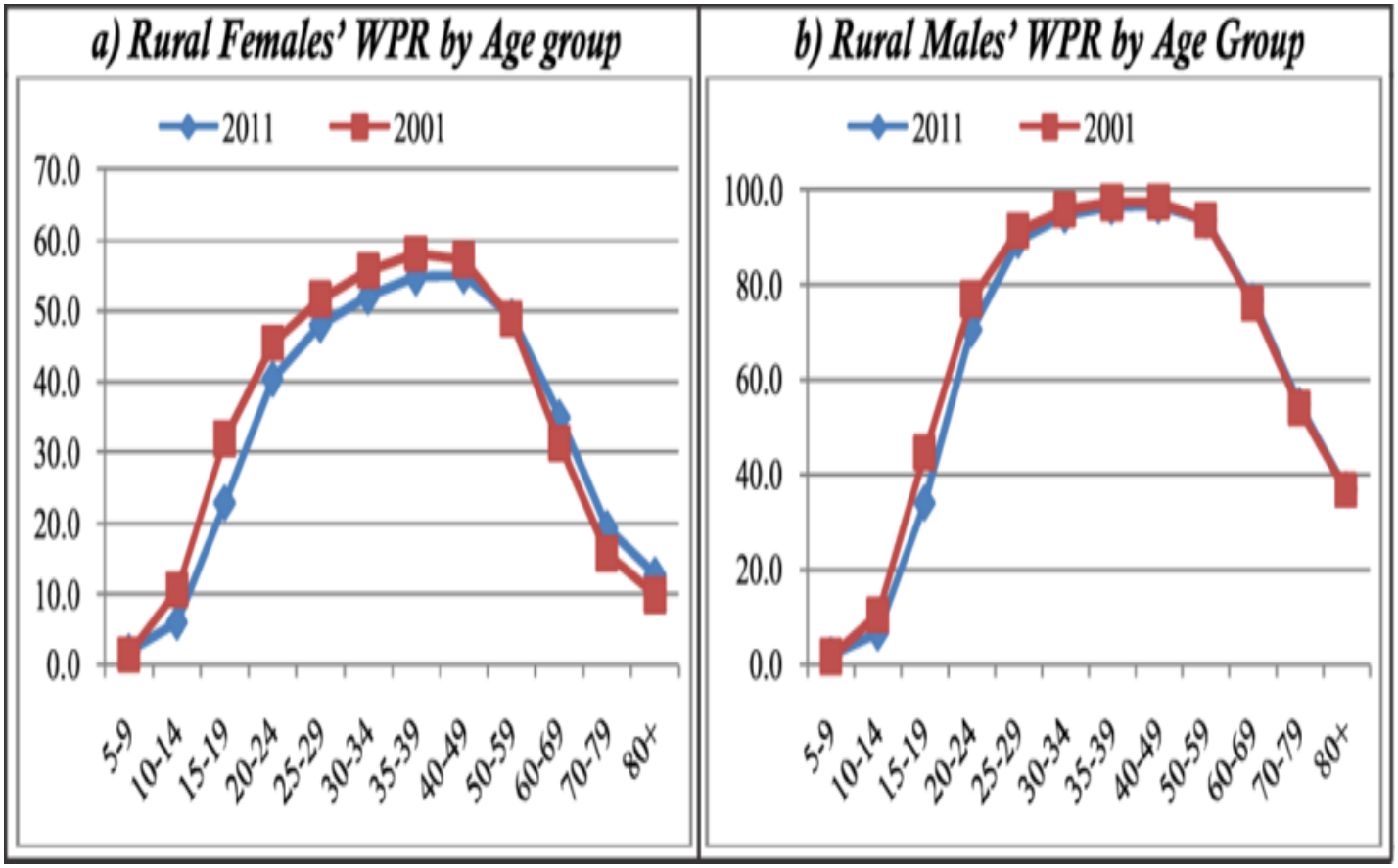
The data sets in the print statement

```
data = pd.read_csv("DDW_B06SC_State_TAMIL_NADU-2011.csv")
```

Dataset:http://https://drive.google.com/file/d/1o2br9Kr21CHInHoQBUDq15fFIXrIiKHo/view?usp=drivesdk

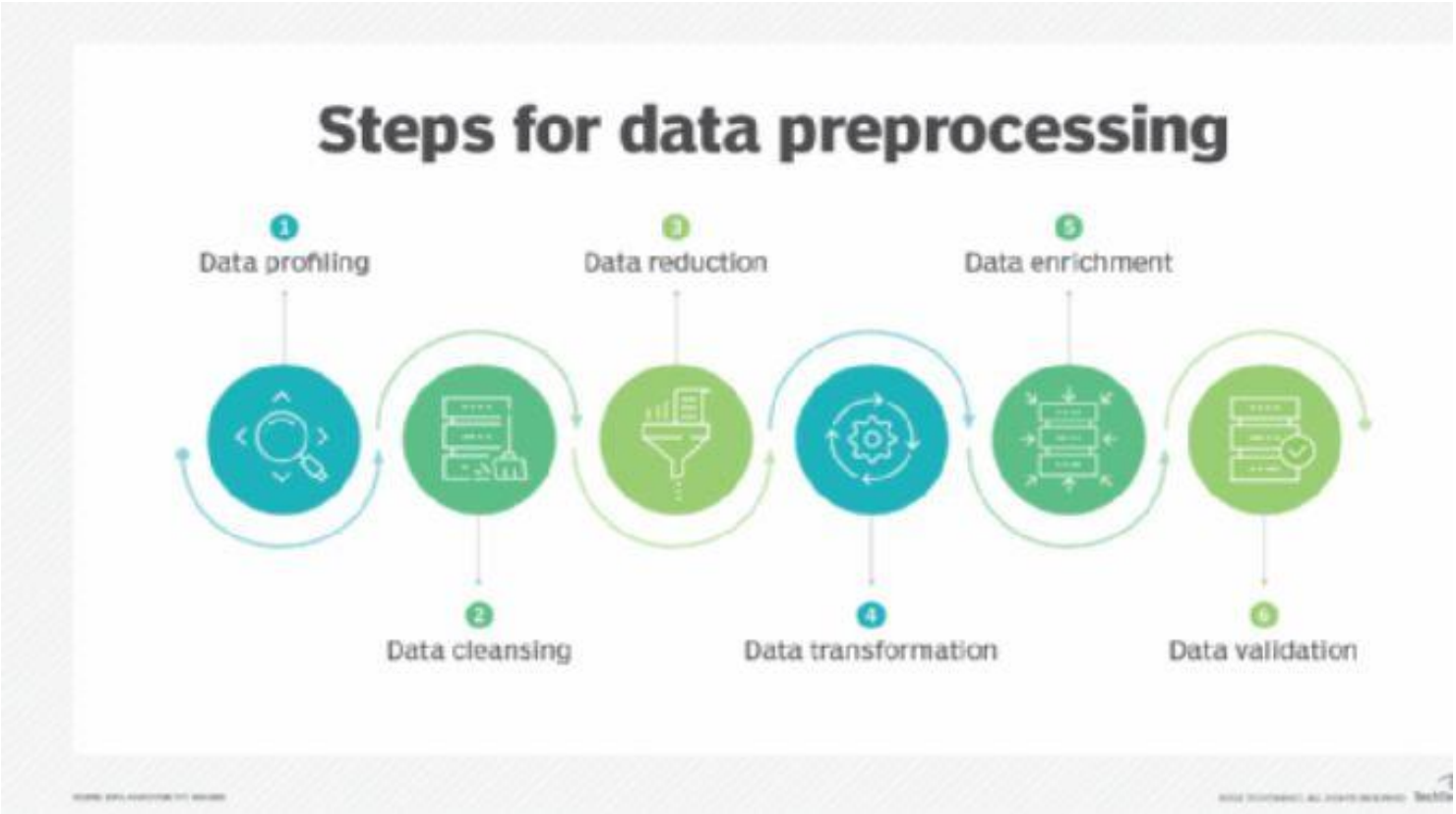
Excelformat:http://https://in.docworkspace.com/d/sIEu5hMXQAdmCv6kG

