Barcelona | The City of Marvels | Neighborhoods, districts, data and machine learning

1. Introduction

Barcelona is well known as one of the most iconic, unique, and beautiful architectural cities in the world. With its cultural value, art, and tradition, it's one of the most visited cities within Europe and all over the globe. It has a population of 1,6 million, and the second-most populous urban after Madrid. According to Statista, the number of passengers at Barcelona airport in 2019 was 52,7 million, and international tourist spending in 2018 of 7,86 Billion USD.

Barcelona it's a hub of new trends in the world of culture, fashion, and cuisine.

Barcelona, through Open Data BCN, exposes information generated or guarded by public entities and gives full data access for the common good and benefit of anyone and any entity interested. Thus, we would examine and analyze this information looking forward to finding the best approach for the project objective.

The objective of this project is to determine what would be the best district to open a new restaurant/coffee shop in Barcelona and take advantage with full potential, as per the international presence in the city.

To evaluate this is important to take into consideration the following items:

- The population of each district
- Demographic information, age, education, university zones.
- The most common type of food
- Location of each district and its restaurants
- Number of restaurants in each district

2. Data

The solution to the problem should be data-driven to mitigate any risk and subjective approach. Besides, it's an opportunity to apply the tools and techniques learned so far:

- Districts of Barcelona vía Wikipedia
- Barcelona demographic information via Open Data Barcelona
- Venues for each Barcelona district via Foursquare API

2.1. Packages

The following libraries would be used in the next steps, as support for gathering, preparing, and describing the data such as plotting and rendering maps, vertical or horizontal bars for descriptive analysis. Last step, evaluating the model with proper recommendations.

Libraries
Pandas, Numpy
Pgeocode to obtain latitude, longitude, point data
Folium for rendering
Sklearn.cluster for clustering algorithm
Matplotlib for plotting

3. Methodology

The main data obtained from the 'postal codes' table, gather the 10 districts of Barcelona, population, and neighborhoods.

|	Number	District	Size km2	Population	Density inhabitants/km2	Neighbourhoods	Councilman[2]	Party
0	1	Ciutat Vella	4.49	111290.0	24786.0	La Barceloneta, El Gòtic, El Raval, Sant Pere,...	Jordi Rabassa i Massons	Barcelona en Comú
1	2	Eixample	7.46	262485.0	35586.0	L'Antiga Esquerra de l'Eixample, La Nova Esque...	Jordi Martí Grau	Barcelona en Comú
2	3	Sants-Montjuïc	21.35	177636.0	8321.0	La Bordeta, la Font de la Guatlla, Hostafrancs...	Marc Serra Solé	Barcelona en Comú
3	4	Les Corts	6.08	82588.0	13584.0	les Corts, la Maternitat i Sant Ramon, Pedralbes	Xavier Marcé Carol	Socialists' Party of Catalonia
4	5	Sarrià-Sant Gervasi	20.09	140461.0	6992.0	El Putget i Farró, Sarrià, Sant Gervasi - la B...	Albert Batlle i Bastardas	Socialists' Party of Catalonia
5	6	Gràcia	4.19	120087.0	28660.0	Vila de Gràcia, el Camp d'en Grassot i Gràcia ...	Eloi Badia i Casas	Barcelona en Comú
6	7	Horta-Guinardó	11.96	169920.0	14217.0	El Baix Guinardó, El Guinardó, Can Baró, El Ca...	Rosa Alarcón	Socialists' Party of Catalonia
7	8	Nou Barris	8.04	164981.0	20520.0	Can Peguera, Canyelles, Ciutat Meridiana, La G...	Marga Marí-Klose	Socialists' Party of Catalonia
8	9	Sant Andreu	6.56	142598.0	21737.0	Baró de Viver, Bon Pastor, El Congrés i els In...	Lucía Martín González	Barcelona en Comú
9	10	Sant Martí	10.80	221029.0	20466.0	El Besòs i el Maresme, el Clot, El Camp de l'A...	David Escudé Rodríguez	Socialists' Party of Catalonia

