Data analytics with cognos phase 2

Clustering Analysis for Assessing Marginal Workers in Tamil Nadu

Name: Lekhasri S

Reg no: 510421205021

Guided by: Moohambigai B

SYNOPSIS

>> Introduction
>> Challenges
>> Clustering
>> Resources
>> Line of code
>> Visualization
>> Analysis
>> Sample result
>> Conclusion

Introduction

Unlocking Insights: Clustering Analysis for Assessing Marginal Workers in Tamil Nadu. This presentation explores the use of clustering analysis to identify patterns among marginal workers in Tamil Nadu. By analyzing demographic and economic factors, we can gain insights into the challenges faced by these workers.

Main Workers			Marginal Workers		
M	F	P	M	F	P
6	7		8	9	10
615742	[illegible]	[illegible]	102244	[illegible]	[illegible]
29637	[illegible]	[illegible]	7110	[illegible]	5370
5816	[illegible]	[illegible]	9723	[illegible]	8840
6550	[illegible]	[illegible]	12314	[illegible]	11634
7623	[illegible]	[illegible]	13490	[illegible]	13010
7158	[illegible]	[illegible]	11774	[illegible]	11404
7217	[illegible]	[illegible]	11736	[illegible]	11540
1238	[illegible]	[illegible]	19487	[illegible]	[illegible]

Challenges Faced by Marginal Workers

Marginal workers in Tamil Nadu face a range of challenges, including low wages, lack of job security, and limited access to credit. By understanding these challenges, we can develop targeted interventions to improve the economic status of these workers.

What is Clustering Analysis?

- **Clustering Analysis** is a statistical technique used to group similar data points together. In this presentation, we will use clustering analysis to group marginal workers in Tamil Nadu based on their demographic and economic characteristics.

Who are Marginal Workers?

Marginal Workers are those who work for less than six months in a year. They are often employed in low-paying jobs with little job security. In Tamil Nadu, many marginal workers are employed in the agricultural sector.

G	H	I	J	K	L	M	N	O	P	
Worked fo	Worked fo	Worked fo	Worked fo	Worked fo	Worked fo	Industrial (	Industrial (	Industrial (	Industrial (	Ind
1200828	589003	611825	221386	99368	122018	64235	34632	29603	907752	
27791	14125	13666	2447	1247	1200	1710	825	885	6398	
514340	259560	254780	92423	43892	48531	24863	12711	12152	345420	
542581	251957	290624	99202	40691	58511	29692	15927	13765	450052	
115103	62833	52270	27165	13465	13700	7930	5151	2779	105325	
1013	528	485	149	73	76	40	18	22	557	
966645	459738	506907	174443	73663	100780	59637	32189	27448	824698	
17239	8713	8526	1977	985	992	1443	684	759	6005	
406847	198575	208272	71974	31917	40057	22933	11766	11167	316885	
444800	199573	245227	77922	29808	48114	27799	14887	12912	406147	
97011	52498	44513	22446	10902	11544	7425	4835	2590	95151	
748	379	369	124	51	73	37	17	20	510	
234183	129265	104918	46943	25705	21238	4598	2443	2155	83054	
10552	5412	5140	470	262	208	267	141	126	393	
107493	60985	46508	20449	11975	8474	1930	945	985	28535	
97781	52384	45397	21280	10883	10397	1893	1040	853	43905	
18092	10335	7757	4719	2563	2156	505	316	189	10174	
265	149	116	25	22	3	3	1	2	47	
74448	39295	35153	15866	8004	7862	3066	1663	1403	42579	
2521	1284	1237	147	82	65	122	56	66	330	
33568	18049	15519	6529	3654	2875	1225	632	593	15591	
32568	16771	15797	7718	3529	4189	1414	792	622	22192	
5716	3147	2569	1465	739	726	305	183	122	4441	
75	44	31	7	0	7	0	0	0	25	
55577	28082	27495	12131	5653	6478	2804	1511	1293	39766	

RESOURCE

SOURCE

https://tn.data.gov.in/resource/marginal-workers-classified-age-industrial-category-and-sex-scheduled-caste-2011-tamil

DATASET NAME

DDW_B06SC_3300_State_TAMIL_NADU-2011.csv

clustering

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
(eg.,cluster centers)
cluster_centers=scaler.inverse_transform(kmeans.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 OF CLUSTER

CLUSTERING ANALYSIS

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

(b) k-means

Sample result

	AGE	INDUSTRIAL_CATEGORY	CLUSTER
0	23.456789	1.234567	1
1	35.678901	2.345678	2
2	45.789012	3.456789	3

```
•[3]: # 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("DDW_B06SC_3300_State_TAMIL_NADU-2011.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
```

Conclusion

Clustering analysis can provide valuable insights into the challenges faced by marginal workers in Tamil Nadu. By understanding these challenges and developing targeted interventions, we can improve the economic status of these workers and promote inclusive economic growth.

thank you