

Capstone Project - The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

Table of contents

- Problem Background
- Targeted Audience
- Data
- Neighborhood Candidates
- Methodology
- Analysis
- Results and Discussion
- Conclusion

Problem Background

The city of Barcelona is the capital and largest city of the autonomous community of Catalonia, as well as the second most populous municipality of Spain. It is economically one of the most feasible of all of the country to open any business of any kind. In this assignment/project we would like to discover the city's neighborhoods and to choose which one would be the correct choice starting the business taking into mind which vendors, hotels and other stores that are close by to the mentioned business.

In our case, we'll choose to start an Italian restaurant and we would want to see where we'd be landing the first branch.

Targeted Audience

The audience of this project would be the people of Barcelona who would like to try a foreign restaurant, in this case an Italian restaurant. The Data Science Team has been appointed to lead this project where a suitable place would strike as a good starting branch for the business.

Data

- According to <https://en.wikipedia.org/wiki/Barcelona>, the city of Barcelona has about 73 neighborhoods. Barcelona also has a population of 1.6 million.

We will be using the information provided in this Wiki page to navigate any information needed to know about the city.

Factors that will influence our decision are:

- Number of existing restaurants in the neighborhood.
- Distance of neighborhood from city center.

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

- Number of restaurants and their location in every neighborhood will be obtained using Foursquare API.
- Centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding.
- Coordinate of Barcelona center will be obtained using Google Maps API geocoding.

Neighborhood Candidates

Let's create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is aprox. 6x6 kilometers centered around Barcelona city center.

We've found the latitude & longitude of Barcelona city center, using specific, well known address and Google Maps geocoding API.

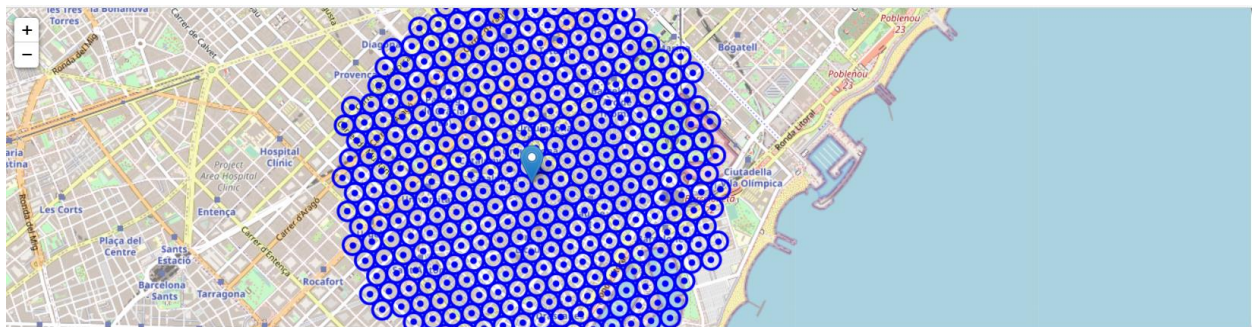
```
Coordinate of Barcelona, Spain: [41.3850639, 2.1734035]
```

```
Out[1]: {'lat': 41.3850639, 'lng': 2.1734035}
```

We've created a grid of area candidates, equally spaced, centered around city center and within ~6km from the city center. Our neighborhoods will be defined as circular areas with a radius of 150 meters, so our neighborhood centers will be 300 meters apart.

To accurately calculate distances we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So let's create functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

Barcelona center longitude=2.1734035, latitude=41.3850639
Barcelona center UTM X=-573518.9890965135, Y=4661662.499852937
Barcelona center longitude=2.1734035000000023, latitude=41.385063899999999



We now have the coordinates of centers of neighborhoods/areas to be evaluated, equally spaced (distance from every point to its neighbors is exactly the same) and within ~6km from city center.

Using Google Maps API to get approximate addresses of those locations.

