

# The Most Tailored Place to Spend your Holidays in Madrid

Ramon F-Reyes

## 0. Background

Madrid is one of the most touristic cities in Europe. Every year, hundreds of thousands of people visit the capital of Spain to enjoy the ambiance there, try delicious and traditional dishes and spend their time going sightseeing while benefiting from the good weather.

Nowadays, traveling to a different country on vacation is not seen as a high-class activity considering the appearance of low-cost airlines and the globalization process since the beginning of the century.

However, it is still difficult for people to optimize their housing expenses when they haven't stayed in that city before. Moreover, those readers that have visited Madrid at some point could agree with me that it is really difficult to visit every place in a short period of time. Apart from that, depending on the neighborhood your place is located on it could significantly affect the amount of money paid. With these considerations, new business opportunities for increasing customer satisfaction are awakening.

## 1. Business Problem

In this context, there is undoubtedly a problem to be solved for customers regarding the linkage between housing price and their requirements. By recommending to customers their most likely choices for accommodation considering their traveling expectations and budget, a solution to the problem of optimizing the housing expenses while enjoying the city may be provided.

Thus, the paramount scope of this project is to determine the most similar neighborhoods based on the facilities, amenities, and restaurants around them providing a defined price the customer would be willing to pay for staying at this neighborhood.

## 2. Machine Learning Solution

The purpose of this capstone project is to implement a well-known application of machine learning to define the most similar neighborhoods. This technique is called **clustering** and consists of developing an algorithm capable of recognizing similar groups in a huge dataset based on their characteristics (features).

## 3. Data Collection

Data on Madrid booking prices will be obtained through a Kaggle project in which Airbnb data is provided. By exploring this database, we are expecting to obtain the booking price per night for each place as well as their coordinates and address information. Apart from this data source, FourSquare API data will be used to explore and target recommended locations across different venues according to the existence of amenities and essential facilities.

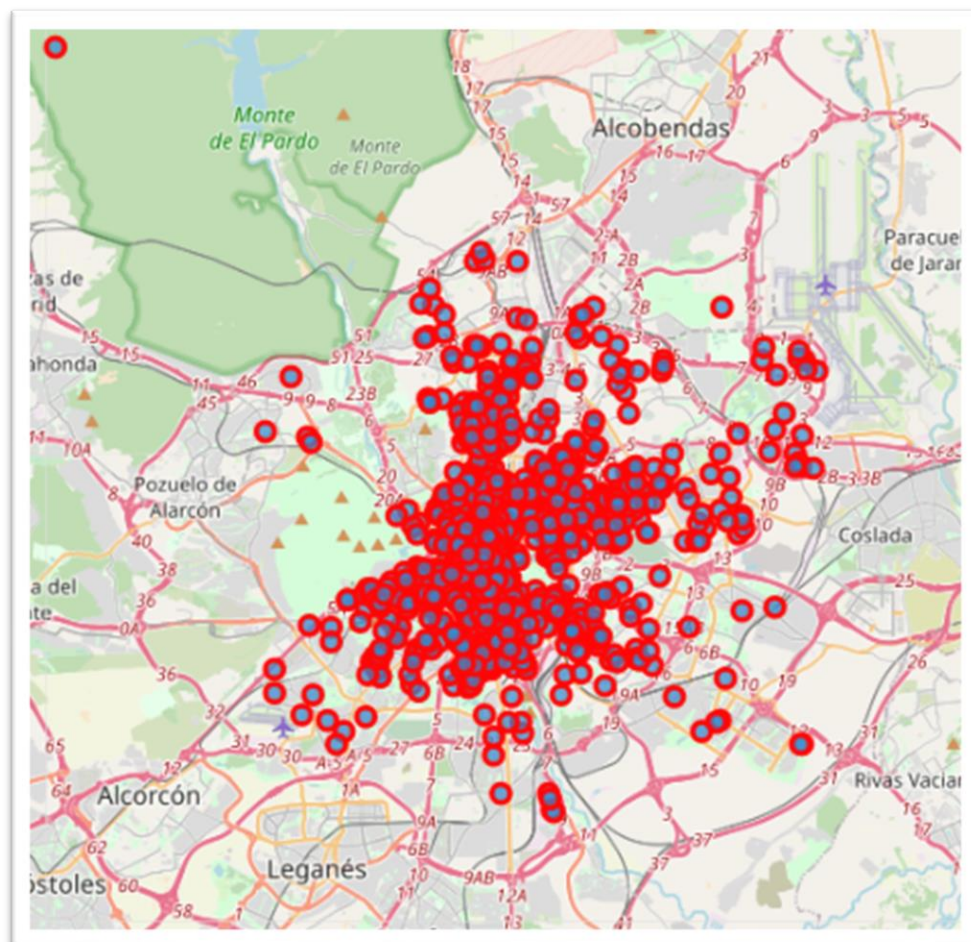
## A. AirBnb Data:

### DataFrame:

	id	name	neighbourhood	latitude	longitude	price
0	6369	Rooftop terrace room with ensuite bathroom, Airc.	Hispanoamérica	40.45628	-3.67763	70
1	21853	Bright and airy room	Cármenes	40.40341	-3.74084	17
2	24805	Gran Vía Studio Madrid	Universidad	40.42202	-3.70395	80
3	24836	Select the Madrid more "cool".	Justicia	40.41995	-3.69764	115
4	26825	Single Room whith private Bathroom	Legazpi	40.38985	-3.69011	25
...	...	...	...	...	...	...
20832	38725567	Acogedor piso de 2 Hab. Hospital La paz 4 Torres	Castilla	40.47847	-3.68395	100
20833	38726351	a 25-minutos del centro de Madrid near Subway	Palomeras Sureste	40.38533	-3.63968	24
20834	38727654	Apartamento Confortable a 300 m de la Gran Vía.	Universidad	40.42182	-3.70336	75
20835	38728727	Estudio en pleno Lavapiés	Embajadores	40.40896	-3.70107	45
20836	38730025	Chueca y gran vía central	Justicia	40.41973	-3.69666	75

20837 rows × 6 columns

### Map Visualization:



On this map are visualized all the potential apartments within a range price of 50-55 € per night.

## B. Nearby Venues obtained from FourSquare Data:

**DataFrame**

	Neighbourhood	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	Abrantes	Pizza Place	Fast Food Restaurant	Metro Station	Bakery	Electronics Store	Food Truck	Food & Drink Shop	Food	Farmers Market	Falafel Restaurant
1	Acacias	Pizza Place	Pub	Tapas Restaurant	Bar	Bookstore	Department Store	Dance Studio	Food Truck	Food & Drink Shop	Food
2	Adelfas	Hotel	Gym	Tapas Restaurant	Spanish Restaurant	Hawaiian Restaurant	Gymnastics Gym	Food & Drink Shop	Food	Fast Food Restaurant	Farmers Market
3	Aeropuerto	Spanish Restaurant	Bar	Grocery Store	Dumpling Restaurant	Food Truck	Food & Drink Shop	Food	Fast Food Restaurant	Farmers Market	Falafel Restaurant
4	Aguilas	Tapas Restaurant	Park	Yoga Studio	Dumpling Restaurant	Food & Drink Shop	Food	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Entertainment Service
5	Alameda de Osuna	Market	Bar	Dog Run	Restaurant	Yoga Studio	Food Truck	Food & Drink Shop	Food	Fast Food Restaurant	Farmers Market
6	Almagro	Spanish Restaurant	Café	Tea Room	Bar	Breakfast Spot	Nightclub	Cocktail Bar	Art Gallery	Spa	Italian Restaurant
7	Almenara	Beer Bar	Bar	Yoga Studio	Entertainment Service	French Restaurant	Food Truck	Food & Drink Shop	Food	Fast Food Restaurant	Farmers Market
8	Almendrales	Spanish Restaurant	Grocery Store	Pub	Fast Food Restaurant	Gastropub	Women's Store	Gym Pool	Dive Bar	Historic Site	Farmers Market
9	Amposta	Falafel Restaurant	Café	Bar	Bakery	Yoga Studio	French Restaurant	Food Truck	Food & Drink Shop	Food	Fast Food Restaurant

