Housing Sales Prices & Venues Data Analysis of Mexico City

Leopoldo Machado

January 5th, 2020

1. Introduction

1.1 Background and Problem

Mexico City's gastronomy is renowned worldwide. The sheer amount of choices, from fine dining to street carts lead to a highly competitive market. Mixed with other factors such as overpopulation (21 million and counting!), traffic, geography and real estate prices pose a challenge for investors. Is a lower priced borough preferable over high venue density or backwards? Considering these factors, we can create a map and information chart where the real estate index is placed on the city and each borough is clustered according to the venue density.

1.2 Data Sources

Data sources are the following:

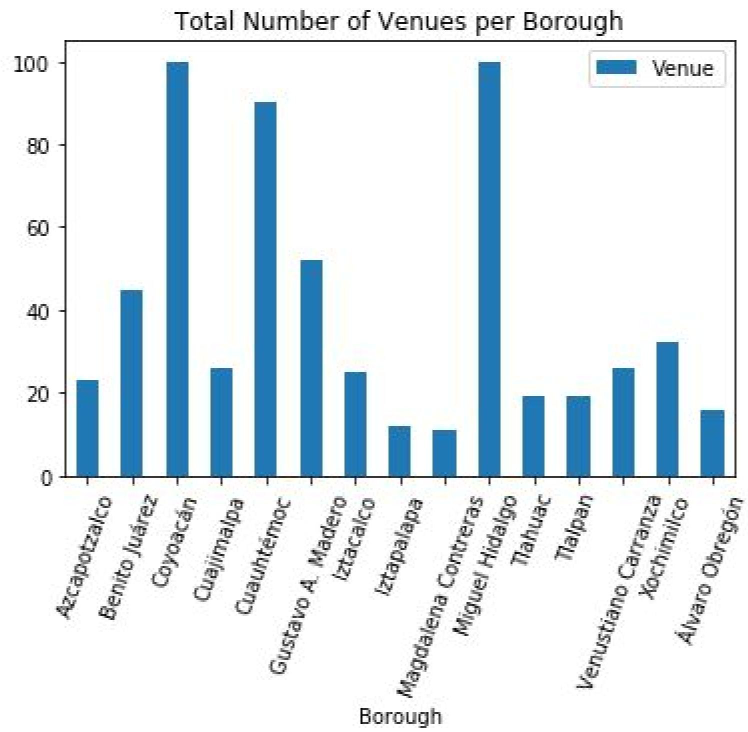
- Real estate price estimates per borough for a commercial property (price per m2) found at https://www.metroscubicos.com/precios/distrito-federal/
- Foursquare API used to find most common venues per borough in Mexico City
- Google maps used to find the coordinates per borough in the city.

2. Methodology

As a database, I prepared a csv file containing each Mexico City borough and its average price per square meter as specified on metroscubicos.com. On the website, the average as well as max and minimum prices are listed for different types of real estate: apartments, houses and commercial terrains. Only commercial terrains was imported as it was the most relevant to our case study. Latitude and longitude were also added to use with Foursquare API in determining most commonly used venues. The csv file was then converted to a dataframe and pandas used for data wrangling.

Folium and madplotlib libraries were used for visualization purposes, maps were created to show the geographical distribution of the boroughs, and following clusterization, to show the distribution of each cluster on the city. Madplotlib helped in better visualizing the amount of venues per burough.

Foursquare API was used to explore the boroughs and segment them. Limit was set at 100 venues and a 500 meter radius. A total of 596 venues were returned by Foursquare, with Miguel Hidalgo, Cuauhtemoc and Benito Juarez boroughs showing the highest number of venues:



Then I created a table to show the top 10 venue categories for each borough:

Borough	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	Azcapotzalco	Mexican Restaurant	Bakery	Seafood Restaurant	Shopping Mall	Lounge	Breakfast Spot	Farmers Market	Ice Cream Shop	Burrito Place
1	Benito Juárez	Mexican Restaurant	Pizza Place	Coffee Shop	Taco Place	Sushi Restaurant	Drugstore	Russian Restaurant	Restaurant	Burrito Place
2	Coyoacán	Mexican Restaurant	Ice Cream Shop	Bar	Coffee Shop	Café	Plaza	Art Gallery	Bookstore	Italian Restaurant
3	Cuajimalpa	Pizza Place	Taco Place	Bakery	Burger Joint	Gym	Soccer Field	Convenience Store	New American Restaura	Dance Studio

									nt	
4	Cuauhtémoc	Mexican Restaurant	Hotel	Taco Place	Bar	Restaurant	Coffee Shop	Art Museum	Pizza Place	Argentinian Restaurant

I then used the k-means algorithm for clusterization, choosing a k of 3 for the number of clusters.