Reverse geocoding check

Address of [41.3850639, 2.1734035] is: Ede ekene, 08002 Barcelona, Spain

The addressed represented in a Pandas DataFrames:

out[10]:

	Address	Latitude	Longitude	X	Y	Distance from center
0	Passatge Vinyeta, 14, 08038 Barcelona, Spain	41.371919	2.170665	-573968.989097	4.660234e+06	1498.123827
1	Passeig de Montjuïc, 23, 08004 Barcelona, Spain	41.372118	2.172414	-573818.989097	4.660234e+06	1460.094175
2	Jardins de les Hortes de Sant Bertran, Passeig...	41.372316	2.174162	-573668.989097	4.660234e+06	1436.793305
3	Ronda Litoral, 100, 08039 Barcelona, Spain	41.372514	2.175911	-573518.989097	4.660234e+06	1428.941916
4	Moll Sant Bertran - Moll Barcelona, 08039 Barc...	41.372713	2.177660	-573368.989097	4.660234e+06	1436.793305
5	MI Barcelona, 2, 08039 Barcelona, Spain	41.372911	2.179409	-573218.989097	4.660234e+06	1460.094175
6	Carrer de Barcelona, 2, 08039 Barcelona, Spain	41.373109	2.181158	-573068.989097	4.660234e+06	1498.123827
7	Carrer Nou de la Rambla, 147, 08004 Barcelona, ...	41.372763	2.167813	-574193.989097	4.660363e+06	1463.941597
8	Carrer de Piquer, 29, 08004 Barcelona, Spain	41.372961	2.169562	-574043.989097	4.660363e+06	1401.115627
9	Carrer de Vila i Vilà, 43, 08004 Barcelona, Spain	41.373160	2.171311	-573893.989097	4.660363e+06	1352.081728

