

The socio-economic analysis of Marginal workers in Tamilnadu

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

★ Housing

★ Education

★ Crime rates

★ Employment

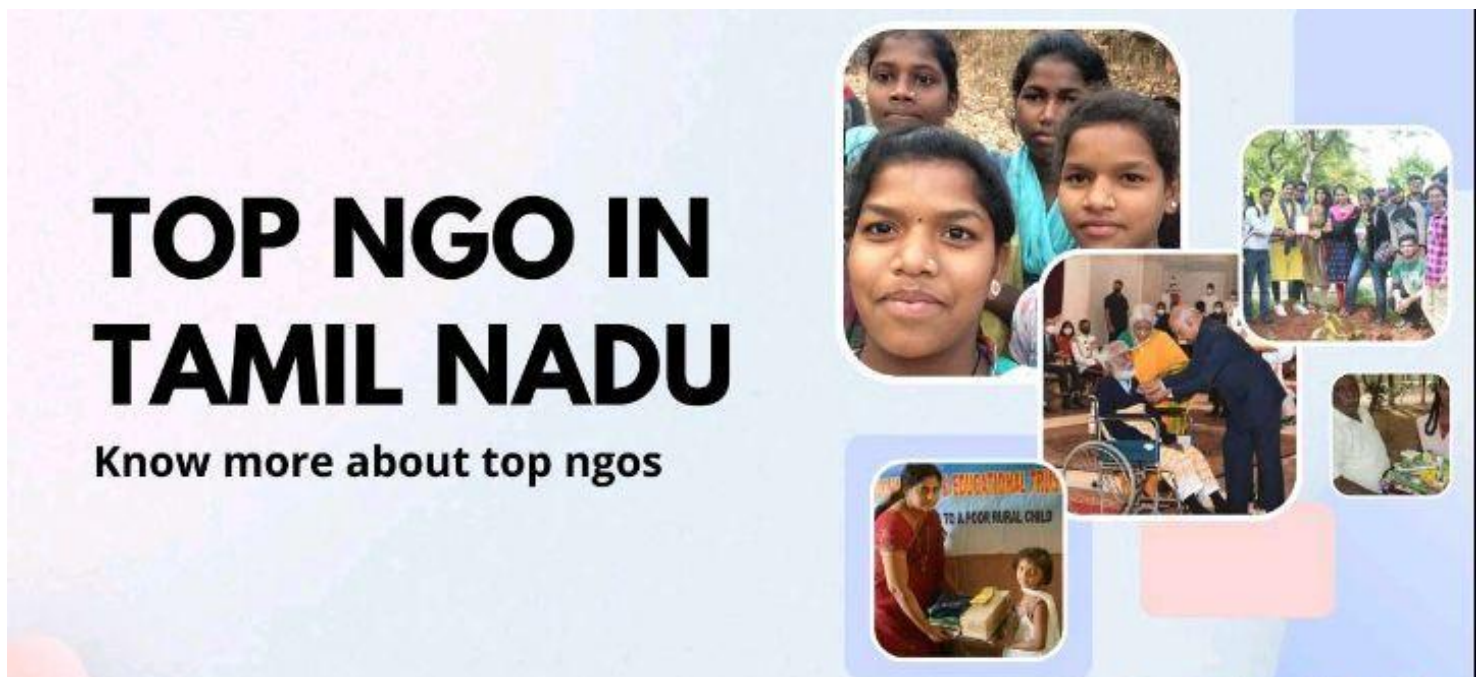
★ Demographics

★ Health

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:

