

Opening a restaurant: Comparing cities, Toronto VS Madrid

Mateo Bonnett Garcia

COURSERA
07/28/2019

1. INTRODUCTION

1.1 Background

“La Primera” is a Spanish restaurant located in the city of Madrid, Spain. With a team with more than 45 years of gastronomical experience, “La Primera” will make your imagination fly, and will surprise you with its local cuisine or their mind-blowing dishes. More information can be found on <http://restaurantelaprimera.com/>

1.2 Problem

Assuming that the owner of the restaurant wants to open a new location, but this time in the City of Toronto, I am faced with the challenge of determining whether or not he should do so, and in case it is appropriate, suggestions on where to open the new restaurant shall be made. This will be accomplished by using the data provided by Foursquare on the restaurant “La Primera” and its surroundings, in order to find a similar neighborhood in Toronto to the one in Madrid where the restaurant is located now.

1.3 Interest

This problem and its solution might concern other owners of restaurants similar to “La Primera” that also want to open new locations in different cities, since it is very important to pick the right place.

1.4 Observations

This restaurant was chosen randomly, and the person writing this report has nothing to do with it. Sorry if any inconveniences might have been caused.

2. DATA

2.2 Source

Most of the data used here came from the city council of Madrid. Including the neighborhood limits, their names and the district they belong to. (Spreadsheet in the first link below). The Latitude and Longitude information was manually retrieved using the GeoHack webpage.

2.3 Cleaning

Data retrieved from both sources was cleaned in order to only include the neighborhood name, the district it belongs to, the longitude and latitude of each neighborhood. However, the original spreadsheet contained the name of the neighborhood, the district it belongs to and some measurements.

All of this was done so I could later retrieve data from Foursquare concerning to restaurants, shops and other locations in the neighborhoods. In the link below you can find the original dataset: <https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=46b55cde99be2410VgnVCM1000000b205a0aRCRD&>

Here is a picture of how the data was read into a DataFrame and how it looks like after being edited.

	Codigo de barrio	Codigo de distrito al que pertenece	Nombre de barrio	Latitude	Longitude
0	1	1	PALACIO	40.415000	-3.713333
1	1	2	IMPERIAL	40.406667	-3.716944
2	1	3	PACIFICO	40.404722	-3.675833
3	1	4	RECOLETOS	40.424167	-3.686389
4	1	5	EL VISO	40.444367	-3.688831

2.3 Comparing data

Data from the last project was used to compare the Madrid information with the Toronto data. This data was first downloaded from the TORONTO_CLUSTERING.ipynb (found in my GitHub repository) using the pandas function to_excel and then uploaded into the ComparingCities.ipynb using the panda function read_excel.

2.4 Observations

Take into account that the data retrieved from: <https://www.madrid.es/UnidadesDescentralizadas/UDCEstadistica/Nuevaweb/Territorio,%20Clima%20y%20Medio%20Ambiente/Territorio/Mapas%20de%20dist%20y%20bar/Centro/Distrito%2001%20-%20Centro.pdf> was used to check for neighborhood limits in the CENTRO district of Madrid, in order to accurately locate ‘La Primera’.

