Author: Włodzimierz Kuczyński

April 2021

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)

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

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,

#### **Mexican Restaurant Location Analysis**

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

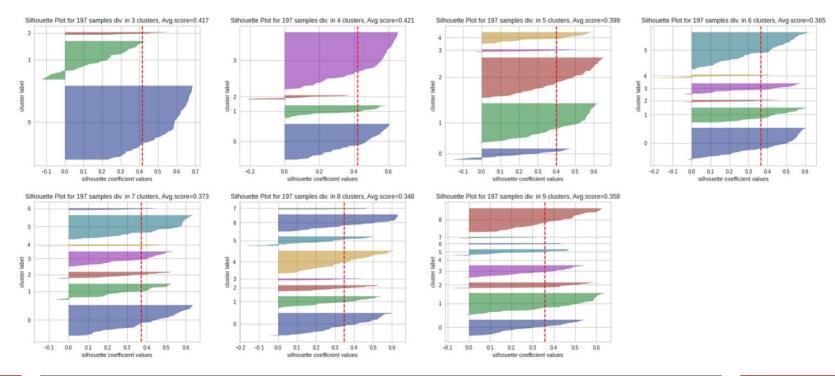
	trctid	avg_hisp	avg_inc_above_100k	avg_30-49_pop land_ai				
0	06075010100	927.280282	5693.425998	5220.121250	0.299801			
1	06075010200	673.152074	15463.014831	8027.589366	0.199064			
2	06075010300	3910.093326	21586.323342	16364.464002	0.103578			
3	06075010400	4204.300087	24951.551587	15598.339407	0.129629			
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			

	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

197 rows × 5 columns

#### **Mexican Restaurant Location Analysis**

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



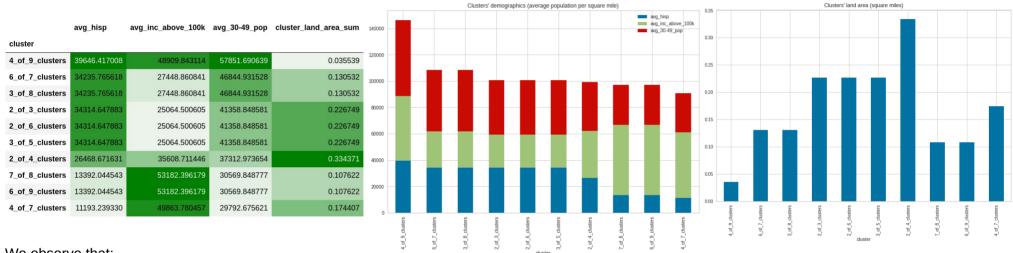
After an initial assessment we can see that most likely we can disregard options with 3, 4 and 5 clusters considering the following:

- 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.

#### **Mexican Restaurant Location Analysis**

### Evaluation of the clusters' demographics and land area:

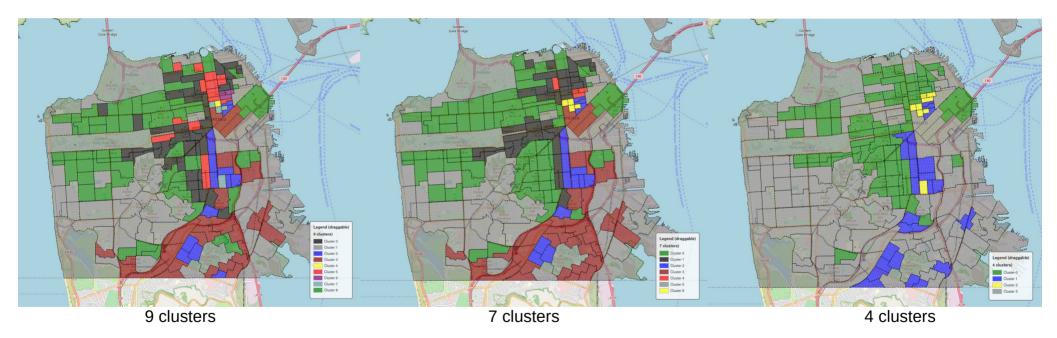


#### We observe that:

- 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

**Mexican Restaurant Location Analysis** 

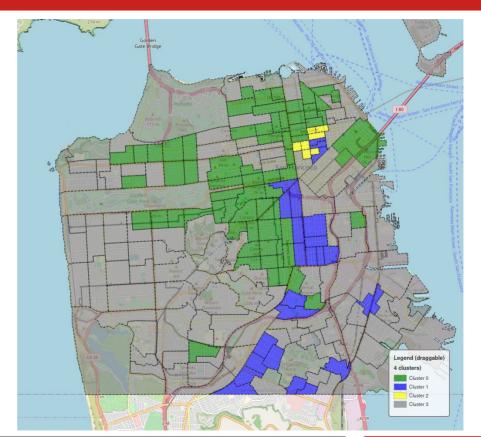
Visualisation of the selected clusters on the San Francisco county map (preferred clusters in yellow)



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.