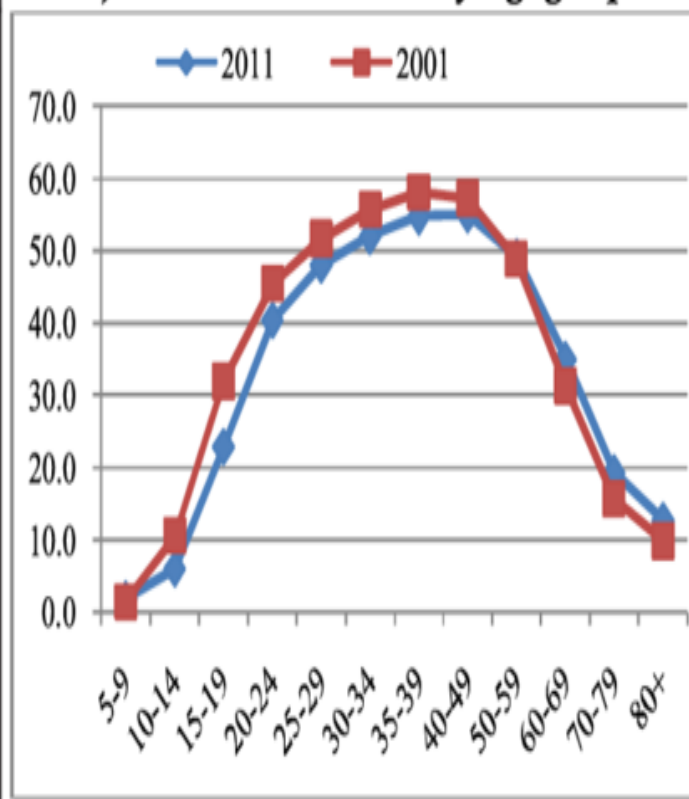
The data sets in the print statement

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

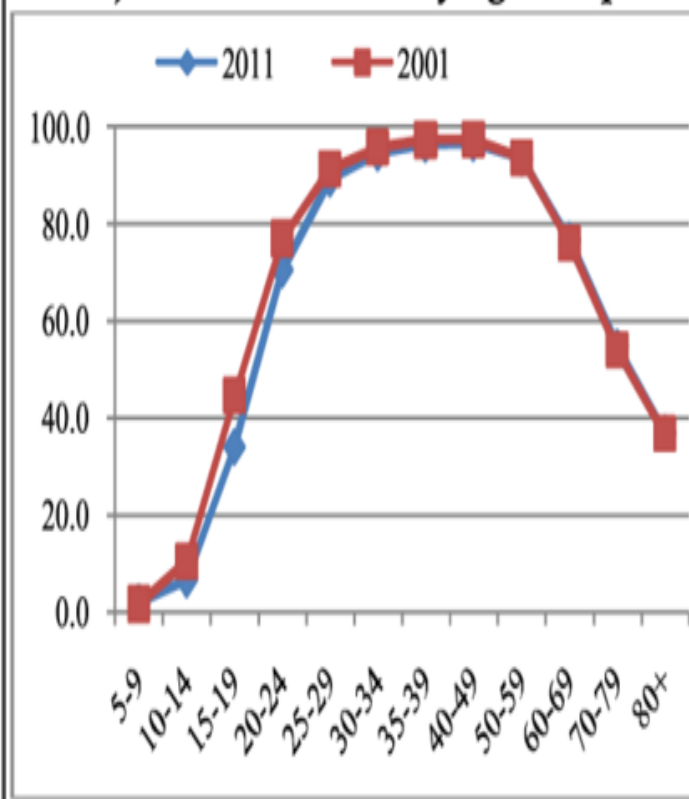
Dataset: <http://https://drive.google.com/file/d/1o2br9Kr21CHInHoQBUDq15fFIXrliKHo/view?usp=drivesdk>

