Location Analysis for New Chinese Restaurant in Barcelona

Hans Kristanto

27th December 2020

1. Introduction

a. About Barcelona

Barcelona is a coastal city located in the southern part of Spain. The city is famous due to its location that sits between seas and mountains, sunny during summer and moderate temperature during winter, food, and its relatively lower cost of living compared to other neighboring countries. These are just a few reasons why Barcelona is one of the tourist destinations in western Europe that attracts millions of visitors every single year.

Being the hottest travel destination in Europe has also impacted Barcelona revenues. According to research, approximately 22 million euros are generated every day from the tourism industry. There is no wonder that the tourism sector will only grow even more in years to come.

b. Business Problem

A businessman named Mr. X saw the opportunity of Barcelona being one of the top tourist destinations, and he would like to open a new Chinese restaurant in the city. The problem here is that he has no idea where he should open it. He would also want to gain maximum potential revenue from this restaurant.

There are couple of points that we know about his plan:
1. He has enough capital to open 1 Chinese restaurant.
2. He believes that due to Barcelona being a place of culture, there will not be many Chinese restaurants available, making the opportunity to be more attractive.
3. Rental fee is not his main concern, and he is willing to pay a premium rental fee, as long as the restaurant is strategically located.
4. He is not interested in competing with other Chinese restaurants. He is willing to compete with other local non-Chinese restaurants.

c. Success Criteria

The success criteria for this project are:
1. List of Barcelona's Neighborhood that he can consider in opening a new Chinese restaurant.
2. Those locations need to pose a maximum opportunity for the restaurant to succeed.

2. Data Acquisition

a. Districts of Barcelona (Wikipedia)

In order to recommend which neighborhood will be best suit for Mr. X to open his Chinese restaurant, we need to understand all districts, along with its neighborhood (for simplification, we will use the terminology of *Barrio* to refer to neighborhood from now on). The information of such districts can be found in this link. Features that are useful for the project:

1. District
2. Barrio within each Barcelona district

b. Venues Information (Foursquare API)

Foursquare is the most trusted, independent location data platform for understanding how people move through the real world. The information will be useful to understand what and how the distribution of place of interest looks like in Barcelona. The information will also be useful to understand what kind of Restaurants are commonly known, along with its frequency. Some of the features that are useful for the project:

1. Venue name
2. Category of venue
3. Geolocation of the venue

c. Nominatim (OpenStreetMap)

OpenStreetMap is a free world map provider. We will be utilizing Nominatim, a tool provided by OpenStreetMap to search geolocation information of a certain point on the map. In this project, we will also be utilizing Geopy, a python library that supports calling nominatim API to get the information we need. Feature that is useful for the project:

1. Geolocation of each Barcelona's Barrio.

3. Methodology

a. Hypothesis

From the “Business Problem” section above, we understand what Mr. X wants to achieve his new restaurant, as well as his level of competitiveness against other restaurants. He also wants to maximize the revenue the new restaurant can generate.
With these information, we can assume that Mr. X’s restaurant will be successful if it satisfies all of these categories:

1. The restaurant is established in a Barrio that has less to no Chinese restaurant in it.
2. The restaurant is established in a Barrio that has a reasonable number of attractions, so that there is a pulling force for tourists and visitors to come.
3. The area in which the restaurant is established should have a market segment that prefers non-Spanish food.

b. Approach

Based on the constructed hypothesis, we now have a better idea on how we should solve the problem. Below are the step-by-step summary on what we will do:

1. Find the current distribution of Chinese restaurants in Barcelona. The information found in this step will confirm Mr. X believes that there are not many Chinese restaurants in the city, as well as marking Barrios that already have Chinese restaurants.
2. Separate Barcelona city into several clusters, based on the characteristics of top restaurants that present in each Barrios. To get this job done, we will be implementing a K-Means clustering algorithm to restaurants dataset that are extracted from foursquare. The result of this step can help us understand the preferred type of food for each of Barrios, and to help Mr. X gauge how high or low the competition that he will face in particular areas.
3. The final step would be to map places of interest from Foursquare into the map, and perform final filtration of Barrios that will be a great potential for Mr. X to open the restaurant.

