

IBM Data Science Professional Certificate Capstone Project

Mexican Restaurant Location Analysis

Author: **Włodzimierz Kuczyński**

April 2021

IBM Data Science Professional Certificate Capstone Project

Mexican Restaurant Location Analysis

Goal of the project: to find the best location for a Mexican restaurant in San Francisco, California.

The restaurant is aimed at the following clientele:

- 30-49 year olds
- with income above 100k\$ per annum
- mainly of Mexican origin (at least at first)

IBM Data Science Professional Certificate Capstone Project

Mexican Restaurant Location Analysis

Steps undertaken to find the optimum restaurant location:

- Analysis of the census demographic data of San Francisco census tracts using K-means clustering algorithm to identify the market area with the optimum clientele base
- Analysis of the direct competition and other restaurants in the selected area with the data obtained with Foursquare API

IBM Data Science Professional Certificate Capstone Project

Mexican Restaurant Location Analysis

The data used to cluster the tracts :

- the average number of the population with income above 100k USD per year per square mile,
- the average number of the Hispanic population per square mile- the majority of Hispanic population in San Francisco are of Mexican origin,
- the average number of the 30-49 year old population per square mile,

IBM Data Science Professional Certificate Capstone Project

Mexican Restaurant Location Analysis

Sample of the data used to cluster the tracts its key statistical measures:

	tractid	avg_hisp	avg_inc_above_100k	avg_30-49_pop	land_area
0	06075010100	927.280282	5693.425998	5220.121250	0.299801
1	06075010200	673.152074	15463.014831	8027.589366	0.199064
2	06075010300	3910.093326	21586.323342	16364.464002	0.103578
3	06075010400	4204.300087	24951.551587	15598.339407	0.129629
4	06075010500	1087.266095	6478.947501	3231.491081	0.263965
...
192	06075980401	0.000000	0.000000	0.000000	0.161902
193	06075980501	240.195549	399.304688	189.542075	0.612001
194	06075980600	108.103664	253.140225	155.474939	0.823284
195	06075980900	41.886091	109.228813	110.492619	1.384708
196	06075990100	0.000000	0.000000	0.000000	0.000000