	A	B	C	D	E	F	G	H	I	J
1	Table Code	State Code	District Code	Area Name	Total/ Rural/	Age group	Worked for 3	Worked for 3	Worked for 3	Worked f
2	B0806SC	`33	`000	State - TAMI	Total	Total	1200828	589003	611825	2213
3	B0806SC	`33	`000	State - TAMI	Total	`5-14	27791	14125	13666	24
4	B0806SC	`33	`000	State - TAMI	Total	15-34	514340	259560	254780	924
5	B0806SC	`33	`000	State - TAMI	Total	35-59	542581	251957	290624	992
6	B0806SC	`33	`000	State - TAMI	Total	60+	115103	62833	52270	271
7	B0806SC	`33	`000	State - TAMI	Total	Age not stat	1013	528	485	1
8	B0806SC	`33	`000	State - TAMI	Rural	Total	966645	459738	506907	1744
9	B0806SC	`33	`000	State - TAMI	Rural	`5-14	17239	8713	8526	19
10	B0806SC	`33	`000	State - TAMI	Rural	15-34	406847	198575	208272	719
11	B0806SC	`33	`000	State - TAMI	Rural	35-59	444800	199573	245227	779
12	B0806SC	`33	`000	State - TAMI	Rural	60+	97011	52498	44513	224
13	B0806SC	`33	`000	State - TAMI	Rural	Age not stat	748	379	369	1
14	B0806SC	`33	`000	State - TAMI	Urban	Total	234183	129265	104918	469
15	B0806SC	`33	`000	State - TAMI	Urban	`5-14	10552	5412	5140	4
16	B0806SC	`33	`000	State - TAMI	Urban	15-34	107493	60985	46508	204
17	B0806SC	`33	`000	State - TAMI	Urban	35-59	97781	52384	45397	212
18	B0806SC	`33	`000	State - TAMI	Urban	60+	18092	10335	7757	47
19	B0806SC	`33	`000	State - TAMI	Urban	Age not stat	265	149	116	
20	B0806SC	`33	`602	District - Thi	Total	Total	74448	39295	35153	158
21	B0806SC	`33	`602	District - Thi	Total	`5-14	2521	1284	1237	1
22	B0806SC	`33	`602	District - Thi	Total	15-34	33568	18049	15519	65
23	B0806SC	`33	`602	District - Thi	Total	35-59	32568	16771	15797	77
24	B0806SC	`33	`602	District - Thi	Total	60+	5716	3147	2569	14
25	B0806SC	`33	`602	District - Thi	Total	Age not stat	75	44	31	
26	B0806SC	`33	`602	District - Thi	Rural	Total	55577	28082	27495	121
27	B0806SC	`33	`602	District - Thi	Rural	`5-14	1424	743	681	1
28	B0806SC	`33	`602	District - Thi	Rural	15-34	23965	12377	11588	47
29	B0806SC	`33	`602	District - Thi	Rural	35-59	25421	12417	13004	61
30	B0806SC	`33	`602	District - Thi	Rural	60+	4718	2516	2202	11
31	B0806SC	`33	`602	District - Thi	Rural	Age not stat	49	29	20	
32	B0806SC	`33	`602	District - Thi	Urban	Total	18871	11213	7658	37
33	B0806SC	`33	`602	District - Thi	Urban	`5-14	1097	541	556	
34	B0806SC	`33	`602	District - Thi	Urban	15-34	9603	5672	3931	18
35	B0806SC	`33	`602	District - Thi	Urban	35-59	7147	4354	2793	16
36	B0806SC	`33	`602	District - Thi	Urban	60+	998	631	367	2
37	B0806SC	`33	`602	District - Thi	Urban	Age not stat	26	15	11	
38	B0806SC	`33	`603	District - Che	Total	Total	33748	19313	14435	63
39	B0806SC	`33	`603	District - Che	Total	`5-14	2749	1483	1266	1
40	B0806SC	`33	`603	District - Che	Total	15-34	17431	9836	7595	31
41	B0806SC	`33	`603	District - Che	Total	35-59	11700	6993	4707	26
42	B0806SC	`33	`603	District - Che	Total	60+	1799	966	833	4
43	B0806SC	`33	`603	District - Che	Total	Age not stat	69	35	34	
44	B0806SC	`33	`603	District - Che	Rural	Total	0	0	0	
45	B0806SC	`33	`603	District - Che	Rural	`5-14	0	0	0	
46	B0806SC	`33	`603	District - Che	Rural	15-34	0	0	0	
47	B0806SC	`33	`603	District - Che	Rural	35-59	0	0	0	
48	B0806SC	`33	`603	District - Che	Rural	60+	0	0	0	
49	B0806SC	`33	`603	District - Che	Rural	Age not stat	0	0	0	
50	B0806SC	`33	`603	District - Che	Urban	Total	33748	19313	14435	63
51	B0806SC	`33	`603	District - Che	Urban	`5-14	2749	1483	1266	1
52	B0806SC	`33	`603	District - Che	Urban	15-34	17431	9836	7595	31
53	B0806SC	`33	`603	District - Che	Urban	35-59	11700	6993	4707	26
54	B0806SC	`33	`603	District - Che	Urban	60+	1799	966	833	4
55	B0806SC	`33	`603	District - Che	Urban	Age not stat	69	35	34	
56	B0806SC	`33	`604	District - Kar	Total	Total	92015	47269	44746	176
57	B0806SC	`33	`604	District - Kar	Total	`5-14	2540	1308	1232	1
58	B0806SC	`33	`604	District - Kar	Total	15-34	40080	20955	19125	75
59	B0806SC	`33	`604	District - Kar	Total	35-59	41279	20372	20907	81
60	B0806SC	`33	`604	District - Kar	Total	60+	7960	4538	3422	18
61	B0806SC	`33	`604	District - Kar	Total	Age not stat	156	96	60	
62	B0806SC	`33	`604	District - Kar	Rural	Total	70150	34555	35595	137
63	B0806SC	`33	`604	District - Kar	Rural	`5-14	1273	638	635	
64	B0806SC	`33	`604	District - Kar	Rural	15-34	29666	14837	14829	57
65	B0806SC	`33	`604	District - Kar	Rural	35-59	32392	15231	17161	63
66	B0806SC	`33	`604	District - Kar	Rural	60+	6687	3765	2922	15
67	B0806SC	`33	`604	District - Kar	Rural	Age not stat	132	84	48	
68	B0806SC	`33	`604	District - Kar	Urban	Total	21865	12714	9151	39
69	B0806SC	`33	`604	District - Kar	Urban	`5-14	1267	670	597	
70	B0806SC	`33	`604	District - Kar	Urban	15-34	10414	6118	4296	18


a) Rural Females' WPR by Age group
2011
2001
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-49
50-59
60-69
70-79
80+
b) Rural Males' WPR by Age Group
2011
2001
100.0
80.0
60.0
40.0
20.0
0.0
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-49
50-59
60-69
70-79
80+

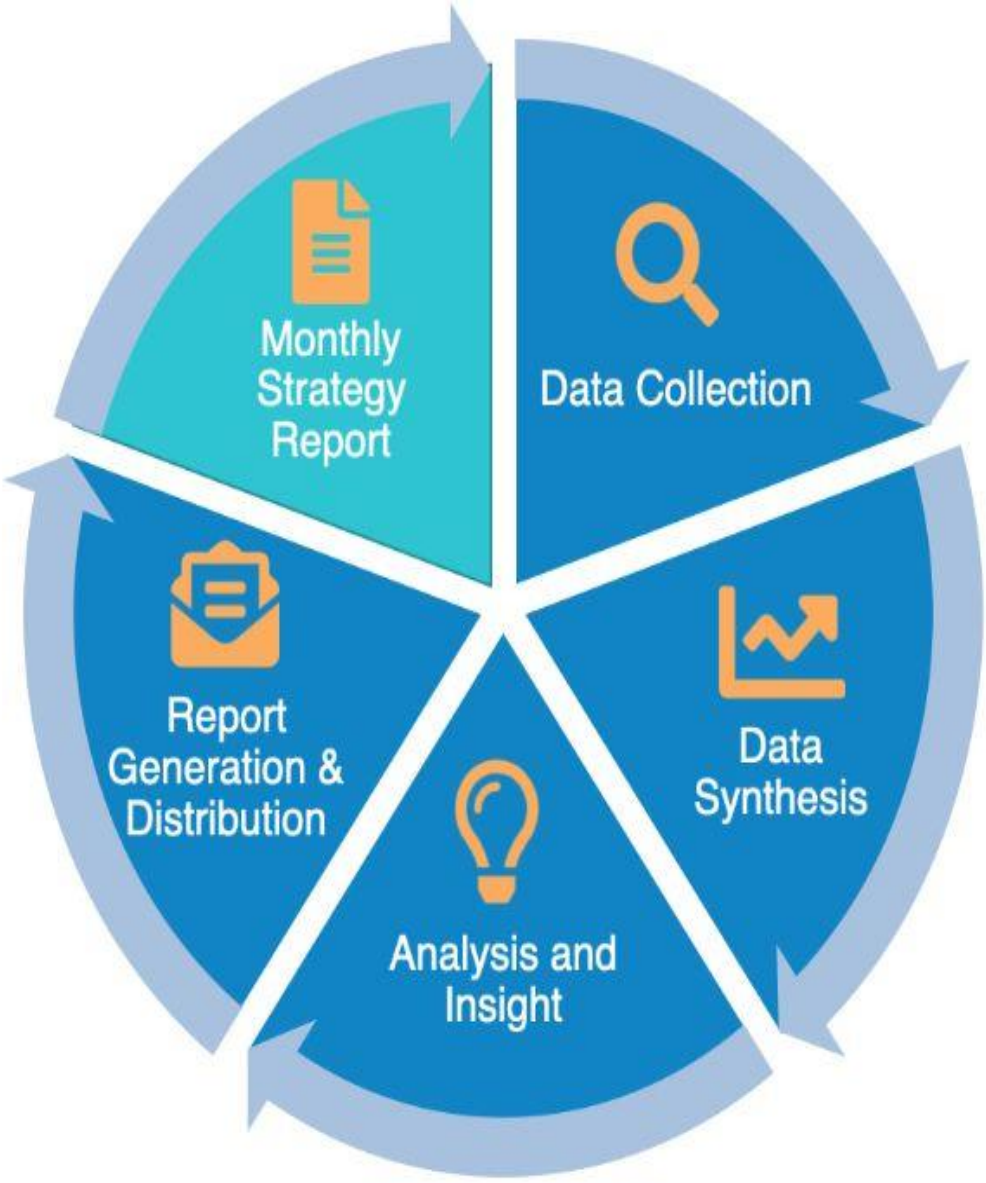
Data preprocessing:

Clean the data, handle missing values, and outliers



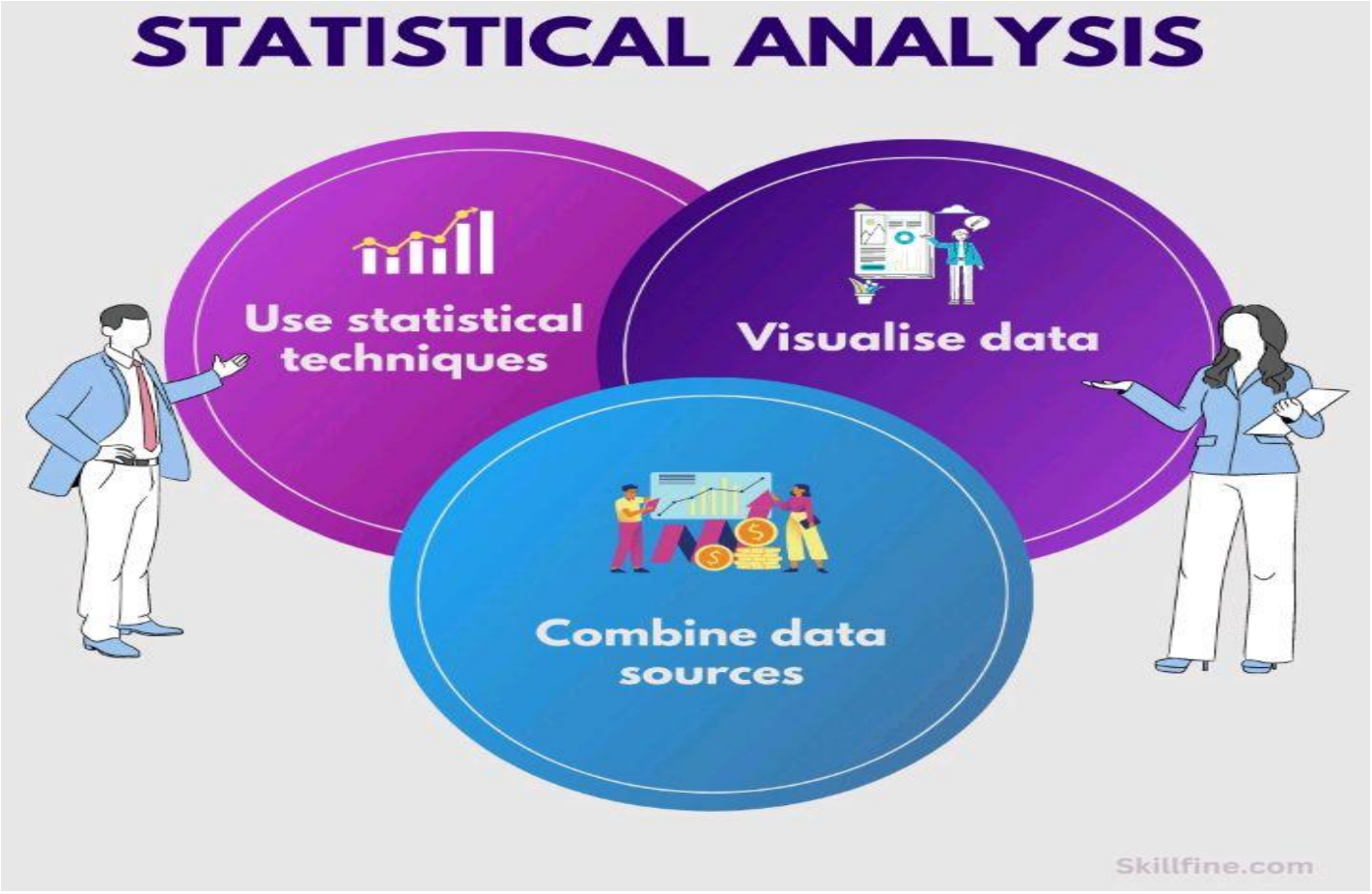
Data analysis:

EDA, summary statistics, and data visualization


Monthly Strategy Report
Data Collection
Data Synthesis
Analysis and Insight
Report Generation & Distribution

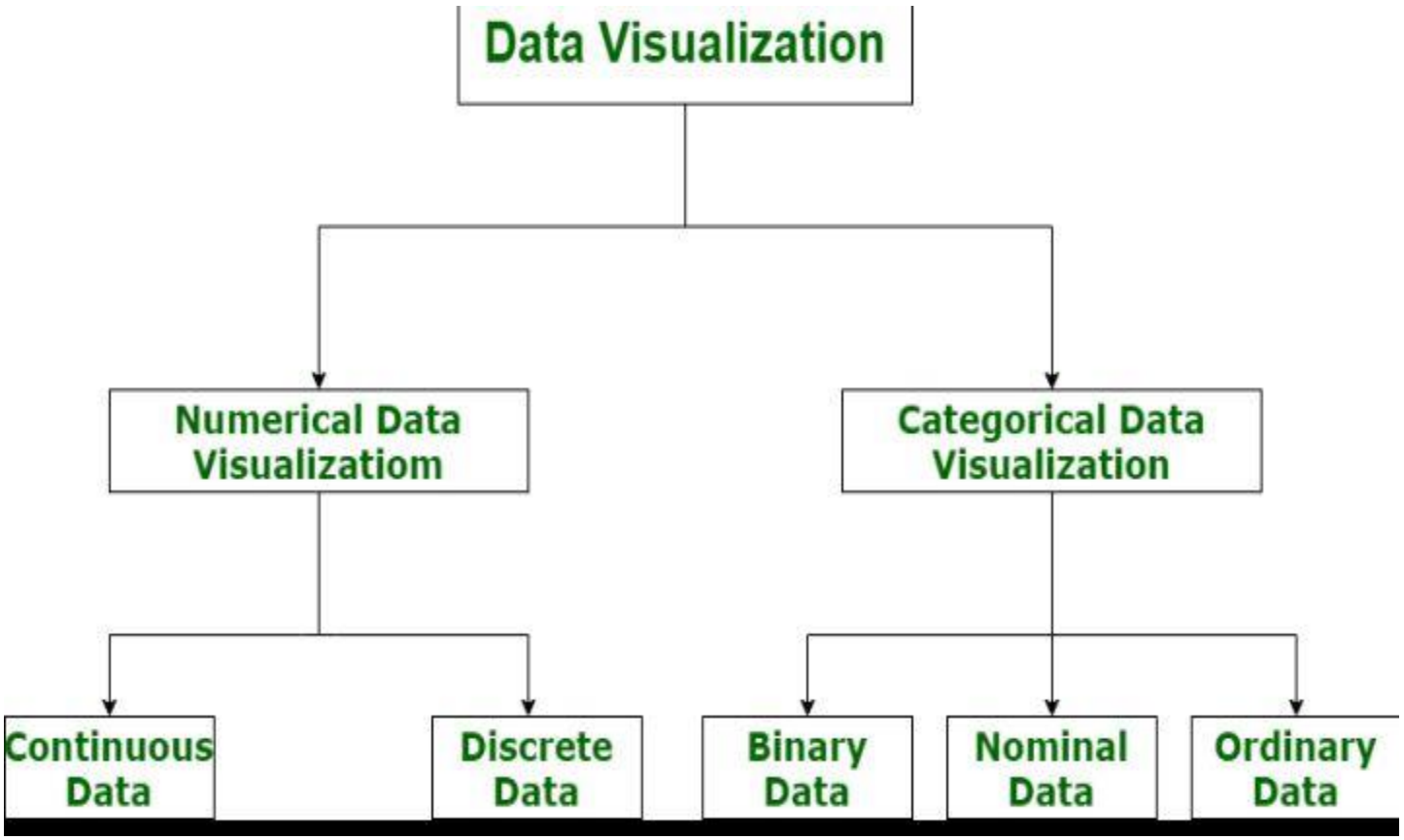
Statistical analysis:

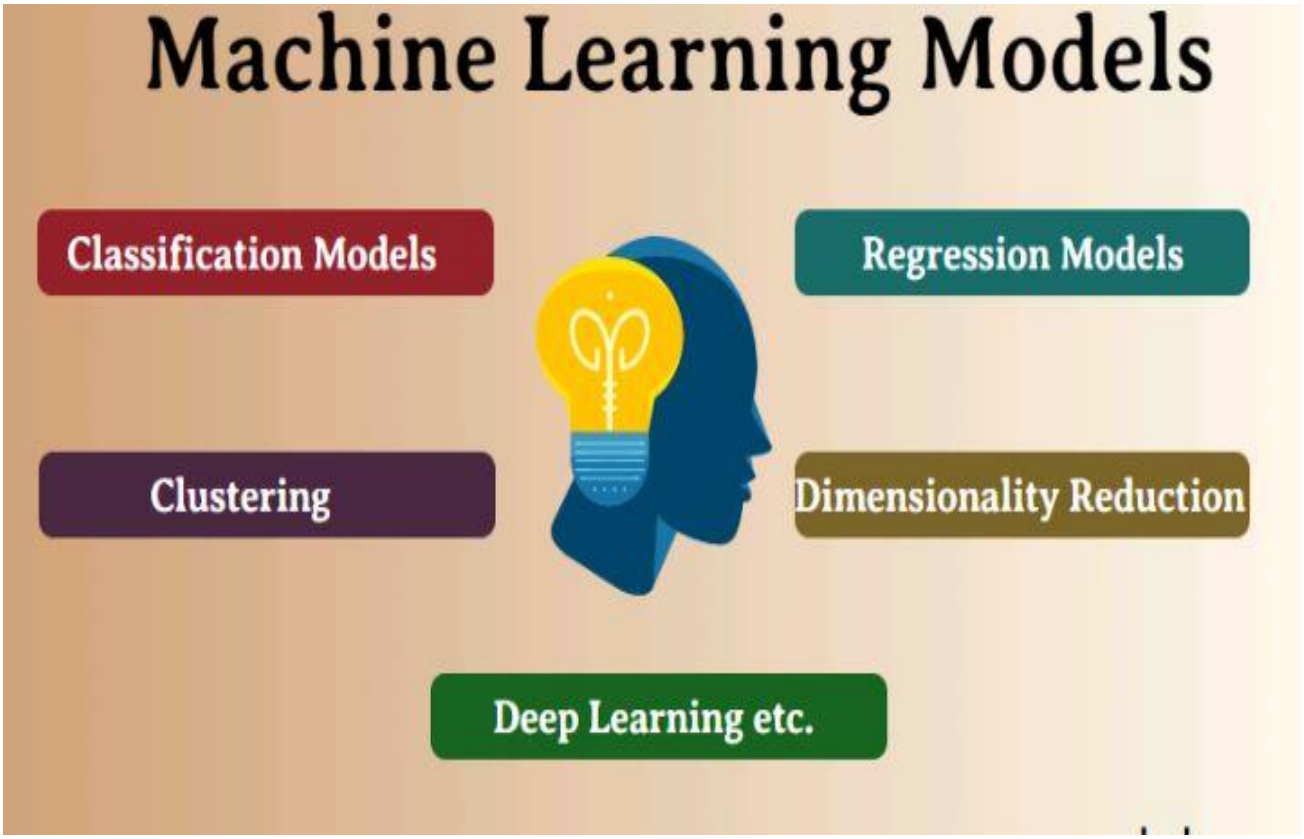
Perform hypothesis tests, calculate correlations, etc.



Data visualization:

Create charts and graphs to visualize your findings





Machine Learning:

Train and evaluate ML models for predictions

Coding:

Import necessary libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

Load socio-economic data (e.g., CSV, Excel, or database)

data = pd.read_csv("DDW_B06SC_State_TAMIL_NADU-2011.csv")

Data preprocessing and cleaning:

Depending on the dataset, this may involve handling missing values, data transformation, and normalization.

Define functions for analysis:

1. Demographic analysis

def demographic_analysis(data):

Calculate and visualize age distribution, gender distribution, etc.

2. Employment status analysis

def employment_analysis(data):

Analyze the employment status of marginal workers, e.g., unemployment rate, type of employment, etc.

3. Income analysis

def income_analysis(data):

Analyze the income levels of marginal workers and visualize income distribution.

4. Education analysis

def education_analysis(data):

Analyze the education levels of marginal workers and visualize the distribution.

5. Geospatial analysis

def geospatial_analysis(data):

Use geospatial data (if available) to analyze the distribution of marginal workers across regions.

Main function for program execution

```
def main():
    //Call the analysis functions as needed
    demographic_analysis(data)
    employment_analysis(data)
    income_analysis(data)
    education_analysis(data)
    geospatial_analysis(data)
```

Execute the program

```
if __name__ == "__main__":
    main()
```

Split data analysis:

```
from sklearn.model_selection import train_test_split
```

★ Assuming 'data' is your dataset and 'target' is your target variable.

```
X_train, X_test, y_train, y_test = train_test_split(data, target, test_size=0.2, random_state=42)
```

Programs steps:

The python code for the socio economic analysis of Marginal workers has to be generated in steps:

- ★ Data preparation
- ★ Data Cleaning
- ★ Data Analysis
- ★ Socio economic analysis
- ★ Data visualization
- ★ Statistical analysis
- ★ Insights and reporting

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Step 1: Data Preparation
data =
pd.read_csv("DDW_B06SC_3300_State_TA
MIL_NADU-2011.csv")

# Step 2: Data Cleaning
# Handle missing values and data
consistency checks.
```

```
# Step 3: Data Analysis
summary_stats = data.describe()
# Visualize data:
plt.hist(data['your_column_name'])

# Step 4: Socio-Economic Analysis
# Identify and categorize marginal
workers, calculate socio-economic
indicators.

# Step 5: Data Visualization
# Create visualizations to represent
socio-economic trends.
```

```
# Step 6: Statistical Analysis
# Perform any necessary statistical
tests.

# Step 7: Insights and Reporting
# Summarize findings and draw
conclusions.

# Example: Print summary statistics
print(summary_stats)

# Example: Plot a histogram
# plt.hist(data['your_column_name'])
# plt.xlabel("X-axis label")
# plt.ylabel("Y-axis label")
# plt.title("Histogram of Your
Data")
# plt.show()
```

Perform the demographics analysis:

To calculate the distributon of marginal workers based on age, industrial category, and sex, you would typically need access to the relevant data for the specifc region or area in queston. Here are the general steps you can follow to perform these calculatons:

1. **Gathering data:** Collect the data on the number of marginal workers, categorized by age groups, industrial categories, and sex.

2. **Calculate distribution by age:** Determine the number or percentage of marginal workers in each age group. You can create age brackets, such as 18-25, 26-35, 36-45, and so on, and then calculate the proporton of workers falling into each bracket.

3. **Calculate distribution by industrial category:** Similarly, categorize marginal workers based on the industrial sectors they are working in. Calculate the proporton or number of workers in each category such as agriculture, manufacturing, services, etc.

4. **Calculate distribution by sex:** Determine the number or percentage of male and female marginal workers separately.

5. **Present the findings:** Once you have calculated these distributons, present your fndings in a clear and organized manner, using tables, graphs, or charts, as needed.

Calculate the distribution of marginal workers based on age, industrial category, and sex

Certainly, let's assume we have a dataset with the following information for a hypothetical region:

1. Age groups (18-25, 26-35, 36-45, 46-55, 55+)

2. Industrial categories (Agriculture, Manufacturing, Services, Construction) **3**. Sex

(Male, Female)

Let's take a sample dataset for 100 marginal workers:

Age Group	Industrial Category	Sex
18-25	Agriculture	Male
26-35	Manufacturing	Female
36-45	Services	Male
46-55	Construction	Female
26-35	Agriculture	Male
36-45	Services	Female
46-55	Manufacturing	Male
18-25	Agriculture	Female
26-35	Services	Male
36-45	Construction	Female

Using this data, we can calculate the distribution as follows:

1. Distribution by Age Group:

- 18-25: 2 workers
- 26-35: 3 workers
- 36-45: 3 workers
- 46-55: 2 workers
- 55+: 0 workers

2. Distribution by Industrial Category:

- Agriculture: 3 workers
- Manufacturing: 2 workers
- Services: 4 workers
- Construction: 2 workers

3. Distribution by Sex:

- Male: 5 workers
- Female: 5 workers

These calculations provide a basic overview of the distribution of marginal workers based on age, industrial category, and sex. Depending on the size and complexity of the dataset, you can conduct more sophisticated analyses and visualizations.