Excel format: <http://https://in.docworkspace.com/d/sIEu5hMXQAdmCv6kG>

a) Rural Females' WPR by Age group



b) Rural Males' WPR by Age Group



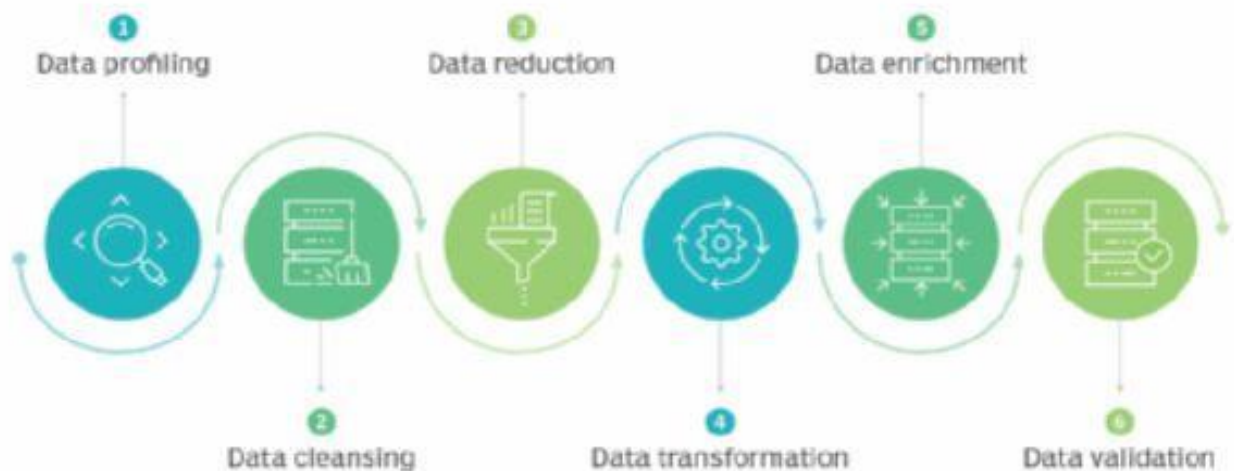
23	B0806SC	'33	'602	District - ThiTotal
24	B0806SC	'33	'602	District - ThiTotal
25	B0806SC	'33	'602	District - ThiTotal
26	B0806SC	'33	'602	District - ThiRural
27	B0806SC	'33	'602	District - ThiRural
28	B0806SC	'33	'602	District - ThiRural
29	B0806SC	'33	'602	District - ThiRural
30	B0806SC	'33	'602	District - ThiRural
31	B0806SC	'33	'602	District - ThiRural
32	B0806SC	'33	'602	District - ThiUrban
33	B0806SC	'33	'602	District - ThiUrban
34	B0806SC	'33	'602	District - ThiUrban
35	B0806SC	'33	'602	District - ThiUrban
36	B0806SC	'33	'602	District - ThiUrban
37	B0806SC	'33	'602	District - ThiUrban
38	B0806SC	'33	'603	District - CheTotal
39	B0806SC	'33	'603	District - CheTotal
40	B0806SC	'33	'603	District - CheTotal
41	B0806SC	'33	'603	District - CheTotal
42	B0806SC	'33	'603	District - CheTotal
43	B0806SC	'33	'603	District - CheTotal
44	B0806SC	'33	'603	District - CheRural
45	B0806SC	'33	'603	District - CheRural
46	B0806SC	'33	'603	District - CheRural
47	B0806SC	'33	'603	District - CheRural
48	B0806SC	'33	'603	District - CheRural
49	B0806SC	'33	'603	District - CheRural
50	B0806SC	'33	'603	District - CheUrban
51	B0806SC	'33	'603	District - CheUrban
52	B0806SC	'33	'603	District - CheUrban
53	B0806SC	'33	'603	District - CheUrban
54	B0806SC	'33	'603	District - CheUrban
55	B0806SC	'33	'603	District - CheUrban
56	B0806SC	'33	'604	District - KarTotal
57	B0806SC	'33	'604	District - KarTotal
58	B0806SC	'33	'604	District - KarTotal
59	B0806SC	'33	'604	District - KarTotal
60	B0806SC	'33	'604	District - KarTotal
61	B0806SC	'33	'604	District - KarTotal
62	B0806SC	'33	'604	District - KarRural
63	B0806SC	'33	'604	District - KarRural
64	B0806SC	'33	'604	District - KarRural
65	B0806SC	'33	'604	District - KarRural
66	B0806SC	'33	'604	District - KarRural
67	B0806SC	'33	'604	District - KarRural
68	B0806SC	'33	'604	District - KarUrban
69	B0806SC	'33	'604	District - KarUrban
70	B0806SC	'33	'604	District - KarUrban