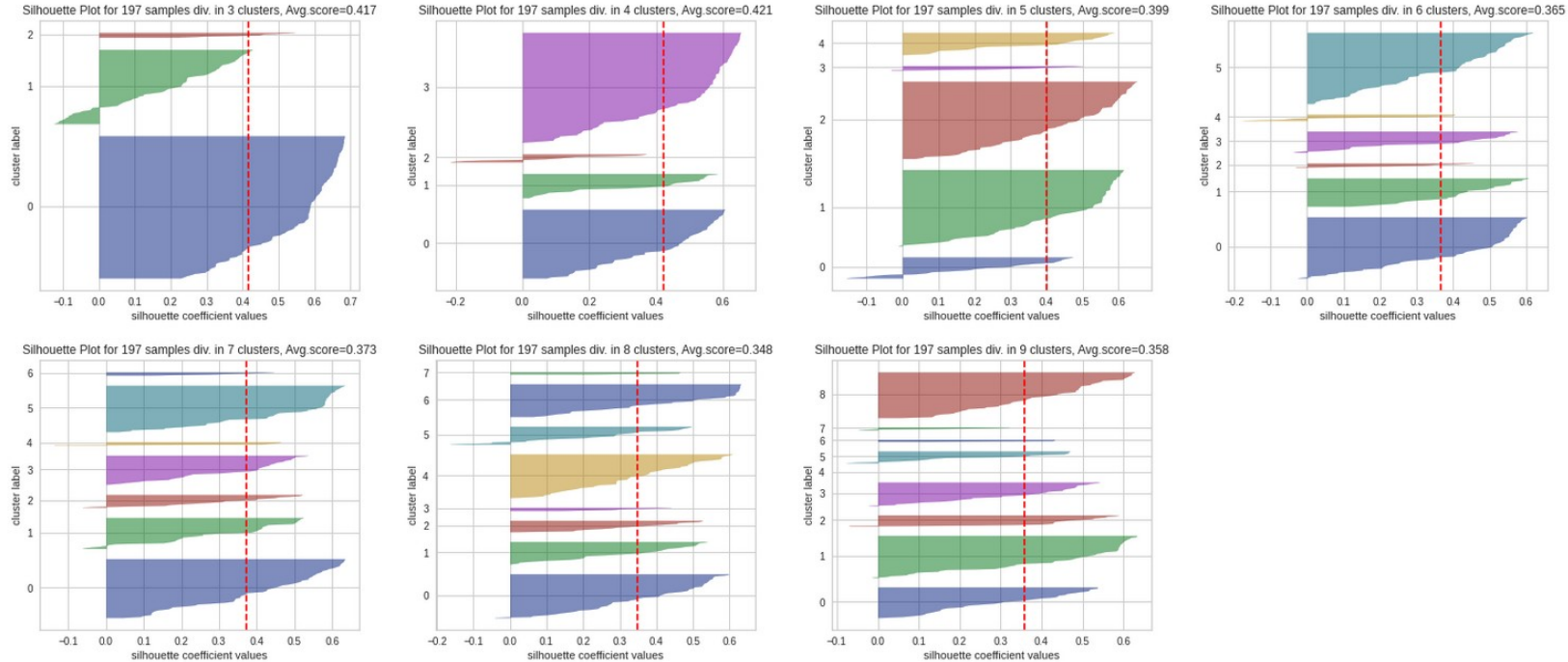
197 rows × 5 columns

	avg_hisp	avg_inc_above_100k	avg_30-49_pop	land_area
count	197.000000	197.000000	197.000000	197.000000
mean	4944.686772	15457.786199	10577.024629	0.237933
std	6420.243833	10152.912164	8075.466483	0.278395
min	0.000000	0.000000	0.000000	0.000000
25%	1450.268496	9008.583807	5873.239622	0.113413
50%	2914.469614	13853.137914	9203.147634	0.158258
75%	5439.174669	20435.035732	13164.907394	0.254182
max	39646.417008	67092.308299	57851.690639	2.358635

IBM Data Science Professional Certificate Capstone Project

Mexican Restaurant Location Analysis

Evaluation of census tracts clustering between the range of 3 and 9 using the Yellowbrick silhouette visualizer:



IBM Data Science Professional Certificate Capstone Project

Mexican Restaurant Location Analysis

After an initial assessment we can see that most likely we can disregard options with 3, 4 and 5 clusters considering the following:

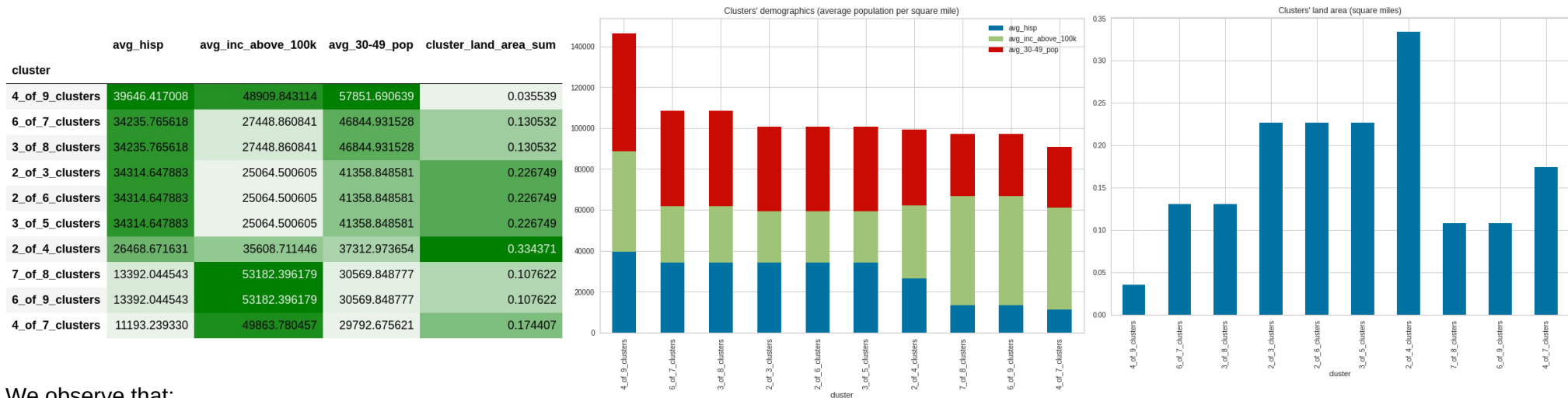
- cluster 1 of the 3 clusters option has a large number of samples with a negative score
- cluster 2 of the 4 clusters option has some number of samples with a negative score and its score is below the average score
- cluster 0 of the 5 clusters option has a significant number of samples with a negative score

Out of the other options the option with 7 clusters looks most promising as it does not have clusters with a significant number of negative scores or clusters with scores below the average score and it has the highest average score out of the options with more clusters than 5.