Then I used a barplot to visualize the number of neighbors per district:
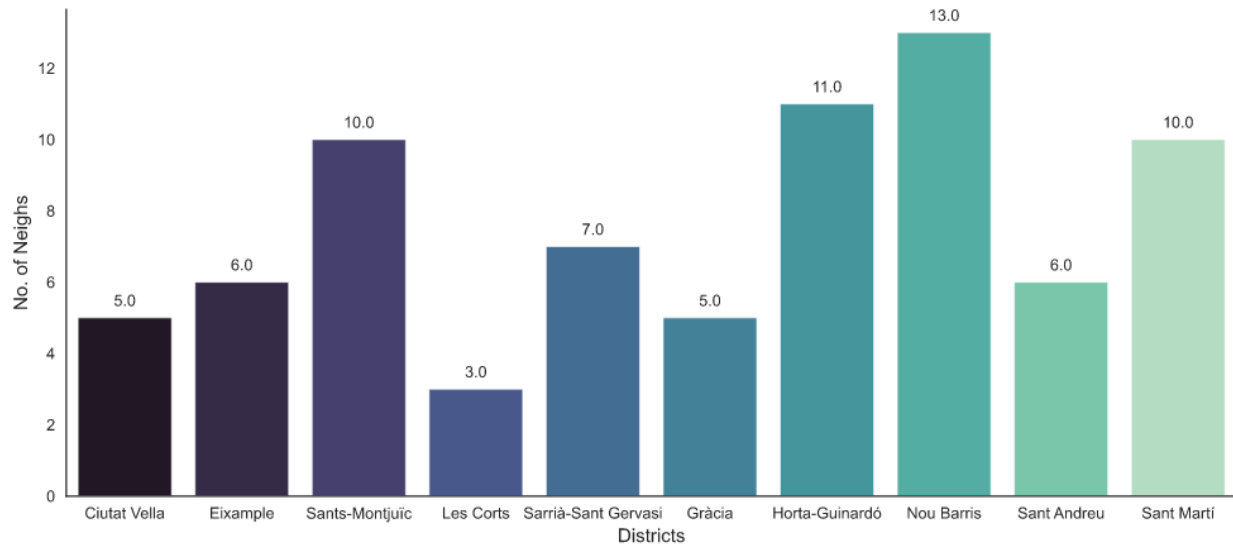
Then after doing some data cleaning and data wrangling and merging tables, I created a new data frame with key information such as population, latitude, longitude, the centroid for each district, plus the data obtained previously.

All data were obtained from Open Data Barcelona. I used the folium library, with a choropleth map to render the districts as per the population density, and markers package to popup the location and names of the neighborhoods.

The result was an interesting map, steps before data analysis and clustering.
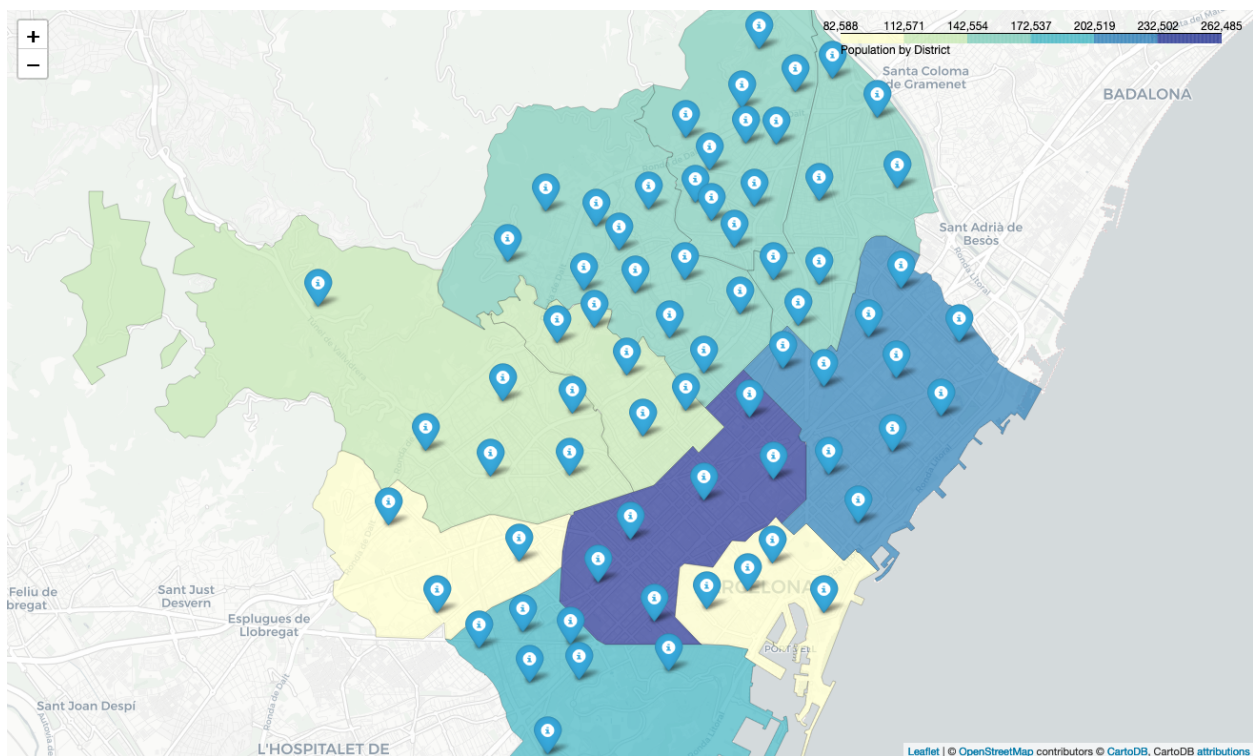
Venues exploration was obtained via Foursquare API with a limit of 100 venues within 500 meters radius. The results returned names of the venue, categories such as cinemas, Mediterranean restaurants, plazas, libraries, and so on.

