Predicting favorable location for a 5 star hotel

Aswin Raj

May 22, 2020

1. Introduction

1.1 Background

Spain is a popular tourist destination for tourists around the world who visit Europe.It is one of the world's leading tourist , economic , trade fair and cultural centers, and its influence in commerce, education, entertainment, media, fashion, science and the arts all contribute to its status as one of the world's major global cities.Particularly renowned are the architectural works of Antoni Gaudí and Lluís Domènech i Montaner, which have been designated UNESCO World Heritage Sites. Therefore it is advantageous for a company to establish a 5 star hotel in Barcelona.

1.2 Problem

We will attempt to find out favourable neighbourhoods to establish our 5 star hotel. As there are a good number of hotels in Barcelona, because it being a popular tourist destination, we will attempt to find the neighbourhoods that have a lesser number of hotels in vicinity. We will also attempt to find a neighbourhood with some popular venues nearby so that it becomes accessible to tourists, which will increase the chances of our hotel being preferred more. We will also be looking for neighbourhoods that are close to Barcelona city centre.

1.2 Interest

Our stakeholders will be interested in knowing what locations will be favourable to establish a 5 star hotel as they would be looking for locations which would attract a good number of tourists to find accommodation.

2. Data acquisition and cleaning

2.1 Data sources

Based on definition of our problem, factors that will influence our decision are:

- number of existing hotels in the neighbourhood (any type of hotel)
- number of venues in a location
- distance of neighbourhood from city centre

Following data sources will be needed to extract/generate the required information:

- List of neighbourhood can be obtained using the Wikipedia link https://en.wikipedia.org/wiki/Districts_of_Barcelona
- centres of candidate areas will be generated algorithmically and approximate addresses of centres of those areas will be obtained using **Google Maps API reverse geocoding**
- number of hotels and their type and location in every neighbourhood and popular venues will be obtained using **Foursquare API**

2.2 Data cleaning

As we obtain the table from Wikipedia, we see that the neighbourhoods we want are listed in a single row separated by commas and listed along districts. We first separate the neighbourhoods. Then we find the coordinates of the separated neighbourhoods using geocoding. We use Google's api to find out the coordinates of the addresses. We obtain the following column.

	Number	District	Neighbourhoods	location	point	latitude	longitude	altitude
0	1	Ciutat Vella	La Barceloneta	(la Barceloneta, Ciutat Vella, Barcelona, Barc...	(41.3806533, 2.1899274, 0.0)	41.380653	2.189927	0.0
0	1	Ciutat Vella	El Gòtic	(el Gòtic, Ciutat Vella, Barcelona, Barcelonès...	(41.3815052, 2.177417944846921, 0.0)	41.381505	2.177418	0.0
0	1	Ciutat Vella	El Raval	(el Raval, Ciutat Vella, Barcelona, Barcelonès...	(41.3795176, 2.1683678, 0.0)	41.379518	2.168368	0.0
0	1	Ciutat Vella	Sant Pere	(Sant Pere, Prat de Sant Pere, Mas Bellsolà, B...	(42.1982061, 2.698957309561404, 0.0)	42.198206	2.698957	0.0
0	1	Ciutat Vella	Santa Caterina i la Ribera	(Sant Pere, Santa Caterina i la Ribera, Ciutat...	(41.3866496, 2.1841940804878095, 0.0)	41.386650	2.184194	0.0

We remove the null values that we obtain. Then we drop the the columns (location, point and altitude) as they are irrelevant for our analysis. We obtain the following column.

