

CasaGrande's Big Launch in Barcelona

The Battle of Neighborhoods
Capstone Project

Introduction

- Barcelona is a city in Spain. It is the capital and largest city of the autonomous community of Catalonia, as well as the second most populous municipality of Spain.
- With 8.3 million visitors in 2015, Barcelona is one of Europe's most popular cities. Barcelona's popularity has in fact been steadily increasing since 2010 (which saw 7.13 million visitors), and tourist numbers have more than doubled since 2000.

Introduction Cont..

1. Economy

The current state of the Spanish economy is a good indicator of the overall health of the Barcelona property market: loan defaults in Spanish banks are down, which indicates that Spanish people are earning more, meaning that the country's economy is gaining momentum; this is backed by the country's predicted GDP increase of 3.2%, which is one of the fastest increase rates in Europe.

2. Destination

Due to the city's popularity as a holiday destination, luxury property is an especially profitable part of the Barcelona real estate market: as the 12th most visited city in the world, and among the top 5 most visited cities in Europe, Barcelona will likely always have a healthy interest from high-net-worth foreigners looking to buy or rent high-end properties in the city.

3. Infrastructure

Barcelona's infrastructure should also be taken into consideration when thinking about investment. Transport and communication links in the city are exemplary, with high-functioning airports, high-speed rail networks, and busy trade ports bolstering the city's economy.

Problem Statement and Target Audience

From all the fun facts, looks like Barcelona seems to be a great place to bloom the real estate business. CasaGrande is a real estate agency based out of Italy and it is looking to expand its horizon in Spain, starting with Barcelona which is the 2nd largest city. Barcelona, the cosmopolitan capital of Spain's Catalonia region, is known for its art and architecture.

For CasaGrande to be successful in Barcelona, it needs to analyse the neighborhood and find its initial launch location. Thus the real estate company has decided to build out a data product and derive its conclusions to help them understand which areas would generate the most profit.

The Approach

- CasaGrande has decided to identify the top 5 neighborhood in the city of Barcelona to ground break it's business. Further then it will use it's best successful advertising and campaign power to build its stable business and retain loyal customers.
- For this the company will be leverage the machine learning techniques and data science to attain its goal which provides a great edge in supporting their decision in every step of the way

Data Description

1. GEONAMES.ORG

Geospatial and zipcode data was pulled from Geonames Geographical Database.

2. FourSquare

For this project we will use the knowledge acquired in course 9 to use the benefits of the FourSquare API to explore data from cities and neighborhoods, and data that is relevant for people who travel such as coffee shops, hotels, restaurants that can visit and be close, theaters and many places for which to have a choice.

3. National Institute of Statistics

We also will be using the data from <http://www.ine.es/dynt3/inebase/index.htm?padre=525> which provides us with population data in each Municipality in Barcelona. This data will help us map the population in each neighborhood and which in return will help us understand a strong neighborhood community where real estate will bloom.

