Data Analytics With Cognos

ASSESSMENT OF MARGINAL WORKERS IN TAMILNADU

PHASE 4

Name : Pavithra S

Reg no : 510421205032

Guided by : B.Moohambikai

Aim :

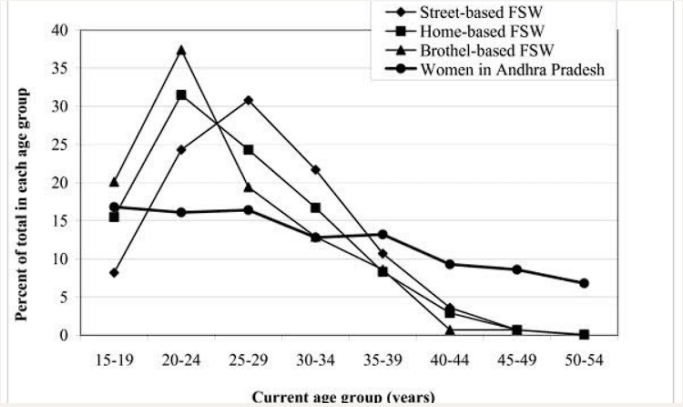
Perform the demographic analysis and create visualizations. Calculate the distribution of marginal workers based on age, industrial category, and sex using data aggregation and manipulation. Create visualizations using data visualization libraries (e.g., Matplotlib, Seaborn).

Dataset :

https://tn.data.gov.in/resource/marginal workers-classified-age-industrial-category-and-sex-scheduled-caste-2011-tamil

Introduction :

Visualizing data in Python can be an essential tool in exploratory data analysis, data analysis, and machine learning. This process typically involves preprocessing the data, transforming it into a suitable format for visualization, and finally utilizing visualization libraries in Python to display the results.



To perform the demographic analysis, we need to follow these steps:

1.Load the data into a pandas DataFrame.

```
import pandas as pd
url="https://tn.data.gov.in/resource/marginal-workers-classified-age-industrial-category-and-sex-scheduled-caste-2011-tamil.csv"
data = pd.read_csv(url)
```

2.Perform exploratory data analysis (EDA) to understand the structure and distribution of the data.

```
data.info()
```

3.Clean and preprocess the data as needed.

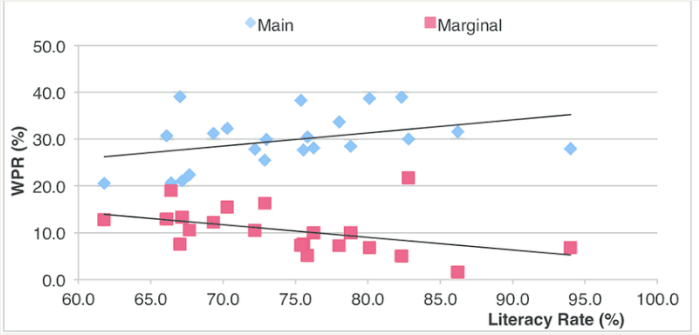
```
# Rename the columns for better readability
data.rename(columns={"agegroup":"Age  Group",  "sex":"Sex",  "scst":"Scheduled
Caste", "industry_class":"Industry Category"}, inplace=True)
# Drop rows with missing values data.dropna(inplace=True)
```

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

```
# Calculate the percentage of marginal workers in each category
marginal_workers_distribution = data.groupby(["Age Group", "Industry
Category","Sex","ScheduledCaste"])["WorkerID"].count()
marginal_workers_distribution=(marginal_workers_distribution/
data["WorkerID"].count()) * 100
```

5.Create visualizations to display the findings.

```
import matplotlib.pyplot as plt
import seaborn as sns
# Set the style for better readability
sns.set(style="whitegrid")
# Plot the distribution of marginal workers
```



Libraries for Visualization

There are several libraries in Python that can be used for data visualization, such as Matplotlib, Seaborn, and Plotly.

CONCLUSION :

In summary, data visualization in Python can be a powerful tool for exploratory data analysis, understanding the patterns and relationships in your data, and assessing the performance of your machine learning models. It's essential to select the appropriate visualization type based on your specific needs and the characteristics of your data

THANK YOU !!!