	Number	District	Neighbourhoods	latitude	longitude
0	1	Ciutat Vella	La Barceloneta	41.380653	2.189927
1	1	Ciutat Vella	El Gòtic	41.381505	2.177418
2	1	Ciutat Vella	El Raval	41.379518	2.168368
3	1	Ciutat Vella	Sant Pere	42.198206	2.698957
4	1	Ciutat Vella	Santa Caterina i la Ribera	41.386650	2.184194
5	2	Eixample	L'Antiga Esquerra de l'Eixample	41.388765	2.156597
6	2	Eixample	La Nova Esquerra de l'Eixample	41.382816	2.149966
7	2	Eixample	Dreta de l'Eixample	41.394124	2.166471
8	2	Eixample	Fort Pienc	41.395925	2.182325
9	2	Eixample	Sagrada Família	41.403479	2.174410
10	2	Eixample	Sant Antoni	41.380052	2.163327
11	3	Sants-Montjuïc	La Bordeta	41.369142	2.137373
12	3	Sants-Montjuïc	la Font de la Guatlla	41.370782	2.144676
13	3	Sants-Montjuïc	Hostafrancs	41.375254	2.143348

Then we go on to find the number of hotels near Barcelona centre and in the neighbourhoods. We use the coordinates of the neighbourhoods to search for the hotels.

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	neighb
0	Hotel Suizo	Hotel	Plaça de l'Àngel 12	ES	Barcelona	España	NaN	143	[Plaça de l'Àngel 12, Barcelona Cataluña, España]	[{'label': 'display', 'lat': 41.38388613223940...	41.383886	2.178524	
1	Hotel Neri	Hotel	Calle Sant Sever 5	ES	Barcelona	España	NaN	184	[Calle Sant Sever 5, 08002 Barcelona Cataluña,...	[{'label': 'display', 'lat': 41.38329530312612...	41.383295	2.175287	
2	Hotel Catalonia Avinyó	Hotel	C. Avinyó, 16	ES	Barcelona	España	NaN	201	[C. Avinyó, 16, 08002 Barcelona Cataluña, España]	[{'label': 'display', 'lat': 41.38125986117526...	41.381260	2.176382	
3	Hotel Colon Barcelona	Hotel	Av. de la Catedral, 7	ES	Barcelona	España	NaN	253	[Av. de la Catedral, 7, 08002 Barcelona Catalu...	[{'label': 'display', 'lat': 41.38489539946631...	41.384895	2.175994	
4	Hotel Colonial	Hotel	Via Laietana, 3	ES	Barcelona	España	NaN	285	[Via Laietana, 3, 08003 Barcelona Cataluña, Es...	[{'label': 'display', 'lat': 41.38234704347800...	41.382347	2.180777	
5	Hotel Barcelona Catedral	Hotel	Capellans, 4	ES	Barcelona	España	Sagristans	300	[Capellans, 4 (Sagristans), 08002 Barcelona Ca...	[{'label': 'display', 'lat': 41.38497150137387...	41.384972	2.175132	

We then obtain the columns name, distance, latitude and longitude for our analysis

	name	distance	lat	lng
0	Hotel Suizo	143	41.383886	2.178524
1	Hotel Neri	184	41.383295	2.175287
2	Hotel Catalonia Avinyó	201	41.381260	2.176382
3	Hotel Colon Barcelona	253	41.384895	2.175994
4	Hotel Colonial	285	41.382347	2.180777

Now we focus on finding popular venues in the neighbourhood. We will be using the location of Barcelona city centre as our reference point to explore neighbourhoods using foursquare api.

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	neighbo
0	Barrio Gótico	Neighborhood	NaN	ES	Barcelona	España	NaN	86	[08002 Barcelona Cataluña, España]	[{'label': 'display', 'lat': 41.38366049614543...	41.383660	2.177290	
1	Gelaaati! di Marco	Ice Cream Shop	C. Llibreteria, 7	ES	Barcelona	España	NaN	32	[C. Llibreteria, 7, 08002 Barcelona Cataluña, ...	[{'label': 'display', 'lat': 41.38318551563699...	41.383186	2.177369	
2	La Colmena	Dessert Shop	Plaça de l'Àngel, 12	ES	Barcelona	España	NaN	148	[Plaça de l'Àngel, 12, Barcelona Cataluña, Esp...	[{'label': 'display', 'lat': 41.3839467278583,...	41.383947	2.178523	
3	Pont del Carrer del Bisbe	Bridge	Carrer del Bisbe	ES	Barcelona	España	NaN	96	[Carrer del Bisbe, Barcelona Cataluña, España]	[{'label': 'display', 'lat': 41.38331035757379...	41.383310	2.176413	
4	La Alcoba Azul	Spanish Restaurant	C. San Domenec del Call, 14	ES	Barcelona	España	NaN	161	[C. San Domenec del Call, 14, 08002 Barcelona ...	[{'label': 'display', 'lat': 41.38283347569902...	41.382833	2.175506	
5	Frankfurt Sant Jaume	Hot Dog Joint	Pl. Sant Jaume, s/n	ES	Barcelona	España	NaN	81	[Pl. Sant Jaume, s/n, 08002 Barcelona Cataluña...	[{'label': 'display', 'lat': 41.382275, 'lng':...	41.382275	2.176912	