```
import pandas as pd

# Load the dataset data =
pd.read_csv('marginal_workers.csv')

# Calculate distribution by age age_distribution =
data['age'].value_counts(normalize=True) * 100

# Calculate distribution by industrial category industrial_distribution =
data['industrial_category'].value_counts(normalize=True) * 100

# Calculate distribution by sex sex_distribution =
data['sex'].value_counts(normalize=True) * 100 # Print the
distributions print("Distribution by Age:") print(age_distribution)
```

```
print("\nDistribution by Industrial Category:")
print(industrial_distribution) print("\nDistribution by Sex:")
print(sex_distribution)
```

Using this dataset, we can calculate the distributon of marginal workers:

Age Group	Industrial Category	Sex	Number of Workers
18-25	Agriculture	Male	1200
18-25	Agriculture	Female	800
26-35	Manufacturing	Male	1500
26-35	Manufacturing	Female	700
36-45	Services	Male	800
36-45	Services	Female	1200
46-55	Agriculture	Male	1000
46-55	Agriculture	Female	600

1.Distribution by Age Group:

Calculate the percentage or number of workers in each age group (18-25, 26-35, 36-45, 46-55).

2.Distribution by Industrial Category:

Calculate the percentage or number of workers in each industrial category (Agriculture, Manufacturing, Services).

3.Distribution by Sex:

Calculate the percentage or number of male and female workers.

Using data aggregaton and manipulaton techniques, you can derive meaningful insights from the dataset. If you have a specifc queston about this example or if you'd like me to perform calculatons

Create visualizations.

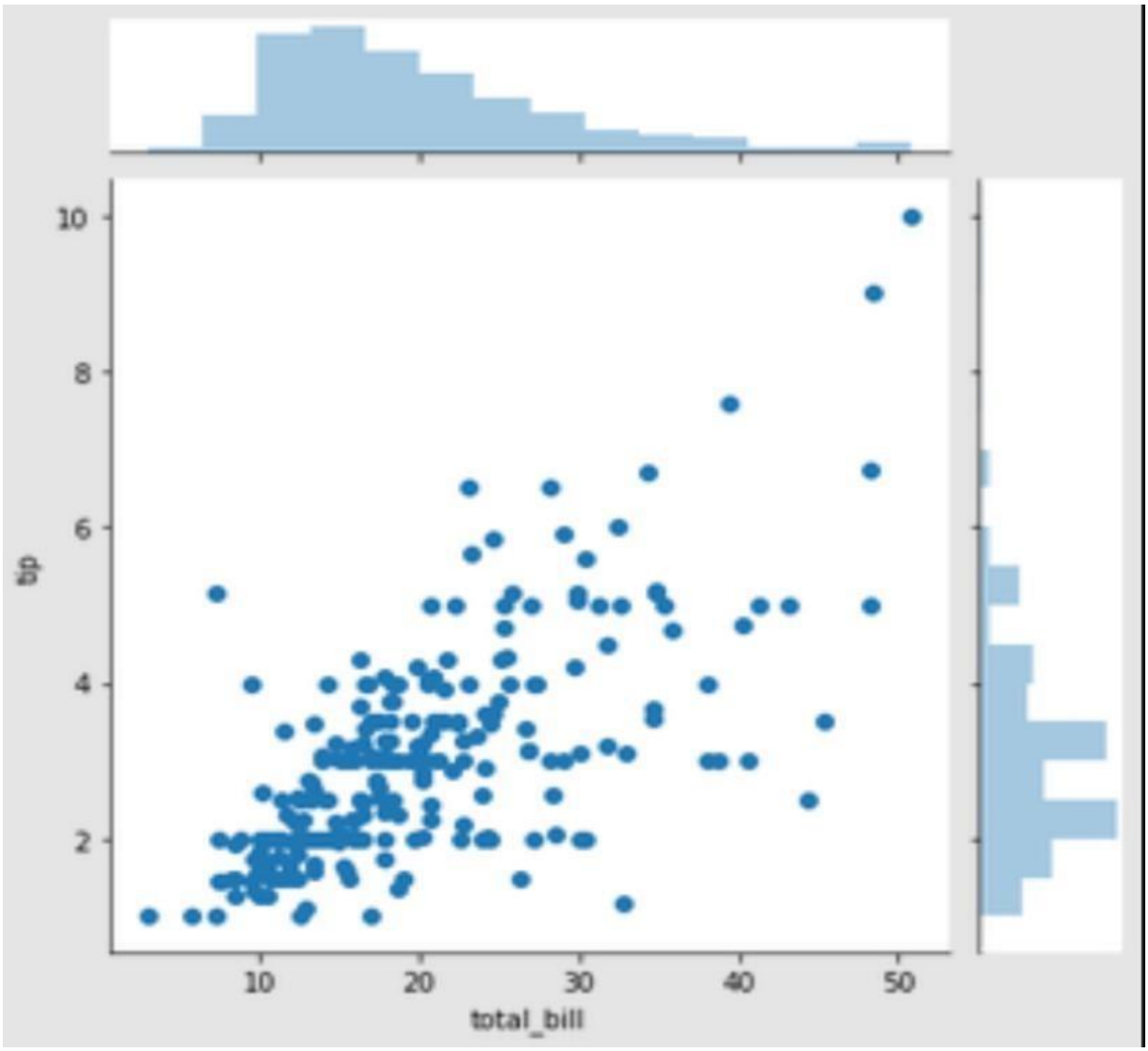
```
import matplotlib.pyplot as plt

# Example data age_groups = ['18-25', '26-35', '36-45', '46-
55'] age_distributon = [3000, 2200, 2000, 1600]  # Example
values

# Creatng a bar plot for age distributon
plt.fgure(fgsize=(8, 6)) plt.bar(age_groups,
age_distributon, color='skyblue') plt.xlabel('Age Groups')
plt.ylabel('Number of Workers') plt.ttle('Distributon of
Marginal Workers by Age Group') plt.show()
```

Import Libraries :

```
# importng and creatng alias for seaborn  import
seaborn as sns

# loading tps dataset  tps =
sns.load_dataset("tps")  #
plotng scaterplot with
histograms for features total
bill and tp.
sns.jointplot(data=tps,
x="total_bill", y="tp")
```

Output:

```
<seaborn.axisgrid.JointGrid at 0x26203152688>
```



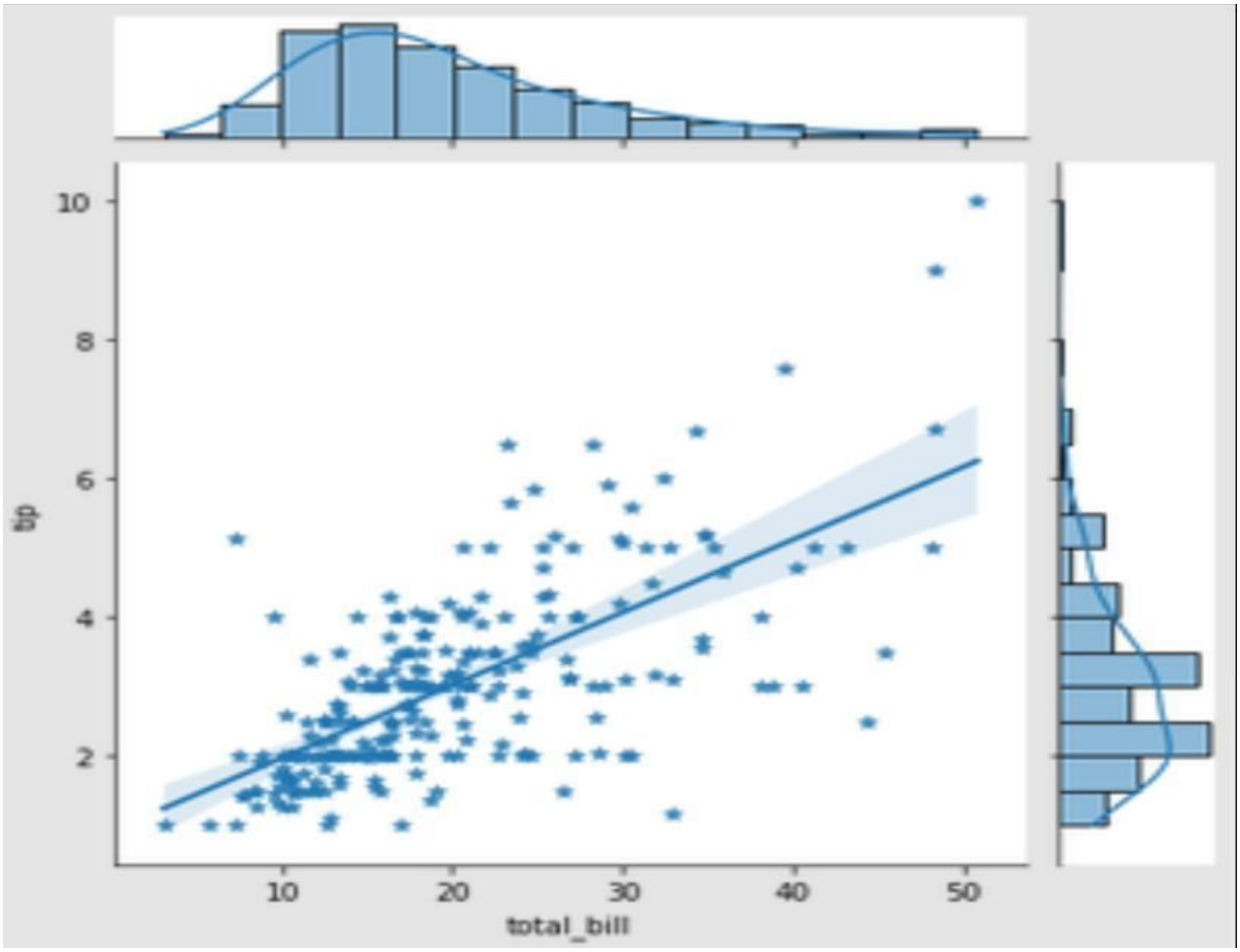
Using kind=”reg” atribute you can add a linear regression ft and univariate KDE curves.

```
import seaborn as sns

tps = sns.load_dataset("tps")
```

here "*" is used as a marker for scaterplot sns.jointplot(data=tps,

x="total_bill", y="tp", kind="reg", marker="*")

Output:

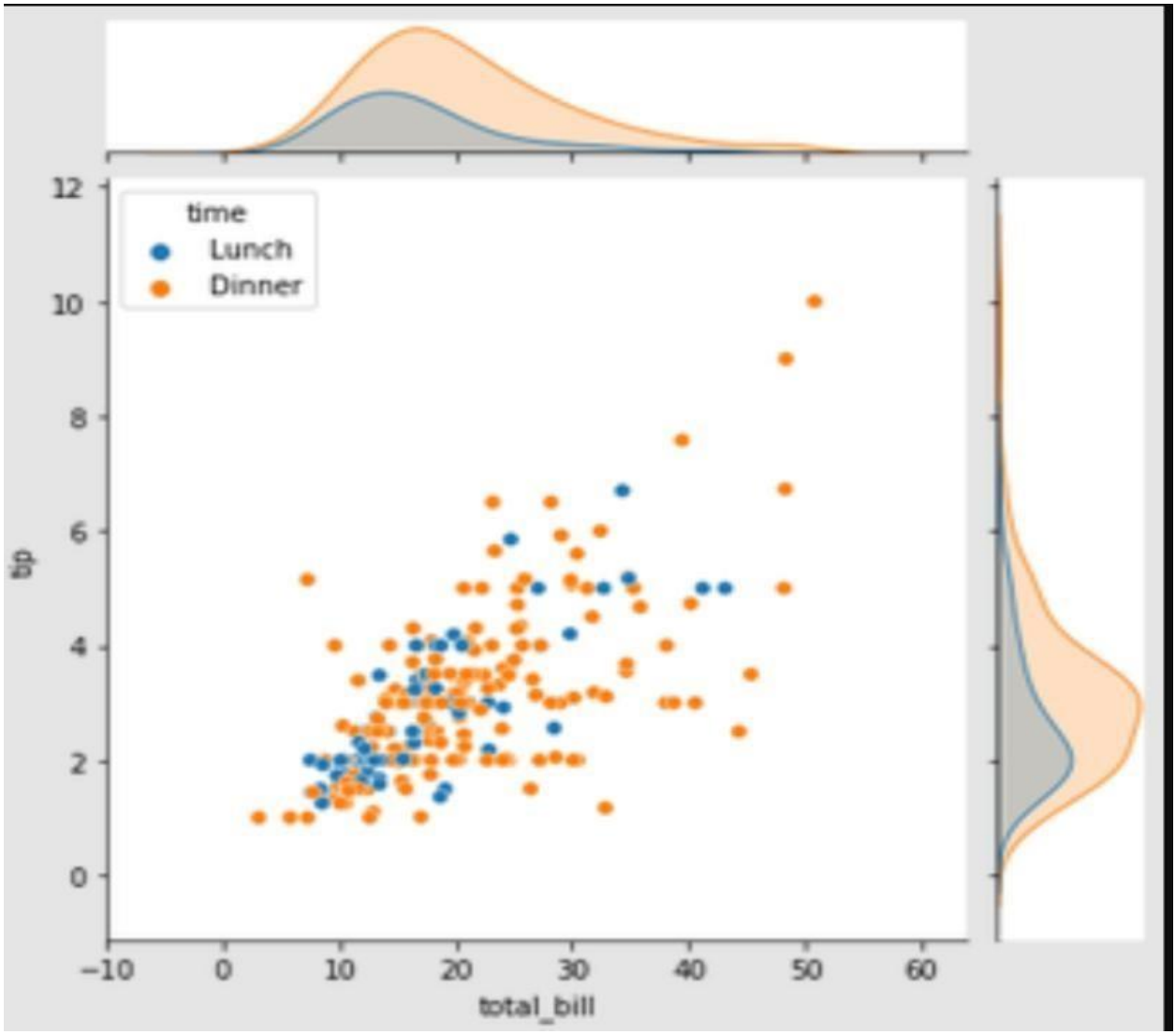


To add conditonal colors to the scaterplot you can use hue atribute but it draws separate density curves (using kdeplot()) on the marginal axes

Import seaborn as sns

Tips = sns.load_dataset(“tps”)

Sns.jointplot(data=tps, x=”total_bill”, y=”tp”, hue=”tme”)



visualizations using data visualization libraries (e.g., Matplotlib, Seaborn).

To understand how variables in a dataset are related to one another and how that relatonship is dependent on other variables, we perform statstcal analysis. This Statstcal analysis helps to visualize the trends and identfy various paterns in the dataset. One of the functons which can be used to get the relatonship between two variables in Seaborn is relplot().

Relplot() combines FacetGrid with either of the two axes-level functons scaterplot() and lineplot(). Scaterplot is default kind of relplot(). Using this we can visualize joint distributon of two variables through a cloud of points. We can draw scaterplot in seaborn using various ways. The most common one is when both the variables are numeric.

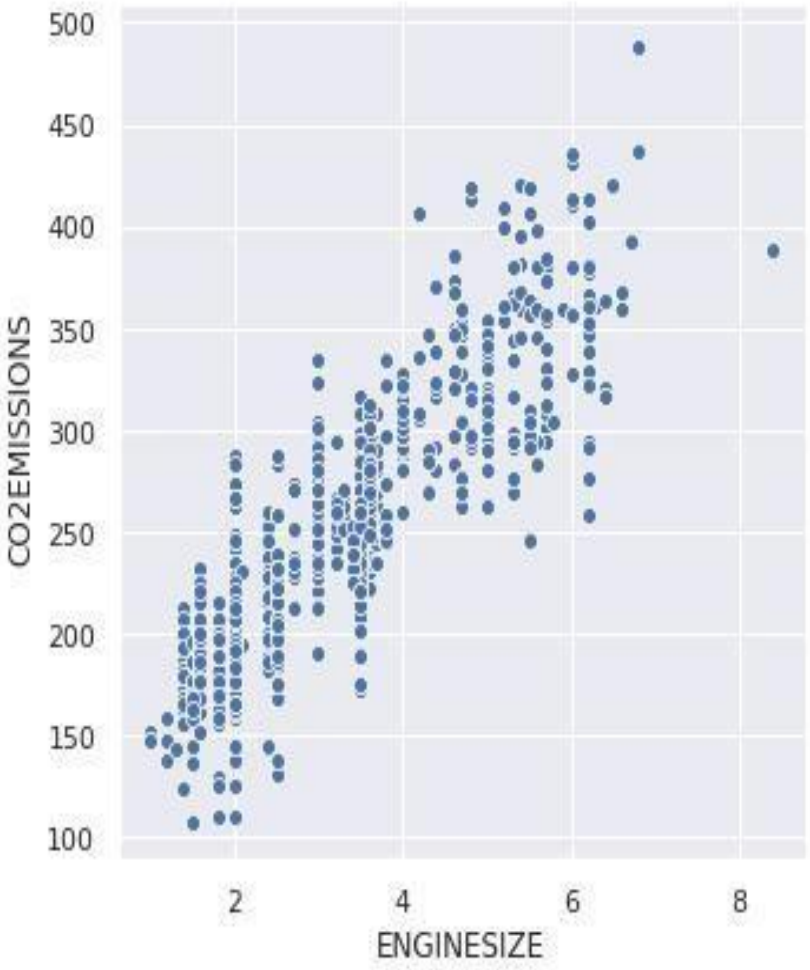
Let's take an example of a dataset that consists a data of CO2 emissions of diferent vehicles. To get the dataset click here.