c. Exploratory Data Analysis

We will first extract a list of districts and Barrios in Barcelona, utilizing Python and BeautifulSoup library:

```
wikipedia_url = 'https://en.wikipedia.org/wiki/Districts_of_Barcelona'
response = requests.get(wikipedia_url)
soup = BeautifulSoup(response.text, 'html.parser')

districts_table = soup.find('table', class_='wikitable')
columns = ['District', 'Barrio']
barrios_list = []

for row in districts_table.findAll('tr'):
    cells = row.findAll('td')

    if cells:
        district = cells[1].find(text=True).rstrip()
        barrios = cells[5].findAll(text=True)
        barrios = [b for b in barrios if b not in (', ', '\n', '.\n', '*\n', '*, ', ' i ')]

        for barrio in barrios:
            barrios_list.append([district, barrio])

barcelona_barrios = pd.DataFrame(barrios_list, columns=columns)
barcelona_barrios.head()
```

The first 5 elements of district barrio mapping can be found in below dataframe:

	District	Barrio
0	Ciutat Vella	La Barceloneta
1	Ciutat Vella	El Gòtic
2	Ciutat Vella	El Raval
3	Ciutat Vella	Sant Pere, Santa Caterina i la Ribera
4	Eixample	L'Antiga Esquerra de l'Eixample

Once we have completed the extraction, we need to know the geographical location for each Barrio. To do this, we will be using Geopy library that is utilizing Nominatim tools:

```
import time

geolocator = Nominatim(user_agent='bcn_explr')
lat_lng = []

for i, row in barcelona_barrios.iterrows():
    barrio = row['Barrio']
    location = geolocator.geocode(barrio)

    if location:
        lat_lng.append([location.latitude, location.longitude])
    else:
        lat_lng.append([np.nan, np.nan])
        print('Can\'t find lat long for barrio {}'.format(barrio))

    # Avoid rate limiting
    time.sleep(0.5)

barcelona_barrios[['Latitude', 'Longitude']] = lat_lng
barcelona_barrios
```

The result will be a complete dataframe that has information about district, barrio, its latitude and longitude, as shown below:

	District	Barrio	Latitude	Longitude
0	Ciutat Vella	La Barceloneta	41.380653	2.189927
1	Ciutat Vella	El Gòtic	41.383395	2.176912
2	Ciutat Vella	El Raval	41.379518	2.168368
3	Ciutat Vella	Sant Pere, Santa Caterina i la Ribera	41.388322	2.177411
4	Eixample	L'Antiga Esquerra de l'Eixample	41.390000	2.155000
...	...	...	...	...
70	Sant Martí	Poblenou	41.400527	2.201729
71	Sant Martí	Provençals del Poblenou	41.411948	2.204125
72	Sant Martí	Sant Martí de Provençals	41.416519	2.198968
73	Sant Martí	La Verneda i la Pau	41.423220	2.202940
74	Sant Martí	la Vila Olímpica del Poblenou	41.389868	2.196846

75 rows × 4 columns

From the above, we can see that 1 district will have more than 1 barrio. However, the information of the district is not useful for this project, and information about barrio will carry more importance, as the restaurant needs to be established in a specific area and place, so we will drop this information from future usage.

Next step, we will try to confirm Mr. X theory that there are not many Chinese restaurants present in Barcelona. To do this, we need to first extract a list of venues from Foursquare from the given barrio above. The result of such extraction can be found below:

	Barrio	Barrio Latitude	Barrio Longitude	Venue	Venue Latitude	Venue Longitude	Category
0	La Barceloneta	41.380653	2.189927	Baluard Barceloneta	41.380047	2.189250	Bakery
1	La Barceloneta	41.380653	2.189927	BRO	41.380214	2.189007	Burger Joint
2	La Barceloneta	41.380653	2.189927	La Cova Fumada	41.379254	2.189254	Tapas Restaurant
3	La Barceloneta	41.380653	2.189927	Plaça de la Barceloneta	41.379739	2.188135	Plaza
4	La Barceloneta	41.380653	2.189927	Somorrostro	41.379156	2.189100	Spanish Restaurant
...	...	...	...	...	...	...	...
2981	la Vila Olímpica del Poblenou	41.389868	2.196846	Restaurant Tapas Locas	41.388650	2.199989	Tapas Restaurant
2982	la Vila Olímpica del Poblenou	41.389868	2.196846	Butyklan	41.391394	2.199898	Sandwich Place
2983	la Vila Olímpica del Poblenou	41.389868	2.196846	Divan	41.388108	2.199298	Turkish Restaurant
2984	la Vila Olímpica del Poblenou	41.389868	2.196846	Hot Music	41.388482	2.199763	Hookah Bar
2985	la Vila Olímpica del Poblenou	41.389868	2.196846	METRO Ciutadella | Vila Olímpica	41.388320	2.193770	Metro Station