We extract the columns name, categories, lat and lng for our analysis.

	name	categories	distance	lat	lng
0	Barrio Gótico	Neighborhood	86	41.383660	2.177290
1	Gelaaati! di Marco	Ice Cream Shop	32	41.383186	2.177369
2	La Colmena	Dessert Shop	148	41.383947	2.178523
3	Pont del Carrer del Bisbe	Bridge	96	41.383310	2.176413
4	La Alcoba Azul	Spanish Restaurant	161	41.382833	2.175506

We have gathered and cleaned the datasets that we will be using for our analysis

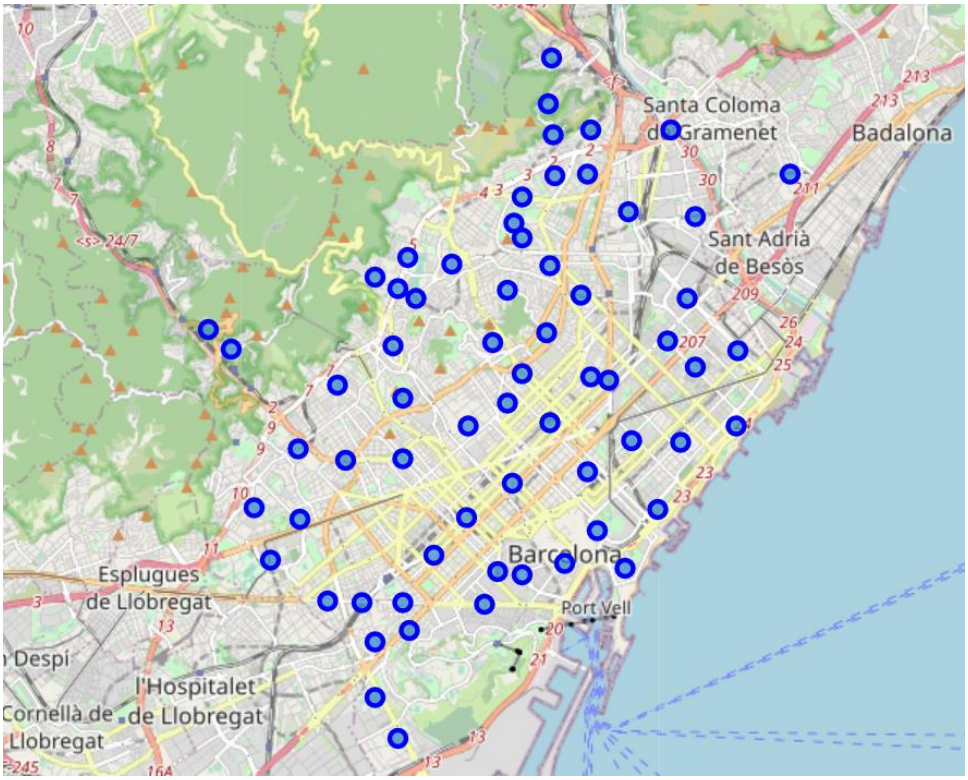
3. Methodology

In this project we will direct our efforts on detecting areas which are favourable for establishing a hotel. In first step we have collected the required data: location of neighborhoods, hotels around barcelona city center and popular venues near city center (according to Foursquare categorization). Second step in our analysis will be exploration of Hotels and popolar venues in a particular area. We will be using folium maps to see how close the hotels are with venues. Our point of reference will be the barcelona city center.