```
# import libraries  import

pandas as pd  import numpy

as np  import

matplotlib.pyplot as plt

import seaborn as sns


# set grid style  sns.set(style

="darkgrid")


# import dataset  dataset =

pd.read_csv('FuelConsumpton.csv')
```

Let's plot the basic scaterplot for visualizing the relaton between the target variable "CO2EMISSIOnS" and "ENGINE SIZE"

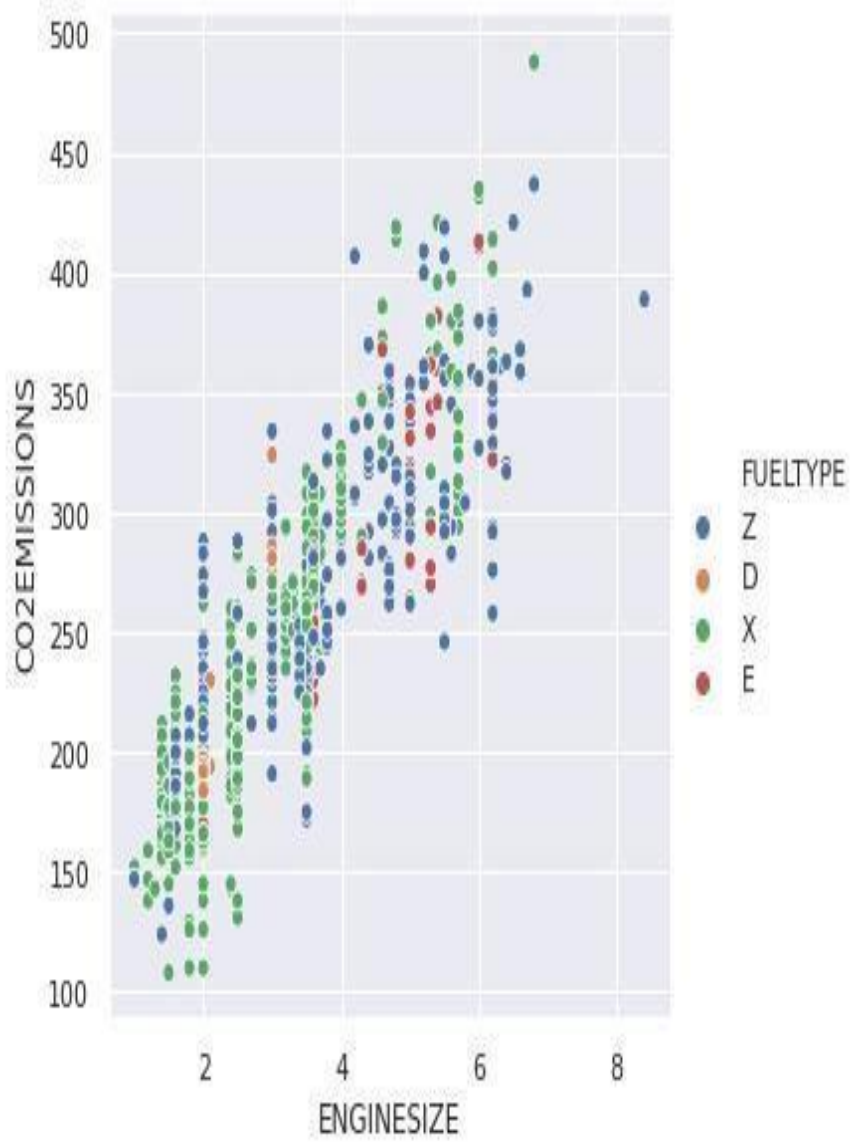
```
sns.relplot(x ="ENGINESIZE", y ="CO2EMISSIONS",

                                data = dataset);
```

Output:



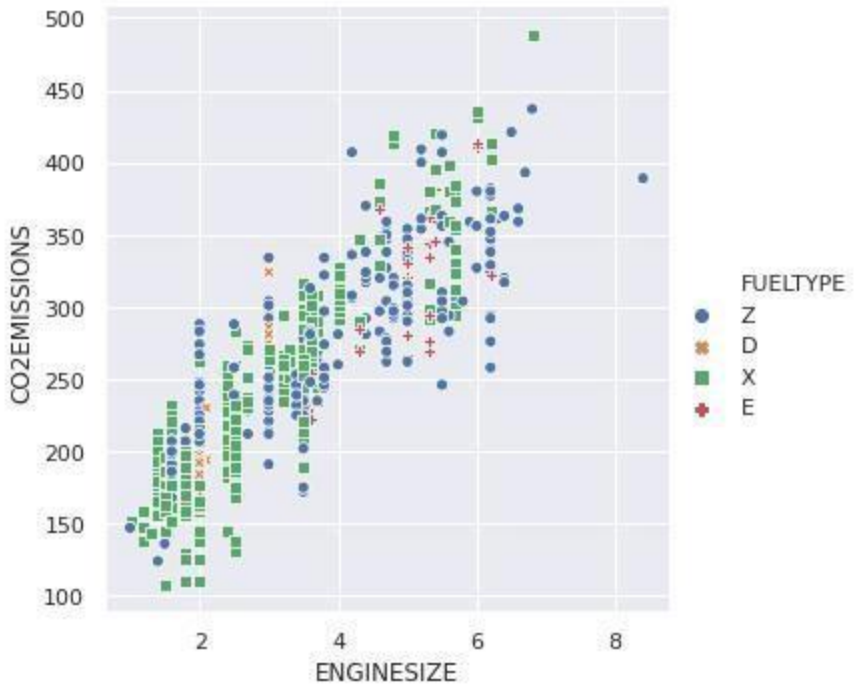
We can add visualize one more variable by adding another dimension to the plot. This can be done by using “hue”, which colors the points of the third variable, thus adding a meaning to it.