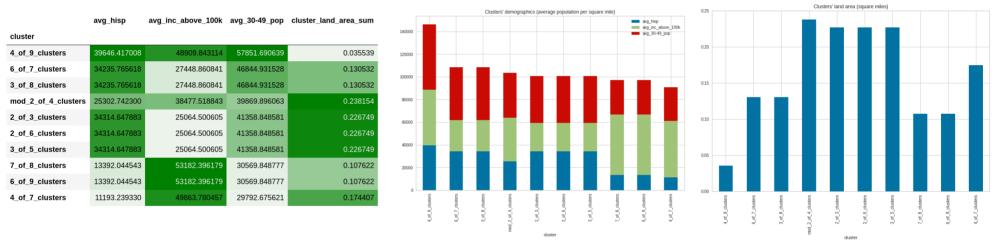
**Mexican Restaurant Location Analysis** 

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



#### **Mexican Restaurant Location Analysis**

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



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.

**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.

						inc	dex trctid	Tract Latitude	Tract Longitude	Venue Id	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Category ID	color
trctid Tract Latitude Tra	ct Longitude Venue Id	Venue Venue	Latitude Venue Longitude	Venue Category	Venue Category ID color	0	4 06075011901	37.790473	-122.413914	59cb288d6c08d172265a8e6d	Chisme	37.788467	-122.414802	Mexican Restaurant	4bf58dd8d48988d1c1941735	red
<b>0</b> 06075011901 37.790473			.789594 -122.413873		4bf58dd8d48988d110941735 blue	1	51 06075011901	37.790473	-122.413914	55ee12f4498e5acd582cd338	El Rincón Yucateco	37.785824	-122.412842	Mexican Restaurant	4bf58dd8d48988d1c1941735	red
<b>1</b> 06075011901 37.790473 <b>2</b> 06075011901 37.790473	-122.413914 44f72e06f964a5204b381fe3 -122.413914 54c2a5e4498ee52a438800f7	· ·	.791490 -122.412393 .788332 -122.414605		4bf58dd8d48988d14e941735 blue 52e81612bcbc57f1066b79fe blue	2	60 06075011901	37.790473	-122.413914	49fa3c25f964a520dc6d1fe3		37.791985	-122.421047	Mexican Restaurant	4bf58dd8d48988d1c1941735	red
<b>3</b> 06075011901 37.790473	-122.413914 5643ce76498ec3c226f039e3		.789807 -122.411347		4bf58dd8d48988d1ca941735 blue	3 1	144 06075012000	37.787965	-122.418527	56b28aa9498ec9031e9f4b0a	Grill Matador	37.788898	-122.411570		4bf58dd8d48988d151941735	purple
4 06075011901 37.790473	-122.413914 59cb288d6c08d172265a8e6d	Chisme 37	122.414802 	Mexican Restaurant	4bf58dd8d48988d1c1941735 red	4 1	155 06075012100	37.788807	-122.411888	4a846e6cf964a52094fc1fe3	Taqueria Castillo B2	37.783778			4bf58dd8d48988d153941735	
<b>210</b> 06075012401 37.783051 <b>211</b> 06075012401 37.783051	-122.415789 4a2c0803f964a5200e971fe3 -122.415789 5c40381f603d2a002c295ff9		7.778375 -122.415426 7.781060 -122.408548		4bf58dd8d48988d147941735 blue 4bf58dd8d48988d1db931735 blue	5 1	164 06075012100	37.788807	-122.411888	4a7b2e4ff964a52037ea1fe3	Taqueria La	37.783328	-122.413703	Taco Place	4bf58dd8d48988d151941735	purple
<b>212</b> 06075012401 37.783051	-122.415789 59a24aa726659b0902fb4d04	•	7.786124 -122.410247		4bf58dd8d48988d1d2941735 blue	6 1	169 06075012201	37.785846	-122.416353	51b28827498ef0351235c618	El Castillito		-122.414737	Mexican	4bf58dd8d48988d1c1941735	red
<b>213</b> 06075012502 37.783933 <b>214</b> 06075012502 37.783933	-122.412595 4dfd856d483b96a3aaa9eb34 -122.412595 4af09c87f964a520a1dd21e3		7.781158 -122.406243 7.780840 -122.405770		4bf58dd8d48988d14e941735 blue 4bf58dd8d48988d14a941735 blue	7 1	172 06075012201	37.785846	-122.416353	49e4f160f964a52074631fe3	Taqueria Cancun	37.781875	-122.410322		4bf58dd8d48988d153941735	
215 rows × 10 columns						8 1	185 06075012201	37.785846	-122.416353	4550e681f964a520e43c1fe3		37.787109	-122.410533	Mexican	4bf58dd8d48988d1c1941735	red
						9 1	190 06075012202	37.785412	-122.419644	4b524b07f964a5205f7527e3	Terresia El	37.781569	-122.416897		4bf58dd8d48988d153941735	
						10 1	191 06075012202	37.785412	-122.419644	568ed7e8498ec5e165503bbe	Taqueria Catillito	37.781752	-122.416873	Taco Place	4bf58dd8d48988d151941735	purple

**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.