Using examination to compare the common features between venues, we can categorize the clusters in the following manner:

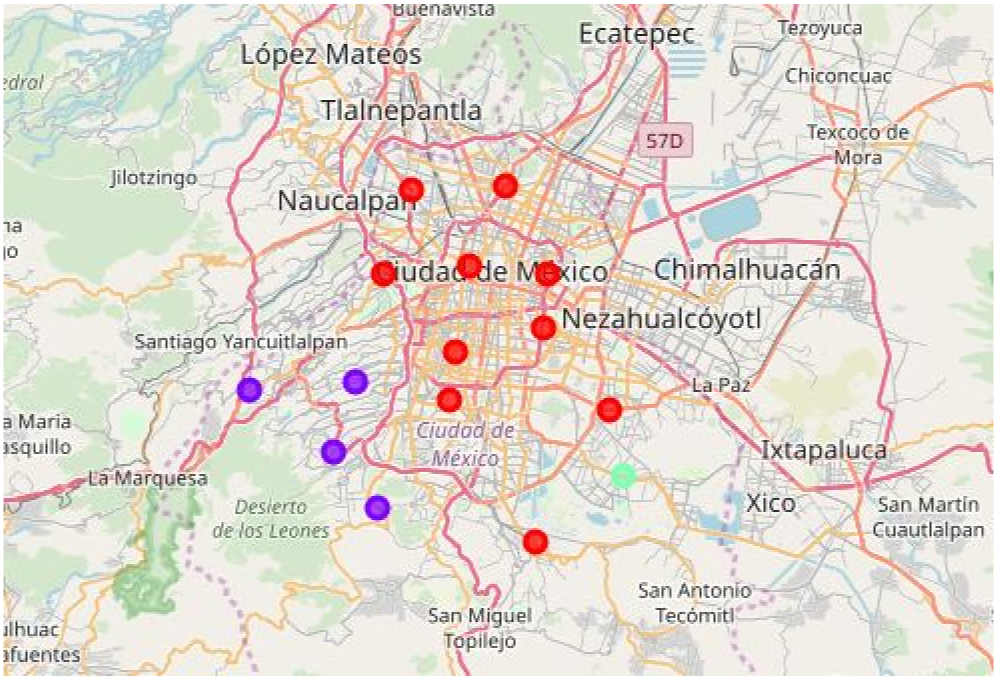
- Cluster 0 - Restaurants and fine dining, mexican food mostly prevalent.
- Cluster 1 - Food carts and informal places (taquerias).
- Cluster 2 - Non- mexican food.

3. Results

The following table was generated showing a clusterized division of venues per average price and density

	Borough	Avg-HousePrice	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
0	Álvaro Obregón	10,676.65	19.3605	-99.2267	1	Pizza Place	Mexican Restaurant	Burger Joint	Breakfast Spot	Stationery Store	
1	Azcapotzalco	13,828.76	19.4847	-99.1887	0	Mexican Restaurant	Bakery	Seafood Restaurant	Shopping Mall	Lounge	
2	Benito Juárez	35,961.82	19.3794	-99.1591	0	Mexican Restaurant	Pizza Place	Coffee Shop	Taco Place	Sushi Restaurant	
3	Coyoacán	19,142.35	19.3487	-99.1629	0	Mexican Restaurant	Ice Cream Shop	Bar	Coffee Shop	Café	
4	Cuajimalpa	5,913.82	19.3558	-99.2994	1	Pizza Place	Taco Place	Bakery	Burger Joint	Gym	S
5	Cuauhtémoc	40,069.73	19.4356	-99.1495	0	Mexican Restaurant	Hotel	Taco Place	Bar	Restaurant	
6	Gustavo A. Madero	17,980.86	19.4873	-99.1236	0	Bakery	Pharmacy	Café	Mexican Restaurant	Seafood Restaurant	

A clusterized map of the venues on Mexico City was also generated:



4. Discussion

The complexity of such a big metropolis as Mexico City makes it difficult to predict or estimate the most preferable place for investors. Based on the variables I chose to generate this study, a pattern emerges based on geographic distribution: boroughs closer geographically to the city center have a higher density of restaurants and fine dining venues, while the further you get to the eastern and western limits of the city, more prevalence is observed for smaller or more informal eateries.

On the referenced site to obtain the average prices per borough, data per colony/neighborhood is also available. Deeper analysis of this more in-depth data could be helpful in finding other trends or clusters that could better aid in determining the most desired place to open a restaurant.

K-means analysis was used as part of this study, as well as visualization, clustering and exploring techniques.

5. Conclusion

In this study, I analyzed the relationship between real estate prices and venue density as factors to determine the best location to open a restaurant in Mexico City. I identified average cost per m2 on each of the cities' boroughs, latitude and longitude coordinates, and venue density as important factors to take into account in planning an investment in this field. The K-means algorithm was used as part of this clustering study. Data exploring and visualization techniques also helped in identifying common traits between the clusters each borough ended up classified as.