Methodology

1. Data Processing

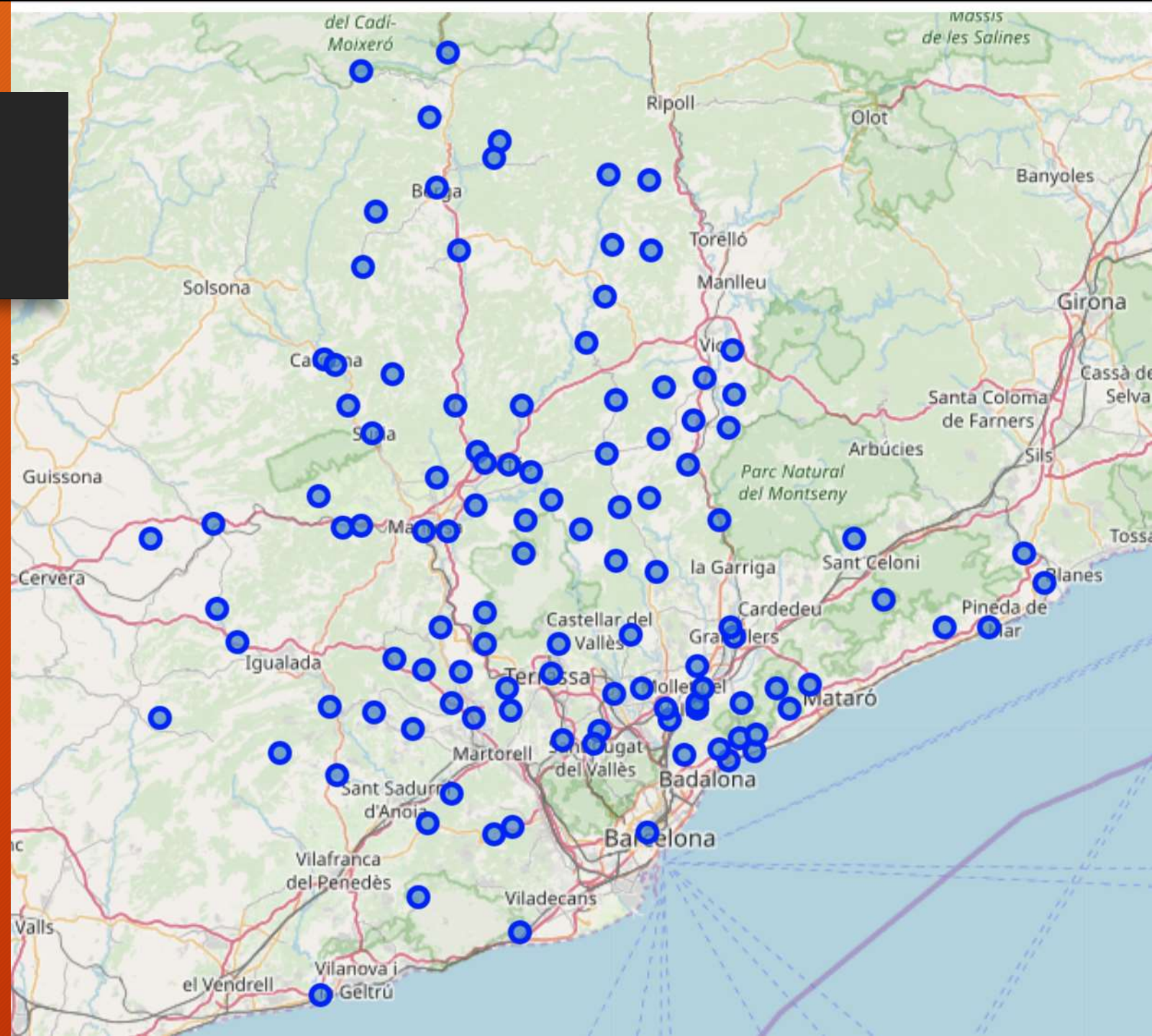
Data was extracted from the above mentioned sources and successfully placed in a DataFrame for further processing. The data from Geonames that contains the latitude and longitude needed further extraction to be placed in their respective columns. The coordinates that are inserted in the form of rows were placed in new columns named as latitude and longitude. Duplicates data was removed. Data from with the population details did not require any transformation and hence fitted perfectly in line with the geonames data.

	Neighborhood	Code	Country	State	Province	Latitude	Longitude
0	Barcelona	08080	Spain	Cataluna	Barcelona	41.389	2.159
1	Gallecs	08104	Spain	Cataluna	Barcelona	41.527	2.232
2	Martorelles	08107	Spain	Cataluna	Barcelona	41.533	2.233
3	La VallenÀsana	08110	Spain	Cataluna	Barcelona	41.476	2.214
4	La Llagosta	08120	Spain	Cataluna	Barcelona	41.514	2.193

Methodology Cont ..

2.Exploring the Data

Using geopy library to getting the latitude and longitude values of Barcelona. Creating a map of Barcelona with neighborhoods superimposed on top show that there is a good spread of points across the county. This done using folium



Methodology Cont ..

