

IBM Data Science

Applied Data Science Project

Peer-graded Assignment:

Capstone Project

The Battle of Neighborhoods (Week 2)

Vasileios Fafoutis

13.12.2020



Index

p.3	Introduction
p.4	Data Source & Description
p.7	Methodology
p.14	Results
p.17	Discussion & Conclusion

1. Introduction

Objective of this project is to **classify Barcelona city (Spain) neighborhoods**, according to its venues, criminality rate, population density and rent prices. This report can be useful to anyone interested to expand his/her business to another neighborhood or simply looking for a new neighborhood to move in.

Barcelona is divided in 7 districts (see image below), each of which contains several neighborhoods. **A total of 73 neighborhoods are defined in Barcelona.**

7 Districts of Barcelona:

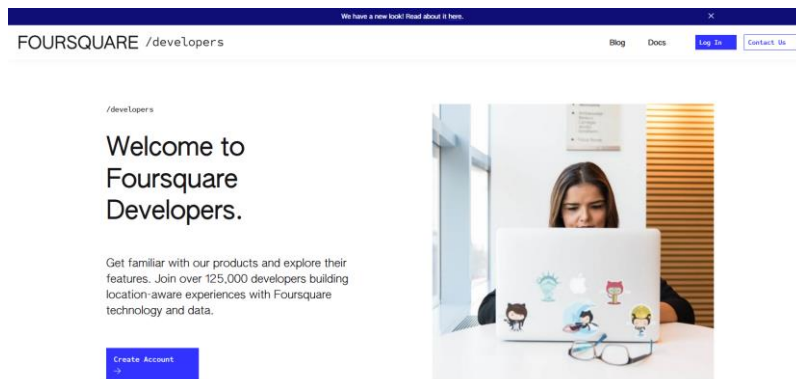


2. Data sources

The data used for this project comes from the following sources:

1. Foursquare / Developer:

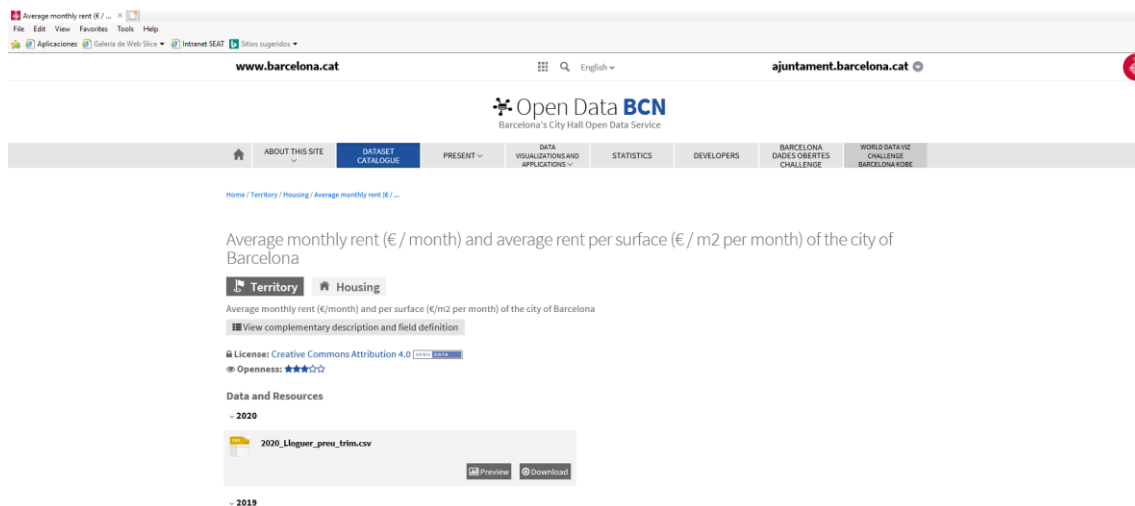
<https://developer.foursquare.com/developer/>



- ➔ Data regarding venues in every neighborhood. A specific account was created during the course and all relevant data is retrieved through API

2. Rent prices - Barcelona City Hall open data:

<https://opendata-ajuntament.barcelona.cat/data/en/dataset/est-mercat-immobiliari-lloguer-mitja-mensual>



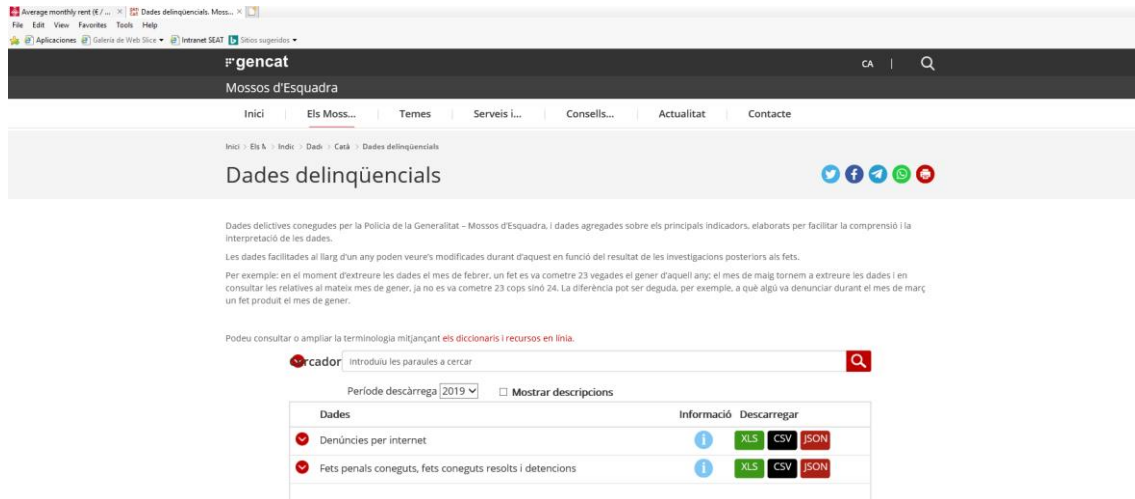
- ➔ Mean rent prices in € per trimester, per neighborhood. During data exploration, it was detected that various records were invalid (rent prices

Capstone Project - The Battle of Neighborhoods

up to x20 times lower than normal). Such values were substituted by the mean value of the dataset (using excel, directly in the retrieved .csv file)

3. Criminality rate per Neighborhood - Police Force of Catalunya:

https://mossos.gencat.cat/ca/els_mossos_desquadra/indicadors_i_qualitat/dades_s_obertes/catalog_dades_obertes/dades-delinquencials/



The screenshot shows the 'Dades delinquencials' page on the Mossos d'Esquadra website. It features a search bar at the top with the text 'introduïu les paraules a cercar'. Below the search bar, there is a table with two columns: 'Dades' and 'Informació'. The table lists two categories: 'Denúncies per internet' and 'Fets penals coneguts, fets coneguts resolts i detencions'. Each category has a download button with options for 'XLS', 'CSV', and 'JSON'. The page also includes a sidebar with navigation links and a footer with social media icons.