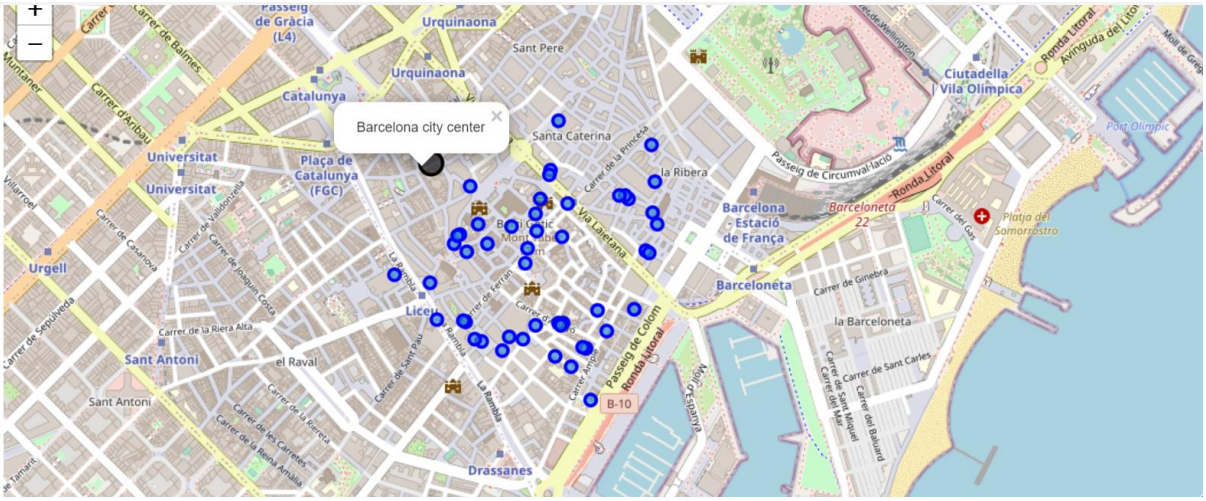
In third and final step we will focus on most promising areas and within those create clusters of venues established in discussion with stakeholders: . We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify geneeral zones which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

We start with plotting maps for each of our datasets.