2986 rows × 7 columns

There are about 2986 venues found from this extraction, ranging from restaurant, cafe, park, etc. There are just too many information here, so let's see what is the top 5 venues, as well as how many categories found:

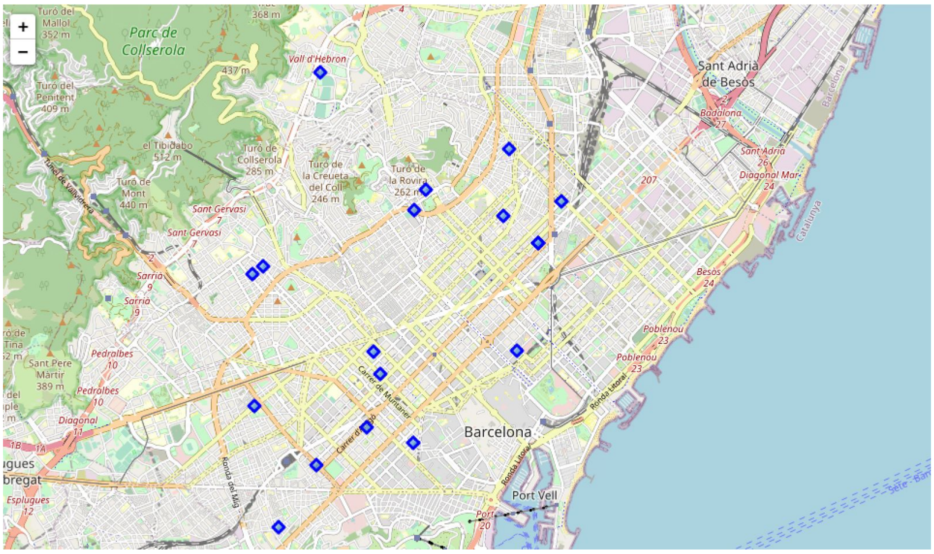
Category	Barrio
Spanish Restaurant	159
Tapas Restaurant	146
Hotel	123
Mediterranean Restaurant	122
Café	121
...	...
Russian Restaurant	1
General Travel	1
General College & University	1
Construction & Landscaping	1
Print Shop	1

280 rows × 1 columns

Spanish restaurants dominate the venues with 159 places, followed by Tapas restaurant. Hotel comes on the third with 123 places, Mediterranean restaurant on the fourth with 122 places, and Cafe with 121 places. So far, Mr. X's hypothesis seems to be correct, at least Chinese restaurants do not appear on the top 5 venues from this list. Let's see how many Chinese restaurants are there in Barcelona:

Category	Barrio
Chinese Restaurant	19

There are only 19 Chinese restaurants in Barcelona. Let's see how the distribution of such restaurants look like:



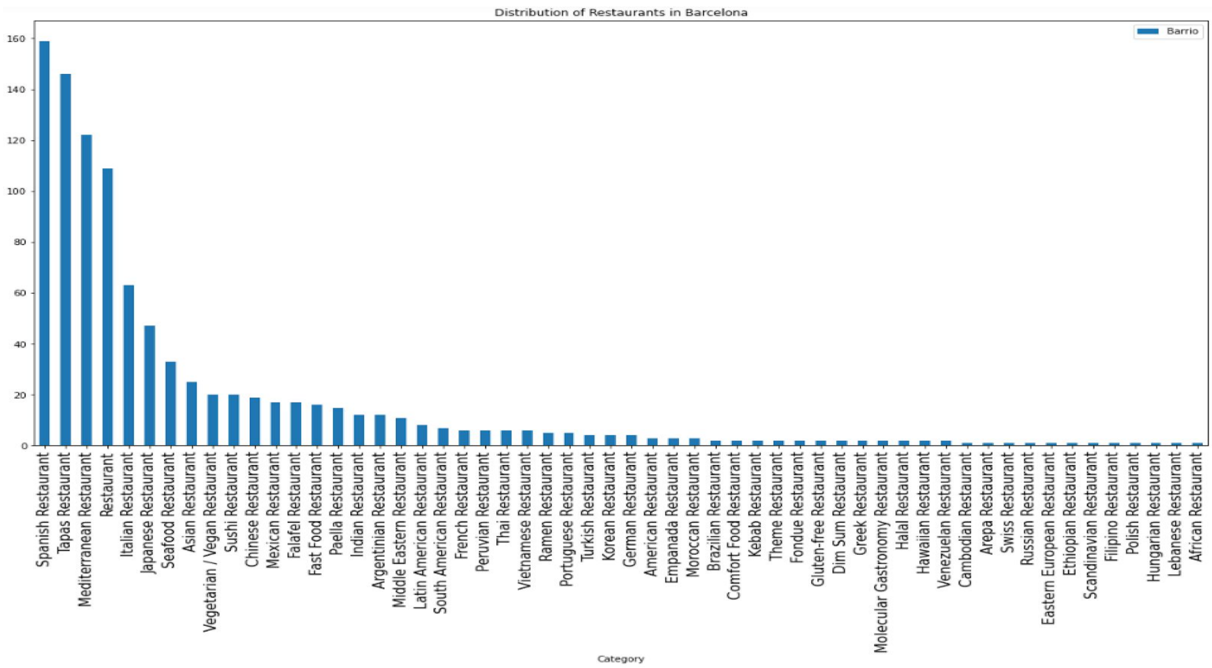
It looks Chinese restaurants are distributed quite evenly throughout the south-western and center part of the city, with 1 restaurant appearing in the northern part of the city.

