

FCM SocioEco

June 30, 2024

1 Fuzzy c-Means on socio-economic dataset

```
[1]: import pandas as pd
from fcmeans import FCM
from sklearn.metrics import silhouette_score
import numpy as np

df = pd.read_excel('Preprocessed_data_standardscaler.xlsx')

# Dropping the ID Column, as it is not required in clustering.

new_df = df.drop(columns=['ID'])
new_df
```

```
[1]:
```

	Normalised Age (Yrs)	Gender (Male)	Education (<Class 10)	\
0	-0.927352	1	1	
1	0.387063	1	0	
2	-0.739578	1	0	
3	0.762611	0	1	
4	-1.678446	1	0	
..	
476	-0.739578	1	0	
477	-1.302899	1	0	
478	2.077026	1	0	
479	-0.551804	1	0	
480	-0.551804	1	0	

	Education (Class 10-12)	Graduate	Higher Edu	Income <20k	\
0	0	0	0	1	
1	0	1	0	0	
2	0	1	0	0	
3	0	0	0	1	
4	1	0	0	0	
..	
476	0	1	0	0	
477	0	1	0	0	
478	0	1	0	0	

479	0	1	0	0
480	0	0	1	0

	Income 20-40k	Income 40-60k	Income 60-80k	Income >80k \
0	0	0	0	0
1	0	1	0	0
2	0	1	0	0
3	0	0	0	0
4	1	0	0	0
..
476	0	1	0	0
477	0	0	1	0
478	0	0	1	0
479	0	1	0	0
480	0	1	0	0

	Experience (<1yr)	Experience (1-2yr)	Experience (2-5yr) \
0	0	0	0
1	0	0	0
2	0	0	0
3	0	0	1
4	1	0	0
..
476	0	0	0
477	0	0	1
478	0	0	0
479	0	0	0
480	0	0	0

	Experience (5-10yr)	Experience (10-20yr)	Experience (20-30yr) \
0	1	0	0
1	0	1	0
2	1	0	0
3	0	0	0
4	0	0	0
..
476	1	0	0
477	0	0	0
478	0	0	1
479	0	1	0
480	0	1	0

	Experience (>30yr)	Helmet (Full-mask)
0	0	1
1	0	0
2	0	1
3	0	0

```

4           0           1
..         ...         ...
476         0           1
477         0           1
478         0           1
479         0           1
480         0           1

```

[481 rows x 19 columns]

```

[2]: def part_coeff(U):
      U_1 = U**2
      U_1 = np.sum(U_1,axis=1)
      U_1 = np.sum(U_1)
      U_1 = U_1/U.shape[0]
      return U_1

```

```

[3]: array_df = new_df.to_numpy()

fuzz = [1.5,2,2.5]

list1,list2,list3 = [],[],[]
coefficients = [list1,list2,list3]

for i in range(0,len(fuzz)):
    for j in range(2,7):
        fcm = FCM(n_clusters=j,m=fuzz[i],random_state=42)
        fcm.fit(array_df)
        coeff = part_coeff(fcm.u)
        coefficients[i].append(coeff)

list_ = list(range(2,7))

part_coefficients = pd.DataFrame({
    'no. of clusters': list_,
    'm = 1.5': list1,
    'm = 2': list2,
    'm = 2.5': list3
})

part_coefficients

```

```

[3]:   no. of clusters   m = 1.5   m = 2   m = 2.5
0           2   0.664656   0.500308   0.500000
1           3   0.506294   0.333606   0.333333
2           4   0.435992   0.250006   0.250000
3           5   0.444104   0.200142   0.200000

```

4 6 0.432245 0.166695 0.166667

```
[4]: scores = {}

for i in range(2,7):
    fcm = FCM(n_clusters=i,m=1.5,random_state=42)
    fcm.fit(array_df)
    score = silhouette_score(array_df,fcm.predict(array_df))
    scores[i] = score

df_scores = pd.DataFrame(list(scores.items()), columns=['no. of clusters',
    ↳ 'silhouette score'])
df_scores
```

```
[4]:
```

	no. of clusters	silhouette score
0	2	0.197952
1	3	0.156246
2	4	0.147711
3	5	0.179893
4	6	0.188116

```
[5]: fcm = FCM(n_clusters=2,m=1.5,random_state=42)
fcm.fit(array_df)

cluster_list = list(fcm.predict(array_df))

column_list = list(new_df.columns)

clustered_data = pd.DataFrame(array_df,columns=column_list)
clustered_data['Cluster'] = cluster_list
clustered_data
```

```
[5]:
```

