**ASSESMENTS OF MARGINAL WORKERS IN TAMILNADU**

**INTRODUCTION:**

**In the realm of data analytics, exploring the dynamics of marginal workers is imperative. This assessment delves into the intricate patterns of individuals on the periphery of the workforce, unraveling insights that transcend traditional employment metrics. Through meticulous data scrutiny, we aim to decipher the unique challenges faced by marginal workers, shedding light on disparities and potential avenues for empowerment. This analysis not only captures the nuances of their professional landscape but also serves as a compass for informed decision-making, fostering inclusivity in economic participation.**

**Project Objectives:**

1. **Analyses Marginal Worker Demographics: Investigate the demographic characteristics of marginal workers within the dataset, such as age, gender, education level, and occupation.**
2. **Understand Age and Gender Distribution: Examine the age and gender distribution of the workforce to identify patterns or imbalances.**
3. **Explore Industrial Categories: Explore the dataset's industrial categories to understand the distribution of workers across different sectors and industries**

**Visualization Selection:**

1. Bar Charts: Utilize bar charts to visualize the distribution of categorical variables like gender, education level, and occupation. Stacked bar charts can show the breakdown of these categories within marginal workers.
2. Pie Charts: Create pie charts to display the proportional distribution of workers across different industrial categories. Each slice of the pie represents a specific sector, making it easy to compare their relative sizes.
3. Heatmaps: Generate heatmaps to visualize correlations between variables, such as age and education level. This can help identify trends or patterns within the data.
4. Scatterplots: Use scatterplots to explore relationships between variables, such as age and occupation. This can reveal potential insights into the workforce composition.
5. Histograms: Create histograms to visualize the age distribution of workers. This can provide a clear view of age demographics.
6. Box Plots: Employ box plots to display the distribution of age within different industrial categories, helping identify variations and outliers.
7. Line Charts: If applicable, use line charts to depict trends in workforce demographics over time, if the dataset includes temporal data.

**INNOVATION:**

Performing clustering analysis to identify patterns among different industrial categories and age groups can provide valuable insights for various purposes, such as marketing strategies, customer segmentation, or policy planning. Here are the steps you can follow to conduct this analysis:

**1. Data Collection:**

Gather data that includes information about individuals or companies, their industrial categories, and age groups. Ensure that your dataset is clean and properly formatted.

**2. Data Preprocessing:**

Handle missing data: Fill in missing values or remove incomplete recordsNormalize or standardize the data: Ensure that all variables are on the same scale if they have different units or ranges. Encode categorical variables: Convert categorical variables (such as industrial categories) into numerical format using techniques like one-hot encoding.

**3. Feature Selection:**

Decide which features (variables) you want to use for clustering. In your case, the industrial category and age group are relevant, but you may have other variables as well.

**4.Choose Clustering Algorithm:**

Select an appropriate clustering algorithm based on your dataset and the nature of the problem. Common clustering algorithms include K-Means, Hierarchical Clustering, and DBSCAN.

**5. Determine the Number of Clusters:**

Decide how many clusters you want to create. You can use methods like the elbow method or silhouette analysis to find the optimal number of clusters.

**Analyzing the Clusters:**

We now know that our data can be divided into 5 clusters, and we can start to interpret them and draw insights. We can sort them like this:

* Green corresponds with an average annual income and average spending score
* Blue corresponds with a low annual income and high spending score
* Red corresponds with a high annual income and low spending score
* Purple corresponds with a high annual income and high spending score

**Python Program:**

import pandas as pd

import numpy as np

from sklearn.preprocessing import StandardScaler

from sklearn.cluster import KMeans

import matplotlib.pyplot as plt

data = pd.read\_csv('data.csv')

selected\_features = data[['IndustrialCategory', 'Age']]

selected\_features = selected\_features.dropna()

selected\_features = pd.get\_dummies(selected\_features, columns=['IndustrialCategory'])

scaler = StandardScaler()

scaled\_data = scaler.fit\_transform(selected\_features)

for i in range(1, 11):

kmeans = KMeans(n\_clusters=i, init='k-means++', max\_iter=300, n\_init=10, random\_state=0)

kmeans.fit(scaled\_data)

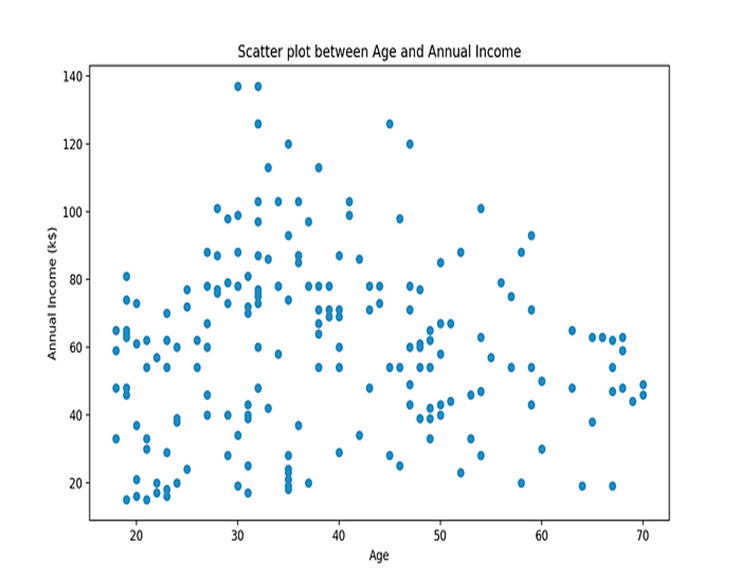
wcss.append(kmeans.inertia\_)

plt.figure(figsize=(8, 6))

plt.plot(range(1, 11), wcss, marker='o', linestyle='--')

plt.title('Elbow Method')