Now we have a better picture on how current prospects of Chinese restaurants look from the location and distribution. Although there are a couple of areas that do not have a presence of Chinese restaurants, the biggest question is, which shall we choose? Before we try to answer that question, let's prepare a dataframe that contains all restaurants (irregardless of type) of Barcelona.

	Barrio	Barrio Latitude	Barrio Longitude	Venue	Venue Latitude	Venue Longitude	Category
0	La Barceloneta	41.380653	2.189927	La Cova Fumada	41.379254	2.189254	Tapas Restaurant
1	La Barceloneta	41.380653	2.189927	Bitácora	41.382070	2.187608	Tapas Restaurant
2	La Barceloneta	41.380653	2.189927	La Bombeta	41.380521	2.187573	Tapas Restaurant
3	La Barceloneta	41.380653	2.189927	Jai-Ca	41.381501	2.187858	Tapas Restaurant
4	La Barceloneta	41.380653	2.189927	el bar del Basko	41.379414	2.190501	Tapas Restaurant
...	...	...	...	...	...	...	...
964	el Parc i la Llacuna del Poblenou	41.400733	2.191342	El Quinto	41.400180	2.196257	Gluten-free Restaurant
965	Poblenou	41.400527	2.201729	El Quinto	41.400180	2.196257	Gluten-free Restaurant
966	Poblenou	41.400527	2.201729	Zozan	41.398316	2.205121	Kebab Restaurant
967	Provençals del Poblenou	41.411948	2.204125	Bar Shawarma i Döner	41.408253	2.204866	Kebab Restaurant
968	Poblenou	41.400527	2.201729	L'autèntic	41.401814	2.200012	Lebanese Restaurant

969 rows × 7 columns

There are about 969 restaurants found in total. Putting it into a perspective, this constitutes 32% of total venues in Barcelona, indicating the demand of restaurants is high. At this point, it is even more convincing that opening a new restaurant may be a good move. In the chart below, you can see how the distribution of restaurant look like, grouped by its food type:



As we already know, Spanish, Tapas and Mediterranean restaurants sit on the top 3 of restaurant types. Generic type of restaurant is sitting in the 4th position. Japanese restaurant seems to be the only Asian food restaurant that has quite a number of places in Barcelona.

We have sufficient information about the distribution of restaurants in Barcelona and we can now move on to the next step, which is to cluster barrios based on characteristics of restaurants.

d. Barrios Clustering

To know the characteristics of each Barrio, we will need to use the K-Means clustering algorithm to get the information we need. We have already extracted all the restaurants in Barcelona. Before we can do that, we need to have a dataframe that is clean and can be processed by K-Means algorithm, which basically includes:

1. Display all possible type of restaurants as a column
2. Score the ratio of each particular restaurants for each barrio
3. Temporarily drop the barrio column, as this information does not have any relevance for the algorithm.