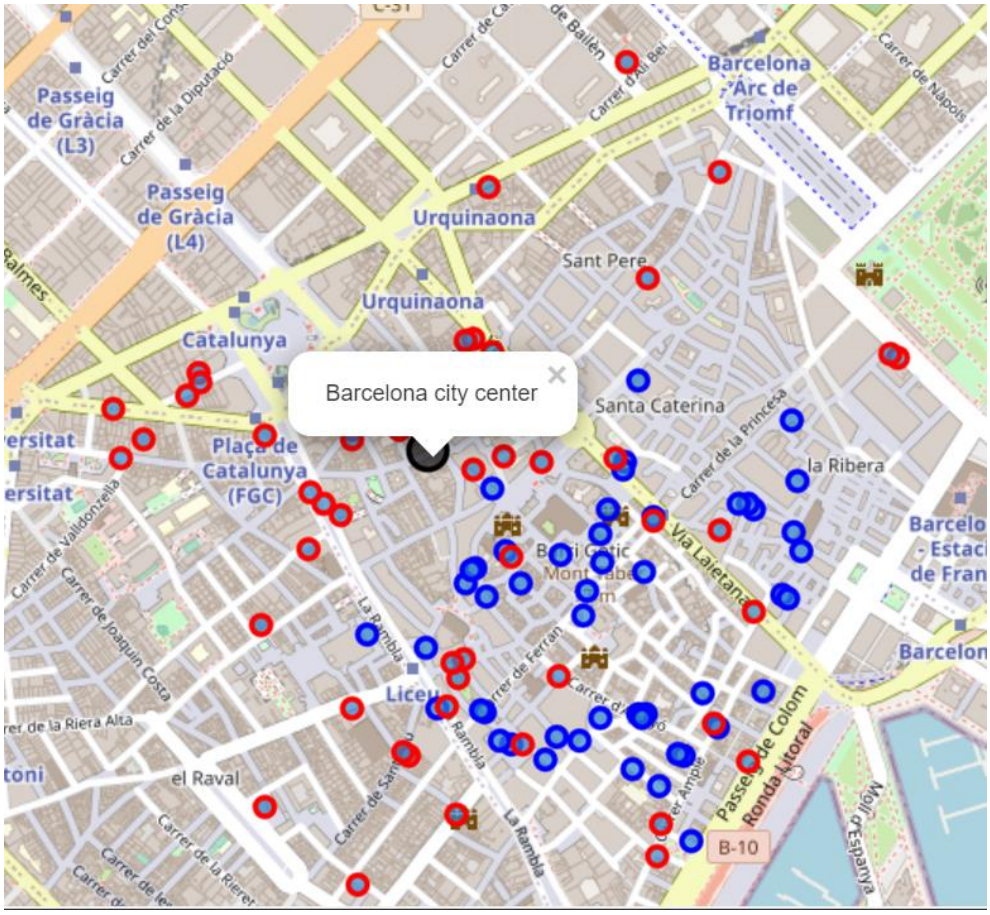
Map for Neighbourhood dataset



Map for venues dataset



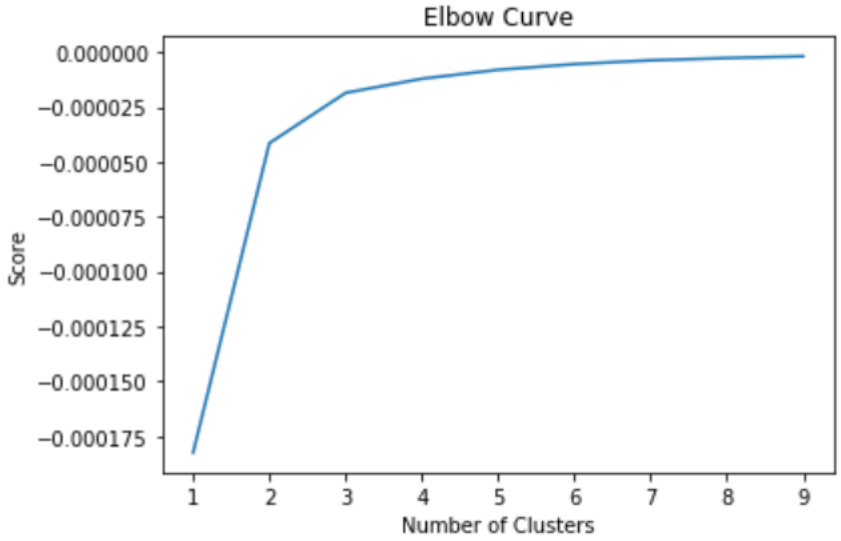
We can add the hotel coordinates to this map



The venues are marked with blue points and hotels are marked with red points.

As we can see there are a lot venues which are concentrated in some part of neighbourhood. So we aim to apply the clustering algorithm of machine learning to form clusters of venues. We can then find the optimum location for establishing our hotel. The factors which we will use to cluster the venues are latitudes and longitudes.

We will be using k means clustering as our clustering algorithm. For that, we will find a suitable k(i.e. number of clusters) for our algorithm. We can find that by plotting an elbow plot.

