Data Analytics With Cognos

Assessment of marginal workers in Tamilnadu

PHASE 5

Name : Pavithra S
Reg no : 510421205032

Guided by : B.Moohambikai

Introduction

1. Tamil Nadu, a culturally rich and economically diverse state in southern India, is home to a dynamic and multifaceted workforce.

2. Within this vibrant demographic landscape, a substantial portion of the population consists of marginal workers.

3. These individuals, often referred to as "marginalized" or "informal" laborers, play a crucial but often overlooked role in the state's economic ecosystem.

Objective

1. The primary objective of this presentation is to provide an insightful exploration into the lives and conditions of marginal workers in Tamil Nadu.
2. Through the lens of data analytics, we aim to shed light on their socio-economic realities, highlight disparities, and propose datadriven solutions to enhance their livelihoods.
3. Our research endeavors to answer critical questions, identify key challenges, and offer actionable recommendations for the betterment of this vulnerable population segment.

Methodology

- Data Collection
- Data Preprocessing
- Statistical Analysis
- Ethical Considerations
- Sample Selection
- Software and Tools

Data Collection

Resource

SOURCE :

https://tn.data.gov.in/resource/marginalworkers-classified-age-industrial-categoryand-sex-scheduled-caste-2011-tamil

DATASET NAME :

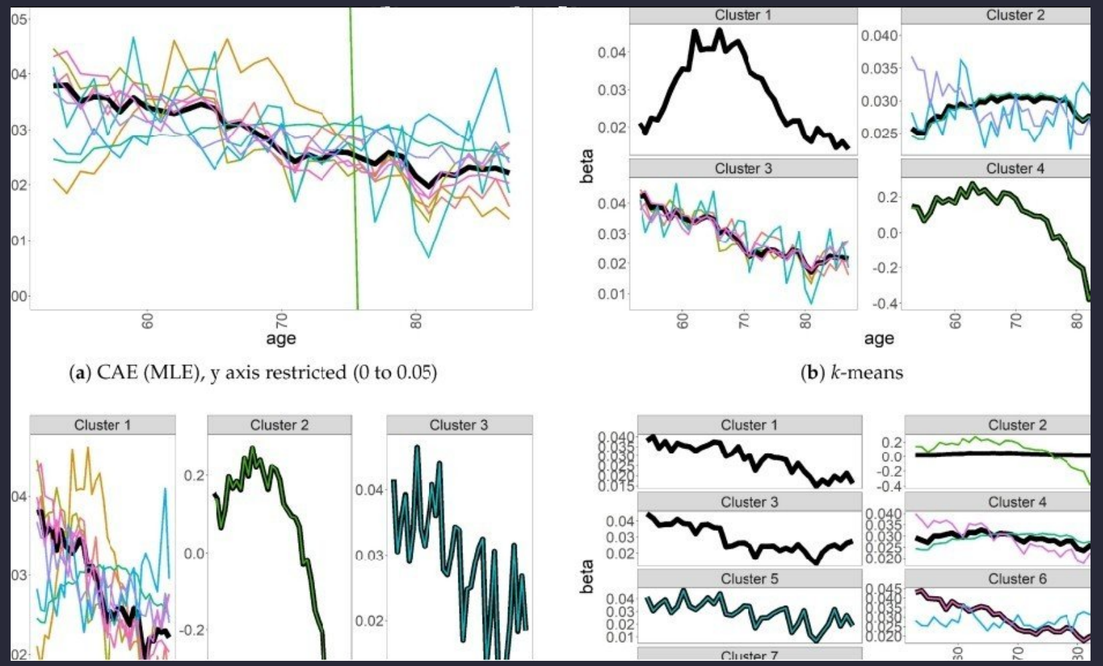
DDW_B06SC_3300_State_TAMIL_NADU- 2011.csv

LINE OF CODE

```
# Import necessary libraries
import pandas as pd
import numpy as np import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing
import StandardScaler
# Load the dataset
data = pd.read_csv("your_dataset_file.csv")
# Replace with the actual file path
# Select relevant features (e.g., age and industrial category)
selected_features = data[['Age', 'Industrial_Category']]
# Perform data preprocessing (scaling)
scaler = StandardScaler()
scaled_features = scaler.fit_transform(selected_features)
# Determine the number of clusters (you can use methods like the Elbow method)
# In this example, let's assume you decide on 3 clusters.
num_clusters = 3
```