IBM Data Science Professional Certificate Capstone Project

Mexican Restaurant Location Analysis

Evaluation of the clusters' demographics and land area:



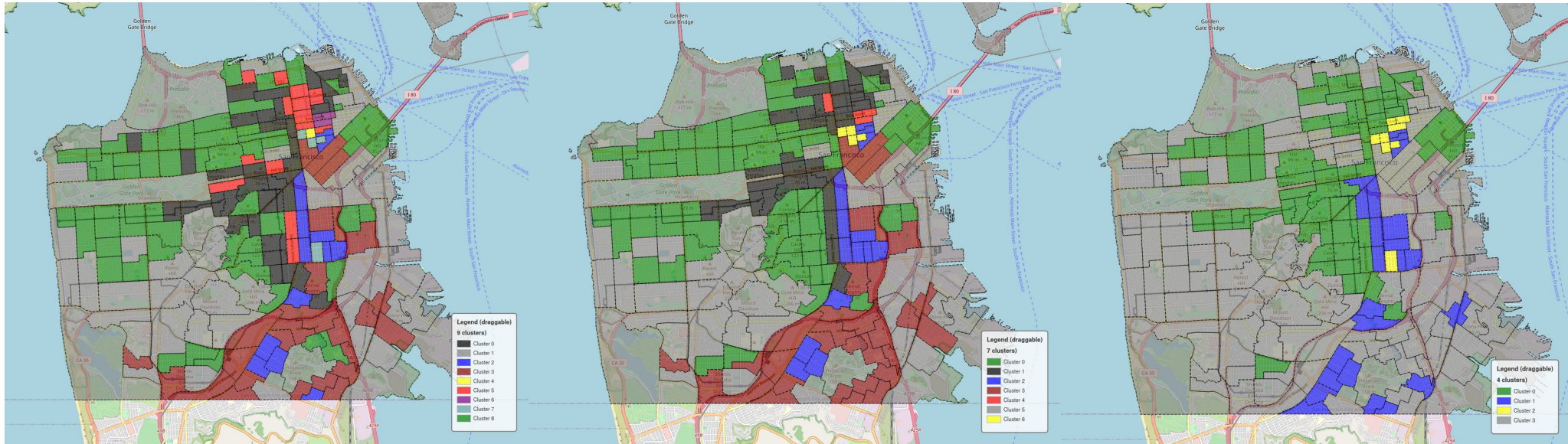
We observe that:

- cluster 4 of 9 has the most beneficial demographics but it has also the smallest area
- cluster 6 of 7 has good demographics particularly the age and ethnicity and significantly larger area then the previous option
- cluster 2 of 4 the largest area but a lower demographics suitability in terms of ethnicity and age of the population

IBM Data Science Professional Certificate Capstone Project

Mexican Restaurant Location Analysis

Visualisation of the selected clusters on the San Francisco county map (preferred clusters in yellow)



9 clusters

7 clusters

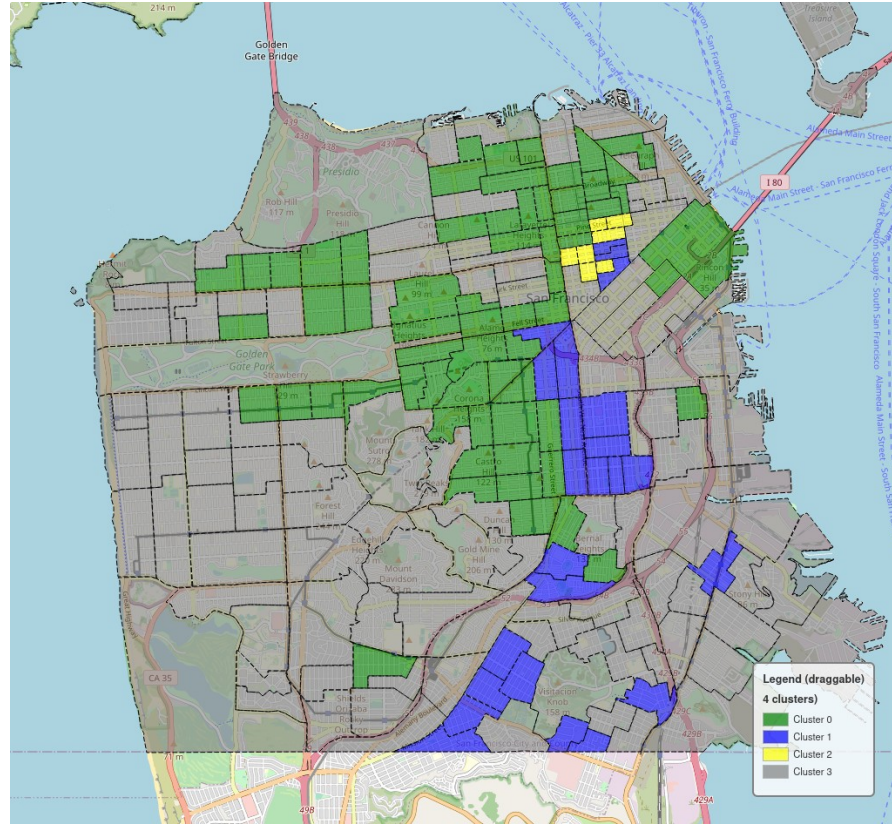
4 clusters

Cluster 2 of 4 incorporates the other 2 selected clusters and includes a tract which not adjacent to the other tracts of this cluster. I removed it from the cluster and analysed the clusters again.