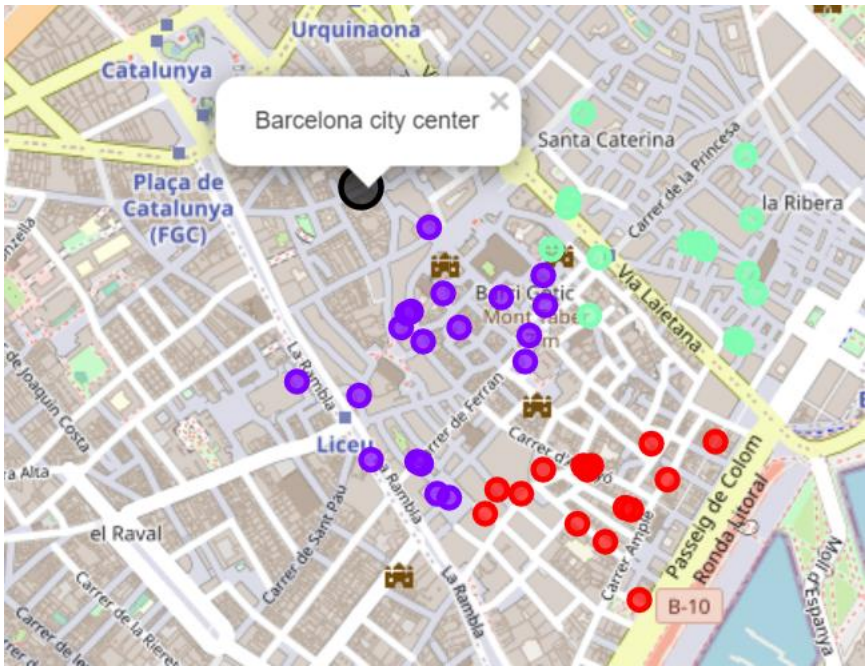


As we can see from the curve, 3 is our flexing point. So we can assign 3 as our k value

Now we apply the k means algorithm to our coordinates. As we obtain our cluster labels, we can merge it with our original dataframe of venues. We obtain the following table.

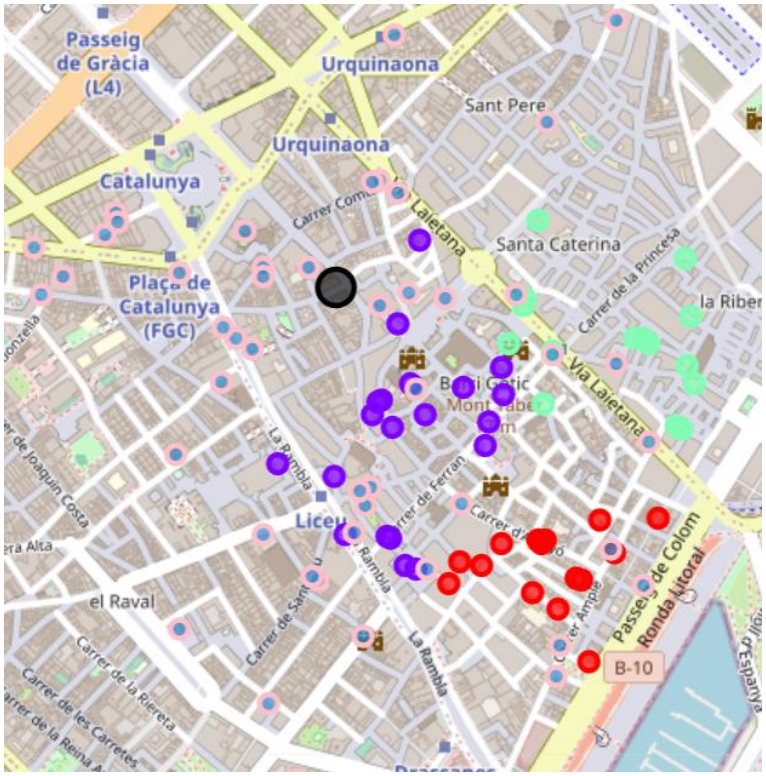
	name	categories	distance	lat	lng	Cluster label
0	Barrio Gótico	Neighborhood	86	41.383660	2.177290	1
1	Gelaaati! di Marco	Ice Cream Shop	32	41.383186	2.177369	1
2	La Colmena	Dessert Shop	148	41.383947	2.178523	2
3	Pont del Carrer del Bisbe	Bridge	96	41.383310	2.176413	1
4	La Alcoba Azul	Spanish Restaurant	161	41.382833	2.175506	1
5	Frankfurt Sant Jaume	Hot Dog Joint	81	41.382275	2.176912	1

Now we can plot a map based on the cluster of venues that we got from our machine clustering algorithm.



