**Project Title: Assessment of Marginal Workers in Tamil Nadu- A Socioeconomic analysis**

**Introduction:**

Certainly, the statement "Assessment of Marginal Workers in Tamil Nadu: A Socioeconomic Analysis" suggests a comprehensive examination of the conditions and challenges faced by the marginalized workforce in the state of Tamil Nadu, India. This topic is of critical importance as it addresses the socioeconomic disparities and vulnerabilities prevalent among workers belonging to marginalized communities.

The project can aim to:

1. **Identify Marginalized Groups:** Clearly delineate the marginalized groups within the labor force, such as agricultural laborers, informal sector workers, and those engaged in low-paying, precarious employment.
2. **Assess Socioeconomic Conditions:** Delve into the socioeconomic realities faced by these workers, including their income levels, access to education, healthcare, and other essential services, and their overall standard of living.
3. **Analyze Employment Patterns:** Investigate the employment patterns prevalent among marginal workers, including seasonal unemployment, underemployment, and the lack of job security, and assess the impact of these factors on their socioeconomic status.
4. **Examine Policy Frameworks:** Evaluate existing government policies and programs aimed at uplifting marginalized workers, highlighting their effectiveness, shortcomings, and potential areas for improvement.
5. **Suggest Interventions:** Propose targeted interventions and policy recommendations that can enhance the socioeconomic status of marginal workers, including measures for skill development, access to financial resources, and improved social protections.
6. **Advocate for Social Inclusion:** Advocate for the creation of a more inclusive and equitable society by promoting awareness, fostering community participation, and encouraging advocacy for the rights and well-being of marginalized workers.

Code:  
import pandas as pd

df = pd.read\_csv("ds.csv")

print(df.head())

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

df=pd.read\_csv("ds.csv")

df

df.plot()

males\_col=[i for i in df.columns if 'Males' in i]

females\_col=[i for i in df.columns if 'Females' in i]

print("----MALES----")

for i in males\_col:

print(i)

print("----FEMALES----")

for i in females\_col:

print(i)

from sklearn.cluster import KMeans

X = df[['Industrial Category - B - Males', 'Industrial Category - C - HHI - Persons','Industrial Category - B - Females']].copy()

n\_clusters = 5

n\_init = 10

kmeans = KMeans(n\_clusters=n\_clusters, n\_init=n\_init)

wcss = []

for i in range(1, 594):

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

kmeans.fit(X)

wcss.append(kmeans.inertia\_)

import matplotlib.pyplot as plt

import seaborn as sns

sns.set()

plt.plot(range(1, 594), wcss)

plt.title('Selecting the Numbeer of Clusters using the Elbow Method')

plt.xlabel('Clusters')

plt.ylabel('WCSS')

plt.show()

from sklearn.mixture import GaussianMixture

n\_clusters = 2

gmm\_model = GaussianMixture(n\_components=n\_clusters)

gmm\_model.fit(X)

cluster\_labels = gmm\_model.predict(X)

X = pd.DataFrame(X)

X['cluster'] = cluster\_labels

import matplotlib.pyplot as plt

# Define a list of colors for each cluster

colors = ['red','green']

# Assuming n\_clusters is defined

data = X[X["cluster"] == k]

plt.scatter(data["Industrial Category - B - Males"], data["Industrial Category - C - HHI - Persons"], c='red', label='Cluster 0')

plt.scatter(data["Industrial Category - B - Females"], data["Industrial Category - C - HHI - Persons"], c='green',label='cluster 1')

plt.title("Clusters Identified by Gaussian Mixture Model")

plt.xlabel("Industrial Category - B - Females")

plt.ylabel("Industrial Category - C - HHI - Persons")

plt.legend()

plt.show()

1. A line plot will be generated representing the WCSS (Within-Cluster Sum of Squares) against the number of clusters. The plot will help in selecting the appropriate number of clusters using the Elbow Method.
2. A scatter plot will be displayed, representing the clusters identified by the Gaussian Mixture Model. The data points will be colored in red and green, corresponding to the two clusters. The plot will have 'Industrial Category - B - Females' on the x-axis and 'Industrial Category - C - HHI - Persons' on the y-axis.

Suppose the dataset looks like this:

Industrial Category - B - Males Industrial Category - C - HHI - Persons Industrial Category - B - Females

0 5 20 10

1 6 25 9

2 3 28 8

3 8 30 7

4 2 35 6

For the Cluster plot, it might look like this:

Cluster 0

0 10

1 9

2 8

3 7

4 6

This is a simplified representation, and your actual data and output may differ based on the specifics of your dataset. Please run the code with your actual data to get a more accurate representation of the output.