➔ Large data set containing criminality data per district in the whole Catalunya area. The data is used to retrieve the overall criminality rate per neighborhood (proportional of the neighborhood population to the district population)

4. Demographics data, surface, population density etc of each neighborhood and district – Wikipedia

https://ca.wikipedia.org/wiki/Districtes_i_barris_de_Barcelona

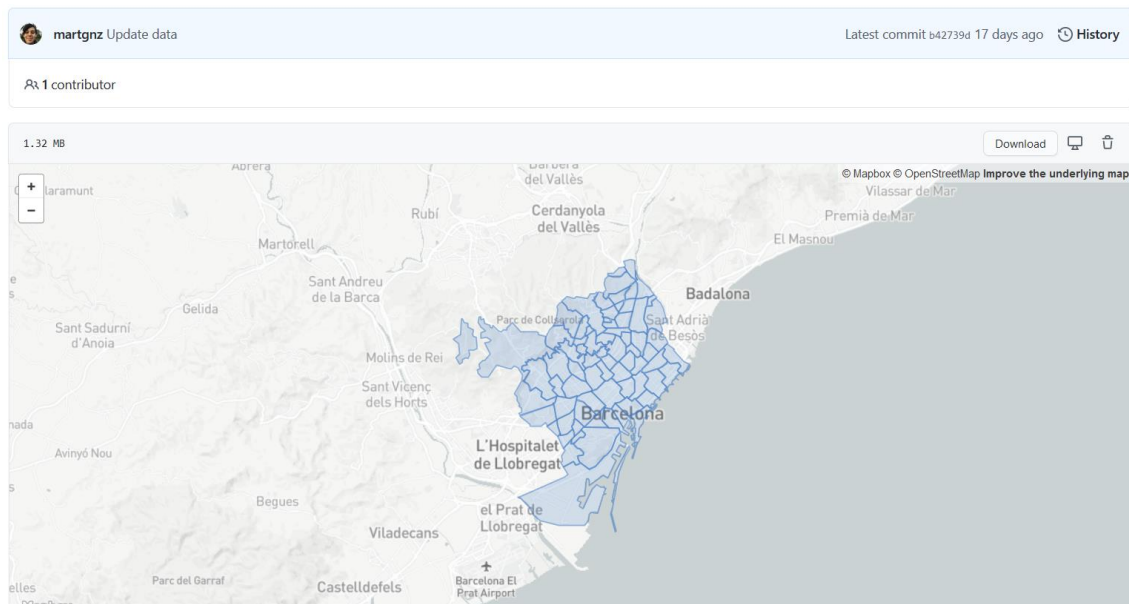


The screenshot shows the Wikipedia page for 'Districtes i barris de Barcelona'. The page title is 'Districtes i barris de Barcelona'. The main text describes the administrative structure of Barcelona, mentioning that the city is divided into districts and neighborhoods. It also includes a map of the city showing the districts. The page has a sidebar with navigation links and a footer with social media icons.

Capstone Project - The Battle of Neighborhoods

5. geoJSON file to create the choropleth map:

<https://github.com/martgnz/bcn-geodata/blob/master/barris/barris.geojson>



3. Methodology

The machine learning algorithm that will be implemented in order to classify the neighborhoods by the desired characteristics, is k means clustering. In order to implement the algorithm, a proper dataframe needs to be prepared first.

In summary the following steps have been followed (for details please see the notebook in Github).

- a. The necessary libraries and tools (pandas, numpy, matplotlib, geopy, requests, json, folium etc) are imported.
- b. Data regarding neighborhoods (district and neighborhood names and numbers..) is imported from Barcelona City Hall data.

	CODI_DISTRICTE	NOM_DISTRICTE	CODI_BARRI	NOM_BARRI
0	1	Ciutat Vella	1	el Raval
1	1	Ciutat Vella	2	el Barri Gòtic
2	1	Ciutat Vella	3	la Barceloneta
3	1	Ciutat Vella	4	Sant Pere, Santa Caterina i la Ribera
4	2	Eixample	5	el Fort Pienc
...
68	10	Sant Martí	69	Diagonal Mar i el Front Marítim del Poblenou
69	10	Sant Martí	70	el Besòs i el Maresme
70	10	Sant Martí	71	Provençals del Poblenou
71	10	Sant Martí	72	Sant Martí de Provençals
72	10	Sant Martí	73	la Verneda i la Pau

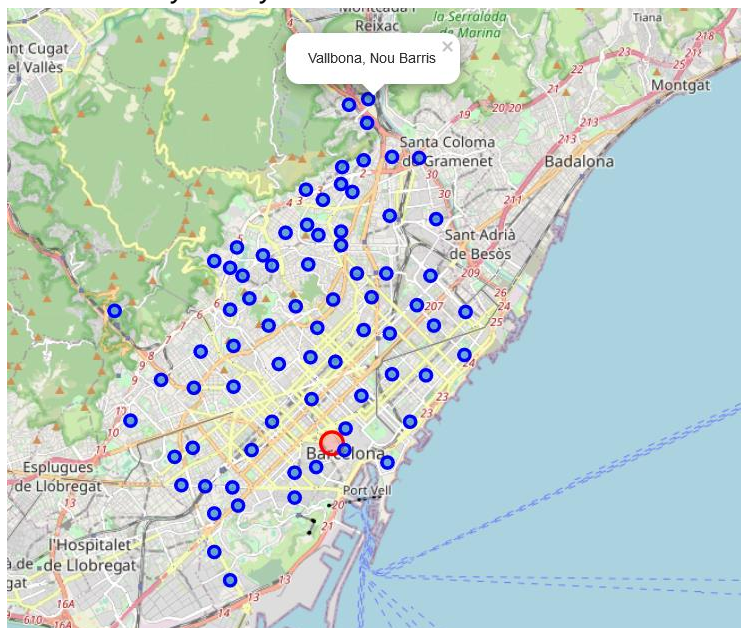
- c. The dataframe is properly preprocessed so that the latitude and longitude of each neighborhood is retrieved using geopy.

Capstone Project - The Battle of Neighborhoods

	Address	Lat	Long
0	el Raval ,Ciutat Vella ,Barcelona, Spain	41.379518	2.168368
1	el Barri Gòtic ,Ciutat Vella ,Barcelona, Spain	41.383395	2.176912
2	la Barceloneta ,Ciutat Vella ,Barcelona, Spain	41.380653	2.189927
3	Sant Pere, Santa Caterina i la Ribera ,Ciutat ...	41.388322	2.177411
4	el Fort Pienc ,Eixample ,Barcelona, Spain	41.395925	2.182325
...
68	Diagonal Mar i el Front Marítim del Poblenou ,...	41.405228	2.213352
69	el Besòs i el Maresme ,Sant Martí ,Barcelona, ...	41.414979	2.213877
70	Provençals del Poblenou ,Sant Martí ,Barcelona...	41.411948	2.204125
71	Sant Martí de Provençals ,Sant Martí ,Barcelon...	41.416519	2.198968
72	la Verneda i la Pau ,Sant Martí ,Barcelona, Spain	41.423220	2.202940

73 rows × 3 columns

- d. The created dataframe is visualized through a folium map to check validity of neighborhoods *(In fact during exploration phase, after checking the map for the 1st time, one of the neighborhoods was found as falsely located, because its name was similar to a city nearby Barcelona. The correction made can be seen in the notebook)*



Capstone Project - The Battle of Neighborhoods

- e. Using the neighborhood dataframe (incl. coordinates) the relevant venues from Foursquare are retrieved.

Example of venues in "el Raval" neighborhood

	name	categories	lat	lng
0	A Tu Bola	Tapas Restaurant	41.380096	2.169054
1	Arume	Spanish Restaurant	41.378953	2.166008
2	La Robadora	Gastropub	41.379500	2.170463
3	Chulapio	Cocktail Bar	41.379264	2.165905
4	Robadors 23	Jazz Club	41.379581	2.170603
...
95	Two schmucks	Cocktail Bar	41.383136	2.165190
96	JOVANI&VINS	Wine Shop	41.376797	2.163708
97	La Isabela (Hotel 1898)	Hotel Pool	41.383388	2.170812
98	Forn Boix	Bakery	41.383571	2.169590
99	M-Store	Gift Shop	41.381429	2.163405

100 rows × 4 columns

- f. The same process is done for all 73 neighborhoods in Barcelona, and a table of 3144 rows (venues) of 244 unique categories is retrieved.