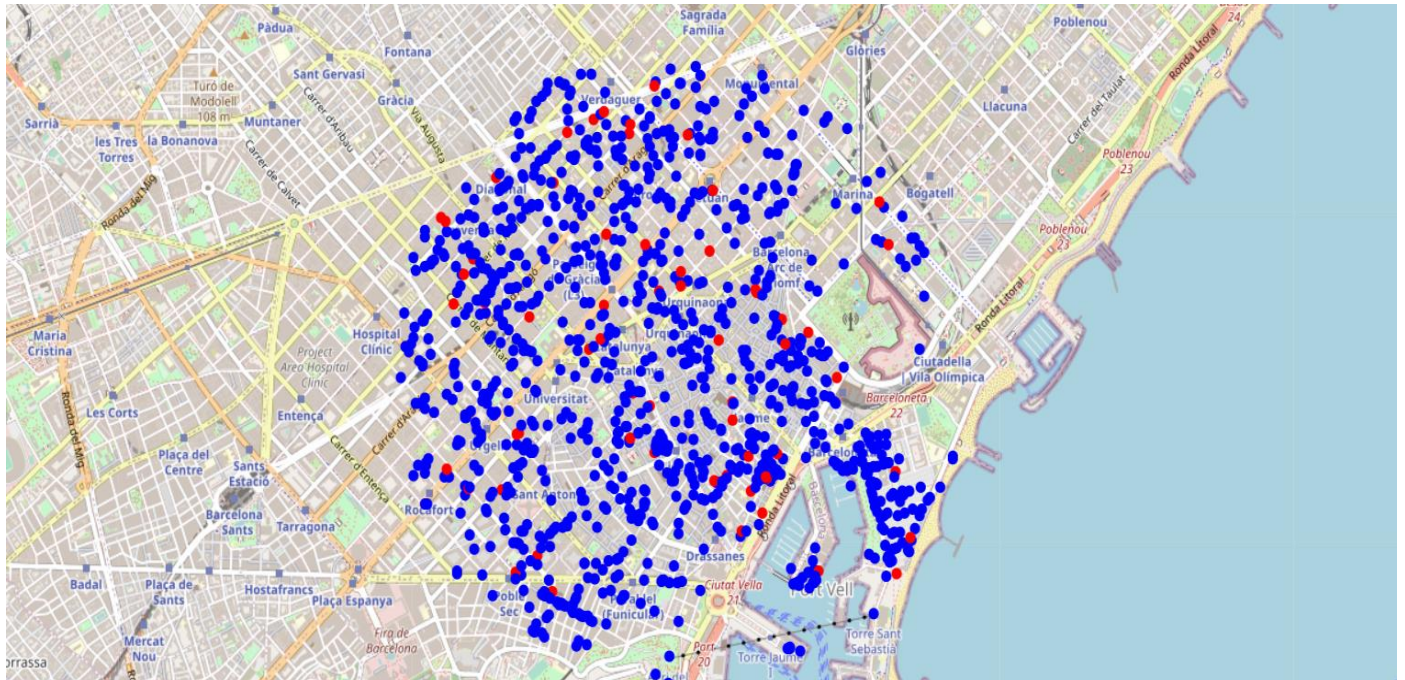
Foursquare

Now that we have our location candidates, we've used Foursquare API to get info on restaurants in each neighborhood.

We're interested in venues in 'food' category, but only those that are proper restaurants - coffee shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So we will include in our list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Italian restaurant' category, as we need info on Italian restaurants in the neighborhood.

```
Total number of restaurants: 1080
Total number of Italian restaurants: 66
Percentage of Italian restaurants: 6.11%
Average number of restaurants in neighborhood: 25.274725274725274
```

We'll now see all the collected restaurants in our area of interest on map, and we'll also show Italian restaurants in different color (red).



We have all the restaurants in the vicinity of a few kilometers from the city center, as well as the Italian ones in a different color.

This concludes the data gathering step and we can proceed with data analysis.

Methodology

We are going to direct our data analysis in the path of finding the optimal place to launch the first branch, and that would be the place where it is least crowded with Italian restaurants. We have acquired the neighborhood names in the vicinity, this case the 6x6 kilometers around city center and we have collected the data of every restaurant and specifically Italian restaurants.

We will then proceed with finding the area and producing heatmaps showing which areas are not crowded with restaurants and which ones of these do not have Italian restaurants.

We will be using k-means clustering to create clusters of neighborhoods and decide the optimal starting point of our restaurant.

Analysis

Let's derive some additional info from our raw data and count the number of restaurants in every area candidate.

Average number of restaurants in every area with radius=300m: 25.274725274725274

Out[20]:

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
0	Passatge Vinyeta, 14, 08038 Barcelona, Spain	41.371919	2.170665	-573968.989097	4.660234e+06	1498.123827	10
1	Passeig de Montjuïc, 23, 08004 Barcelona, Spain	41.372118	2.172414	-573818.989097	4.660234e+06	1460.094175	7
2	Jardins de les Hortes de Sant Bertran, Passeig...	41.372316	2.174162	-573668.989097	4.660234e+06	1436.793305	6
3	Ronda Litoral, 100, 08039 Barcelona, Spain	41.372514	2.175911	-573518.989097	4.660234e+06	1428.941916	2
4	Moll Sant Bertran - Moll Barcelona, 08039 Barc...	41.372713	2.177660	-573368.989097	4.660234e+06	1436.793305	1
5	MI Barcelona, 2, 08039 Barcelona, Spain	41.372911	2.179409	-573218.989097	4.660234e+06	1460.094175	1
6	Carrer de Barcelona, 2, 08039 Barcelona, Spain	41.373109	2.181158	-573068.989097	4.660234e+06	1498.123827	2
7	Carrer Nou de la Rambla, 147, 08004 Barcelona,...	41.372763	2.167813	-574193.989097	4.660363e+06	1463.941597	30
8	Carrer de Piquer, 29, 08004 Barcelona, Spain	41.372961	2.169562	-574043.989097	4.660363e+06	1401.115627	6
9	Carrer de Vila i Vilà, 43, 08004 Barcelona, Spain	41.373160	2.171311	-573893.989097	4.660363e+06	1352.081728	5

We'll calculate the distance to nearest Italian restaurant from every area candidate center.