3. Analysing Neighborhoods in Barcelona

Lets use foursquare to let's get the top 100 venues that are in Barcelona within a radius of 500 meters and print each neighborhood along with the top 5 most common venues. In order to avoid overlapping, we decide to limit our search of businesses to 500 meters around each of these locations. We use the Foursquare API to create a dataframe of which includes, the business name, the type of business, and neighborhood in which it resides.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barcelona	41.389	2.159	Tandem	41.389533	2.157835	Cocktail Bar
1	Barcelona	41.389	2.159	La Patisserie	41.389442	2.161291	Pastry Shop
2	Barcelona	41.389	2.159	Robata	41.390551	2.158549	Japanese Restaurant
3	Barcelona	41.389	2.159	Foc i Oli	41.390034	2.156935	Sandwich Place
4	Barcelona	41.389	2.159	Sant Jordi Hostel Rock Palace	41.390297	2.160753	Hostel

Let's find out how many unique categories can be curated from all the returned venues and Let's check the size of the resulting dataframe

```
print('There are {} uniques categories.'.format(len(barca_venues['Venue Category'].unique())))
```

There are 147 uniques categories.

Cont..

Grouping the data by Neighbours and summarizing the data allows us to see the frequency of most popular categories of Venues.

Final step in this process would be to merge the data with Barcelona Data and render a new dataframe with top 10 venues for each neighborhood. This will serve as the input for K Means Clustering

3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
Sandwich Place	Supermarket	Bakery	Falafel Restaurant	Grocery Store
Food	Wine Shop	Café	Plaza	Japanese Restaurant
Bar	Flea Market	Fried Chicken Joint	Fountain	Food Truck
Plaza	Tapas Restaurant	Bus Station	Dog Run	Electronics Store
Coffee Shop	Spanish Restaurant	Gastropub	Café	Tapas Restaurant

```
---Barcelona---
      venue  freq
0  Cocktail Bar  0.08
1 Mediterranean Restaurant  0.06
2           Hotel  0.06
3 Spanish Restaurant  0.06
4  Tapas Restaurant  0.06
```

```
---Bellaterra---
      venue  freq
0 Mediterranean Restaurant  0.23
1           Trail  0.15
2           Café  0.15
3           Plaza  0.08
4 Gym / Fitness Center  0.08
```

```
---Berga---
      venue  freq
0           Plaza  0.25
1           Park  0.15
2           Café  0.15
3 Italian Restaurant  0.05
4       Grocery Store  0.05
```


K-means Clustering

K-means Clustering Algorithm is applied to partition the final data frame into $k(=10)$ partitions (clusters). The data was normalized using the MinMaxScaler from the sklearn library in Python and then fed to the KMeans cluster object to be fitted. We create another dataframe from this which shows the average amount of each business type per Neighborhood. Finally, we use the common types of businesses around each Neighborhood to put them into 10 clusters. In other words, each group contains Neighborhood with similar types of business.

