**Marginal Workers in Tamil Nadu**

**Table of Contents:**

**1. Introduction**

**2. Data Collection**

**3. Variables and Indicators**

**4. Methodology**

**5. Descriptive Analysis**

**6. Employment and Work Patterns**

**7. Income and Living Standards**

**8. Assessment of marginal workers**

**9. Conclusion**

**10. References**

**1. Introduction:**

* The "Marginal Workers in Tamil Nadu" project aims to analyze and understand the characteristics, distribution, and socio-economic factors of marginal workers in the Tamil Nadu region of India. This documentation provides an overview of the project, its goals, and the analysis process.

**2. Data Used:**

* The project uses a dataset sourced from [Provide the data source or origin], which contains information on marginal workers in Tamil Nadu. The dataset includes information on age, gender, education, income, and industrial categories.

**3. Variables and Indicators:**

* To analyze the age distribution of marginal workers in Tamil Nadu.
* To examine the distribution of marginal workers based on their industrial category and gender.
* To identify any correlations or patterns within the dataset.
* To build a predictive model for classifying marginal workers.

**4. Methodology:**

* A Random Forest Classifier was selected as the machine learning model. The data was split into training and testing sets, the model was trained, and its performance was evaluated using accuracy and a classification report.

**5. Descriptive Analysis:**

* Data analysis included the calculation of descriptive statistics and the creation of cross-tabulations. The analysis explored the dataset's characteristics and provided an initial understanding of the marginal workers' attributes.

**6. Employment and Work Patterns:**

* Analyze the types of employment that marginal workers are engaged in, their working conditions, and whether they experience job insecurity or seasonality in employment.

**7. Income and Living Standards:**

* Examine the income levels of marginal workers, their access to basic amenities, and the quality of their living conditions. This section can also discuss any regional disparities.

**8. Assessment of marginal workers:**

* Evaluate the socio-economic status of marginal workers in Tamil Nadu. Consider factors such as their access to education, healthcare, social welfare programs, and whether they face discrimination or exploitation.

**9. Conclusion:**

* The "Marginal Workers in Tamil Nadu" project provided valuable insights into the characteristics and distribution of marginal workers in the Tamil Nadu region. The documentation summarizes the project's goals, the dataset used, data analysis, visualization, feature engineering, model training, and notable findings. This information can be used for further research and policymaking related to marginalized labor in Tamil Nadu.

**10.Reference:**

**Program:**

**#importing the libraries in python**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**#import seaborn for using piechart**

**import seaborn as sns**

**import numpy as np**

# Importing csv file data sets

**data = pd.read\_csv("C:/Users/balur/Downloads/marginalworkers.csv")**

**#checking the dataset given is null or not**

**df.isnull()**

**Fetching and describe the data**

**l1=tuple([df['Worked for 3 months or more but less than 6 months - Females'],df['Worked for 3 months or more but less than 6 months - Males']])**