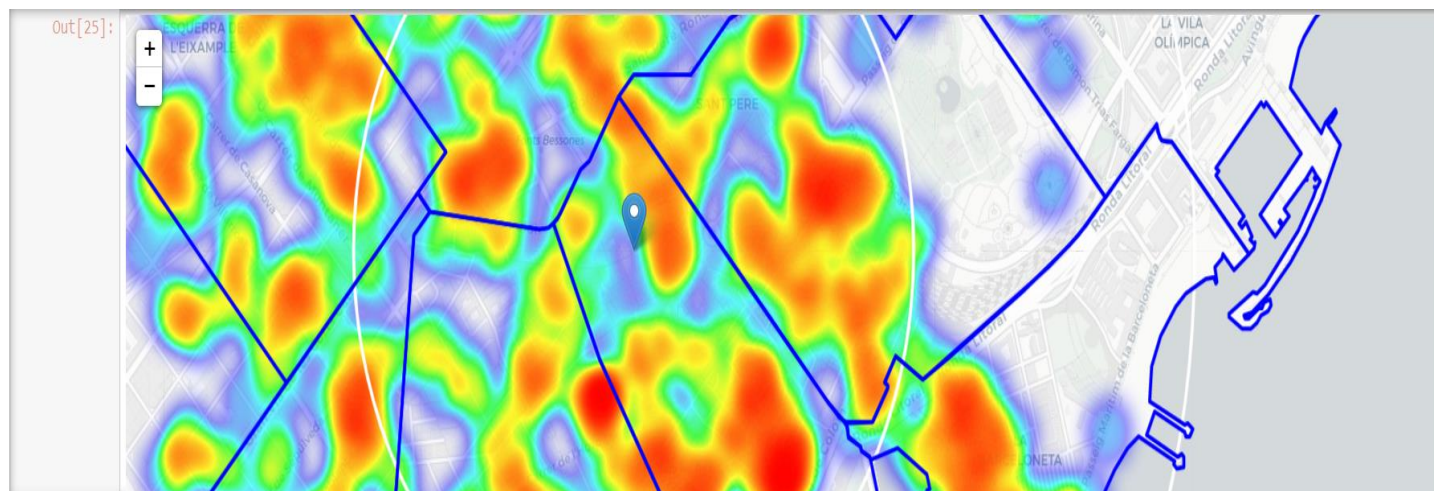
Out[21]:

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to Italian restaurant
0	Passatge Vinyeta, 14, 08038 Barcelona, Spain	41.371919	2.170665	-573968.989097	4.660234e+06	1498.123827	10	644.756189
1	Passeig de Montjuic, 23, 08004 Barcelona, Spain	41.372118	2.172414	-573818.989097	4.660234e+06	1460.094175	7	757.055691
2	Jardins de les Hortes de Sant Bertran, Passeig...	41.372316	2.174162	-573668.989097	4.660234e+06	1436.793305	6	655.931683
3	Ronda Litoral, 100, 08039 Barcelona, Spain	41.372514	2.175911	-573518.989097	4.660234e+06	1428.941916	2	576.506223
4	Moll Sant Bertran - Moll Barcelona, 08039 Barc...	41.372713	2.177660	-573368.989097	4.660234e+06	1436.793305	1	528.651566
5	MI Barcelona, 2, 08039 Barcelona, Spain	41.372911	2.179409	-573218.989097	4.660234e+06	1460.094175	1	461.928208
6	Carrer de Barcelona, 2, 08039 Barcelona, Spain	41.373109	2.181158	-573068.989097	4.660234e+06	1498.123827	2	338.364930
7	Carrer Nou de la Rambla, 147, 08004 Barcelona,...	41.372763	2.167813	-574193.989097	4.660363e+06	1463.941597	30	385.814198
8	Carrer de Piquer, 29, 08004 Barcelona, Spain	41.372961	2.169562	-574043.989097	4.660363e+06	1401.115627	6	514.714442
9	Carrer de Vila i Vilà, 43, 08004 Barcelona, Spain	41.373160	2.171311	-573893.989097	4.660363e+06	1352.081728	5	652.693893

Average distance to closest Italian restaurant from each area center: 218.9931736284998

On average Italian restaurant can be found within ~220m from every area center candidate, so we need to filter our areas very carefully.