## 4. Methodology

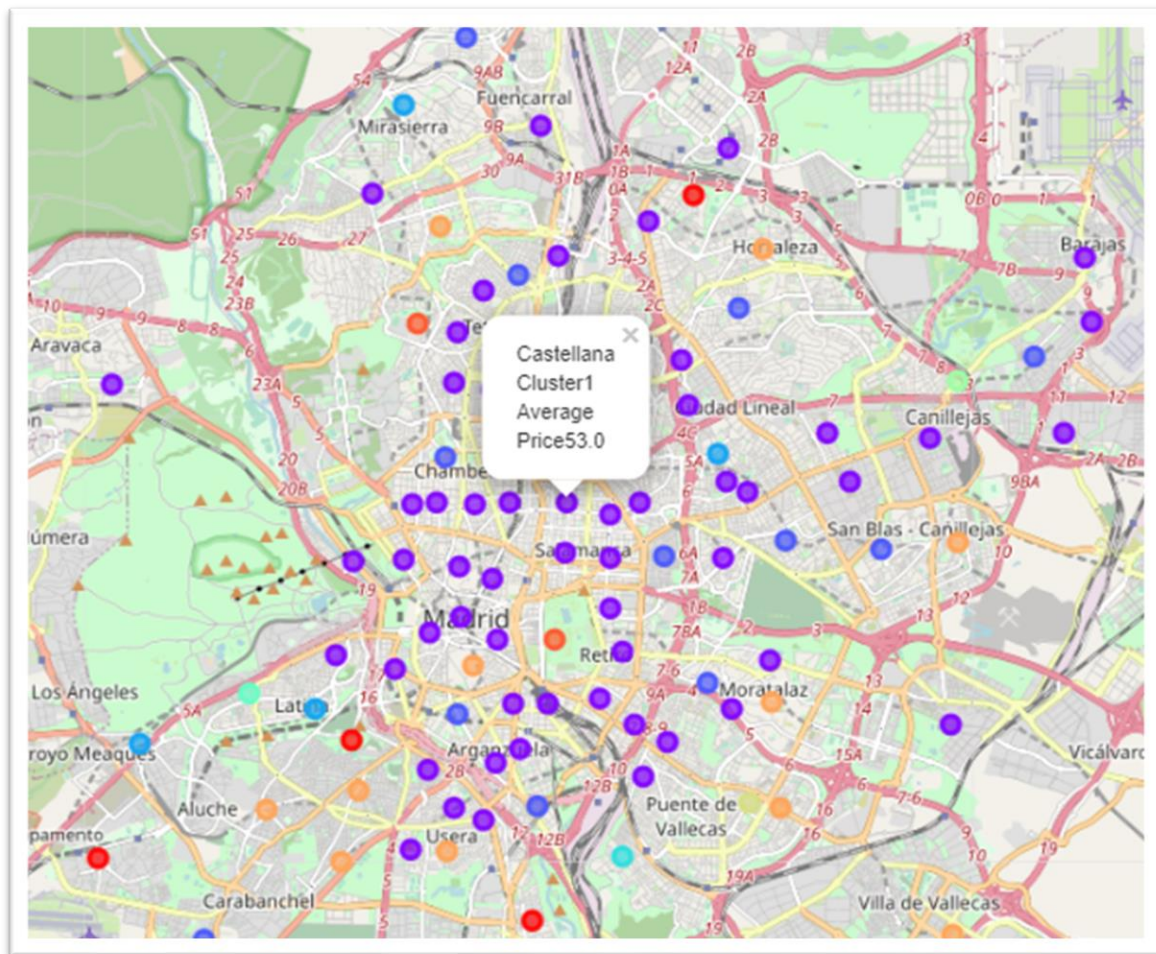
After exploring the dataset and gaining insights into it, we are ready to use the clustering methodology to analyze accommodation choices. We will use the k-means clustering technique as it is fast and efficient in terms of computational cost, is highly flexible to account for eventual mutations.