For the Gràcia district, some of the venues:

	name	categories	lat	lng
0	A Casa Portuguesa	Pastry Shop	41.403977	2.155947
1	Viblioteca	Wine Bar	41.402957	2.157109
2	Karakala	Mediterranean Restaurant	41.402777	2.156505
3	Verdi Park	Indie Movie Theater	41.404396	2.157736
4	Cinemes Verdi	Indie Movie Theater	41.403944	2.156805

The analysis was applied to all Barcelona neighborhoods within districts. Each neighborhood was grouped and evaluated by the frequency of occurrence of each category.

Data frame resultant was the 73 neighborhoods of Barcelona with the 10th most common venues, as follows *[showing 5 first neighbors]*:

	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	Bar� de Viver	Park	Supermarket	Deli / Bodega	Restaurant	Dessert Shop	Asian Restaurant	Women's Store	Filipino Restaurant	Fabric Shop	Falafel Restaurant
1	Can Bar�	Spanish Restaurant	Grocery Store	Plaza	Scenic Lookout	Chinese Restaurant	Park	Café	Tapas Restaurant	Big Box Store	Bar
2	Can Peguera	Park	Hostel	Metro Station	Food & Drink Shop	Fast Food Restaurant	Escape Room	Restaurant	German Restaurant	Supermarket	Plaza
3	Canyelles	Plaza	Grocery Store	Soccer Field	Mediterranean Restaurant	Skate Park	Market	Café	Brewery	Dessert Shop	Deli / Bodega
4	Ciutat Meridiana	Metro Station	Park	Grocery Store	Plaza	Train Station	Supermarket	Mediterranean Restaurant	Women's Store	Event Space	Fabric Shop

4. Machine learning and Results

The algorithm used to do data segmentation was K-means, which groups unsupervised data as per the similarity of neighbors to each other.

After running the model with 5 clusters, each one was grouped by neighborhood and enumerated from 1 to the 10th most common venue. The resulting data frame below *[Barri as the number of the neighbor, N_Barri as the name, 'C_Distri' as the number of the District]*

|	BARRI	Latitude	Longitude	C_Distri	N_Barri	Cluster Labels	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	1	41.379877	2.168835	1	el Raval	0	Cocktail Bar	Spanish Restaurant	Tapas Restaurant	Bar	Pizza Place	Mediterranean Restaurant	Café	Burger Joint	Bookstore	Mexican Restaurant
1	2	41.382370	2.176181	1	el Barri G�tic	0	Spanish Restaurant	Tapas Restaurant	Plaza	Ice Cream Shop	Italian Restaurant	Mediterranean Restaurant	Café	Cocktail Bar	Coffee Shop	Bar
2	3	41.379373	2.189934	1	la Barceloneta	0	Paella Restaurant	Spanish Restaurant	Tapas Restaurant	Mediterranean Restaurant	Seafood Restaurant	Bar	Restaurant	Burger Joint	Beer Bar	Board Shop
3	4	41.386124	2.180597	1	Sant Pere, Santa Caterina i la Ribera	0	Tapas Restaurant	Hotel	Cocktail Bar	Café	Spanish Restaurant	Plaza	Bar	French Restaurant	Coffee Shop	Ice Cream Shop
4	5	41.397414	2.180783	2	el Fort Pienc	0	Restaurant	Hotel	Spanish Restaurant	Coffee Shop	Chinese Restaurant	Beer Bar	Tapas Restaurant	Bistro	Breakfast Spot	Soccer Stadium

Exploring each cluster let's visualize the map for better understanding.