We'll create a map showing heatmap / density of restaurants and try to extract some meaningful info from that. Also, let's show borders of Barcelona neighborhoods on our map and a few circles indicating distance of 1km, 2km and 3km from city center.



Areas of low restaurant density can be found in the east, southeast of the city center.

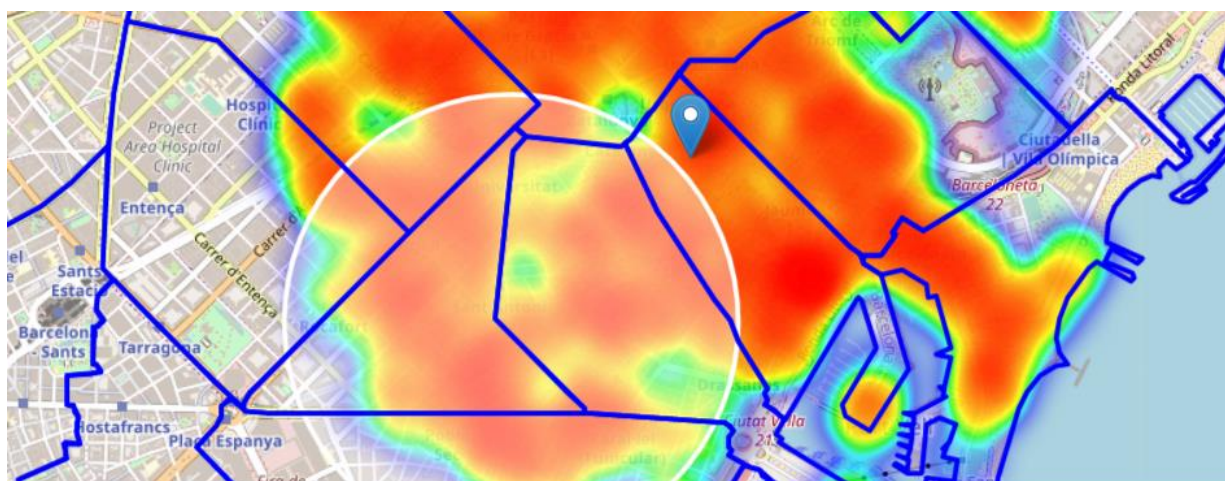
We'll create another heatmap map showing heatmap/density of Italian restaurants only.



Based on this heatmap, the density of Italian restaurants are positioned within east of city center. So, we'd move our area of interest to launch our branch west, north-west and south-west of that center, and we're more likely to move into the south-west.

This will put us in either "El Raval" or "Sant Antoni" neighborhoods, since they are south-west of the city where it is less crowded with Italian restaurants and are considered places which tourists in the city usually visit.

Let's define new, more narrow region of interest, which will include low-restaurant-count parts of "El Raval" and "Sant Antoni" closest to city center.



This covers mostly of the two neighborhoods. We'll now create new, more dense grid of location candidates restricted to our new region of interest (we'll make our location candidates 100m apart).