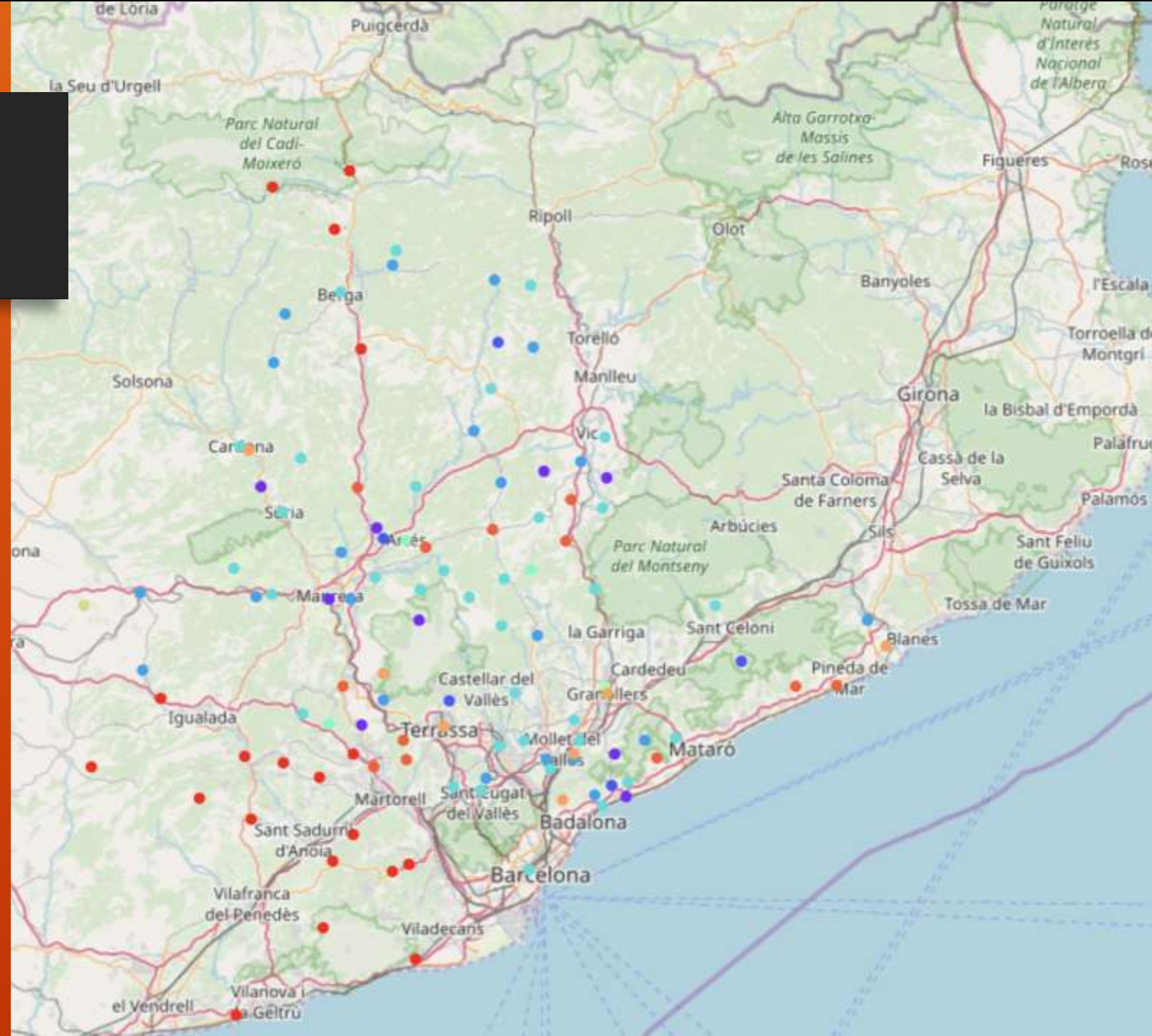
K	Count

0	18
1	9
2	5
3	21
4	34
5	3
6	1
7	1
8	7
9	12
10	1

	Province	Code	Latitude	Longitude	Cluster Labels	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	Barcelona	08552	41.875	2.287	1.0	Talamanca	Restaurant	Hotel	Mediterranean Restaurant	Fish & Chips Shop	Fountain	Food Truck	Food & Drink Shop	Food	Flower Shop	Flea Market
1	Barcelona	08188	41.533	2.300	1.0	Can Domenec (Tordera)	Restaurant	Spanish Restaurant	Fast Food Restaurant	Fountain	Food Truck	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish & Chips Shop
2	Barcelona	08650	41.813	1.906	1.0	Vilada	Restaurant	Women's Store	Fish & Chips Shop	Fountain	Food Truck	Food & Drink Shop	Food	Flower Shop	Flea Market	Fast Food Restaurant
3	Barcelona	08297	41.569	1.881	1.0	Martorelles	Restaurant	Women's Store	Fish & Chips Shop	Fountain	Food Truck	Food & Drink Shop	Food	Flower Shop	Flea Market	Fast Food Restaurant
4	Barcelona	08278	41.699	1.976	1.0	Granera	Restaurant	Women's Store	Fish & Chips Shop	Fountain	Food Truck	Food & Drink Shop	Food	Flower Shop	Flea Market	Fast Food Restaurant
5	Barcelona	08243	41.725	1.827	1.0	Copons	Restaurant	Paintball Field	Women's Store	Fish & Chips Shop	Food Truck	Food & Drink Shop	Food	Flower Shop	Flea Market	Fast Food Restaurant
6	Barcelona	08320	41.480	2.319	1.0	Monistrol De Calders	Restaurant	Mediterranean Restaurant	Women's Store	Fish & Chips Shop	Fountain	Food Truck	Food & Drink Shop	Food	Flower Shop	Flea Market
7	Barcelona	08269	41.864	1.714	1.0	Fonollosa	Restaurant	Performing Arts Venue	Women's Store	Fast Food Restaurant	Food Truck	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish & Chips Shop