Fo rendering the map, I used the data frame from above with folium and stamen toner layer. For each color:

- Cluster 1: Green
- Cluster 2: Blue
- Cluster 3: Purple
- Cluster 4: Red
- Cluster 5: Dark green
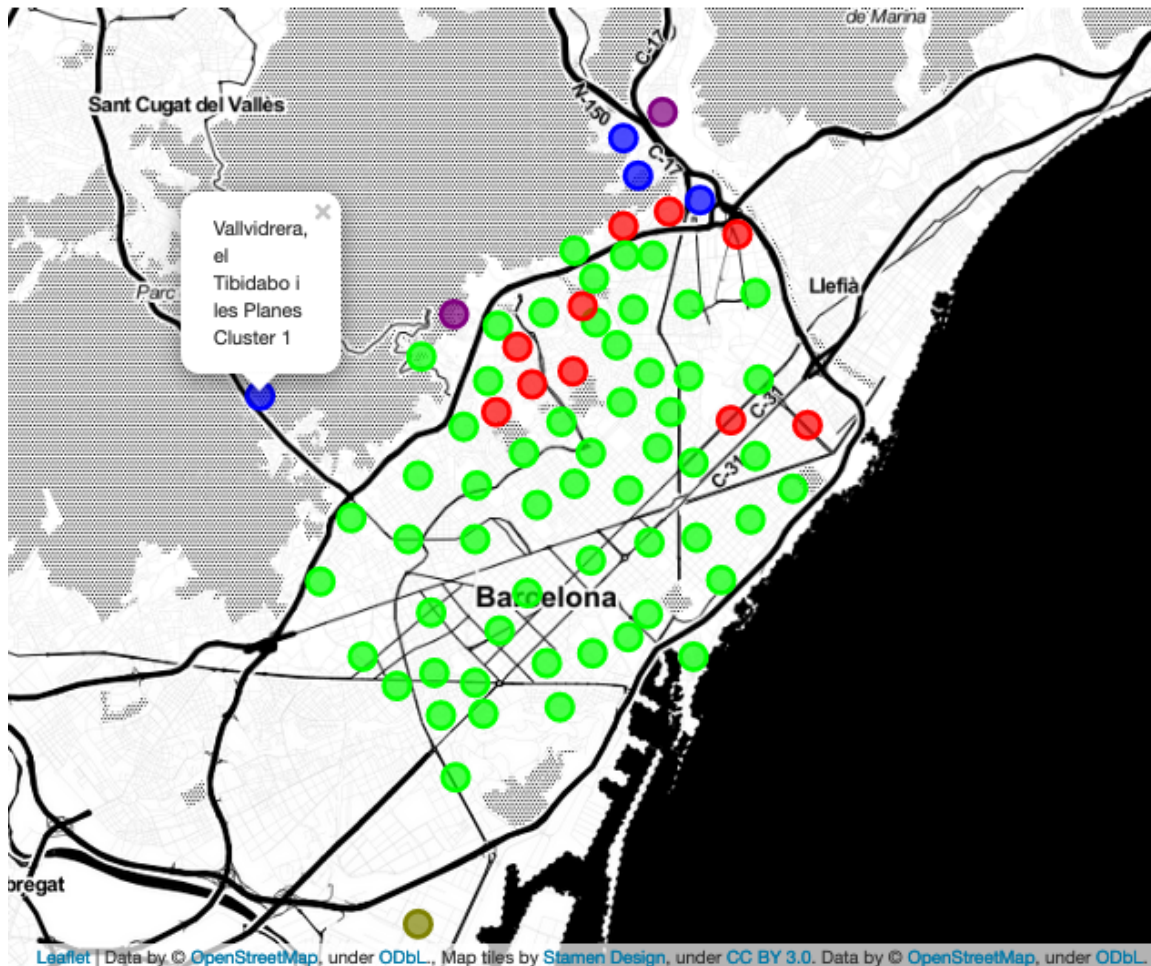
The neighborhood with the largest variety and most common venues was Cluster 1. If we examine the categories we found the top 4 venues are restaurants and coffee shops:
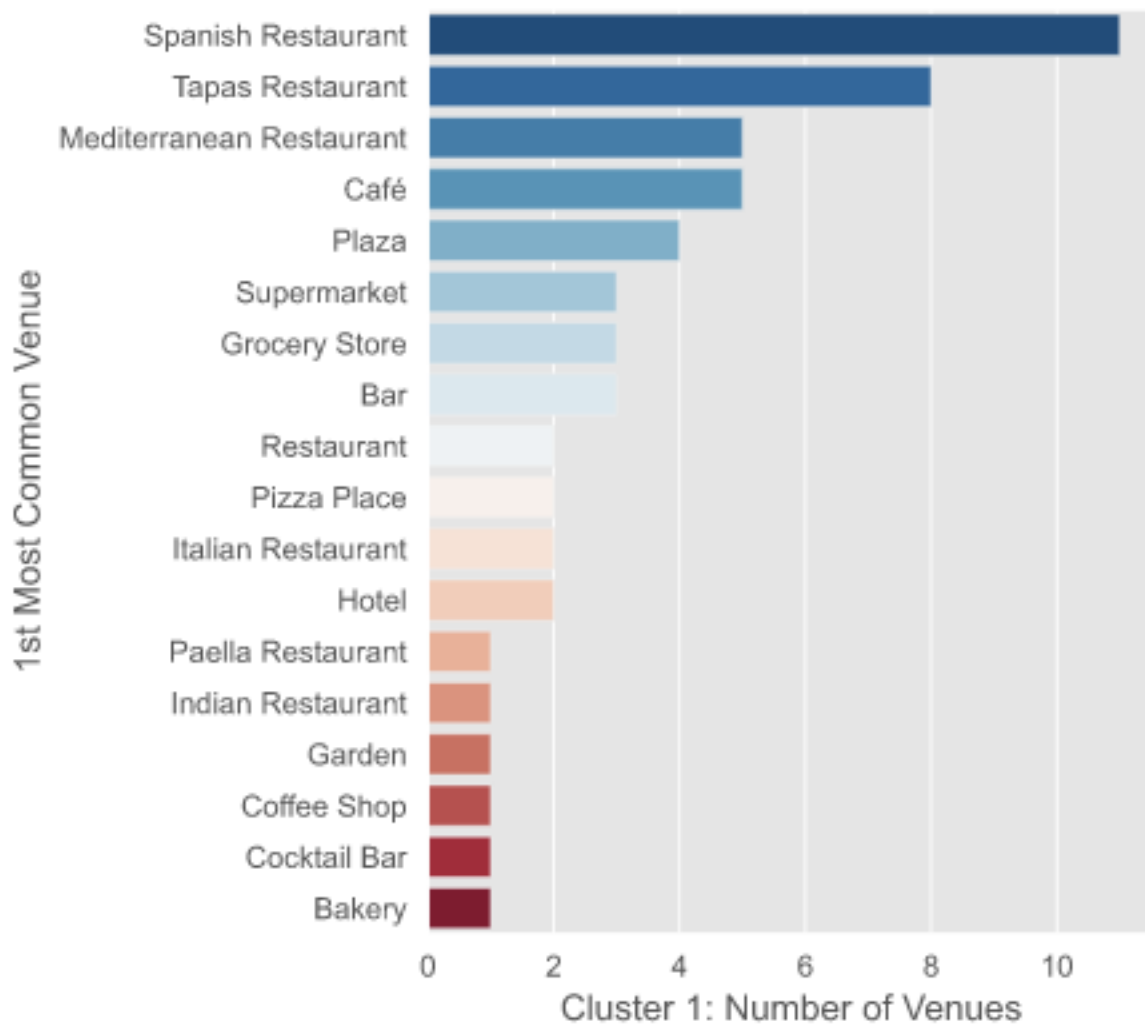
5. Discussion