**Dataframe** merging Airbnb Data, Cluster Labels and Top 10 common venues.

	Neighbourhood	latitude	longitude	price	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	Abrantes	40.380030	-3.728640	52.500000	8	Pizza Place	Fast Food Restaurant	Metro Station	Bakery	Electronics Store	Food Truck	Food & Drink Shop	Food	Farmers Market	Falafel Restaurant
1	Acacias	40.402128	-3.705378	52.000000	2	Pizza Place	Pub	Tapas Restaurant	Bar	Bookstore	Department Store	Dance Studio	Food Truck	Food & Drink Shop	Food
2	Adelfas	40.400598	-3.670360	52.000000	1	Hotel	Gym	Tapas Restaurant	Spanish Restaurant	Hawaiian Restaurant	Gymnastics Gym	Food & Drink Shop	Food	Fast Food Restaurant	Farmers Market
3	Aeropuerto	40.461305	-3.580210	51.000000	1	Spanish Restaurant	Bar	Grocery Store	Dumpling Restaurant	Food Truck	Food & Drink Shop	Food	Fast Food Restaurant	Farmers Market	Falafel Restaurant
4	Aguilas	40.380547	-3.776523	51.666667	0	Tapas Restaurant	Park	Yoga Studio	Dumpling Restaurant	Food & Drink Shop	Food	Fast Food Restaurant	Farmers Market	Historic Site	Entertainment Service
5	Alameda de Osuna	40.455970	-3.591565	51.500000	2	Market	Bar	Dog Run	Restaurant	Yoga Studio	Food Truck	Food & Drink Shop	Food	Fast Food Restaurant	Farmers Market
6	Almagro	40.434014	-3.695172	51.666667	1	Spanish Restaurant	Café	Tea Room	Bar	Breakfast Spot	Nightclub	Cocktail Bar	Art Gallery	Spa	Italian Restaurant
7	Almenara	40.468239	-3.693510	51.714286	2	Beer Bar	Bar	Yoga Studio	Entertainment Service	French Restaurant	Food Truck	Food & Drink Shop	Food	Fast Food Restaurant	Farmers Market
8	Almendrales	40.386250	-3.700474	54.428571	1	Spanish Restaurant	Grocery Store	Pub	Fast Food Restaurant	Gastropub	Women's Store	Gym Pool	Dive Bar	Historic Site	Farmers Market
9	Amposta	40.427170	-3.621720	50.000000	2	Falafel Restaurant	Café	Bar	Bakery	Yoga Studio	French Restaurant	Food Truck	Food & Drink Shop	Food	Fast Food Restaurant

The number of clusters range from 0 to 10 providing the great number of different neighborhoods and in order to prevent underfitting.

### Map Visualization of the results:



## 5. Conclusion

Finally, the result of the project gives us a classification for every neighborhood by labeling them in 10 different clusters. Those more similar based on the type of facilities nearby are tagged as the same cluster. Moreover, the customer can set the range of price per night she is willing to pay for accommodation, so the average price of the neighborhood's apartment is also displayed.

In conclusion, this Capstone Project is intended to provide the most tailored place for customers to stay when spending their holidays in Madrid. One result directly obtained from this project is that in the center of Madrid most of the places count with the same kind of venues around. As we move down to the south, another type of neighborhood exists that the customer can choose based on their holiday requirements and paying a similar amount of money.