	Normalised Age (Yrs)	Gender (Male)	Education (<Class 10)	\
0	-0.927352	1.0	1.0	
1	0.387063	1.0	0.0	
2	-0.739578	1.0	0.0	
3	0.762611	0.0	1.0	
4	-1.678446	1.0	0.0	
..	
476	-0.739578	1.0	0.0	
477	-1.302899	1.0	0.0	
478	2.077026	1.0	0.0	
479	-0.551804	1.0	0.0	
480	-0.551804	1.0	0.0	

	Education (Class 10-12)	Graduate	Higher Edu	Income <20k	\
0	0.0	0.0	0.0	1.0	

1	0.0	1.0	0.0	0.0
2	0.0	1.0	0.0	0.0
3	0.0	0.0	0.0	1.0
4	1.0	0.0	0.0	0.0
..
476	0.0	1.0	0.0	0.0
477	0.0	1.0	0.0	0.0
478	0.0	1.0	0.0	0.0
479	0.0	1.0	0.0	0.0
480	0.0	0.0	1.0	0.0

	Income 20-40k	Income 40-60k	Income 60-80k	Income >80k \
0	0.0	0.0	0.0	0.0
1	0.0	1.0	0.0	0.0
2	0.0	1.0	0.0	0.0
3	0.0	0.0	0.0	0.0
4	1.0	0.0	0.0	0.0
..
476	0.0	1.0	0.0	0.0
477	0.0	0.0	1.0	0.0
478	0.0	0.0	1.0	0.0
479	0.0	1.0	0.0	0.0
480	0.0	1.0	0.0	0.0

	Experience (<1yr)	Experience (1-2yr)	Experience (2-5yr) \
0	0.0	0.0	0.0
1	0.0	0.0	0.0
2	0.0	0.0	0.0
3	0.0	0.0	1.0
4	1.0	0.0	0.0
..
476	0.0	0.0	0.0
477	0.0	0.0	1.0
478	0.0	0.0	0.0
479	0.0	0.0	0.0
480	0.0	0.0	0.0

	Experience (5-10yr)	Experience (10-20yr)	Experience (20-30yr) \
0	1.0	0.0	0.0
1	0.0	1.0	0.0
2	1.0	0.0	0.0
3	0.0	0.0	0.0
4	0.0	0.0	0.0
..
476	1.0	0.0	0.0
477	0.0	0.0	0.0
478	0.0	0.0	1.0

479	0.0	1.0	0.0
480	0.0	1.0	0.0

	Experience (>30yr)	Helmet (Full-mask)	Cluster
0	0.0	1.0	0
1	0.0	0.0	1
2	0.0	1.0	0
3	0.0	0.0	1
4	0.0	1.0	0
..
476	0.0	1.0	0
477	0.0	1.0	0
478	0.0	1.0	1
479	0.0	1.0	0
480	0.0	1.0	0

[481 rows x 20 columns]

[6]: *# Now that the clustering is completed, the 'Normalised Age (Yrs)' column is*
↳dropped and replaced with the 'Age (Yrs)' column.

```
final_df = clustered_data.copy()

old_df = pd.read_excel('Asansol socio-economic data 1.xlsx')
age = list(old_df['Age (Yrs)'])

final_df['Age (Yrs)'] = age

final_df = final_df.drop(columns = ['Normalised Age (Yrs)'])

list2 = list(final_df.columns)
new_order = [list2[-1]] + list2[0:-1]

final_df = final_df[new_order]
final_df
```

[6]:

	Age (Yrs)	Gender (Male)	Education (<Class 10)	Education (Class 10-12)	\
0	26	1.0	1.0	0.0	
1	40	1.0	0.0	0.0	
2	28	1.0	0.0	0.0	
3	44	0.0	1.0	0.0	
4	18	1.0	0.0	1.0	
..	
476	28	1.0	0.0	0.0	
477	22	1.0	0.0	0.0	
478	58	1.0	0.0	0.0	
479	30	1.0	0.0	0.0	