"Head" of new dataframe

(3144, 7)

Out[14]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	el Raval	41.379518	2.168368	A Tu Bola	41.380096	2.169054	Tapas Restaurant
1	el Raval	41.379518	2.168368	Arume	41.378953	2.166008	Spanish Restaurant
2	el Raval	41.379518	2.168368	La Robadora	41.379500	2.170463	Gastropub
3	el Raval	41.379518	2.168368	Chulapio	41.379264	2.165905	Cocktail Bar
4	el Raval	41.379518	2.168368	Robadors 23	41.379581	2.170603	Jazz Club

- g. The dataframe is then grouped and the categorical data to dummies.

Capstone Project - The Battle of Neighborhoods

	Neighborhood	Yoga Studio	Accessories Store	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepas Restaurant	Argentinian Restaurant	Art Gallery	...	Vacation Rental	Vegetarian / Vegan Restaurant	Venezuelan Restaurant	Video Game Store	Video Store	Vietnamese Restaurant	Wine Bar	Wine Shop	Winery	Women's Store
0	Baró de Viver	0.00	0.0	0.0	0.00	0.0	0.00	0.0	0.0	0.0	...	0.0	0.000000	0.00	0.0	0.00	0.0	0.00	0.00	0.0	0.0
1	Can Baró	0.00	0.0	0.0	0.00	0.0	0.00	0.0	0.0	0.0	...	0.0	0.000000	0.00	0.0	0.00	0.0	0.00	0.00	0.0	0.0
2	Can Peguera	0.00	0.0	0.0	0.00	0.0	0.00	0.0	0.0	0.0	...	0.0	0.000000	0.00	0.0	0.00	0.0	0.00	0.00	0.0	0.0
3	Canyelles	0.00	0.0	0.0	0.00	0.0	0.00	0.0	0.0	0.0	...	0.0	0.000000	0.00	0.0	0.00	0.0	0.00	0.00	0.0	0.0
4	Ciutat Meridiana	0.00	0.0	0.0	0.00	0.0	0.00	0.0	0.0	0.0	...	0.0	0.000000	0.00	0.0	0.00	0.0	0.00	0.00	0.0	0.0
...
68	la Vila Olímpica del Poblenou	0.01	0.0	0.0	0.01	0.0	0.01	0.0	0.0	0.0	...	0.0	0.000000	0.00	0.0	0.00	0.0	0.00	0.00	0.0	0.0
69	la Vila de Gràcia	0.00	0.0	0.0	0.00	0.0	0.00	0.0	0.0	0.0	...	0.0	0.030000	0.01	0.0	0.01	0.0	0.03	0.02	0.0	0.0
70	les Corts	0.00	0.0	0.0	0.00	0.0	0.00	0.0	0.0	0.0	...	0.0	0.051724	0.00	0.0	0.00	0.0	0.00	0.00	0.0	0.0
71	les Roquetes	0.00	0.0	0.0	0.00	0.0	0.00	0.0	0.0	0.0	...	0.0	0.000000	0.00	0.0	0.00	0.0	0.00	0.00	0.0	0.0
72	les Tres Torres	0.00	0.0	0.0	0.00	0.0	0.00	0.0	0.0	0.0	...	0.0	0.000000	0.00	0.0	0.00	0.0	0.00	0.00	0.0	0.0

73 rows × 284 columns

h. Using the frequency of each venue, a dataframe with the top 10 most common venues per neighborhood is retrieved (I will be used later for the understanding of the clusters).

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 Baró de Viver	Plaza	Metro Station	Café	Track Stadium	Burger Joint	Supermarket	Park	Women's Store	Fast Food Restaurant	Fabric Shop
1 Can Baró	Spanish Restaurant	Grocery Store	Restaurant	Chinese Restaurant	Scenic Lookout	Soccer Stadium	Tapas Restaurant	Big Box Store	Fried Chicken Joint	Breakfast Spot
2 Can Peguera	Park	Food & Drink Shop	Supermarket	Basketball Stadium	Escape Room	Tapas Restaurant	Café	Grocery Store	Gym / Fitness Center	German Restaurant
3 Canyelles	Soccer Field	Market	Food & Drink Shop	Café	Grocery Store	Tapas Restaurant	Mediterranean Restaurant	Skate Park	Plaza	Filipino Restaurant
4 Ciutat Meridiana	Metro Station	Park	Plaza	Grocery Store	Supermarket	Train Station	Women's Store	Farmers Market	Escape Room	Ethiopian Restaurant

i. The rest of the dataframes (with rent prices, surface, population, crimes etc) are inserted in the notebook, properly preprocessed (partially using excel), normalized and then merged with the dummy data dataframe.

	Neighborhood	Surface_(ha)	Price	Population_Density_(hab/ha)	Crime_Proportion(Nr)
0	el Raval	109.8	923.0	434.0	23962.0
1	el Barri Gòtic	84.2	1139.7	181.0	7684.0
2	la Barceloneta	131.4	924.0	114.0	7566.0
3	Sant Pere, Santa Caterina i la Ribera	111.4	1020.2	200.0	11224.0
4	el Fort Pienc	92.9	1064.4	357.4	7345.0
...
68	Diagonal Mar i el Front Marítim del Poblenou	120.3	1498.6	112.0	1557.0
69	el Besòs i el Maresme	118.5	923.0	208.0	2845.0
70	Provençals del Poblenou	108.6	1013.6	193.0	2422.0
71	Sant Martí de Provençals	73.4	923.0	355.0	3005.0
72	la Verneda i la Pau	112.7	923.0	255.0	3320.0

73 rows × 5 columns

Normalized values:

Capstone Project - The Battle of Neighborhoods

	Surface_(ha)	Price	Population_Density_(hab/ha)	Crime_Proportion(Nr)
0	-0.077607	-0.512343	1.139035	5.430410
1	-0.260635	0.529486	-0.494437	1.200080
2	0.076823	-0.507536	-0.927017	1.169415
3	-0.066167	-0.045034	-0.371765	2.120056
4	-0.198434	0.167466	0.644474	1.111981
...
68	-0.002537	2.254972	-0.939930	-0.392205
69	-0.015406	-0.512343	-0.320114	-0.057480
70	-0.086186	-0.076765	-0.416960	-0.167409
71	-0.337850	-0.512343	0.628978	-0.015899
72	-0.056873	-0.512343	-0.016663	0.065963