35-39	32388	18771	13797	77
60+	5716	3147	2569	14
Age not stat	75	44	31	
Total	55577	28082	27495	121
5-14	1424	743	681	1
15-34	23965	12377	11588	47
35-59	25421	12417	13004	61
60+	4718	2516	2202	11
Age not stat	49	29	20	
Total	18871	11213	7658	37
5-14	1097	541	556	
15-34	9603	5672	3931	18
35-59	7147	4354	2793	16
60+	998	631	367	2
Age not stat	26	15	11	
Total	33748	19313	14435	63
5-14	2749	1483	1266	1
15-34	17431	9836	7595	31
35-59	11700	6993	4707	26
60+	1799	966	833	4
Age not stat	69	35	34	
Total	0	0	0	
5-14	0	0	0	
15-34	0	0	0	
35-59	0	0	0	
60+	0	0	0	
Age not stat	0	0	0	
Total	33748	19313	14435	63
5-14	2749	1483	1266	1
15-34	17431	9836	7595	31
35-59	11700	6993	4707	26
60+	1799	966	833	4
Age not stat	69	35	34	
Total	92015	47269	44746	176
5-14	2540	1308	1232	1
15-34	40080	20955	19125	75
35-59	41279	20372	20907	81
60+	7960	4538	3422	18
Age not stat	156	96	60	
Total	70150	34555	35595	137
5-14	1273	638	635	
15-34	29666	14837	14829	57
35-59	32392	15231	17161	63
60+	6687	3765	2922	15
Age not stat	132	84	48	
Total	21865	12714	9151	39
5-14	1267	670	597	
15-34	10414	6118	4296	18

Data preprocessing:

Clean the data, handle missing values, and outliers

Steps for data preprocessing



Data analysis:

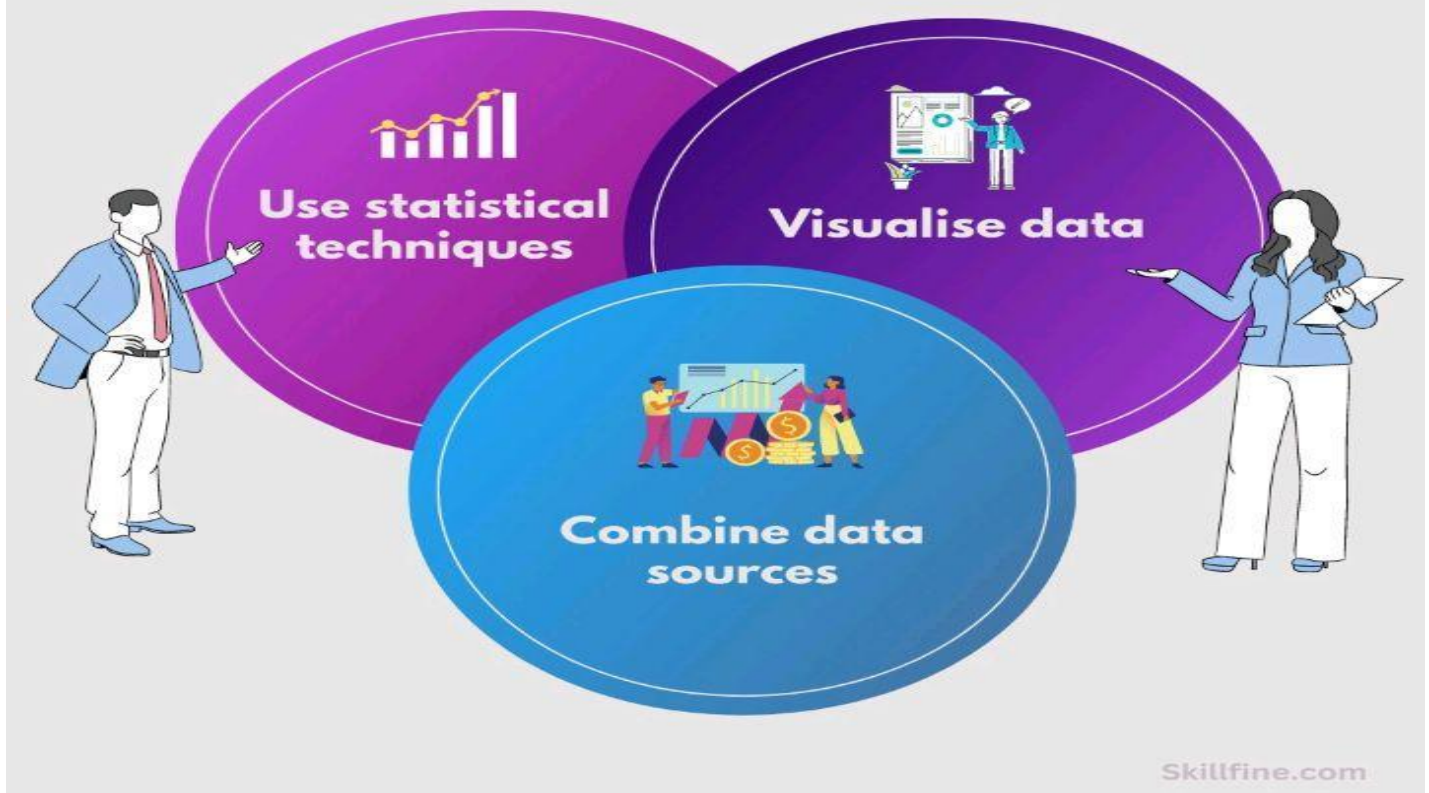
EDA, summary statistics, and data visualization