To do this, we can execute below python code:

```
# 1. Convert unique category into column (Onehot)
restaurant_onehot = pd.get_dummies(barcelona_restaurants[['Category']], prefix='', prefix_sep='')
restaurant_onehot['Barrio'] = barcelona_restaurants['Barrio']
columns = [restaurant_onehot.columns[-1]] + list(restaurant_onehot.columns[:-1])
restaurant_onehot = restaurant_onehot[columns]

# 2. Score each of restaurant type for each of Barrio
restaurant_onehot = restaurant_onehot.groupby('Barrio').mean().reset_index()

# 3. Drop the 'Barrio' column
restaurant_onehot.drop('Barrio', axis=1)
```

The result of above operation is this dataframe:

	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Brazilian Restaurant	Cambodian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	...	Spanish Restaurant	Sushi Restaurant	R
0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.000000	
1	0.0	0.000000	0.0	0.0	0.000000	0.0	0.1	0.200000	0.0	0.0	...	0.400000	0.000000	
2	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.000000	
3	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.000000	
4	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.000000	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
62	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.000000	
63	0.0	0.027027	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.081081	0.027027	
64	0.0	0.000000	0.0	0.0	0.038462	0.0	0.0	0.038462	0.0	0.0	...	0.269231	0.076923	
65	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.000000	
66	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.181818	0.090909	

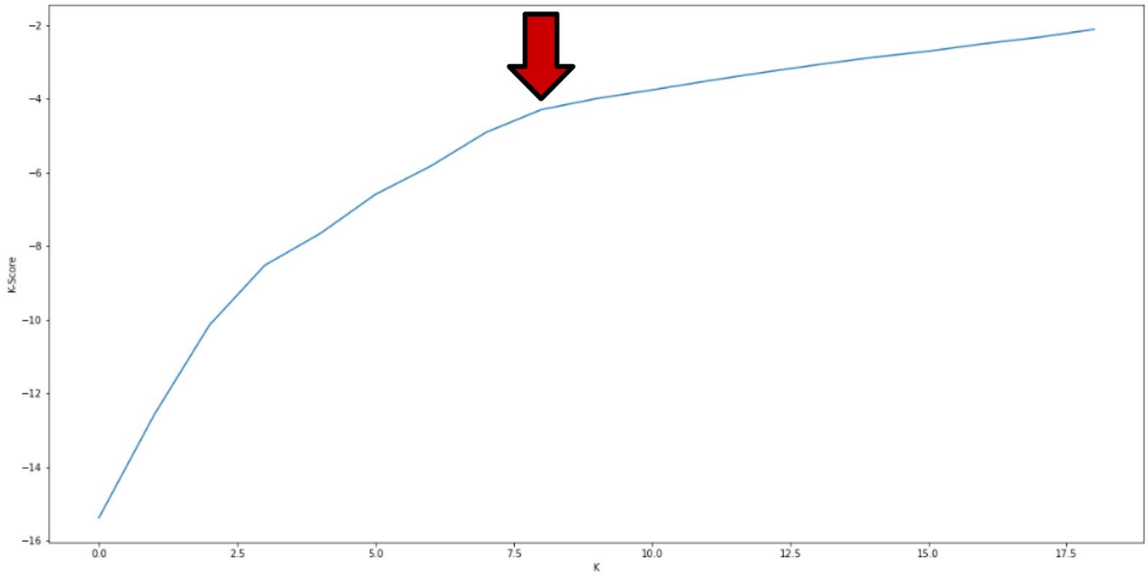
67 rows × 56 columns

There are 57 columns in this dataframe, and it is possible to display all the columns in this document. As you can see, there are some numbers in this dataframe, indicating the ratio of such restaurants to appear in a region. With this information, we can use the K-Means algorithm to cluster areas that have some similarity.

We are almost ready to start clustering the data. 1 challenge that we need to overcome next is to identify the number of clusters. The problem with our use case is that there is no logical number that we can use. Therefore we need to utilize the *elbow method* to identify the best number of clusters based on the dataset that we have. What we are going to do next is to score our clustering models from 1 to 20 clusters. Using this code below:

```
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

scores = []

restaurant_clustering = restaurant_onehot.drop('Barrio', axis=1)

for k in range(1, 20):
    scores.append(KMeans(n_clusters=k, random_state=1).fit(restaurant_clustering).score(restaurant_clustering))

plt.plot(scores)
plt.ylabel('K-Score')
plt.xlabel('K')
plt.show()
```

We can see the movement of score in the chart below:



The *elbow method* describes the best number of clusters found when there are no significant changes of the score observed. From the chart above, we can see that there are about 10 points move (~ -15 to -4) when the number of clusters move from 0 to 8, and 2 points move (-4 to -2) when the number of clusters move from 8 to 20. Based on the elbow *method*, this is the best number of clusters we can use.