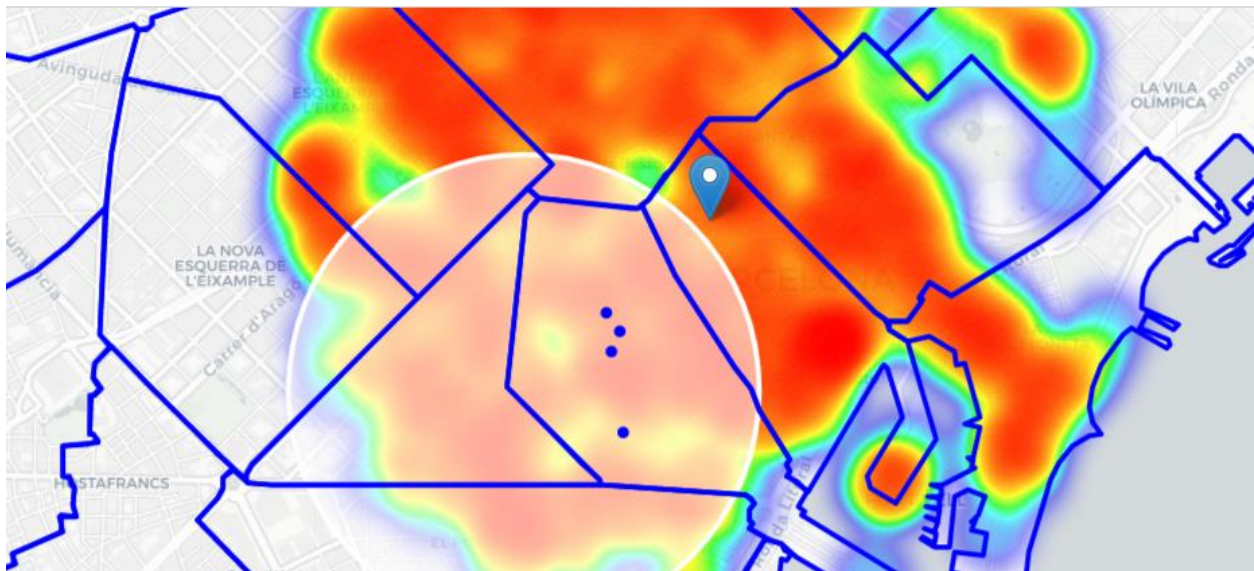
Out[64]:

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to Italian restaurant
0	41.377024	2.169048	-574018.989097	4.660822e+06	2	523.055797
1	41.377851	2.169479	-573968.989097	4.660909e+06	7	423.089933
2	41.378545	2.168744	-574018.989097	4.660996e+06	6	385.801465
3	41.379372	2.169175	-573968.989097	4.661082e+06	6	290.276585
4	41.380066	2.168440	-574018.989097	4.661169e+06	5	239.962890
5	41.380893	2.168871	-573968.989097	4.661255e+06	4	140.247144
6	41.381587	2.168136	-574018.989097	4.661342e+06	5	134.559938

We'll now filter those locations. We're interested only in locations with no more than five restaurants in radius of 100 meters, and no Italian restaurants in radius of 100 meters.

```
Locations with no more than five restaurants nearby: 4
Locations with no Italian restaurants within 200m: 7
Locations with both conditions met: 4
```

Let's see how this looks on a map.



We now have four potential locations, all located in "El Rava" neighborhood. No further clustering or analysis is needed since they are fairly close to each other and each one of them meet the conditions where no more than five restaurants and no Italian restaurant are within 100 m of these four locations.

Results and Discussion

We started by identifying the number of restaurants in Barcelona city and calculated how many of them are Italians, which we found that the city has X restaurant and X of them are Italian. The percentage of Italian restaurants are calculated and we've identified using a heatmap which area of the city is dense with restaurant and with Italian restaurants. Based on the maps shown in the analysis section we've identified two potential neighborhood for our restaurant and they are "El Rava" and "Sant Antoni". These two neighborhoods are considered good places for tourists and are also one of the lively places in the city.

We've specified two conditions to be met: no more than five restaurant should be in the vicinity of 100m and no Italian restaurant should be present in this distance as well. We've arrived at four locations that meet the specified conditions and no further analysis was needed since they both are fairly close and are located in "El Rava" neighborhood.

Conclusion

Using maps of Barcelona city with the help of Google and Foursquare we've identified, as a data science team, which locations should be a good starting point for the restaurant that would be constructed. We started our analysis from Barcelona city center and went down to find one specific neighborhood with four specific location that could be a potential point of interest for us. We've made a few conditions to reach this result and we didn't need more data analysis since we've found them fairly close to each other and are present in only one neighborhood "El Rava".

Final decision will be made by the stakeholders on which location should be chosen.