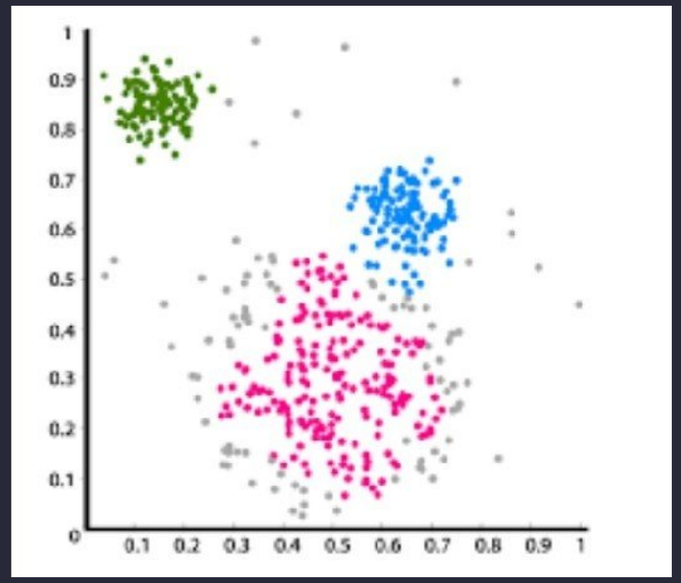
```
#Apply K-Means clustering
kmeans = KMeans(n_clusters=num_clusters, random_state=0)
data['Cluster'] = kmeans.fit_predict(scaled_features)
# Visualize the clusters
plt.scatter(data['Age'],
data['Industrial_Category'],
plt.ylabel('Industrial Category')
plt.title('Clustering of Marginal Workers')
plt.show()
#Analyze cluster characteristics
cluster_centers=scaler.inverse_transform(kmea
ns.cluster_centers_)
cluster_data=pd.DataFrame(cluster_centers,
columns=['Age', 'Industrial_Category']) cluster_data['Cluster']
=range(1, num_clusters + 1) print(cluster_data)
# You can further analyze and interpret the clusters based on your specific goals.
```