480	30	1.0	0.0	0.0
-----	----	-----	-----	-----

	Graduate	Higher Edu	Income <20k	Income 20-40k	Income 40-60k \
0	0.0	0.0	1.0	0.0	0.0
1	1.0	0.0	0.0	0.0	1.0
2	1.0	0.0	0.0	0.0	1.0
3	0.0	0.0	1.0	0.0	0.0
4	0.0	0.0	0.0	1.0	0.0
..
476	1.0	0.0	0.0	0.0	1.0
477	1.0	0.0	0.0	0.0	0.0
478	1.0	0.0	0.0	0.0	0.0
479	1.0	0.0	0.0	0.0	1.0
480	0.0	1.0	0.0	0.0	1.0

	Income 60-80k	Income >80k	Experience (<1yr)	Experience (1-2yr) \
0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0
4	0.0	0.0	1.0	0.0
..
476	0.0	0.0	0.0	0.0
477	1.0	0.0	0.0	0.0
478	1.0	0.0	0.0	0.0
479	0.0	0.0	0.0	0.0
480	0.0	0.0	0.0	0.0

	Experience (2-5yr)	Experience (5-10yr)	Experience (10-20yr) \
0	0.0	1.0	0.0
1	0.0	0.0	1.0
2	0.0	1.0	0.0
3	1.0	0.0	0.0
4	0.0	0.0	0.0
..
476	0.0	1.0	0.0
477	1.0	0.0	0.0
478	0.0	0.0	0.0
479	0.0	0.0	1.0
480	0.0	0.0	1.0

	Experience (20-30yr)	Experience (>30yr)	Helmet (Full-mask)	Cluster
0	0.0	0.0	1.0	0
1	0.0	0.0	0.0	1
2	0.0	0.0	1.0	0
3	0.0	0.0	0.0	1
4	0.0	0.0	1.0	0

```

..          ...          ...          ...          ...
476          0.0          0.0          1.0          0
477          0.0          0.0          1.0          0
478          1.0          0.0          1.0          1
479          0.0          0.0          1.0          0
480          0.0          0.0          1.0          0

```

[481 rows x 20 columns]

```
[7]: # now we seperate out the various clusters from the dataset
```

```

cluster_1 = final_df.loc[final_df['Cluster']==0]
cluster_1
cluster_1.describe()

```

```

[7]:      Age (Yrs)  Gender (Male)  Education (<Class 10)  \
count  272.000000      272.000000      272.000000
mean    28.470588        0.948529        0.198529
std       4.892107        0.221363        0.399628
min       16.000000        0.000000        0.000000
25%       25.750000        1.000000        0.000000
50%       28.000000        1.000000        0.000000
75%       32.000000        1.000000        0.000000
max       39.000000        1.000000        1.000000

      Education (Class 10-12)  Graduate  Higher Edu  Income <20k  \
count      272.000000      272.000000      272.000000      272.000000
mean          0.312500      0.455882      0.033088      0.297794
std          0.464367      0.498968      0.179197      0.458131
min           0.000000      0.000000      0.000000      0.000000
25%           0.000000      0.000000      0.000000      0.000000
50%           0.000000      0.000000      0.000000      0.000000
75%           1.000000      1.000000      0.000000      1.000000
max           1.000000      1.000000      1.000000      1.000000

      Income 20-40k  Income 40-60k  Income 60-80k  Income >80k  \
count      272.000000      272.000000      272.000000      272.000000
mean          0.393382      0.246324      0.047794      0.014706
std          0.489401      0.431663      0.213724      0.120595
min           0.000000      0.000000      0.000000      0.000000
25%           0.000000      0.000000      0.000000      0.000000
50%           0.000000      0.000000      0.000000      0.000000
75%           1.000000      0.000000      0.000000      0.000000
max           1.000000      1.000000      1.000000      1.000000

      Experience (<1yr)  Experience (1-2yr)  Experience (2-5yr)  \
count      272.000000      272.000000      272.000000

```


mean	0.047794	0.128676	0.202206
std	0.213724	0.335459	0.402385
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.000000	0.000000	0.000000
max	1.000000	1.000000	1.000000

	Experience (5-10yr)	Experience (10-20yr)	Experience (20-30yr)	\
count	272.000000	272.000000	272.000000	
mean	0.481618	0.136029	0.003676	
std	0.500583	0.343452	0.060634	
min	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	
50%	0.000000	0.000000	0.000000	
75%	1.000000	0.000000	0.000000	
max	1.000000	1.000000	1.000000	