3. METHODOLOGY

3.1 Visualizing the data

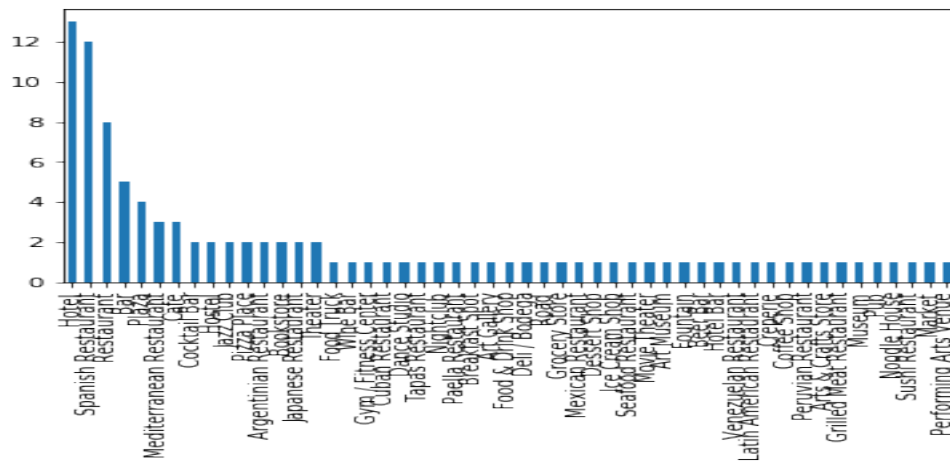
Maps were constantly used in order to visualize the data found in the DataFrames. They allow me to show where the neighborhoods of Madrid are located, and later on, they allow me to visually compare the clusters created by the K-means algorithm.

3.2 Exploratory data analysis

Some exploratory analysis was done on the restaurant’s neighborhood and the restaurant itself, including:

- Obtaining the restaurant’s coordinates
- Counting the amount of venues in the Cortes neighborhood retrieved from the Foursquare API
- Retrieving the amount of venues in the Centro district, so comparisons could be made with the venues located in Downtown Toronto

- A bar plot was constructed to make everyone an idea of the kind of venues one could find in Cortes. Plotted using `value_counts` and the `plot` function for pandas DataFrames.



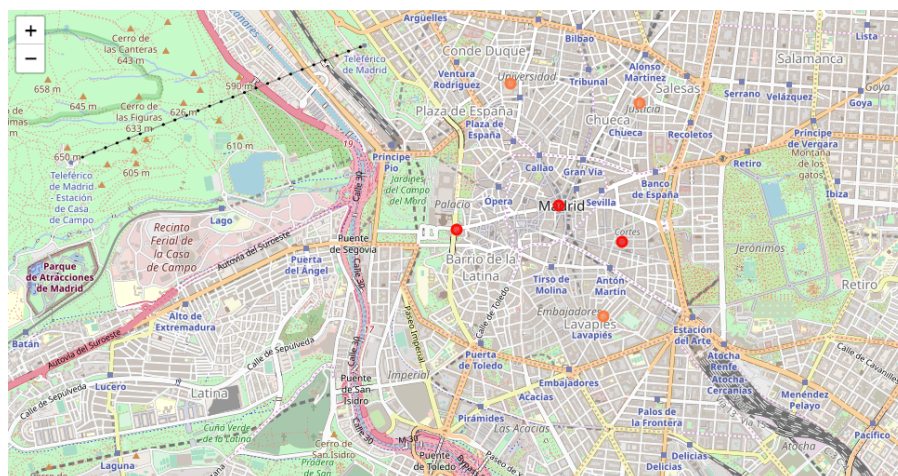
3.3 Machine learning algorithm

The K-means algorithm was used to determine whether or not there was a similar neighborhood in Downtown Toronto to Cortes. By clustering the data into 7 clusters, one can see what neighborhoods are similar to Cortes.

4. RESULTS AND DISCUSSION

After running the algorithm with different number of clusters, I found that Cortes is not similar to any neighborhood in Downtown Toronto. In fact, no neighborhood in the Centro district is similar or grouped in the same cluster with any neighborhood in Downtown Toronto.

However, Cortes was found to be similar to Sol. This might suggest that the owner of the restaurant should consider opening the new location in the vicinity of its restaurant and forget about the idea of opening it in Toronto. Another option for the new store could be Palacio, which was proposed by the algorithm when creating 7 clusters for the whole data.



Other algorithms should be tried, including: KNN and other clustering algorithms.

5. CONCLUSIONS

In this report I analyzed how venues can help to detect similarity between neighborhoods, in order to suggest new locations for opening restaurants. I identified the 10 most common venues in each neighborhood and then developed the clustering model to cluster them into 7 groups. This kind of models could help identify potential locations for starting a business or to open a new location of a previously existing one.

