Segmenting Madrid

Fabián Torres

Applied Data Science Capstone

The Battle of Neighborhoods

Agenda

- Introduction
- Data
- Methodology
- Results
- Discussion
- Conclusion

Introduction

"La Milonguita" is a chain of restaurants located in Buenos Aires, Argentina, it offers typical foods of that country.

The owners wants to expand their chain to Europe and they are very interested in establish a new restaurant in Madrid, Spain.

Madrid is a big city with more than 3 millions inhabitants and a metropolitan area population of about 6.5 millions. It's composed by more than one hundred neighborhoods.

In order to get directions about where to establish their first restaurant in Madrid, the owners of "La Milonguita" has asked us to do an study of different places of the city.

Marketing consultants has determined that we must pay special attention to entertainment services, because they are considered attractive to potential customers.

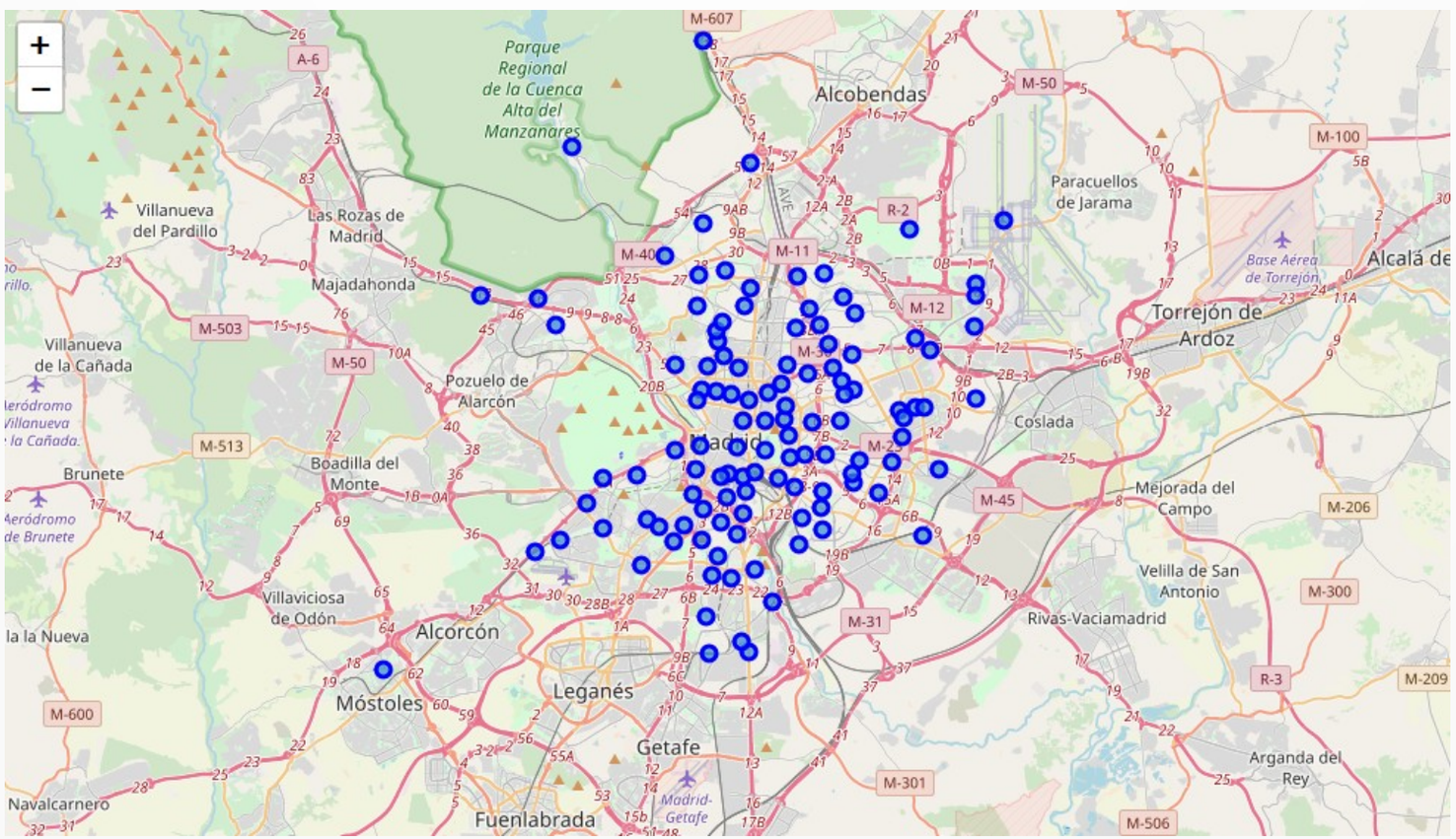
Data

Various sources of data are used:

- Wikipedia, Madrid page give a list of neigborhoods.
- Geocoder provides services for geolocations.
- Foursquare provides information about places in given areas.

Methodology

- **Neighborhood geolocation:** It's needed to geographically locate every neighborhood in order to get information about venues. It also permit us to visualize their distribution in a map, like the one show in the figure.



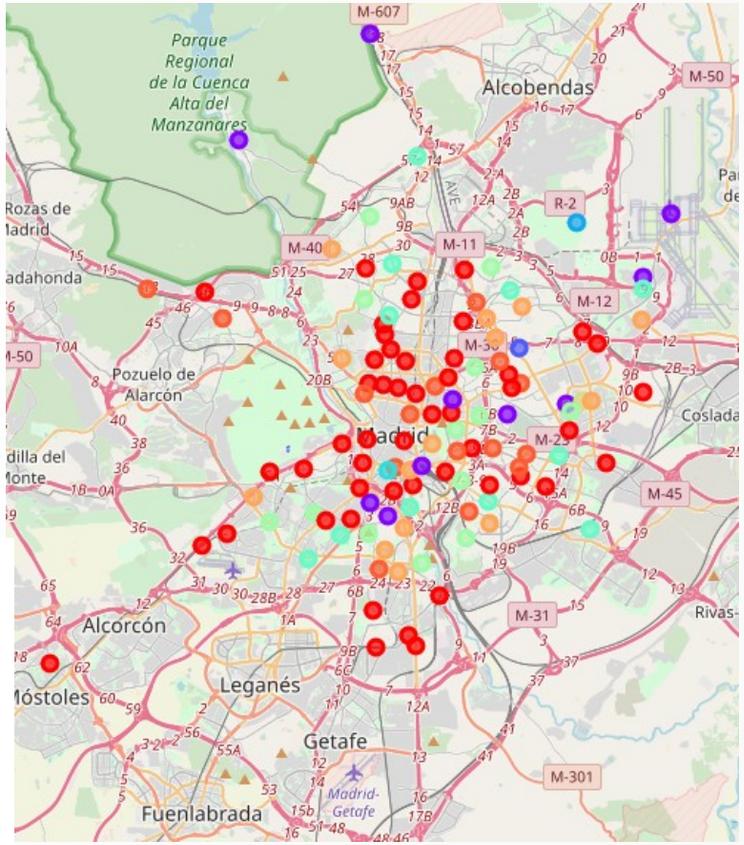
- **Venues Compilation:** Using Foursquare services, we can found venues for every neighborhood. This is our main input in order to be able to classify them in segments.

  As result we obtain a list like the following.