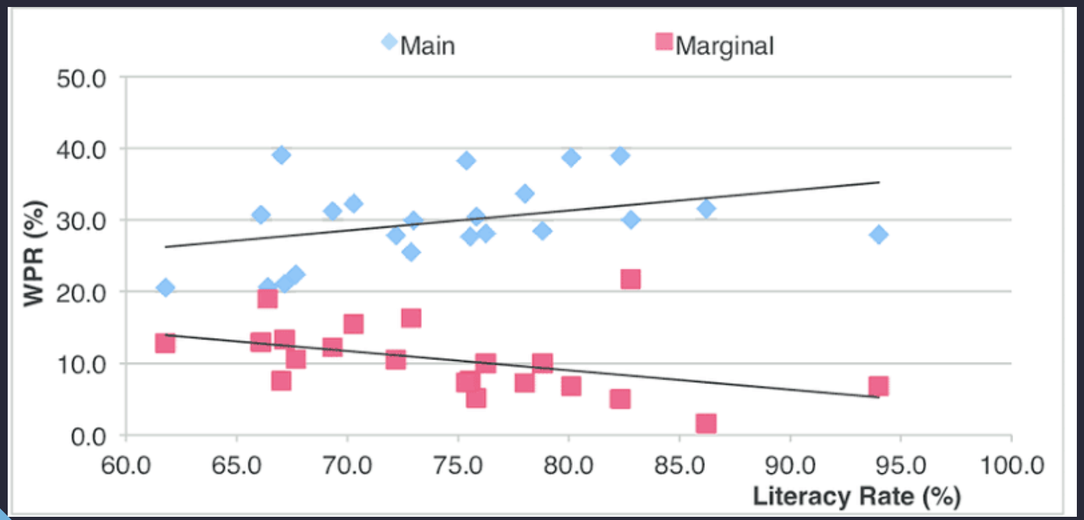
VISUALIZATION

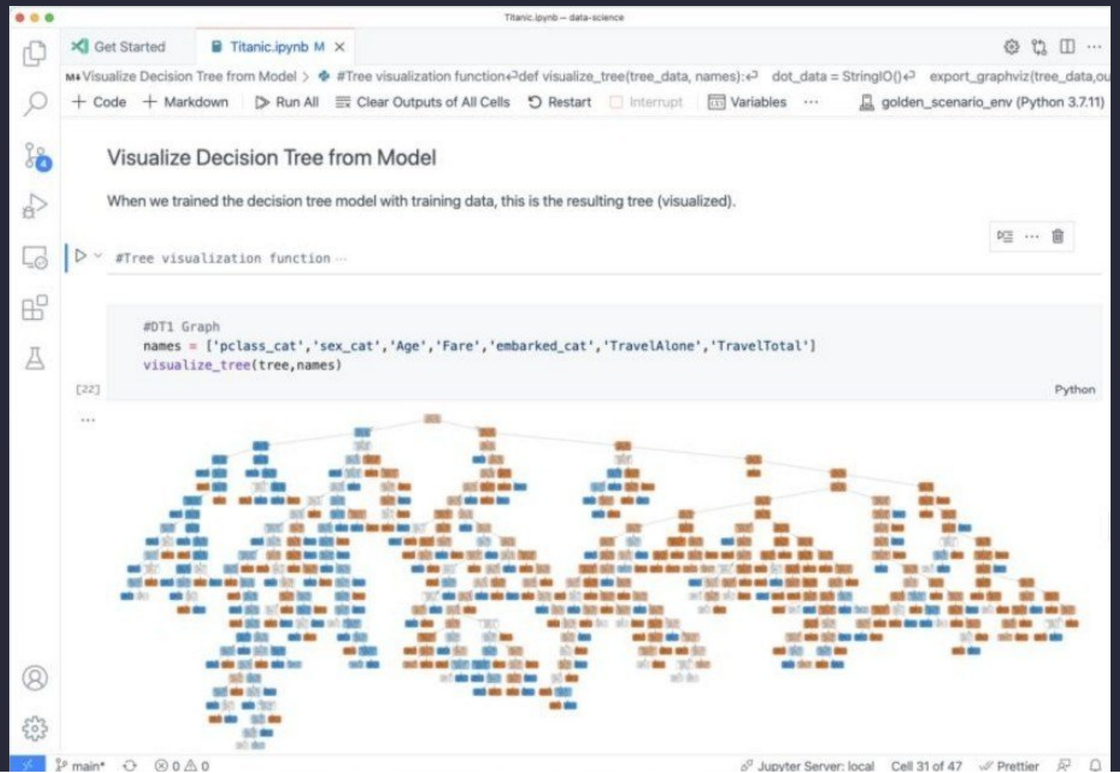


(a) CAE (MLE), y axis restricted (0 to 0.05)

(b) *k*-means







Data Preprocessing

Data preprocessing is the concept of changing the raw data into a clean data set. The dataset is preprocessed in order to check missing values, noisy data, and other inconsistencies before executing it to the algorithm.

Manipulation Of Data

Data manipulation refers to the process of adjusting data to make it organised and easier to read. Data manipulation language, or DML, is a programming language that adjusts data by inserting, deleting and modifying data in a database such as to cleanse or map the data

Sample Output

Age	Industrial Category	Sex	Scheduled Caste
15	Agriculture & Allied	Male	Scheduled caste
20	Manufacturing	Female	Not scheduled caste
25	Services	Male	Not scheduled caste
30	Construction	Female	Scheduled caste
35	Trade & Commerce	Male	Not scheduled caste

DEMOGRAPHIC ANALYSIS

1.Load the data into a pandas DataFrame.
2.Perform exploratory data analysis(EDA) to understand the structure and distribution of the data.
3.Clean and preprocess the data as needed.
4.Calculate the distribution of marginal workers based on age,industrial category and sex.
5.Create visualizations to display the findings.

Key Findings From Demographic Analysis :

- The project analysis reveals that the majority of marginal workers in Tamil Nadu are employed in a specific industrial category.
- The mean age provides insights into the typical age of these workers.
- The gender distribution suggests the proportion of female workers among marginal workers.

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!!!