Let’s build our model using this code:

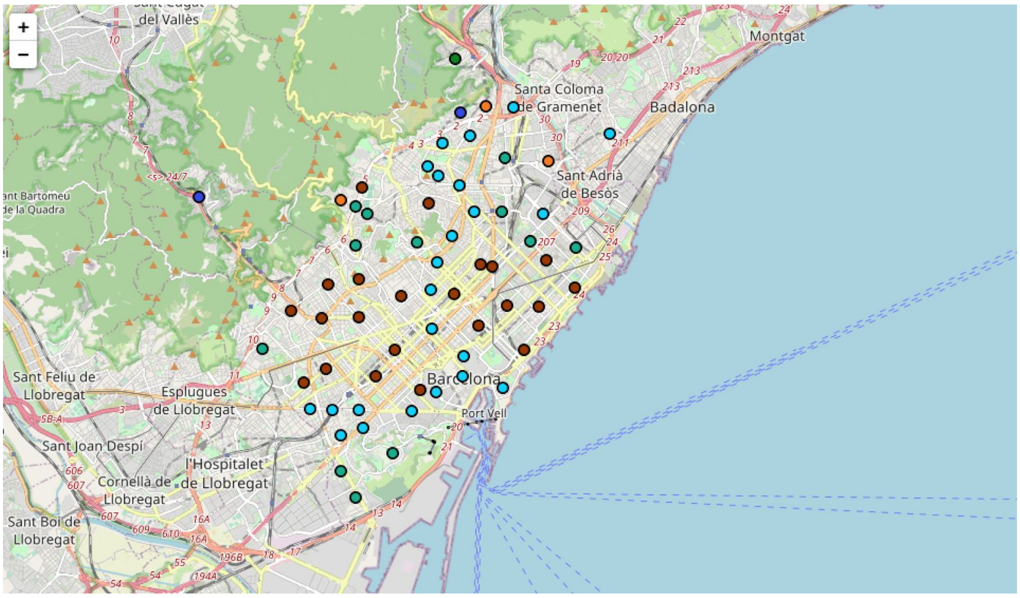
```
kclusters = 8
n = restaurant_onehot['Barrio']
barrio = restaurant_onehot[['Barrio']]
restaurant_clustering = restaurant_onehot.drop('Barrio', 1)
kmeans = KMeans(n_clusters=kclusters, random_state=1).fit(restaurant_clustering)
clusters_label = pd.DataFrame([x for x in kmeans.labels_], columns=['Cluster'])

barrio_cluster = pd.concat([clusters_label, barrio], axis=1)
barrio_cluster.groupby('Cluster').count()
```

The result of such clustering is:

Cluster	0	1	2	3	4	5	6	7
Barrio	22	3	3	1	12	2	1	23

It does not really give us good information on what this information entails, but 1 thing for sure is that the data are heavily distributed within 3 clusters, namely cluster-0, cluster-4, and cluster-7. If we merge this information with the original barrio dataset and plot the visualization over the map, this is what we get. As reference, these are the color code that I used to visualize the map:

Cluster-0	**Cluster-1**	**Cluster-2**	**Cluster-3**
Cluster-4	**Cluster-5**	**Cluster-6**	**Cluster-7**



From the visualization above, cluster-0 and cluster-7 are dominating the majority of Barcelona city center, pushing away other clusters to the side.

4. Result and Recommendations

We now have a good vision on how the clustering looks like, as the outcome of the K-Means algorithm. Let us now see what are the characteristics of each cluster, by analyzing the top restaurant category on each cluster, as well as giving a preliminary label whether the restaurant should be opened in any of the barrio within the cluster:

		Barrio
Cluster	**Category**	
0	**Spanish Restaurant**	1
	Sushi Restaurant	1
	Tapas Restaurant	8
	Thai Restaurant	1
	Turkish Restaurant	1
	Vegetarian / Vegan Restaurant	6
	Venezuelan Restaurant	1
	Vietnamese Restaurant	3
1	**Tapas Restaurant**	3
2	**Mediterranean Restaurant**	3
3	**Portuguese Restaurant**	1
4	**Spanish Restaurant**	7
	Tapas Restaurant	4
	Thai Restaurant	1
5	**Restaurant**	2
6	**French Restaurant**	1
7	**Tapas Restaurant**	15
	Theme Restaurant	1
	Turkish Restaurant	1
	Vegetarian / Vegan Restaurant	3
	Vietnamese Restaurant	3

Cluster-0

Barrios in this cluster have Tapas and Vegetarian listed as the top restaurant, with a small number of Vietnamese and other restaurants. There seems to be quite a variety of restaurants here, making it a good candidate for recommendations.