  Every venue has a “category”. It is the base of the classification process.

	neighborhood	neighborhood_latitude	neighborhood_longitude	venue	venue_latitude	venue_longitude	venue_category
0	Centro,Palacio	40.41512925	-3.71561799839909	Santa Iglesia Catedral de Santa María la Real ...	40.41576693264202	-3.7145161628723145	Church
1	Centro,Palacio	40.41512925	-3.71561799839909	Park de La Almudena	40.41631960737639	-3.713776929870609	Park
2	Centro,Palacio	40.41512925	-3.71561799839909	Cervecería La Mayor	40.41521786102789	-3.7121938520878386	Bar
3	Centro,Palacio	40.41512925	-3.71561799839909	Taberna Rayuela	40.41317891156541	-3.7134962971934686	Tapas Restaurant
4	Centro,Palacio	40.41512925	-3.71561799839909	Corral de la Morería	40.412619019854375	-3.7142493289306913	Art Gallery

- **Neighborhood segmentation:** In order to classify the neighborhoods a machine learning algorithm is used.
- Classification is based on the categories of the found venues.
- Results are shown in a map, like the one shown, among detailed list for venues categories discussed in separated section in this presentation.
- Following an small example of the list, where the ten most common venues are shown for every neighborhood.



	neighborhood	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
0	Arganzuela,Acacias	Bar	Theater	Spanish Restaurant	Tapas Restaurant	Coffee Shop	Art Gallery	Park	Market	Pizza	Sporting Goods Shop
1	Arganzuela,Atocha	Spanish Restaurant	Gym	Fast Food Restaurant	Hotel	Park	Bar	Tapas Restaurant	Museum	Coffee Shop	Grocery Store
2	Arganzuela,Delicias	Tapas Restaurant	Spanish Restaurant	Bar	Market	Coffee Shop	Chinese Restaurant	Museum	Dessert Shop	Park	Gym
3	Arganzuela,Imperial	Spanish Restaurant	Coffee Shop	Park	Hotel	Gym	Japanese Restaurant	Pizza	Grocery Store	Spa	Chinese Restaurant
4	Arganzuela,La Chopera	Park	Spanish Restaurant	Bar	Gym	Food & Drink Shop	Italian Restaurant	Tapas Restaurant	Coffee Shop	Pool	Theater

Results

As a result of segmenting Madrid, ten clusters where defined.

- **Cluster 1:** It's a big cluster where Spanish Restaurants together with "Tapas" Restaurants (typical of Spain) and Bars has a dominant presence. Although there are many other interesting places, like Theaters and pub in example, too many rivalry exists, therefore it doesn't appear to be the best option for an starter business.

- **Cluster 3, 4, 5 and 8:** Clustering algorithm produces just one record for each one. Not much useful information for making a decision.

- **Cluster 2, 7 and 10:** These neighborhoods has various kind of restaurants, Spanish, tapas, Chinese and Greek. It has also some pub and bars, but the lack of entertainments makes them not much interesting for our case.

- **Cluster 9:** This cluster has an interesting variety of restaurants, Spanish, Chinese, French, Greek and fast food. Additionally there are bars and theaters. It results interesting to consider for establishing an Argentinian Restaurant.

- **Cluster 6:** This medium sized cluster results very interesting. There are Spanish Restaurants, Fast Foods and Pizza. And also Theaters and Soccer Fields can be found. Soccer is very related to Argentina, because many Argentinian soccer stars are currently playing in Spanish League.

  This is definitively the best finding.

Discussion

- The objective of this project is found places in Madrid City for establishing the first Argentinian Restaurant of the chain in such city.

- Main requirements are that other kind of restaurants exists and also entertainments for potential customers.

- Applying a given machine learning clustering algorithm was possible to segment neighborhoods based on their venues and, most important, found a group of them that have high potential.

Conclusion

- A group of eleven neighborhood has been selected from more that on hundred that Madrid has.
- In such neighborhoods there are Spanish Restaurants, Fast Foods and Pizza. And also Theaters and Soccer Fields can be found.
- We consider that in one of them will be able to start the company the next endeavor.

  - Arganzuela,Legazpi
  - Tetuán,Berruguete
  - Fuencarral-El Pardo,Barrio del Pilar
  - Fuencarral-El Pardo,Valverde
  - Carabanchel,Buenavista
  - Carabanchel,Abrantes
  - Puente de Vallecas,Portazgo
  - Moratalaz,Horcajo
  - Hortaleza,Pinar del Rey
  - Villa de Vallecas,Santa Eugenia
  - Barajas,Corralejos