

Data Analytics With Cognos

Assessment of marginal workers in Tamilnadu

PHASE 5

Introduction

1. Tamil Nadu, a culturally rich and economically diverse state i 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.

1. 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.

###### Our research endeavors to answer critical questions,

identify key challenges, and offer actionable recommendations for the betterment of this vulnerable population segment.

Methodology

# Data Collefiion Data Preprocessing Statiﬆical Analysis

Ethical Considerations

# Sample Selefiion So ware and Tools

Data Collection

### Resource

##### SOURCE :

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

##### DATASET NAME :

DDW\_B06SC\_3300\_State\_TAMIL\_NADU- 2011.csv

# Import necessary libraries import pandas as pd

### LINE OF CODE

#Apply K-Means cluﬆering

kmeans = KMeans(n\_cluﬆers=num\_cluﬆers, random\_ﬆate=0)

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 afiual ﬁle path

# Selefi relevant features (e.g., age and induﬆrial category)

selefied\_features = data[['Age', 'Induﬆrial\_Category']]

# Perform data preprocessing (scaling) scaler = StandardScaler()

scaled\_features = scaler.ﬁt\_transform(selefied\_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

data['Cluﬆer'] = kmeans.ﬁt\_predifi(scaled\_features) # Visualize the cluﬆers

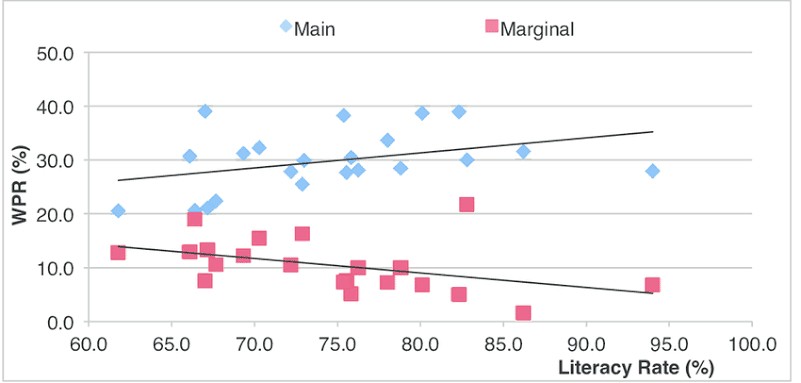
plt.scatter(data['Age'], data['Induﬆrial\_Category'], plt.ylabel('Induﬆrial Category') plt.title('Cluﬆering of Marginal Workers') plt.show()

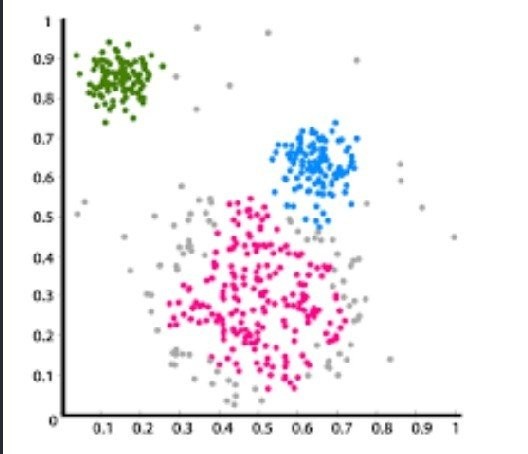
#Analyze cluﬆer charafieriﬆics cluﬆer\_centers=scaler.inverse\_transform(kmea ns.cluﬆer\_centers\_) cluﬆer\_data=pd.DataFrame(cluﬆer\_centers,