**Output:**

****

**Start the data analysis by loading and preprocessing the dataset.**

**Program**

Import pandas as pd

data = pd.read\_csv('data.csv')

print(data. Head())

data = data. dropna()

data = data. drop duplicates()

print (data. Describe ()) data.to\_csv ("cleaned\_data.csv", index=False

**Output:**

Table Code State Code District Code Area Name Total/

Rural/ Urban \

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 0 | B0806SC | `33 | `000 | State | - TAMIL | NADU |
| Total |  |  |  |  |  |  |
| 1 | B0806SC | `33 | `000 | State | - TAMIL | NADU |
| Total |  |  |  |  |  |  |
| 2 | B0806SC | `33 | `000 | State | - TAMIL | NADU |
| Total |  |  |  |  |  |  |
| 3 | B0806SC | `33 | `000 | State | - TAMIL | NADU |
| Total |  |  |  |  |  |  |
| 4 | B0806SC | `33 | `000 | State | - TAMIL | NADU |
| Total |  |  |  |  |  |  |

Age group Worked for 3 months or more but less than 6 months - Persons \

|  |  |  |
| --- | --- | --- |
| 0 | Total | 1200828 |
|  |  |  |
| 1 | `5-14 | 27791 |
|  |  |  |
| 2 | 15-34 | 514340 |
|  |  |  |
| 3 | 35-59 | 542581 |
|  |  |  |
| 4 | 60+ | 115103 |
|  |  |  |
|  |  |  |

Worked for 3 months or more but less than 6 months - Males \

|  |  |
| --- | --- |
| 0 | 589003 |
| 1 | 14125 |
| 2 | 259560 |
| 3 | 251957 |
| 4 | 62833 |

Worked for 3 months or more but less than 6 months - Females \

|  |  |
| --- | --- |
| 0 | 611825 |
| 1 | 13666 |
| 2 | 254780 |
| 3 | 290624 |
| 4 | 52270 |
| Worked for less than 3 months - Persons ... \ | |
| 0 | 221386 ... |
| 1 | 2447 ... |
| 2 | 92423 ... |
| 3 | 99202 ... |
| 4 | 27165 ... |

|  |  |  |  |
| --- | --- | --- | --- |
| Worked for 3 Industrial Category - N to O - Females \ | | | |
| 0 | 3565 | | |
| 1 | 11 | | |
| 2 | 1754 | | |
| 3 | 1619 | | |
| 4 | 175 | | |
| Industrial Category - P to Q - Persons \ | | | |
| 0 | 11080 | |  |
| 1 | 122 | |  |
| 2 | 7536 | |  |
| 3 | 3205 | |  |
| 4 | 211 | |  |
| Industrial Category - P to Q - Males \ | | | |
| 0 | 4019 |  | |
| 1 | 71 |  | |
| 2 | 2718 |  | |
| 3 | 1131 |  | |
| 4 | 93 |  | |
| Industrial Category - P to Q - Females \ | | | |
| 0 | 7061 | | |
| 1 | 51 | | |
| 2 | 4818 | | |
| 3 | 2074 | | |
| 4 | 118 | | |
|  |  | | |
|  |  | | |
|  |  | | |

**Perform the demographic analysis and create visualizations.**

**Step 1: Import Libraries**

**Step 2: Load the CSV File**

**Step 3: Data Exploration**

**Before creating visualizations, it's essential to explore your data to understand its structure and the demographic variables it contains.**

**You can use functions like head(), info(), and describe() to get a sense of the data.**

**Step 4: Demographic Analysis**

**Depending on your dataset and the demographic variables you want to analyze,**

**you can perform various operations.**

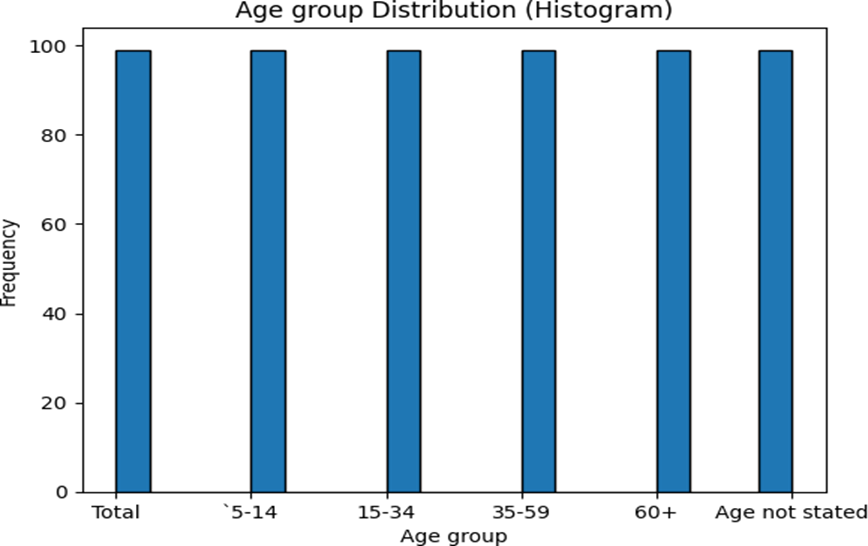
**For example, to analyze the age distribution, you can create a histogram**