Clustering

For the final piece of visualization, I leveraged the folium package by inputting the necessary details like the desired number of clusters, Neighborhood, lat and long as shown below



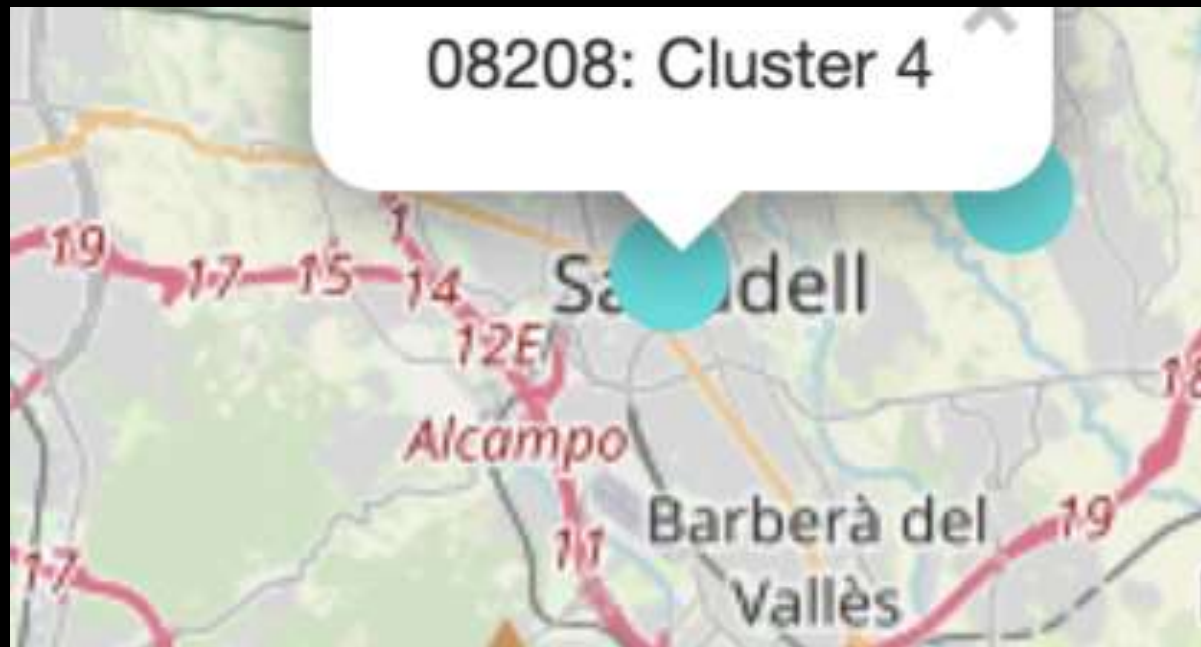
Examine the Clusters- Results

All clusters have restaurants in common, we going to pick a few in particular and check the features associated

- Cluster 0 - Features : Restaurants, Women's store, Fountain, Flee Market
- Cluster 4 - Features : High population, Near by neighborhoods like restaurants, super market, movie theatres, soccer field, Rock climbing
- Cluster 8 - Features : 2nd Highest population, Bus station, Restaurants, Plaza, Flee Market

It is seen that Barcelona and Sandell fall under the same cluster while Badalona and Terrassa fall under the same cluster. And these take 1 - 4 position in terms of population as well.