Statistical analysis:

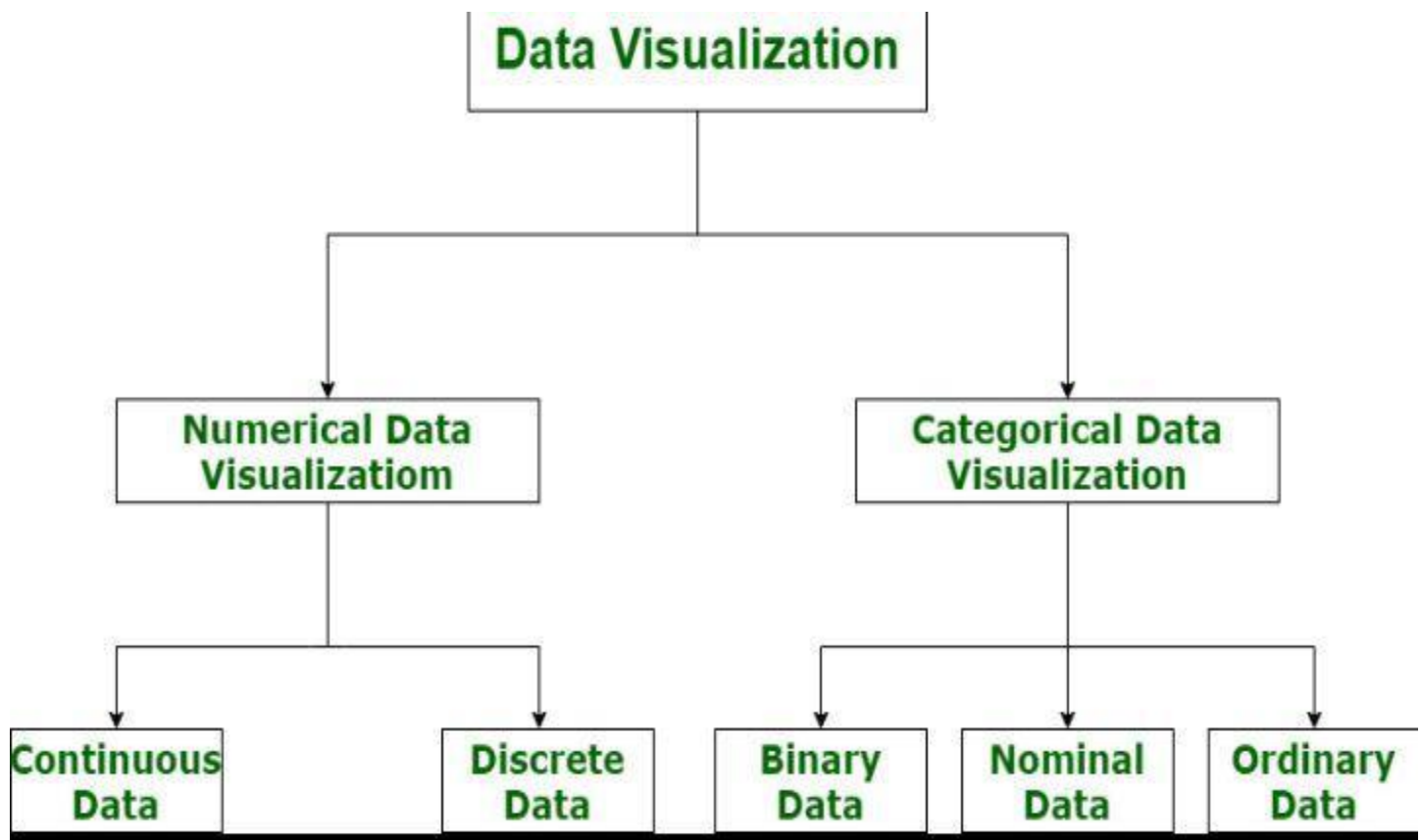
Perform hypothesis tests, calculate correlations, etc.