**l2=tuple([df['Industrial Category - N to O - Females'],df['Industrial Category - P to Q - Persons']])**

**df.describe()**

**l1**

**l2**

**# Data Visualization**

**# assigning the csv data to variable of piechart**

**pie\_chart\_data = df['Worked for 3 months or more but less than 6 months - Persons']**

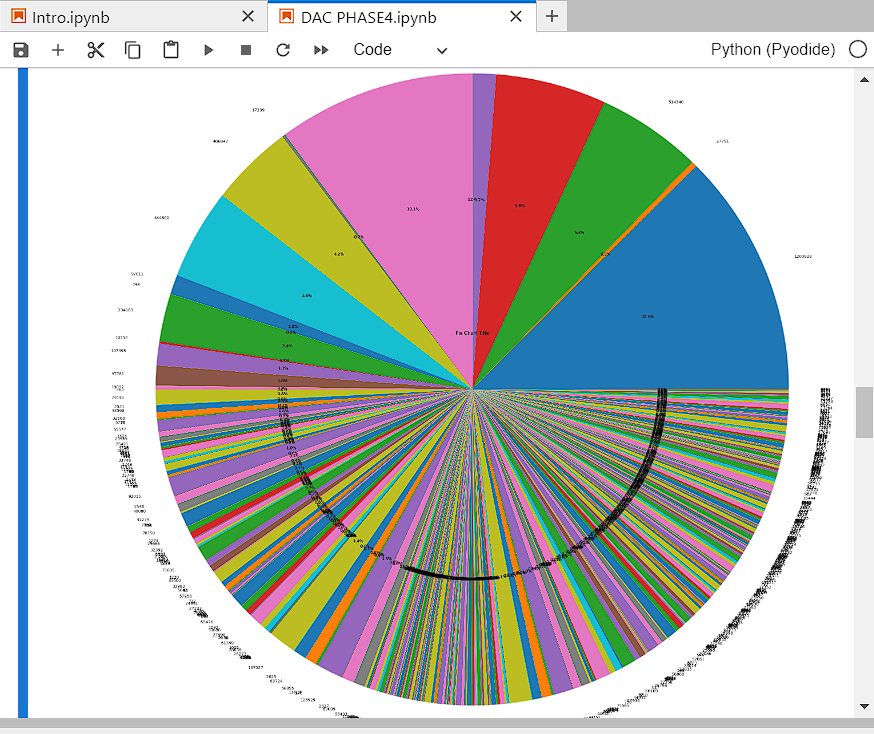
**#asigning values to the pie chart**

**plt.pie(pie\_chart\_data, labels=df['Worked for 3 months or more but less than 6 months - Persons'], autopct='%1.1f%%',radius=7.5)**