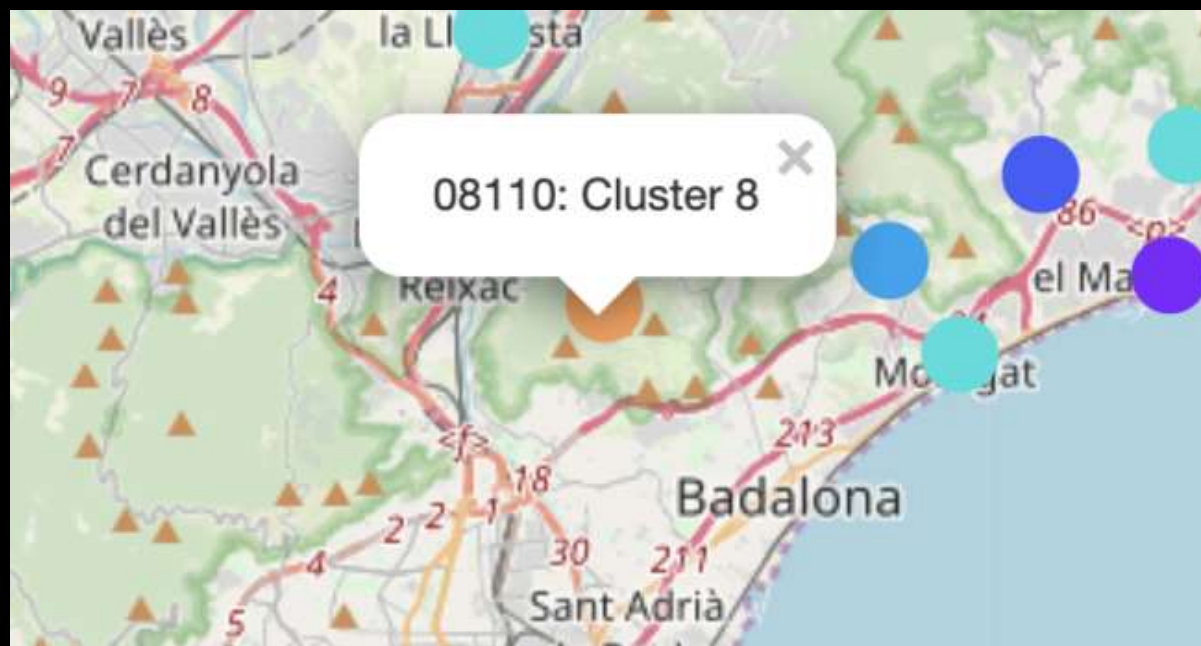
08208: Cluster 4



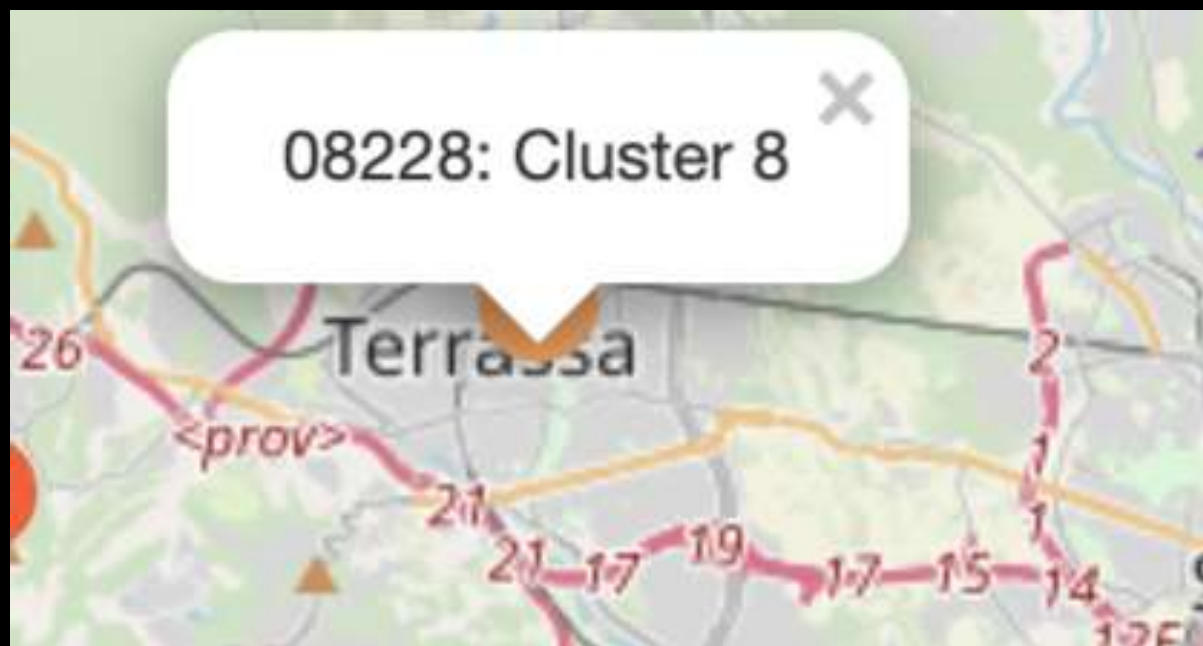
08080: Cluster 4



08110: Cluster 8

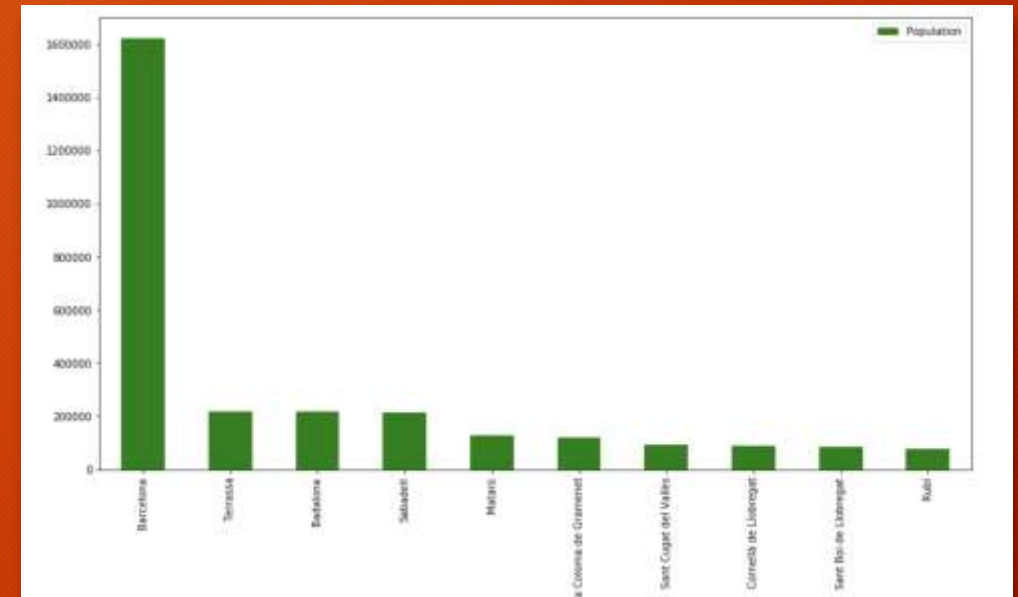


08228: Cluster 8



The Population Dataset

Exploring the population Dataset, its obvious that Barcelona city sweeps the floor with the highest population. Followed by Terrassa and Badalona. This gives us an edge in determining which place should be given a better priority.



Area	Population	Code
Barcelona	1620343.0	8019
Terrassa	218535.0	8279
Badalona	217741.0	8015
Sabadell	211734.0	8187
Mataró	126988.0	8121
Santa Coloma de Gramenet	118821.0	8245
Sant Cugat del Vallès	90664.0	8205
Cornellà de Llobregat	87173.0	8073
Sant Boi de Llobregat	82904.0	8200
Rubí	76423.0	8184

Discussion

Barcelona is a very touristy place, it is seen that Cluster 8 has the highest population and almost all clusters includes restaurants. If CasaGrande would like to start a real estate business in Barcelona it should focus on areas like Barcelona City, Terrassa, Sandell, Badalona and Castelledefels. But we need to take into consideration other factors like schools, income statistics, safety before entering. Machine Learning is an ocean if we can get hold of the other important factoring data like education, income etc, we can perform better analysis and come to a better conclusion

Conclusion Cont..

However the company can start investing in these locations in the form of listing like airBnB or Hotels since it looks like most of the population is floating and the venues are very touristy.

This data allows us to not only see the types of business around each neighborhood, but the larger areas in which those businesses are concentrated. Therefore we can determine the most beneficial places to advertise depending on our company's goals and in what way we should proceed in doing so.

Thank
you!