STATISTICAL ANALYSIS



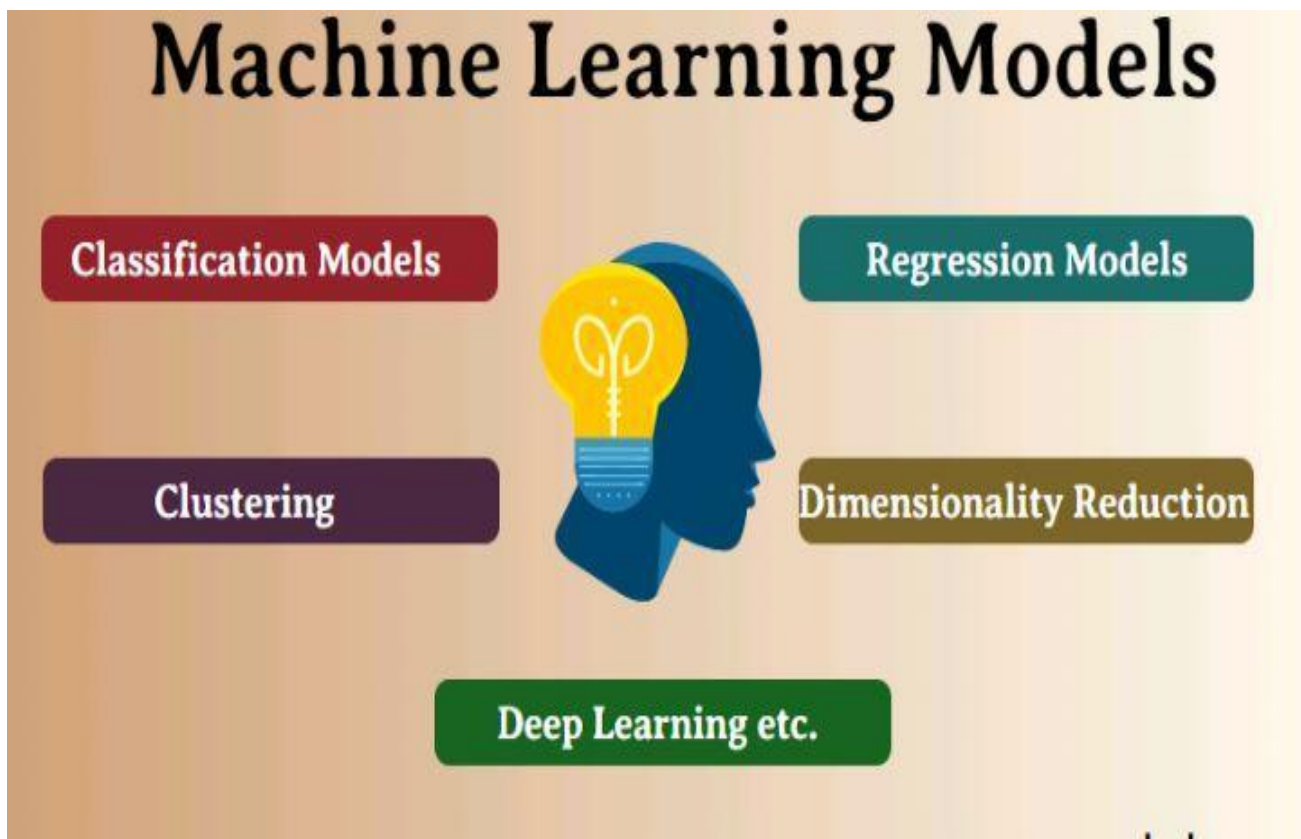
Data visualization:

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_TAMIL_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()
```