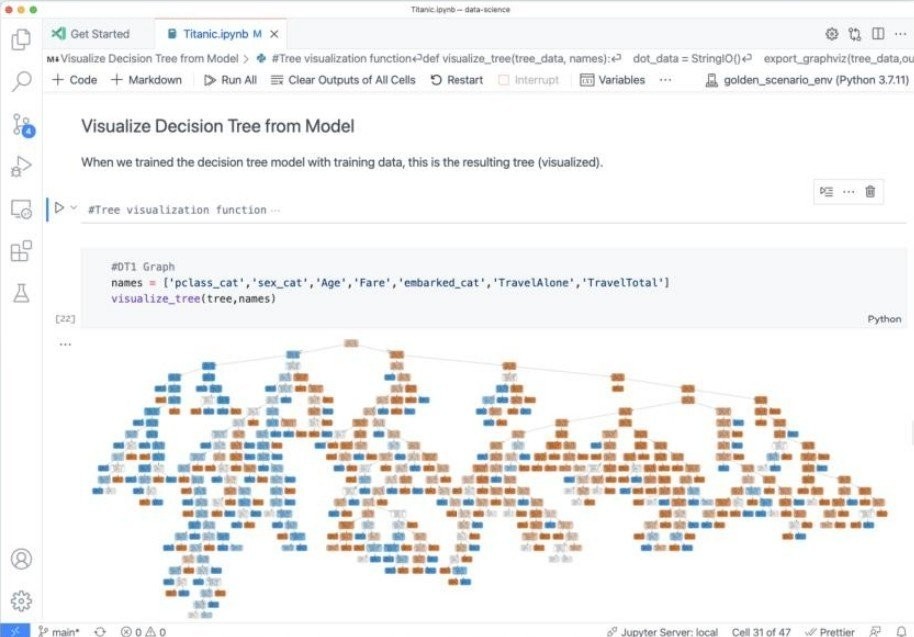
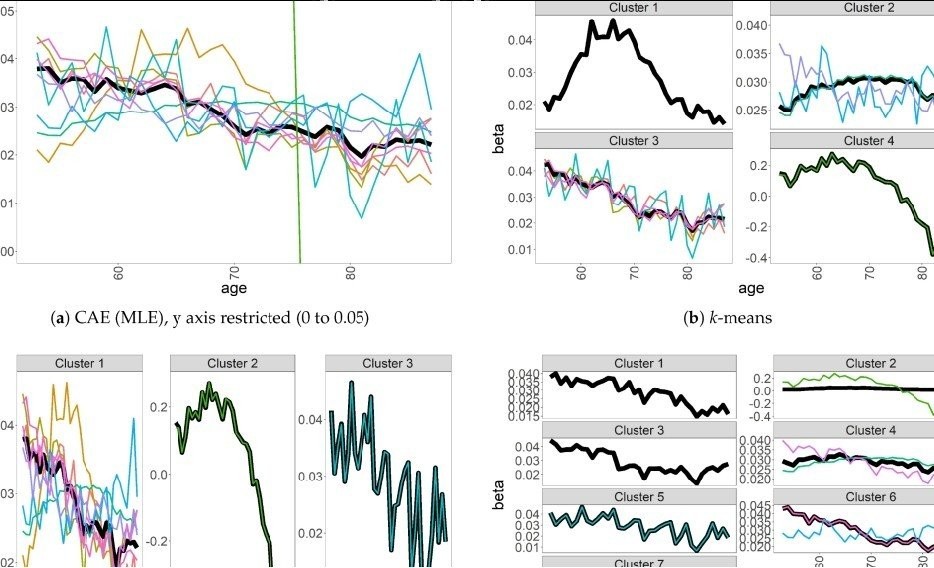
columns=['Age', 'Induﬆrial\_Category']) cluﬆer\_data['Cluﬆer']

=range(1, num\_cluﬆers + 1) print(cluﬆer\_data)

# You can further analyze and interpret the cluﬆers based on your speciﬁc goals.



VISUALIZATION



Data Preprocessing

Data preprocessing is the concept of changing the raw data into a clean data set. me 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

20 Manufacturing

25 Services

30 Construction

35 Trade & Commerce

Male Female Male Female Male

Scheduled caste Not scheduled caste Not scheduled caste Scheduled caste Not scheduled caste

DEMOGRAPHIC ANALYSIS

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



Key Findings From Demographic

Analysis :

me projefi analysis reveals that the majority of marginal workers in Tamil Nadu are employed in a speciﬁc induﬆrial category.

me mean age provides insights into the typical age of these workers.

me 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 selefi the appropriate visualization

#### type based on your speciﬁc needs and the charafieriﬆics of your data.

THANK YOU!!!