73 rows × 4 columns

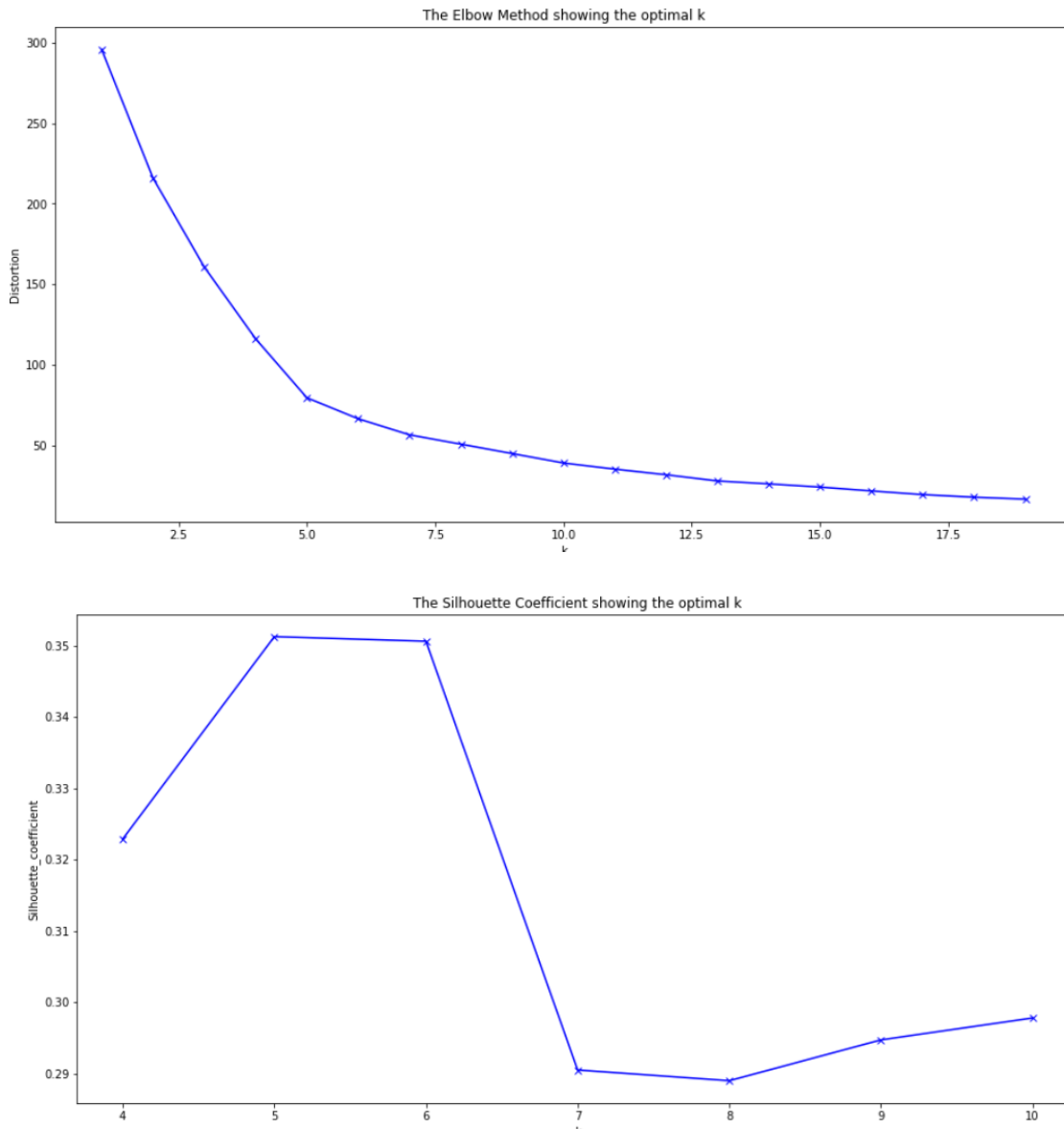
Final dataframe to be used for k means clustering (the neighborhood column will be removed):

	Surface_(ha)	Price	Population_Density_(hab/ha)	Crime_Proportion(Nr)	Neighborhood	Yoga Studio	Accessories Store	African Restaurant	American Restaurant	Antique Shop	...	Vacation Rental	Vegetarian / Vegan Restaurant	Venezuelan Restaurant	Video Game Store	Video Store	Vietnamese Restaurant	Wine Bar
0	-0.077607	-0.512343	1.139035	5.430410	el Raval	0.000000	0.00	0.0	0.0	0.0	...	0.0	0.02	0.0	0.00	0.000000	0.01	0.010000
1	-0.260635	0.529486	-0.494437	1.200080	el Barri Gòtic	0.000000	0.00	0.0	0.0	0.0	...	0.0	0.03	0.0	0.00	0.000000	0.00	0.020000
2	0.076823	-0.507536	-0.927017	1.169415	la Barceloneta	0.000000	0.00	0.0	0.0	0.0	...	0.0	0.00	0.0	0.00	0.000000	0.00	0.000000
3	-0.066167	-0.045034	-0.371765	2.120056	Sant Pere, Santa Caterina i la Ribera	0.010000	0.02	0.0	0.0	0.0	...	0.0	0.03	0.0	0.01	0.000000	0.00	0.000000
4	-0.198434	0.167466	0.644474	1.111981	el Fort Pienc	0.027778	0.00	0.0	0.0	0.0	...	0.0	0.00	0.0	0.00	0.000000	0.00	0.000000
...
68	-0.002537	2.254972	-0.939930	-0.392205	Diagonal Mar i el Front Marítim del Poblenou	0.000000	0.00	0.0	0.0	0.0	...	0.0	0.00	0.0	0.00	0.000000	0.00	0.000000
69	-0.015406	-0.512343	-0.320114	-0.057480	el Besòs i el Maritim	0.000000	0.00	0.0	0.0	0.0	...	0.0	0.00	0.0	0.00	0.000000	0.00	0.000000
70	-0.086186	-0.076765	-0.416960	-0.167409	Provençals del Poblenou	0.000000	0.00	0.0	0.0	0.0	...	0.0	0.00	0.0	0.00	0.000000	0.00	0.035714
71	-0.337850	-0.512343	0.628978	-0.015899	Sant Martí de Provençals	0.000000	0.00	0.0	0.0	0.0	...	0.0	0.00	0.0	0.00	0.000000	0.00	0.000000
72	-0.056873	-0.512343	-0.016663	0.065963	la Verneda i la Pau	0.000000	0.00	0.0	0.0	0.0	...	0.0	0.00	0.0	0.00	0.052632	0.00	0.000000

73 rows × 288 columns

- j. After the final dataframe is checked for any inconsistencies (ex NaN values) the optimum k mean clusters are determined, using the elbow method and Silhouette Coefficient. – clusters=5

Capstone Project - The Battle of Neighborhoods



- k. The cluster labels per neighborhood are added in the original non standardized dataframe and then merged with the most common venues dataframe as well:

Capstone Project - The Battle of Neighborhoods

	Neighborhood	District	Lat	Long	Cluster Labels	Surface_(ha)	Price	Population_Density_(hab/ha)	Crime_Proportion(Nr)	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	el Raval	Ciutat Vella	41.379518	2.168368	0	109.8	923.0	434.0	23962.0	Cocktail Bar	Tapas Restaurant	Spanish Restaurant	Bar	Mediterranean Restaurant	Café	Pizza Place	Restaurant	Bookstore	Hotel
1	el Barri Gòtic	Ciutat Vella	41.383395	2.176912	0	84.2	1139.7	181.0	7684.0	Tapas Restaurant	Ice Cream Shop	Plaza	Hotel	Bar	Italian Restaurant	Coffee Shop	Vegetarian / Vegan Restaurant	Market	Falafel Restaurant
2	la Barceloneta	Ciutat Vella	41.380553	2.189927	4	131.4	924.0	114.0	7566.0	Tapas Restaurant	Mediterranean Restaurant	Paelia Restaurant	Bar	Spanish Restaurant	Ice Cream Shop	Restaurant	Pizza Place	Burger Joint	Beach
3	Sant Pere, Santa Caterina i la Ribera	Ciutat Vella	41.388322	2.177411	0	111.4	1020.2	200.0	11224.0	Hotel	Tapas Restaurant	Japanese Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Bistro	Café	Burger Joint	Spanish Restaurant	Hostel
4	el Fort Pienc	Eixample	41.385925	2.182325	0	92.9	1064.4	357.4	7345.0	Coffee Shop	Chinese Restaurant	Restaurant	Spanish Restaurant	Yoga Studio	Smoke Shop	Sushi Restaurant	Supermarket	Stadium	Bistro
...
68	Diagonal Mar i el Front Marítim del Poblenou	Sant Martí	41.405228	2.213352	2	120.3	1498.6	112.0	1557.0	Mediterranean Restaurant	Restaurant	Italian Restaurant	Hot Dog Joint	Hotel	Beach	Beach Bar	Football Stadium	Bistro	Buffet
69	el Besòs i el Marçnes	Sant Martí	41.414979	2.213877	4	118.5	923.0	208.0	2845.0	Hotel	Diner	Bakery	Supermarket	Spanish Restaurant	Gym / Fitness Center	Climbing Gym	Big Box Store	Restaurant	Fish & Chips Shop
70	Provençals del Poblenou	Sant Martí	41.411948	2.204125	4	108.6	1013.6	193.0	2422.0	Spanish Restaurant	Café	Asian Restaurant	Garden Center	Sports Club	Restaurant	Falafel Restaurant	Mediterranean Restaurant	Auto Workshop	Liquor Store
71	Sant Martí de Provençals	Sant Martí	41.416519	2.198968	3	73.4	923.0	355.0	3005.0	Spanish Restaurant	Supermarket	Soccer Field	Pizza Place	Coffee Shop	Grocery Store	Gym / Fitness Center	Tapas Restaurant	Tennis Court	Mediterranean Restaurant
72	la Verneda i la Pau	Sant Martí	41.423220	2.202940	3	112.7	923.0	255.0	3320.0	Grocery Store	Bakery	Coffee Shop	Café	Tapas Restaurant	Burger Joint	Fast Food Restaurant	Supermarket	Sandwich Place	Electronics Store

73 rows x 19 columns

4. Results visualization

The retrieved overall dataframe with the cluster labels will be:

- a. Divided into one dataframe per cluster
- b. Summarized in one new dataframe
- c. Used to plot markers on a folium map
- d. Then used (with a geojson file) to draw a choropleth map.