Level of recommendations: <u>**MEDIUM**</u>

Cluster-1

The top restaurant in this cluster is Tapas and the barrio that is in this cluster seems to be located quite far from Barcelona city center.

Level of recommendations: <u>**LOW**</u>

Cluster-2

The top restaurant in this cluster is Mediterranean. Like cluster-1, barrios in this cluster seems to be located far from Barcelona city center.

Level of recommendations: <u>**LOW**</u>

Cluster-3

There is only 1 top restaurant here which is Portuguese restaurant and it is located even further away from the city center.

Level of recommendations: <u>**NOT RECOMMENDED**</u>

Cluster-4

There are various types of restaurants here, but Spanish and Tapas restaurants dominate this cluster. Although there are a couple of barrios in this cluster, they are located quite on the side of the city.

Level of recommendations: <u>**MEDIUM**</u>

	Cluster-5 Barrios in this cluster are located very far from the city center. Level of recommendations: **<u>NOT RECOMMENDED</u>**
	Cluster-6 Barrios in this cluster are located very far from the city center. Level of recommendations: **<u>NOT RECOMMENDED</u>**
	Cluster-7 There seems to be a clear winner here. Tapas restaurants are dominating this cluster with 15 out of 23. It may not be a good idea to compete against local Spanish restaurants. Level of recommendations: **<u>NOT RECOMMENDED</u>**

From this table, we can now focus our attention to cluster-0, cluster-1, cluster-2, and cluster-4.

Our 2nd hypothesis says that a new restaurant will be likely to succeed if it is located within reach from attractions. Let's prepare a dataframe that contains a list of attractions using below code:

```
potential_attractions = ['Bookstore','Boat or Ferry','Beach',\
                         'Bridge', 'Hotel','Plaza','Museum','Castle',\
                         'Station','University','Mall',\
                         'Stadium','Park','Monument','Theater','Gallery'\
                         'Boutique', 'Circus', 'Cultural Center', 'Department Store'\
                         'Electronics Store', 'Escape Room', 'Historic Site', \
                         'Hot Spring', 'Music Venue', 'Multiplex', 'Nightclub', \
                         'Wine']
barcelona_attractions = barcelona_full_venues[barcelona_full_venues['Category'].\
                                              str.contains('|'.join(potential_attractions))].reset_index()
barcelona_attractions
```

Note: the list of attractions above is subjective and based on my own personal experience on what tourists would likely to visit if they come to Barcelona.

From the code above, there are about 471 places found:

	index	Barrio	Barrio Latitude	Barrio Longitude	Venue	Venue Latitude	Venue Longitude	Category
0	3	La Barceloneta	41.380653	2.189927	Plaça de la Barceloneta	41.379739	2.188135	Plaza
1	9	La Barceloneta	41.380653	2.189927	Vinoteca Voramar	41.380726	2.188315	Wine Shop
2	12	La Barceloneta	41.380653	2.189927	Platja de la Barceloneta (Playa de la Barcelon...	41.379610	2.193038	Beach
3	28	La Barceloneta	41.380653	2.189927	Passeig Marítim de la Barceloneta	41.380125	2.192824	Beach
4	54	La Barceloneta	41.380653	2.189927	Museu d'Història de Catalunya	41.380684	2.186023	History Museum
...	...	...	...	...	...	...	...	...
466	2966	la Vila Olímpica del Poblenou	41.389868	2.196846	Zich	41.387227	2.198070	Nightclub
467	2967	la Vila Olímpica del Poblenou	41.389868	2.196846	TRAM Ciutadella | Vila Olímpica	41.387483	2.192041	Tram Station
468	2971	la Vila Olímpica del Poblenou	41.389868	2.196846	Mokaï Beach Bar	41.390832	2.202319	Beach Bar
469	2980	la Vila Olímpica del Poblenou	41.389868	2.196846	Amalur	41.390301	2.201289	Beach Bar
470	2985	la Vila Olímpica del Poblenou	41.389868	2.196846	METRO Ciutadella | Vila Olímpica	41.388320	2.193770	Metro Station