```
sns.relplot(x ="ENGINESIZE", y ="CO2EMISSIONS",
                                    hue ="FUELTYPE", data = dataset);
```



To highlight the diferent classes, we can add marker styles

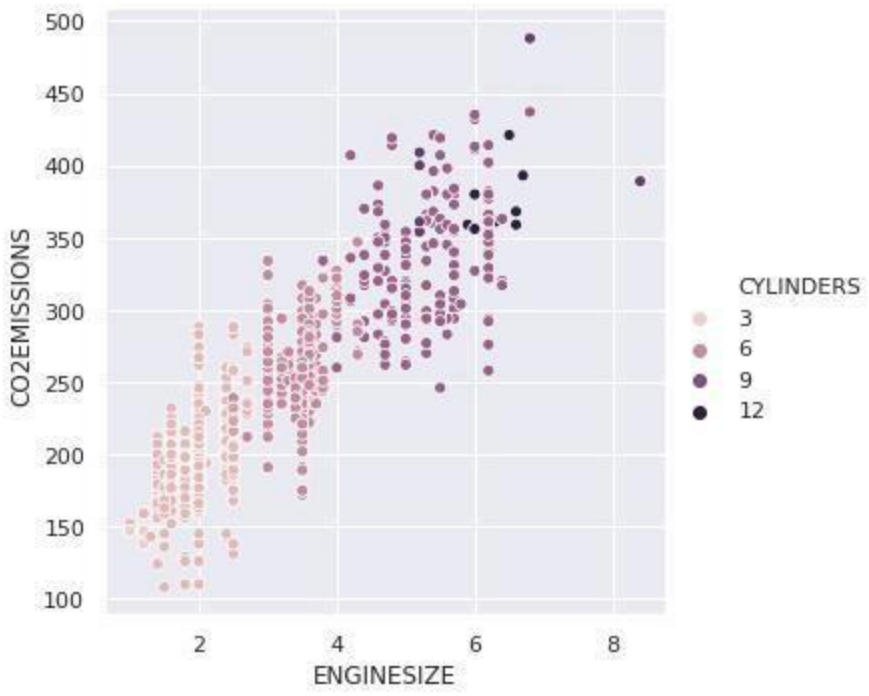
```
sns.relplot(x ="ENGINESIZE", y ="CO2EMISSIONS",
    hue ="FUELTYPE", style ="FUELTYPE", data = dataset);
```



In the previous example, hue semantc was for a categorical variable, so it had a default qualitatve palete. But if we use a numerical variable instead of categorical, then the default palete used is sequental, which can be modifed too.

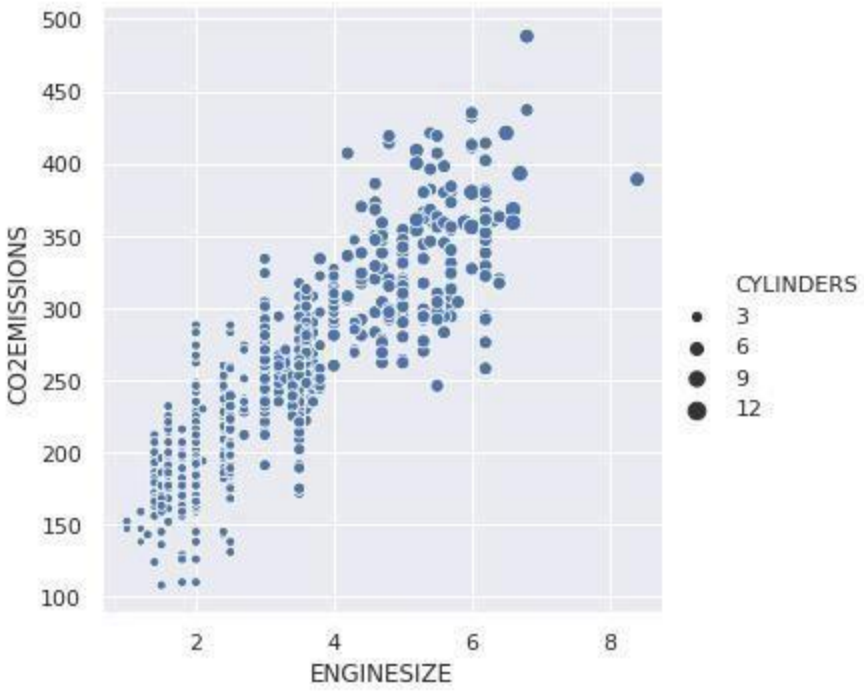
Sns.relplot(x =”ENGINESIZE”, y =”CO2EMISSIONS”,

Hue =”CYLINDERS”, data = dataset);



We can also change the size of points for the third variable.

```
sns.relplot(x ="ENGINESIZE", y ="CO2EMISSIONS",
                                size ="CYLINDERS", data = dataset);
```


500
450
400
350
300
250
200
150
100
CO2EMISSIONS
2
4
6
8
ENGINESIZE
CYLINDERS
3
6
9
12

Advantages:

- ***Identifying Demographic Patterns:*** *The first step involves understanding the demographic characteristics of the marginal workers, including their age, gender, education level, and regional distribution within Tamil Nadu.*

- ***Assessment of Economic Conditions:*** *Analyzing the economic status of these workers by studying their income levels, employment patterns, and job stability can help in understanding their financial vulnerabilities and challenges.*

- ***Evaluation of Social Welfare Measures:*** *This step involves assessing the existing social welfare programs and policies in place for marginal workers in Tamil Nadu to*

identify their effectiveness, gaps, and potential areas for improvement.

- ***Mapping Health and Education Indicators:*** *Analyzing the health and education indicators among these workers can shed light on the accessibility and quality of healthcare and educational facilities available to them, allowing for the identification of potential disparities.*

- ***Exploring Livelihood Opportunities:*** *Understanding the available livelihood opportunities and the constraints faced by the marginal workers can help in formulating targeted policies and interventions to uplift their socio-economic conditions.*

- ***Policy Formulation and Implementation:*** *Utilizing the insights gained from the analysis to formulate comprehensive policy interventions and implementation strategies tailored to the specific needs and challenges of the marginal workers can effectively address their socio-economic concerns.*

Disadvantage:

- ***Data Limitations:*** *Obtaining comprehensive and accurate data on marginal workers might be challenging, leading to incomplete or biased analyses.*

- ***Sample Bias:*** *The sample selection process could inadvertently exclude certain groups, leading to skewed or inaccurate results.*

- ***Contextual Factors:*** *Failing to consider local customs, traditions, and cultural nuances might result in an oversimplified or inaccurate understanding of the socio-economic dynamics.*

- ***Interpretation Challenges:*** *Interpreting the data without a thorough understanding of the local context could lead to misrepresentation or misinterpretation of the findings.*

- ***Ethical Concerns:*** *Ensuring the privacy and well-being of the participants while conducting the study is crucial, and any oversight might lead to ethical controversies.*

- ***Policy Implementation:*** *Overlooking the potential challenges in implementing policies based on the analysis may lead to ineffective or unsustainable interventions, impacting the target population negatively.*

BENEFITS:

- ***Understanding Demographics:*** *Identify the demographic characteristics of marginal workers, including age, gender, education level, and location, to gain insights into their specific needs and challenges.*

- ***Assessing Employment Patterns:*** *Analyze the employment patterns of marginal workers, such as the sectors they are employed in, the nature of their jobs, and their income levels, to understand their economic contribution to the state.*

- ***Evaluating Livelihood Challenges:*** *Identify the challenges faced by marginal workers, such as lack of access to education, healthcare, and social security, which can help in designing targeted interventions to improve their quality of life.*

- ***Assessing Economic Impact:*** *Evaluate the economic impact of marginal workers on the state's economy, including their contribution to various sectors and their role in the overall development of the region.*

- ***Policy Formulation:*** *Use the analysis to formulate policies and programs that cater to the specific needs of marginal workers, ensuring their inclusion in various welfare schemes and initiatives for their social and economic upliftment.*

- ***Creating Awareness:*** *Raise awareness about the issues faced by marginal workers among policymakers, stakeholders, and the general public to garner support for implementing effective measures to improve their socio-economic conditions.*

- ***Monitoring and Evaluation:*** *Continuously monitor the effectiveness of interventions and policies implemented for the betterment of marginal workers, and regularly assess the progress made to ensure sustained development and improvement in their living standards.*

REFERENCE:

- ***Define Research Objectives:*** *Clearly outline the purpose and goals of the analysis.*

- ***Literature Review:*** *Study existing research on socio-economic conditions, marginal workers, and related topics in Tamil Nadu.*

- ***Data Collection:*** *Gather data through surveys, interviews, and secondary sources, focusing on income, education, employment, living conditions, and social factors.*

- ***Data Analysis:*** *Utilize statistical methods to analyze the collected data, including descriptive statistics, regression analysis, and other relevant techniques.*

- ***Identify Marginal Workers:*** *Define the criteria for identifying marginal workers and categorize them based on occupation, income levels, and other relevant factors.*

- ***Socio-Economic Assessment:*** *Evaluate the socio-economic status of marginal workers, considering factors such as income disparity, access to education, healthcare, and other social services.*

- ***Challenges and Opportunities:*** *Identify the challenges faced by marginal workers and analyze potential opportunities for their socio-economic empowerment.*

- ***Policy Recommendations:*** *Propose actionable policy recommendations to improve the socio-economic conditions of marginal workers in Tamil Nadu.*

CONCLUSION:

Summarize Findings: *Recap the key findings and trends observed in the data, highlighting the social and economic challenges faced by marginal workers in Tamil Nadu.*

Assess Implications: *Discuss the implications of the findings on policy formulation, suggesting potential areas for intervention and improvement in the livelihoods of marginal workers.*

Recommendations: *Propose actionable recommendations that can address the identified challenges, considering factors like access to education, healthcare, and skill development, as well as employment opportunities and social security measures.*

Future Research Directions: *Suggest potential areas for future research to deepen the*

understanding of the socio-economic conditions of marginal workers in Tamil Nadu, including specific aspects that require further investigation.

Policy Implications: *Outline the potential policy changes and interventions that could be implemented by the government and relevant organizations to uplift the living standards of marginal workers and ensure their socio-economic well-being.*

Concluding Remarks: *Provide a concise and impactful closing statement that emphasizes the importance of addressing the issues faced by marginal workers in Tamil Nadu and the significance of implementing the recommended measures for their socio-economic empowerment.*