Cluster 0:

	Neighborhood	Surface_(ha)	Price	Population_Density_(hab/ha)	Crime_Proportion(Nr)	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	
0	el Raval	109.8	923.0		434.0	23962.0	Cocktail Bar	Tapas Restaurant	Spanish Restaurant	Bar	Mediterranean Restaurant	Café	Pizza Place	Restaurant	Bookstore	Hotel
1	el Barri Gòtic	84.2	1139.7		181.0	7684.0	Tapas Restaurant	Ice Cream Shop	Plaza	Hotel	Bar	Italian Restaurant	Coffee Shop	Vegetarian / Vegan Restaurant	Market	Falafel Restaurant
3	Sant Pere, Santa Caterina i la Ribera	111.4	1020.2		200.0	11224.0	Hotel	Tapas Restaurant	Japanese Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Bistro	Café	Burger Joint	Spanish Restaurant	Hotel
4	el Fort Pienc	92.9	1064.4		357.4	7345.0	Coffee Shop	Chinese Restaurant	Restaurant	Spanish Restaurant	Yoga Studio	Smoke Shop	Sushi Restaurant	Supermarket	Stadium	Bistro
5	la Sagrada Família	105.1	980.0		503.2	11701.0	Restaurant	Burger Joint	Hotel	Italian Restaurant	Wine Bar	Plaza	Bakery	Pizza Place	Sandwich Place	Café
6	la Dreta de l'Eixample	212.3	1348.9		205.4	9648.0	Hotel	Boutique	Mediterranean Restaurant	Hotel	Bistro	Tapas Restaurant	Japanese Restaurant	Bakery	Roof Deck	Sporting Goods Shop
7	l'Àngel Esquerra de l'Eixample	123.5	1194.4		341.7	9332.0	Hotel	Spanish Restaurant	Cocktail Bar	Mediterranean Restaurant	Japanese Restaurant	Bakery	Restaurant	Pizza Place	Tapas Restaurant	Peruvian Restaurant
8	la Nova Esquerra de l'Eixample	133.8	1083.7		437.8	12955.0	Mediterranean Restaurant	Indian Restaurant	Café	Japanese Restaurant	Supermarket	Korean Restaurant	Bakery	Gym	Tapas Restaurant	Bar
9	Sant Antoni	80.1	1018.4		483.7	8571.0	Cocktail Bar	Mediterranean Restaurant	Bar	Tapas Restaurant	Café	Hotel	Sandwich Place	Burger Joint	Coffee Shop	Spanish Restaurant

Cluster 1:

	Neighborhood	Surface_(ha)	Price	Population_Density_(hab/ha)	Crime_Proportion(Nr)	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
21	Valldirera, el Tibidabo i les Planes	1152.4	1355.5	4.0	349.0	Ice Cream Shop	BBQ Joint	Light Rail Station	National Park	Farmers Market	Plaza	Soccer Field	Building	Convenience Store	Art Gallery

Cluster 2:

	Neighborhood	Surface_(ha)	Price	Population_Density_(hab/ha)	Crime_Proportion(Nr)	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	
20	Pedralbes	270.2	2023.4		43.1	1207.0	Hot Dog Joint	Park	Spanish Restaurant	Mediterranean Restaurant	Garden	Hotel	Bar	Print Shop	College Cafeteria	Palace
22	Sant Martí	303.0	1449.5		81.0	1855.0	Bakery	Mediterranean Restaurant	Breakfast Spot	Café	Tapas Restaurant	Burger Joint	Pedestrian Plaza	Dessert Shop	Diner	Restaurant
23	les Tres Torres	78.8	1745.0		211.0	1253.0	Hotel	Mexican Restaurant	Restaurant	Japanese Restaurant	Café	Burger Joint	Spanish Restaurant	Sports Club	Gastropub	Sushi Restaurant
24	Sant Gervasi - la Bonanova	223.5	1361.6		114.0	1923.0	Seabood Restaurant	Mediterranean Restaurant	Bakery	Spanish Restaurant	Japanese Restaurant	Restaurant	Pizza Place	Bar	Grocery Store	Burger Joint
25	Sant Gervasi - Galvany	165.9	1455.9		283.0	3540.0	Mediterranean Restaurant	Deli / Bodega	Japanese Restaurant	Café	Restaurant	Grocery Store	Hotel	Sandwich Place	Spanish Restaurant	Italian Restaurant
66	la Vila Olímpica del Poblenou	92.5	1444.6		100.0	1068.0	Café	Hookah Bar	Mediterranean Restaurant	Restaurant	Pasta Restaurant	Cocktail Bar	Nightclub	Italian Restaurant	Bar	Middle Eastern Restaurant
68	Diagonal Mar i el Front Marítim del Poblenou	120.3	1498.6		112.0	1557.0	Mediterranean Restaurant	Restaurant	Italian Restaurant	Hot Dog Joint	Hotel	Beach	Beach Bar	Football Stadium	Bistro	Buffet

Cluster 3:

	Neighborhood	Surface_Nh	Price	Population_Density_Nh	Crime_Proportion(Nr)	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	
19	el Poble Sec	84.3	923.0		424.8	5703.0	Tapas Restaurant	Bar	Theater	Café	Mediterranean Restaurant	Cocktail Bar	Pizza Place	Italian Restaurant	Gastropub	Burger Joint
13	la Font de la Guella	33.3	923.0		342.1	1437.0	Spanish Restaurant	Plaza	Hotel	Nightclub	Tapas Restaurant	Pizza Place	Café	Mediterranean Restaurant	Burger Joint	Breakfast Spot
14	Horta-Guinardó	41.0	923.0		385.1	2253.0	Tapas Restaurant	Spanish Restaurant	Bar	Hotel	Pizza Place	Bakery	Asian Restaurant	Breakfast Spot	Hot Dog Joint	
15	la Barceloneta	97.1	923.0		328.5	2617.0	Restaurant	Park	Mediterranean Restaurant	Supermarket	Gym / Fitness Center	Burger Joint	Hotel	Pizza Place	Performing Arts Venue	Coffee Shop
16	Santa Catalina	41.1	923.0		586.9	3427.0	Pizza Place	Tapas Restaurant	Café	Bakery	Supermarket	Bar	Hotel	Restaurant	Escape Room	Falafel Restaurant
17	Sant Martí	109.0	923.0		388.0	5900.0	Mediterranean Restaurant	Bar	Tapas Restaurant	Hotel	Plaza	Middle Eastern Restaurant	Wine Bar	Pie Shop	Italian Restaurant	Peruvian Restaurant
18	les Corts	141.3	1195.2		327.4	4558.0	Supermarket	Spanish Restaurant	Restaurant	Bakery	Café	Vegetarian / Vegan Restaurant	Tapas Restaurant	Asian Restaurant	Cocktail Bar	Italian Restaurant
26	el Poble Nou	84.0	1195.7		345.0	2187.0	Bakery	Burger Joint	Wine Bar	Italian Restaurant	Café	Mediterranean Restaurant	Tapas Restaurant	Supermarket	Spanish Restaurant	Breakfast Spot
27	la Vila de Gràcia	132.1	991.5		379.0	3999.0	Mediterranean Restaurant	Plaza	Pizza Place	Bar	Cocktail Bar	Tapas Restaurant	Ice Cream Shop	Arts & Crafts Store	Café	Bakery
31	el Camp de l'Enxaneta	65.1	1017.6		537.0	2127.0	Tapas Restaurant	Bakery	Restaurant	Japanese Restaurant	Bar	Spanish Restaurant	Mediterranean Restaurant	Hotel	Plaza	Lounge
32	el Barri Vell	96.0	923.0		486.9	1058.0	Japanese Restaurant	Café	Tapas Restaurant	Breakfast Spot	Bar	Tapas Restaurant	Bar	Grocery Store	Restaurant	Park
34	el Clot Nou	103.0	923.0		273.9	2883.0	Tapas Restaurant	Breakfast Spot	Gym	Pizza Place	Restaurant	Playground	Park	Coffee Shop	Café	Farmers Market
35	el Carmel	84.2	923.0		351.3	1823.0	Grocery Store	Coffee Shop	Supermarket	Café	Gym	Food Court	Plaza	Cupcake Shop	Escape Room	Food & Drink Shop
37	la Ribera	33.7	923.0		341.1	858.0	Metro Station	Spanish Restaurant	Mediterranean Restaurant	Bakery	Grocery Store	Farmers Market	Bar	Soccer Field	Playground	Flower Shop
42	Vilapetrera	85.4	923.0		457.0	1386.0	Tapas Restaurant	Grocery Store	Café	Bar	Bakery	Food & Drink Shop	Asian Restaurant	Bookstore	Spanish Restaurant	
43	la Piedad	63.7	923.0		316.3	1403.0	Tapas Restaurant	Grocery Store	Plaza	Coffee Shop	Supermarket	Chef	Japanese Restaurant	Bookstore	Café	Spanish Restaurant
45	el Turó de la Penya	354.0	923.0		446.0	849.0	Pizza Place	Bar	Café	Supermarket	Tapas Restaurant	Burger Joint	German Restaurant	Pharmacy	Food	Italian Restaurant
50	Venúcia	23.7	923.0		877.0	877.0	Grocery Store	Park	Tapas Restaurant	Spanish Restaurant	Bar	Restaurant	Café	Sushi Restaurant	Supermarket	Baby Store
51	la Prosperitat	59.1	923.0		453.0	1448.0	Grocery Store	Spanish Restaurant	Tapas Restaurant	Park	Pizza Place	Athletics & Sports	Bakery	Building	Sports Club	Sushi Restaurant
54	Ciutat Vella	37.1	923.0		267.0	594.0	Metro Station	Park	Plaza	Grocery Store	Supermarket	Train Station	Women's Store	Farmers Market	Escape Room	Ethiopian Restaurant
58	Sant Andreu	180.7	923.0		310.3	4043.0	Spanish Restaurant	Bar	Grocery Store	Plaza	Coffee Shop	Restaurant	Café	Mexican Restaurant	Spanish Restaurant	Thai Restaurant
60	la Sagrada Família	98.0	923.0		286.0	2254.0	Tapas Restaurant	Grocery Store	Tapas Restaurant	Restaurant	Bar	Bakery	Burger Joint	Plaza	Pizza Place	Performing Arts Venue
61	el Camp de l'Enxaneta	40.9	923.0		354.0	1052.0	Tapas Restaurant	Spanish Restaurant	Plaza	Restaurant	Pizza Place	Grocery Store	Bakery	Bakery	Furniture / Home Store	
62	Naveas	42.4	923.0		523.0	1779.0	Supermarket	Spanish Restaurant	Tapas Restaurant	Bakery	Grocery Store	Plaza	Restaurant	Italian Restaurant	BBC Joint	Mexican Restaurant
63	el Camp de l'Enxaneta	74.1	923.0		522.0	4458.0	Bakery	Hotel	Café	Restaurant	Grocery Store	Spanish Restaurant	Pizza Place	Mediterranean Restaurant	Burger Joint	Italian Restaurant
64	el Clot	69.7	923.0		388.0	3118.0	Spanish Restaurant	Café	Tapas Restaurant	Mediterranean Restaurant	Restaurant	Coffee Shop	Hotel	Supermarket	Dessert Shop	Cocktail Bar
71	Sant Martí	14.0	923.0		395.0	3005.0	Spanish Restaurant	Supermarket	Soccer Field	Pizza Place	Coffee Shop	Grocery Store	Gym / Fitness Center	Tapas Restaurant	Tennis Court	Mediterranean Restaurant
72	la Verneda i la Pau	112.7	923.0		255.0	3320.0	Grocery Store	Bakery	Coffee Shop	Café	Tapas Restaurant	Burger Joint	Fast Food Restaurant	Supermarket	Sandwich Place	Electronics Store

Capstone Project - The Battle of Neighborhoods

Cluster 4:

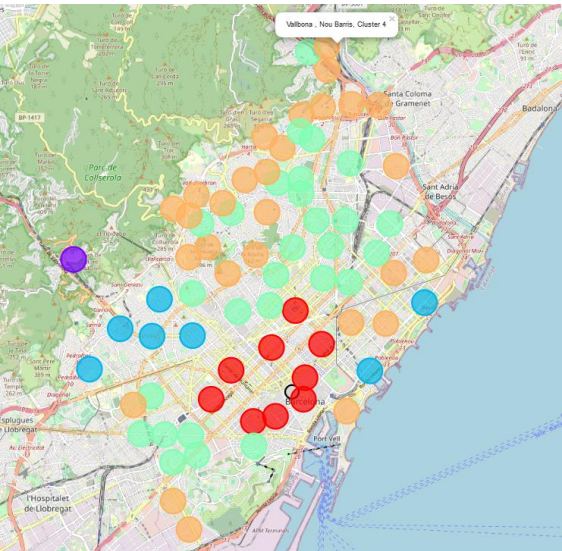
	Neighborhood	Surface_ha	Price	Population_Density_hab/ha	Crime_Proportion(h)	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	la Barceloneta	131.4	524.0	14.0	1085.0	Tapas Restaurant	Mediterranean Restaurant	Pasta Restaurant	Bar	Spanish Restaurant	Ice Cream Shop	Restaurant	Pizza Place	Burger Joint	Beach
11	la Marina del Port Vell	79.3	523.0	14.4	151.0	Spanish Restaurant	Restaurant	Food	Supermarket	Furniture / Home Store	Mediterranean Restaurant	Bus Station	Pizzeria Restaurant	Fabrics Shop	Faithful Restaurant
12	la Marina de Port	125.5	523.0	240.1	4189.0	Café	Spanish Restaurant	Bakery	Grocery Store	Hotel	Cheer	Farmers Market	Latin American Restaurant	Supermarket	Kids Store
19	la Marimar / San Ramon	180.3	1076.1	125.9	2462.0	Spanish Restaurant	Pizza Place	Italian Restaurant	Supermarket	Mediterranean Restaurant	Restaurant	Café	Park	Bakery	Becker Stadium
27	Vallbona / les Penitents	125.1	1028.9	128.9	2776.0	Café	Hotel	Grocery Store	Place	Supermarket	Supermarket	Bakery	Smoke Shop	Snack Place	Burger Joint
28	el Coll	35.3	523.0	211.0	985.0	Hotel	Park	Boat/Lookout	Mediterranean Restaurant	Supermarket	Mountain	Farmers Market	Escape Room	Ethiopian Restaurant	Food Court
29	la Salut	84.8	554.4	208.0	1055.0	Bar	Sandwich Place	Historic Site	Park	Spanish Restaurant	Boat/Lookout	Restaurant	Supermarket	Bakery	Museum
33	Can Bori	38.4	523.0	238.7	532.0	Spanish Restaurant	Grocery Store	Restaurant	Chinese Restaurant	Sports Stadium	Tease Restaurant	Tease Restaurant	Big Box Store	Fried Chicken Joint	Breakfast Spot
35	la Font del Pagadet	65.7	1044.0	145.2	599.0	Pizza Place	Supermarket	Restaurant	Italian Restaurant	Hot Dog Joint	Coffee Shop	Spanish Restaurant	Pizza	Tapas Restaurant	Del / Bodega
38	Sant Gervasi dels Abadells	171.6	523.0	42.1	422.0	Bakery	Farmers Market	Tennis Court	Café	Tease Restaurant	Grocery Store	Pizza Market	Fish & Chips Shop	Farmers Restaurant	Empire Restaurant
39	Montbau	254.7	523.0	25.7	358.0	Place	Park	Gym	Restaurant	Hotel	Breakfast Spot	Tennis Court	Stadium	Mediterranean Restaurant	Metro Station
40	la Vall d'Hebron	71.9	554.5	75.3	327.0	Spanish Restaurant	Bakery	Place	Athletics & Sports	Soccer Stadium	Soccer Field	Stadium	Café	Farmers Market	Bar
41	la Ciutat	17.8	523.0	28.9	28.0	Food Court	Supermarket	Café	Park	Grocery Store	Gym	Coffee Shop	Hotel	Spanish Restaurant	Place
42	Horta	358.2	523.0	80.9	1622.0	Tapas Restaurant	Spanish Restaurant	Pizza Place	Mediterranean Restaurant	Sandwich Place	Burger Joint	Supermarket	Sports Club	Bevairy	Metro Station
46	Can Peguera	91.9	523.0	193.0	124.0	Park	Food & Drink Shop	Supermarket	Sandwich Place	Escape Room	Tapas Restaurant	Café	Grocery Store	Gym / Fitness Center	German Restaurant
47	la Guineueta	81.2	523.0	251.0	630.0	Place	Grocery Store	Seafood Restaurant	Tapas Restaurant	Bakery	Spanish Restaurant	Cheer	Soccer Field	Snack Place	Breakfast Spot
48	Canyelles	79.0	523.0	87.0	372.0	Soccer Field	Market	Food & Drink Shop	Café	Grocery Store	Tapas Restaurant	Mediterranean Restaurant	Skate Park	Place	Pizzeria Restaurant
49	les Roquetes	84.3	523.0	251.0	573.0	Place	Restaurant	Grocery Store	Baby Store	Gym	Faithful Restaurant	Gym Pool	Diner	Castle	Supermarket
52	la Trinitat Nova	57.7	523.0	152.0	410.0	Grocery Store	Metro Station	Breakfast Spot	Music Venue	Place	Spanish Restaurant	Pharmacy	Supermarket	Fabrics Shop	Pizzeria Restaurant
53	Torre Baró	174.0	523.0	17.9	158.0	Metro Station	Construction & Landscaping	Moving Target	Train Station	Fish & Chips Shop	Fabrics Shop	Faithful Restaurant	Farmers Market	Fast Food Restaurant	Farmers Restaurant
55	Vallbona	80.8	523.0	23.0	76.0	Park	Place	Food	Light Rail Station	Women's Store	Faithful Restaurant	Empire Restaurant	Escape Room	Ethiopian Restaurant	
56	la Trinitat Vella	80.8	523.0	127.0	625.0	Park	Soccer Field	Spanish Restaurant	Metro Station	Business Service	Tapas Restaurant	Women's Store	Fast Food Restaurant	Fabrics Shop	Faithful Restaurant
57	Baró de Viver	23.0	523.0	113.0	258.0	Place	Metro Station	Café	Track Stadium	Burger Joint	Supermarket	Park	Women's Store	Fast Food Restaurant	Fabrics Shop
58	el Bar Pastor	185.5	523.0	70.0	1049.0	Café	Bakery	Restaurant	Athletics & Sports	Metro Station	Food	Power Shop	Pizza Market	Fish & Chips Shop	Food & Drink Shop
60	el Parc de la Llibertat del Poblenou	162.2	1072.0	188.9	1783.0	Hotel	Bar	Restaurant	Spanish Restaurant	Clothing Store	Mediterranean Restaurant	Coffee Shop	Café	Music Venue	Tease Restaurant
67	el Besós / el Maritim	198.7	1061.3	218.0	3805.0	Spanish Restaurant	Mediterranean Restaurant	Restaurant	Bakery	Italian Restaurant	Pizza Place	Tease Restaurant	Breakfast Spot	Café	Place
69	el Besós / el Maritim	118.5	523.0	208.0	2845.0	Hotel	Diner	Bakery	Supermarket	Spanish Restaurant	Gym / Fitness Center	Clothing Gym	Big Box Store	Restaurant	Fish & Chips Shop
70	Provençals del Poblenou	108.6	1013.6	193.0	2422.0	Spanish Restaurant	Café	Asian Restaurant	Garden Center	Sports Club	Faithful Restaurant	Mediterranean Restaurant	Auto Workshop	Liquor Store	

The main characteristics of each cluster are summarized in the following table:

Cluster Nr	Mean Rent Price (€)	Mean Population Density (hab/ha)	Mean Crime Nr	1st most frequent venue	2nd most frequent venue	3rd most frequent venue	
0	0	1085.855556	349.355556	11380.222222	Hotel	Tapas Restaurant	Bar
1	1	1355.500000	4.000000	349.000000	Ice Cream Shop	BBQ Joint	Light Rail Station
2	2	1568.371429	134.871429	1771.857143	Mediterranean Restaurant	Mediterranean Restaurant	Bakery
3	3	942.607143	393.082143	2559.250000	Tapas Restaurant	Spanish Restaurant	Tapas Restaurant
4	4	952.092857	132.314286	1321.357143	Spanish Restaurant	Bakery	Restaurant

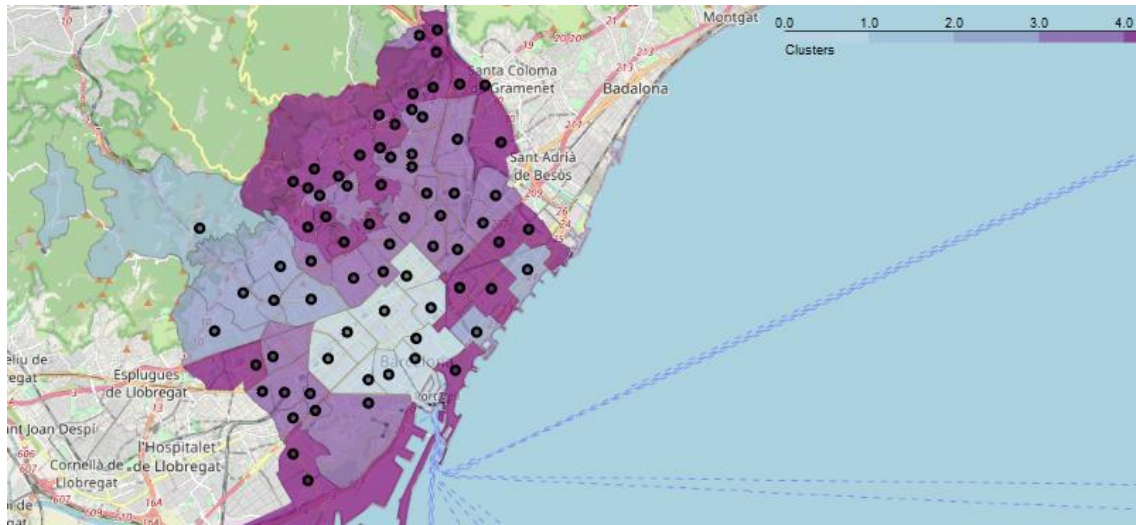
And two versions of maps are plotted:

Marker map:



Capstone Project - The Battle of Neighborhoods

plus a choropleth:



5. Discussion and conclusion

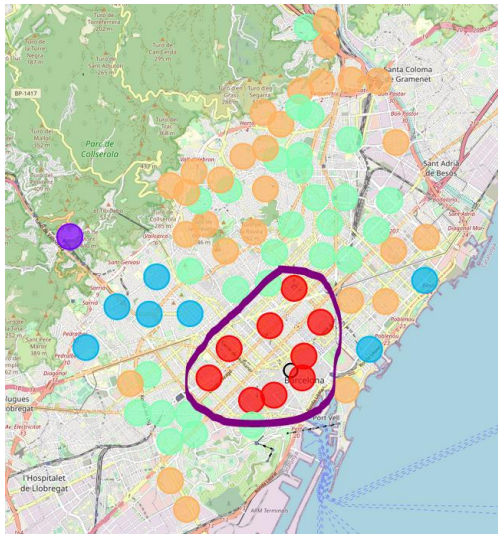
Though the implementation of the k means clustering algorithm on a properly prepared dataframe with barcelona neighborhoods data, a proposal of 5 main clusters (based on surface, population, crimes, rent prices and venues info) are returned.

The relevant question here would be: Do these clusters make sense, especially using a non supervised learning algorithm as kmeans?

The answer is yes, without of course excluding further model and input data improvement possibilities (please see conclusions). Let's try to explain the results through the point of view of a barcelona citizen.

Cluster Nr	Mean Rent Price (€)	Mean Population Density (hab/ha)	Mean Crime Nr	1st most frequent venue	2nd most frequent venue	3rd most frequent venue
0	1085.855556	349.355556	11380.222222	Hotel	Tapas Restaurant	Bar
1	1355.500000	4.000000	349.000000	Ice Cream Shop	BBQ Joint	Light Rail Station
2	1568.371429	134.871429	1771.857143	Mediterranean Restaurant	Mediterranean Restaurant	Bakery
3	942.607143	393.082143	2559.250000	Tapas Restaurant	Spanish Restaurant	Tapas Restaurant
4	952.092857	132.314286	1321.357143	Spanish Restaurant	Bakery	Restaurant

Cluster 0:

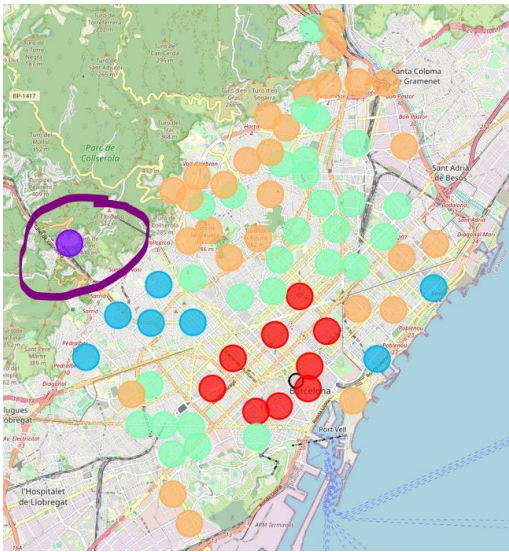


The area around the city center is defined as cluster “0”. It is densely populated, one of the main touristic areas and characterized mainly of old buildings and monuments (from 1500s up to early 1900s “modernistic”). It is observed that the results show exacty this, eg. Hotel and restaurants as the main venues. The higher criminality rate can be explained due to small scale tourism related thefts, frequent in some neighborhoods only (eg. in el Raval). The price rate is generally high due to the unique architecture of the buildings and high residence market demand.

Capstone Project - The Battle of Neighborhoods



Cluster 1:

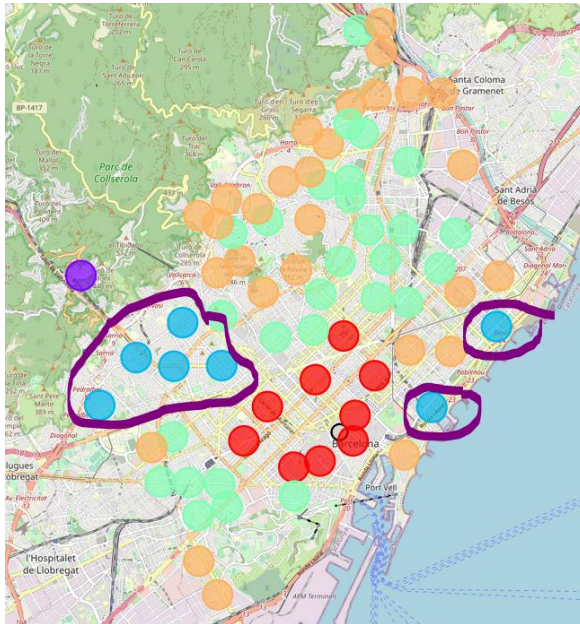


The neighborhood of Tibidabo, Vallvidrera and les Planes is forming one cluster (cluster "1"). This also makes sense. This neighborhood is more isolated from the rest of the city, located on a hill next to Barcelona. It has a low population density, expensive flats with an astonishing view over Barcelona. Its main attraction is a theme park next to an iconic church.



Capstone Project - The Battle of Neighborhoods

Cluster 2:

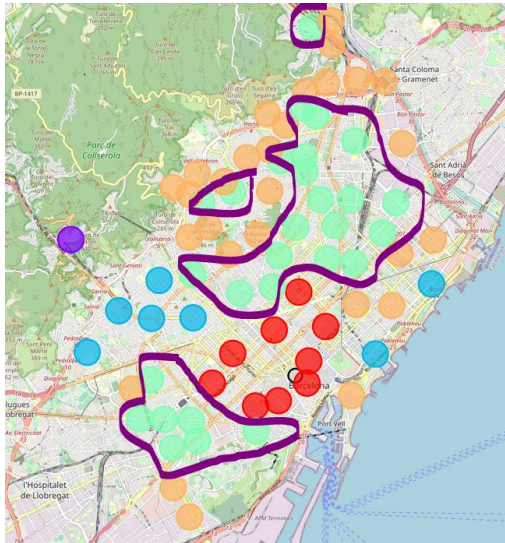


The cluster nr “2” includes the most expensive and “classy” areas of barcelona. It is mainly residential (plus some important companies HQ offices as well) with a lower population density than the rest of the city. High rent prices and venues such as restaurants and bakeries are expected as such.



Capstone Project - The Battle of Neighborhoods

Cluster 3:



Cluster 3 contains the residential neighborhoods of Barcelona with the highest population density and a low-moderate touristic interest. As expected the rent prices are somehow lower than the rest of the clusters and the most common venues are local restaurants.



Cluster 4:

The final cluster contains residential areas with a lightly lower population density, where also some small industrial polygons are located. The neighborhoods in this area are also characterized by a large extension of building rehabilitation and

Capstone Project - The Battle of Neighborhoods

modernization activities. A little bit higher rent prices are observed versus cluster 3 and the venues are mostly “resident focused” (restaurants, markets etc).



General conclusion

73 Barcelona neighborhoods were clustered using an unsupervised machine learning algorithm (k – means) applied on a dataframe combining data from various sources.

The results are considered plausible from the point of view of a resident. Nevertheless, the following points could be improved in a future repetition of the analysis:

- Crime rate was originally available at district level and was extrapolated to neighborhood level.
- Several rent price data values were missing or were incomplete and were substituted by mean values.
- The population data was from 2015 and did not include further demographic elements, such as age, profession, education, unemployment rate or origin of population.
- Other considerations such as schools or hospitals in the neighborhood were absent from the database.

Thank you for your attention.

P.S. Images were downloaded from Google, from various sources.