As we can see, the blue cluster has more venues in comparison to other clusters and is fairly closer to the City Centre.

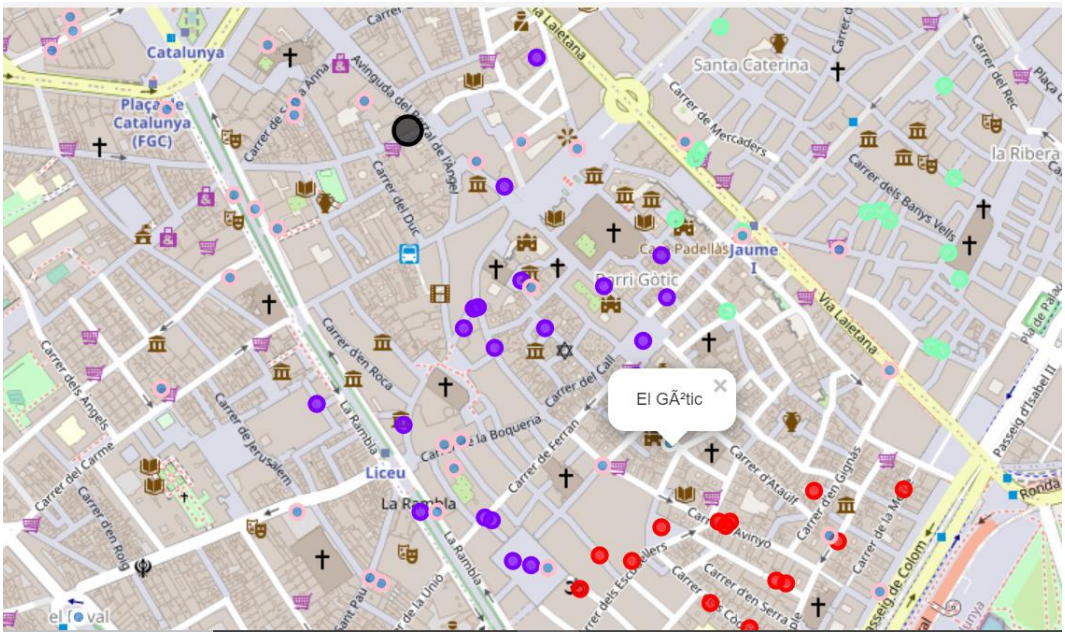
We now add the coordinates of hotels to this map



As we can see from the above map, there are some hotels who come in the vicinity of our favorable cluster of venues. But if we explore in the street level, we can in find locations with lesser number of hotels nearby.

We can now add neighborhood coordinates to our map

The neighborhood coordinates are represented by white points.



As we can see, we have found El gotic' to be our preferred neighborhood for establishing our 5 star hotel.

This concludes our analysis.

4. Results

We can derive from the above maps that the venues falling in cluster 1 are favorable for locating are 5 star hotel as they are more in number and fairly nearer to Barcelona city Centre. We also found El Gòtic to be the favorable neighborhood.

5. Discussions

The 3 factors were considered while plotting the map. So the stakeholders can come to a final decision based on location that we have obtained by considering the three factors. A street level exploration can be carried out to find the optimal location in the El Gòtic neighborhood.

6. Conclusion

In this project we have successfully analysed what area would be favourable to setup a 5 star hotel. The factors we based our analysis were

- Lesser number of hotels nearby
- Good number of venues near the hotel
- Closer to city centre

We started out by extracting data for our neighbourhoods. We used web scraping to find out the names of neighbourhoods in Barcelona. We then used geocoding to obtain coordinates of the neighbourhoods. Then we got the info of hotels and venues near Barcelona city centre by foursquare api. We plotted the points on a folium map to show their locations. We then created clusters of venues based on their coordinates. We plotted the clusters on a folium map along with hotels and neighbourhood points. We have obtained our desired neighbourhood to setup our hotel. The neighbourhood is El Gòtic. The stakeholders can now explore on street level based on our analysis. They can view the folium map for reference.