DATA ANALYTICS WITH COGNOS

**PHASE 4:** DEVELOPMENT PART 2

***PROJECT TITLE: MARGINAL WORKERS ASSESSMENT IN TN SOCIO ECONOMIC ANALYSIS***

**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).**

**DATA VISUALIZATION**

Data visualization is important for many analytical tasks including data summarization, exploratory data analysis and model output analysis. One of the easiest ways to communicate your findings with other people is through a good visualization. Fortunately, Python features many libraries that provide useful tools for gaining insights from data.

**DATA VISUALIZATION WITH PYTHON**

Data visualization is a key part of communicating your research to others. Whether via histograms, scatter plots, bar charts or pie charts, a good visualization helps unlock insights from your data. Fortunately, Python makes creating visualizations easy with Matplotlib and Seaborn.

The most well-known of these data visualization libraries in Python, Matplotlib, enables users to generate visualizations like histograms, scatter plots, bar charts, pie charts and much more.

Seaborn is another useful visualization library that is built on top of Matplotlib. It provides data visualizations that are typically more aesthetic and statistically sophisticated. Having a solid understanding of how to use both of these libraries is essential for any data scientist or data analyst as they both provide easy methods for visualizing data for insight.

**MATPLOTLIB VS. SEABORN**

* Matplotlib is a library in Python that enables users to generate visualizations like histograms, scatter plots, bar charts, pie charts and much more.
* Seaborn is a visualization library that is built on top of Matplotlib. It provides data visualizations that are typically more aesthetic and statistically sophisticated.

**Advantages of Data Visualization:**

* Easier representation of compels data.
* Highlights good and bad performing areas.
* Explores relationship between data points.
* Identifies data patterns even for larger data points.

**PROGRAM:**

**1. Import necessary libraries:**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

**2. Load the dataset:**

data = pd.read\_csv(r"C:\\Users\\ranji\\Desktop\\finalmarginal.csv")

**3. Calculate the distribution of marginal workers based on age, industrial category, and sex using data aggregation and manipulation. You can use pandas for this purpose:**

age\_distribution = data.groupby('Age group').size()

industrial\_category\_distribution = data.groupby('Industrial\_Category').size()

sex\_distribution = data.groupby('Sex').size()

**4. Create visualizations using Matplotlib and Seaborn:**

**For age distribution, you can create a bar chart:**

plt.figure(figsize=(10, 5))

age\_distribution.plot(kind='bar')

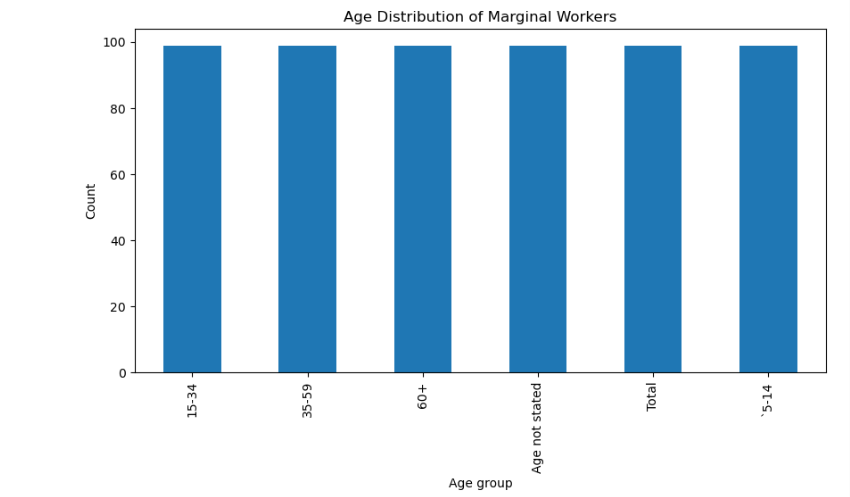
plt.xlabel('Age')

plt.ylabel('Count')

plt.title('Age Distribution of Marginal Workers')

plt.show()

**OUTPUT :**



**For industrial category and sex distributions, you can create pie charts:**

plt.figure(figsize=(10, 5))

industrial\_category\_distribution.plot(kind='pie', autopct='%1.1f%%')

plt.title('Industrial Category Distribution of Marginal Workers')

plt.show()

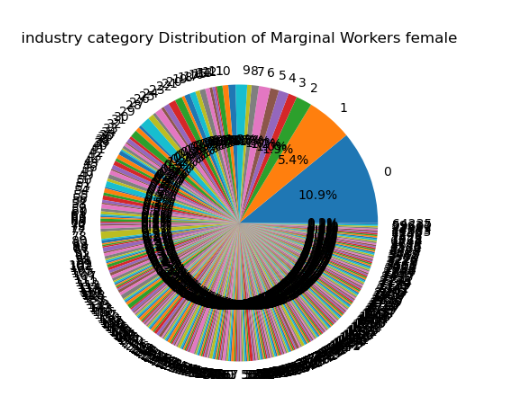
plt.figure(figsize=(10, 5))

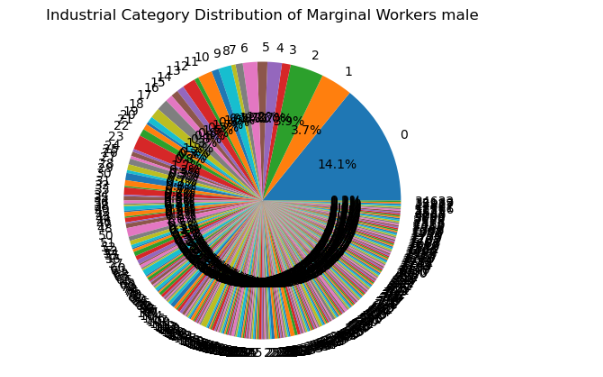
sex\_distribution.plot(kind='pie', autopct='%1.1f%%')

plt.title('Sex Distribution of Marginal Workers')

plt.show()

**OUTPUT:**





**TEAMMATES:**

* + 1. M KEERTHIGA
    2. S RANJITHA
    3. K VASUKI
    4. K KAYALVIZHI