471 rows × 8 columns

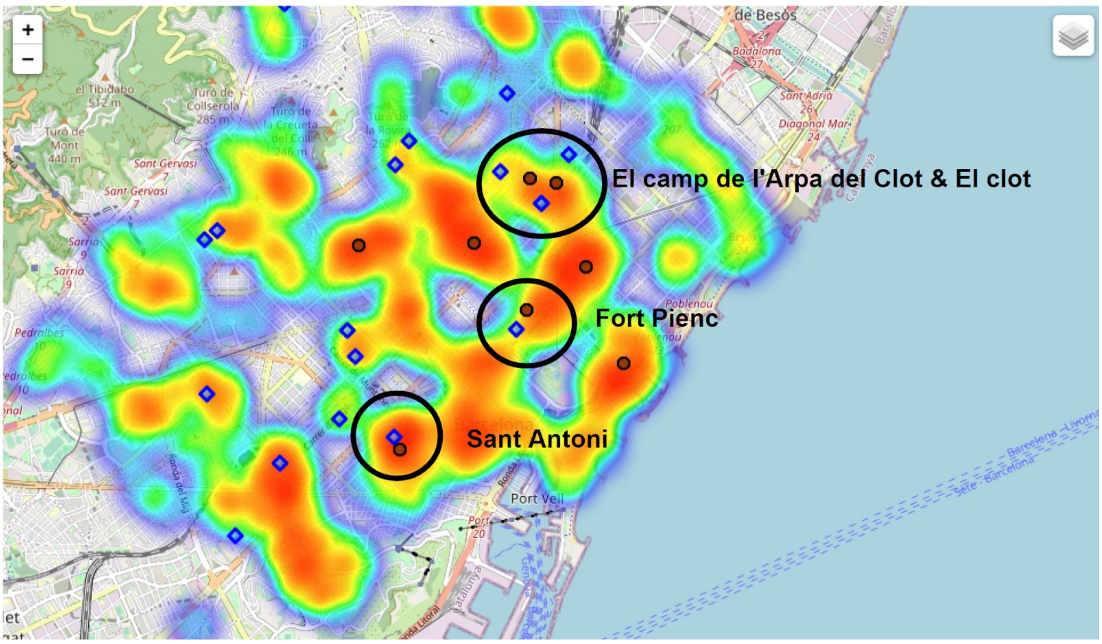
We need to do the same exercise of visualizing this information. However, it is not going to be ideal if we use point markers in the map. The idea of this exercise is to see how dense the attractions are within a certain area, and to accomplish this goal, we will be using heatmap instead. From the above dataframe, we will get below map:

It is expected that attractions are concentrated in the city center of Barcelona, with some places detected in the south-eastern part of the city. There are not many attractions found on the west, north and the north-eastern part of the city, which, according to our 2nd hypothesis, is not an ideal area to open a new restaurant.

We have now plotted 2 maps, cluster map and attractions heat map. Let's combine both of the map and see what we get:

As expected, cluster-0 and cluster-7 (2 of the biggest clusters) are dominating areas that have a wider coverage of heatmap. However, we agreed that cluster-7 will not be part of the recommendation, because of the domination of Spanish restaurants. Let us now focus on cluster-0.

The next process is to identify how distribution of Chinese restaurants look like, in relation to all barrios in cluster-0. Let's visualize that:



From the map above, clearly there are 4 barrios that already have Chinese restaurants present, or quite close to it. These areas are:

1. El camp de l'Arpa del Clot & El clot with 3 restaurants surrounding these Barrios
2. Fort Pienc with 1 restaurant
3. Sant Antoni with 1 restaurant

And the remaining 4 areas are:

1. Vila de Gràcia
2. Sagrada Familia
3. El Parc i la Llacuna del Poblenou
4. La Vila Olimpica del Poblenou

These areas are the final barrios that we are going to recommend to Mr. X, and from the analysis above, all of these 4 fits our hypothesis, as well as Mr. X's requirements.

5. Discussion

There are a couple of further research that we can conduct to further enhance the recommendation. Barcelona is such a big city, and it will be interesting if we can correlate the size of Barrio with what kind of restaurants that are famous in that area. Perhaps Barrio may prefer a specific type of food, due to its demographic and ethnicity.

Also clustering Barrio based on characteristics of its public transport and cross check against the Restaurant clusters can be an interesting idea. Unfortunately Foursquare does not have complete information of public transport, and due to the limited amount of time, I have to exclude this information from the study.

Improving the model can also be part of improvement we can explore. The scoring of the K-Means model is not optimal, mainly due to the data distribution on the restaurant. Normalizing the data and removing some outliers from may improve the clustering model moving forward.

6. Conclusion

Apart from being proven by data that those 4 barrios hold a certain level of importance, those areas are considered to be places where not only tourists, but local residents will go to hang out. With the vast amount of attractions and considering Barcelona's easy access with public transport, there will be a chance that new business being setup here may reap a positive outcome in the future.