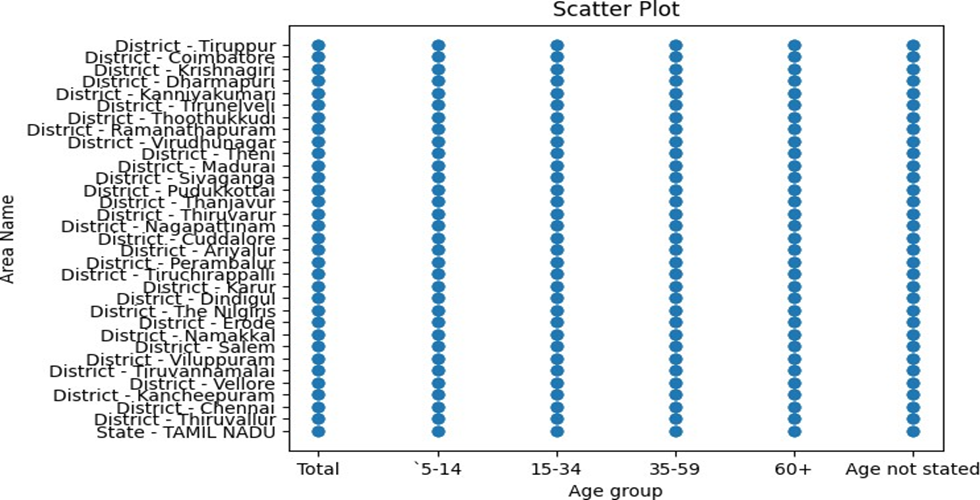
**Step 5: Additional Visualizations**

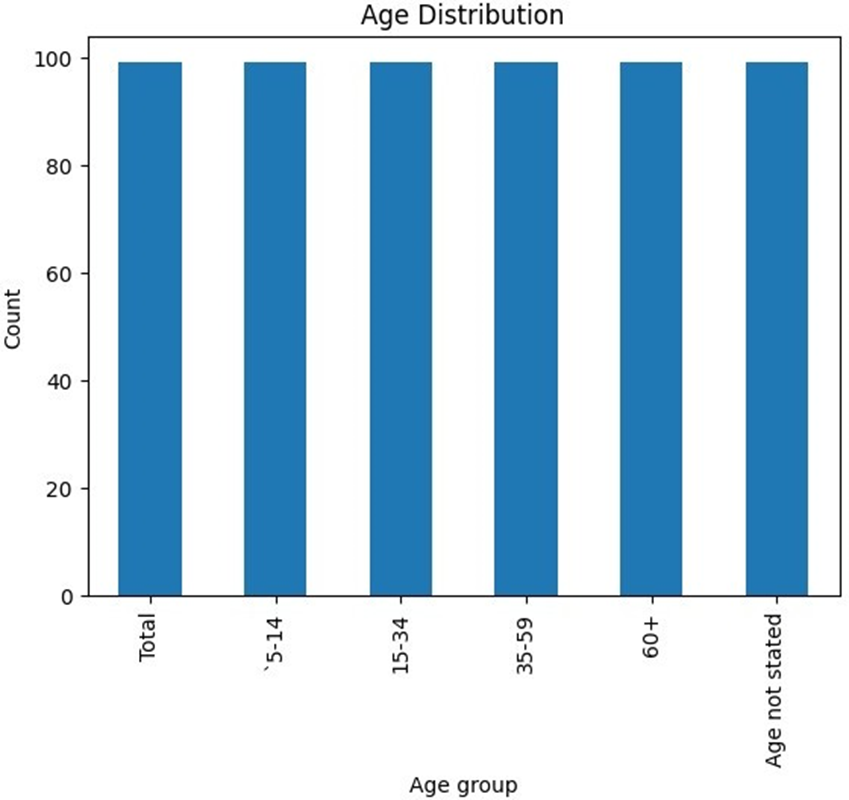
**Depending on your dataset and analysis goals, you can create various types of visualizations, including bar charts, pie charts, scatter plots, and more.**

**Step 6: Export or Display Visualizations**

**You can display the visualizations in your Jupyter Notebook or export them to files using**





****

**CONCLUSION:**

In conclusion, the data analytics on the assessment of marginal workers in Tamil Nadu provides valuable insights into their socio-economic conditions. By leveraging this information, targeted interventions and policies can be devised to uplift this segment, fostering inclusive growth and sustainable development