Most of the restaurants and coffee shops locations are in Barcelona downtown, also in neighbors such as Ciutat Vella, Eixample, and Gràcia. As cluster 1 shows, having mostly Mediterranean restaurants, fusion, international, and so on. Coffee shops are a pleasant and typical venue due to the Spanish culture to have an espresso early in the morning, after lunch, and in the evening.

- Based on the criteria, descriptive statistics, and applying the K-means algorithm, the best neighborhood to start a new project restaurant or Coffee Shop in Barcelona is Eixample.
- Eixample has one of the largest populations with 262.485,0, a well-looking zone, safe, beautiful facades, very well connected through the public transport network.

- The second neighborhood is near the beach, in Poblenou, San Martí District.
- San Martí has the second-highest population with 221.029,0. Often is crowded by tourists or people hanging out with their families and friends near the beach.

6. Conclusions

Barcelona would keep its pace expanding and delivering new gastronomy experiences for all kinds of tastes for all the people who visit or live in it. While some Coffee shops are emerging, it is an attractive investment opportunity for a new startup project.

Based on the neighborhood analysis:

- Barcelona is the most multicultural and diverse city.
- Barcelonians are starting to acquiescence new places, different from downtown, non-traditional.
- Eixample and Sant Martí districts are the best neighborhoods to start a new project. Yet Barcelona has tons of secret places, like the Gràcia district with pedestrian-only roads, activities in plazas, and bohemian places.
- Eixample has the most diverse in terms of restaurants and coffee shops, at walking distance.

References

- [Open Data Barcelona](#)
- [Districts of Barcelona](#)
- [Foursquare API](#)