**plt.title('Pie Chart Title')**

**#printing the pie chart**

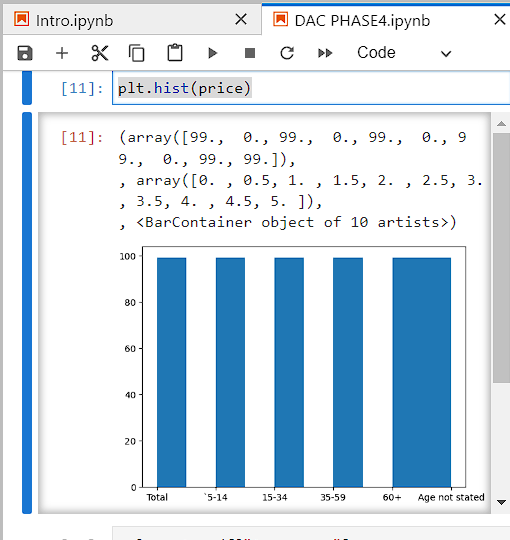
**plt.show()**

****

# printing the pie chart using the given csv file data sets

**price = df["Age group"]**

# visualizing the data sets column in the form of of histogram

**plt.hist(price)**

**column\_1 = df["Age group"]**

**column\_2 = df["Industrial Category - A - Cultivators - Persons"]**

**# Create the histogram**

**fig, axs = plt.subplots(1, 2)**

**axs[0].hist(column\_1)**

**axs[1].hist(column\_2)**

**# Add a title and axis labels for each subplot**

**axs[0].set\_title("Histogram of {} Column".format(column\_1.name))**

**axs[1].set\_title("Histogram of {} Column".format(column\_2.name))**

**axs[0].set\_xlabel(column\_1.name)**

**axs[1].set\_xlabel(column\_2.name)**

**axs[0].set\_ylabel("Count")**

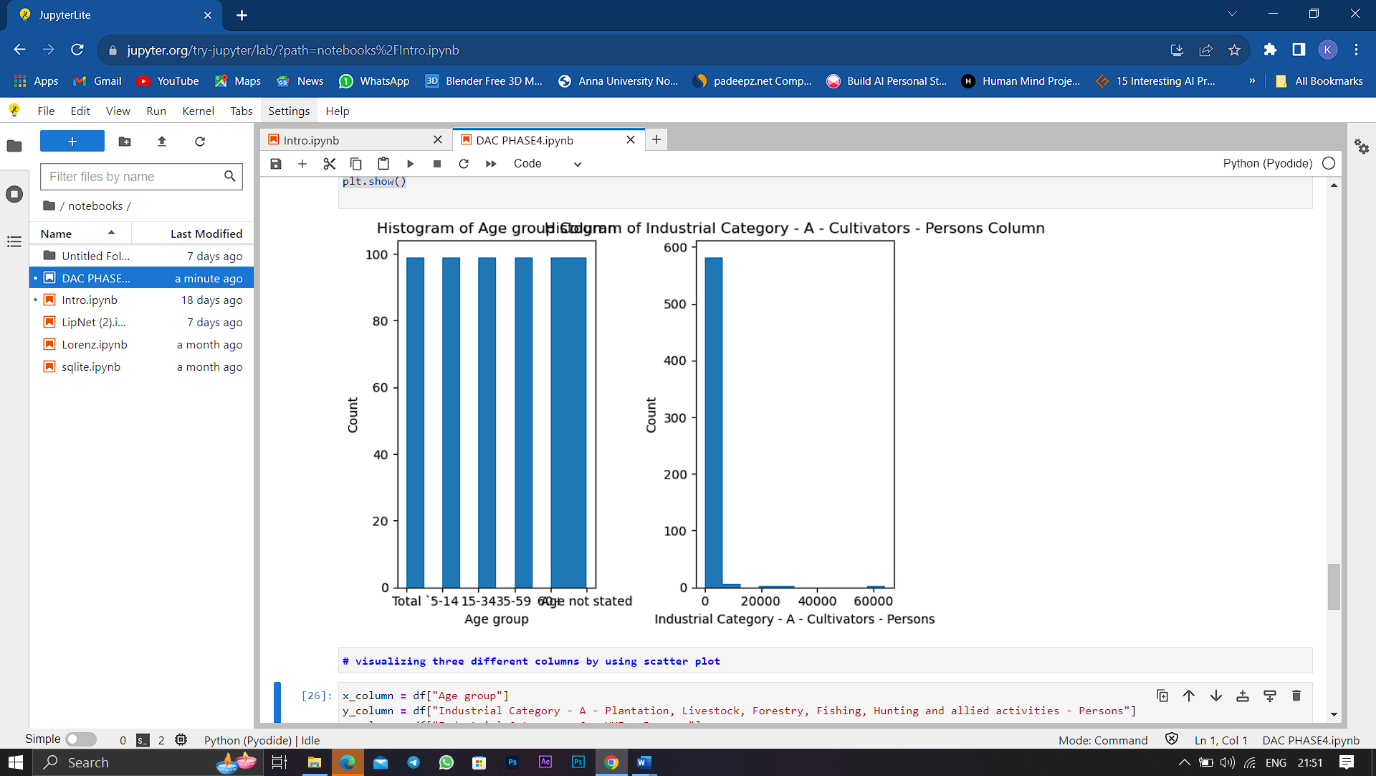