IBM Data Science Professional Certificate Capstone Project

Mexican Restaurant Location Analysis

Modified 4 clusters option
visualised on the San Francisco
county map (preferred
Cluster 2 in yellow)

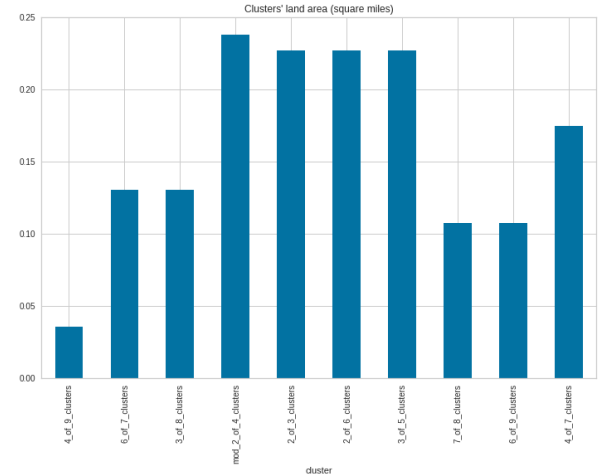
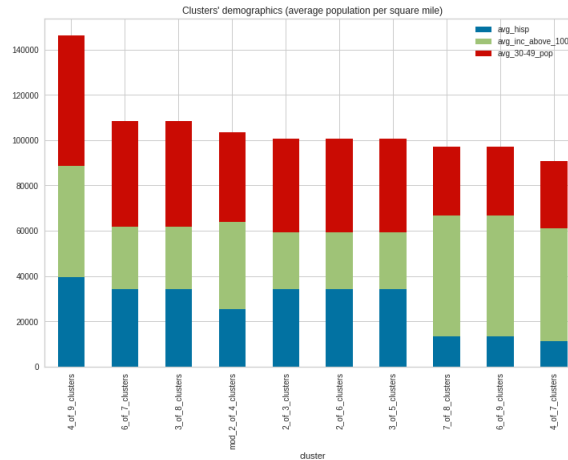


IBM Data Science Professional Certificate Capstone Project

Mexican Restaurant Location Analysis

Evaluation of the clusters' demographics and land area after the modification of cluster 2 of 4:

cluster	avg_hisp	avg_inc_above_100k	avg_30-49_pop	cluster_land_area_sum
4_of_9_clusters	39646.417008	48909.843114	57851.690639	0.035539
6_of_7_clusters	34235.765618	27448.860841	46844.931528	0.130532
3_of_8_clusters	34235.765618	27448.860841	46844.931528	0.130532
mod_2_of_4_clusters	25302.742300	38477.518843	39869.896063	0.238154
2_of_3_clusters	34314.647883	25064.500605	41358.848581	0.226749
2_of_6_clusters	34314.647883	25064.500605	41358.848581	0.226749
3_of_5_clusters	34314.647883	25064.500605	41358.848581	0.226749
7_of_8_clusters	13392.044543	53182.396179	30569.848777	0.107622
6_of_9_clusters	13392.044543	53182.396179	30569.848777	0.107622
4_of_7_clusters	11193.239330	49863.780457	29792.675621	0.174407



The cluster 2 of 4 after modification significantly improved its demographics, become continuous but of course at the cost of its size although still remaining the largest cluster.

The modified cluster 2 of 4 is a good balance between demographics and size of the cluster. It is the largest and third out of the top ten clusters in terms of the demographic characteristics. It contains both 4 of 9 and 6 of 7 clusters.

IBM Data Science Professional Certificate Capstone Project

Mexican Restaurant Location Analysis

Sample of restaurants data and the Mexican restaurants, taco and burrito places list in cluster 2 of 4 and its surroundings downloaded from the Foursquare API.

																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					</
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

215 rows × 10 columns

IBM Data Science Professional Certificate Capstone Project

Mexican Restaurant Location Analysis

Visualisation of restaurants in cluster 2 of 4 and its vicinity on the map.

Conclusion

I propose to locate the Mexican restaurant in area highlighted with the red circle next to the cluster 4 of 9 (marked with the yellow and gray hatch) which is the best in terms of demographic data. This location is within not a far distance away from the rest of the cluster and far enough from the direct competitors.