	Experience (>30yr)	Helmet (Full-mask)	Cluster
count	272.0	272.000000	272.0
mean	0.0	0.882353	0.0
std	0.0	0.322784	0.0
min	0.0	0.000000	0.0
25%	0.0	1.000000	0.0
50%	0.0	1.000000	0.0
75%	0.0	1.000000	0.0
max	0.0	1.000000	0.0

```
[8]: cluster_2 = final_df.loc[final_df['Cluster']==1]
      cluster_2.describe()
```

	Age (Yrs)	Gender (Male)	Education (<Class 10)	\
count	209.000000	209.000000	209.000000	
mean	45.516746	0.952153	0.177033	
std	8.127645	0.213955	0.382613	
min	34.000000	0.000000	0.000000	
25%	40.000000	1.000000	0.000000	
50%	44.000000	1.000000	0.000000	
75%	50.000000	1.000000	0.000000	
max	73.000000	1.000000	1.000000	

	Education (Class 10-12)	Graduate	Higher Edu	Income <20k	\
count	209.000000	209.000000	209.000000	209.000000	
mean	0.296651	0.46890	0.057416	0.167464	
std	0.457878	0.50023	0.233195	0.374286	
min	0.000000	0.00000	0.000000	0.000000	
25%	0.000000	0.00000	0.000000	0.000000	

50%	0.000000	0.000000	0.000000	0.000000
75%	1.000000	1.000000	0.000000	0.000000
max	1.000000	1.000000	1.000000	1.000000

	Income 20-40k	Income 40-60k	Income 60-80k	Income >80k \
count	209.000000	209.000000	209.000000	209.000000
mean	0.306220	0.315789	0.186603	0.023923
std	0.462029	0.465946	0.390528	0.153178
min	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000	0.000000
75%	1.000000	1.000000	0.000000	0.000000
max	1.000000	1.000000	1.000000	1.000000

	Experience (<1yr)	Experience (1-2yr)	Experience (2-5yr) \
count	209.000000	209.000000	209.000000
mean	0.009569	0.019139	0.043062
std	0.097588	0.137342	0.203485
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.000000	0.000000	0.000000
max	1.000000	1.000000	1.000000

	Experience (5-10yr)	Experience (10-20yr)	Experience (20-30yr) \
count	209.000000	209.000000	209.000000
mean	0.124402	0.521531	0.229665
std	0.330832	0.500736	0.421627
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	1.000000	0.000000
75%	0.000000	1.000000	0.000000
max	1.000000	1.000000	1.000000

	Experience (>30yr)	Helmet (Full-mask)	Cluster
count	209.000000	209.000000	209.0
mean	0.052632	0.755981	1.0
std	0.223833	0.430535	0.0
min	0.000000	0.000000	1.0
25%	0.000000	1.000000	1.0
50%	0.000000	1.000000	1.0
75%	0.000000	1.000000	1.0
max	1.000000	1.000000	1.0

[9]: # The following dataset contains the mean value for each parameter of each
↳ cluster.

```
overall_mean = final_df.groupby('Cluster').mean()

overall_mean
```

```
[9]:      Age (Yrs)  Gender (Male)  Education (<Class 10)  \
Cluster
0      28.470588      0.948529      0.198529
1      45.516746      0.952153      0.177033

      Education (Class 10-12)  Graduate  Higher Edu  Income <20k  \
Cluster
0      0.312500  0.455882      0.033088      0.297794
1      0.296651  0.468900      0.057416      0.167464

      Income 20-40k  Income 40-60k  Income 60-80k  Income >80k  \
Cluster
0      0.393382      0.246324      0.047794      0.014706
1      0.306220      0.315789      0.186603      0.023923

      Experience (<1yr)  Experience (1-2yr)  Experience (2-5yr)  \
Cluster
0      0.047794      0.128676      0.202206
1      0.009569      0.019139      0.043062

      Experience (5-10yr)  Experience (10-20yr)  Experience (20-30yr)  \
Cluster
0      0.481618      0.136029      0.003676
1      0.124402      0.521531      0.229665

      Experience (>30yr)  Helmet (Full-mask)
Cluster
0      0.000000      0.882353
1      0.052632      0.755981
```

```
[10]: # Exporting the datasets to Excel:

cluster_1.to_excel('Cluster_1_FCM.xlsx')
cluster_2.to_excel('Cluster_2_FCM.xlsx')

overall_mean.to_excel('Means_of_parameters_FCM.xlsx')
final_df.to_excel('clustered_data_FCM.xlsx')
```