**axs[1].set\_ylabel("Count")**

**# Adjust the subplot layout**

**plt.tight\_layout()**

**# Show the plot**

**plt.show()**

****

**x\_column = df["Age group"]**

**y\_column = df["Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons"]**

**z\_column = df["Industrial Category - C - HHI - Persons"]**

**# Create the scatter plot**

**plt.scatter(x\_column, y\_column, z\_column)**

**# Add a title and axis labels**

**plt.title("Scatter Plot of {} vs. {}".format(x\_column.name, y\_column.name, z\_column.name))**

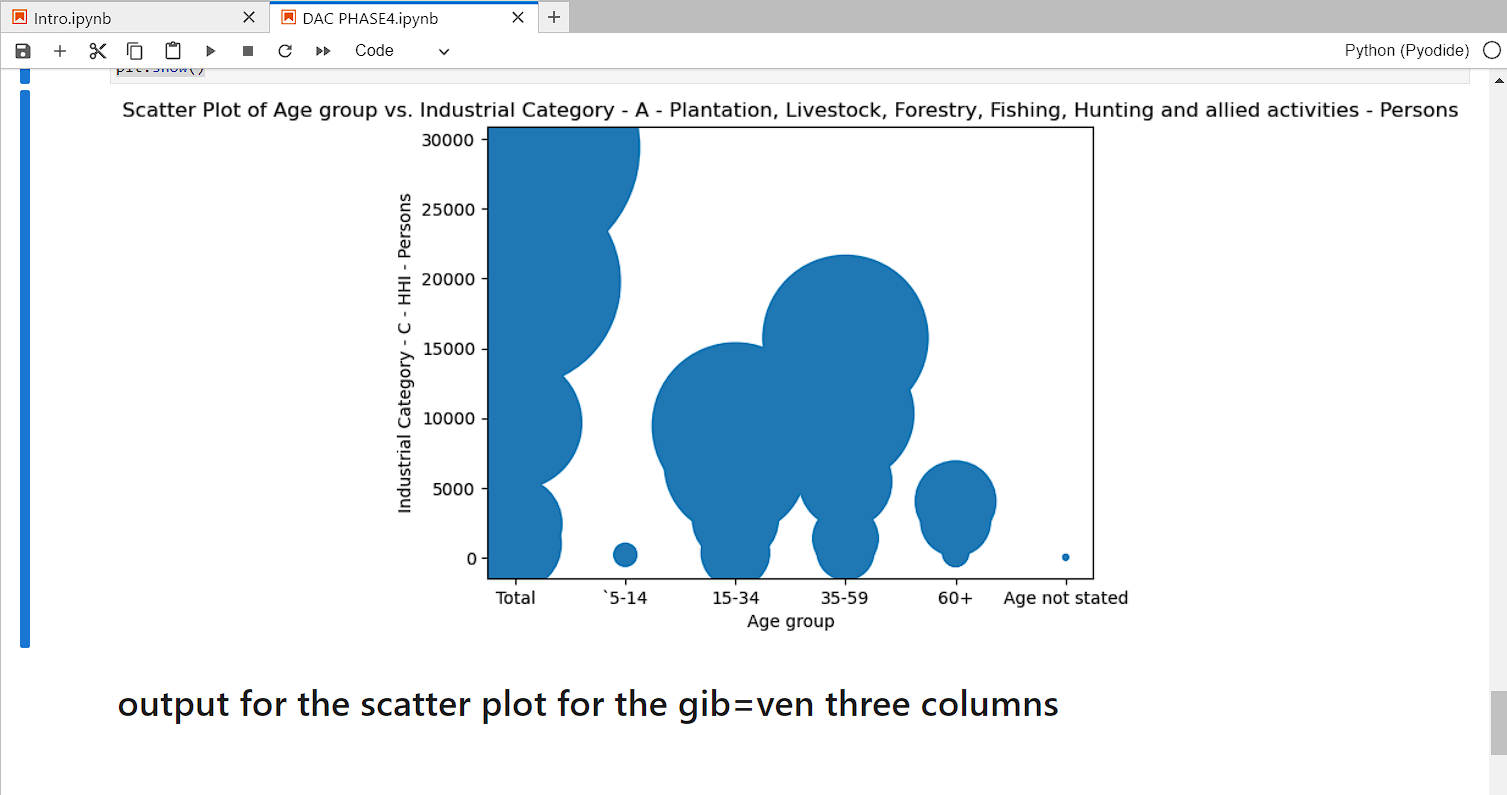
**plt.xlabel(x\_column.name)**

**plt.ylabel(y\_column.name)**

**plt.ylabel(z\_column.name)**

**# Show the plot**

**plt.show()**

**output:**