

# Coursera Capstone project

## Data

Science

Data

Opening new Restaurants in Mexico City  
By Juan Manuel Escoto Lopez  
January 2020

## Introduction

For many people, visiting restaurants is a great way to relax and enjoy their loving ones during the year and also create new business relationships. For retailers, the central location and the large crowd provides a great distribution channel to market their services. Property developers are also taking advantage of this trend to build more business buildings or malls to cater to the demand. As a result, there are many restaurants in the Mexico City and many more are being built. Opening restaurants allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new business requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the restaurants is one of the most important decisions that will determine whether the mall will be a success or a failure.

## Business Problem

The objective of this capstone project is to analyze and select the best locations in the city of Mexico City, Mexico to open a new restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the Mexico City, if a property developer is looking to open a new restaurant, where would you recommend that they open it?

## Target Audience of this project

This project is particularly useful to property developers and investors looking to open or invest in new restaurants in the capital city of Mexico i.e. Mexico City. This project is timely as the city is currently suffering from oversupply of restaurants.

## Data

To solve the problem, we will need the following data:

- List of neighborhoods in Mexico City. This defines the scope of this project which is confined to Mexico City, the capital city of the country Mexico
- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to restaurants. We will use this data to perform clustering on the neighborhoods.

## Sources of data and methods to extract them

This Wikipedia page ("[https://en.wikipedia.org/wiki/Category:Neighborhoods\\_in\\_Mexico\\_City](https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Mexico_City)") contains a list of neighborhoods in Mexico City, with a total of 80 neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and BeautifulSoup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers.

Foursquare API will provide many categories of the venue data, we are particularly interested in the Restaurant category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.

## Methodology

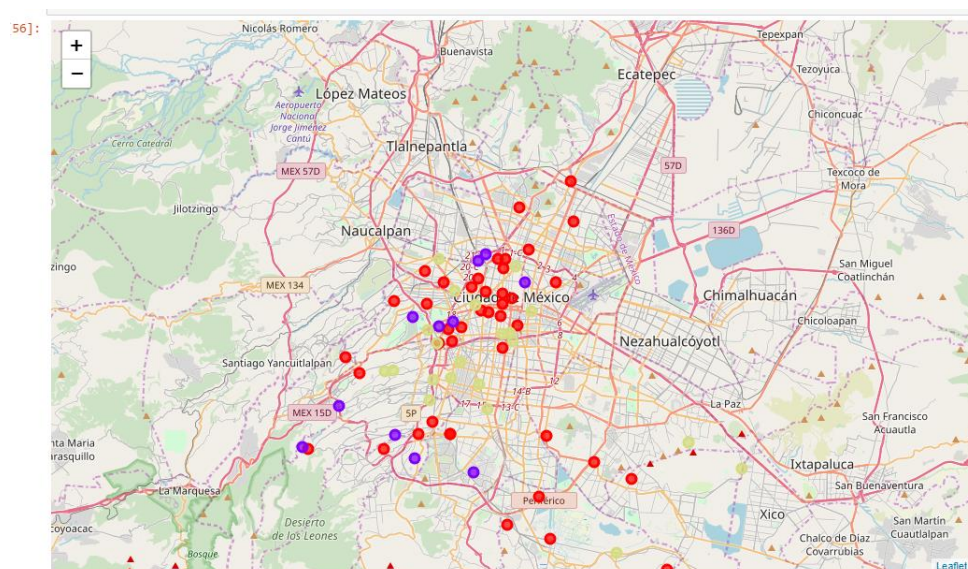
Data scrapping from the Wikipedia page that contains the up-to-date population statistics of Mexico City neighborhoods has been used. This is critical to understand the population of each neighborhood which is one of the key elements in the neighborhood of choice in this project.

Furthermore, we need to know the coordinates and locations of this neighborhoods, and therefore the geocoder API has been used for achieving this objective. This is important so that we can input this information into the location information provider such as Foursquare.com to obtain venue information in these neighborhoods, and this is precisely what we have done for it in this project.

We will also use machine learnings techniques such as the K-Means to segment and cluster these neighborhoods so that we can group them together to understand their similarities. This is critical as we need to recommend the best place to open a restaurant.

## Results

With K-Means clustering technique, the top 4 clusters the neighborhoods have been apparent in the result, see below. These clusters are group together based on the similar nearby venues in each of the neighborhoods. This information is critical so that we can target on the cluster that offer the best option to open a new restaurant with the best opportunity to grow.



With Foursquare.com API, we are also able to leverage on the data to find out the top common nearby venues and their categories in each of these neighborhoods. This is critical as we want to recommend a neighborhood whereby the supply is low (lower competition).

## Cluster 1

```
In [59]: kl_merged.loc[kl_merged['Cluster Labels'] == 1]
```

Out[59]:

	Neighborhood	Restaurant	Cluster Labels	Latitude	Longitude
63	San Miguel Chapultepec	0.060000	1	19.412450	-99.186750
51	Lomas de Chapultepec	0.060000	1	19.418773	-99.206001
75	University City, Mexico	0.080000	1	19.357320	-99.260160
46	Jardines del Pedregal	0.080000	1	19.320940	-99.204560
24	Colonia Maza	0.080000	1	19.336789	-99.219128
42	Condesa	0.060000	1	19.415460	-99.176020
65	San Pedro Atocpan	0.068966	1	19.200620	-99.050070
11	Colonia Atlampa	0.060000	1	19.457220	-99.157890
12	Colonia Buenavista	0.064516	1	19.328920	-99.286818
34	Colonia Santa María Insurgentes	0.060000	1	19.461800	-99.152610
69	Santa Úrsula, Mexico City	0.080000	1	19.311325	-99.161135
73	Tepito	0.060000	1	19.442880	-99.123240

## Cluster 2

```
In [60]: kl_merged.loc[kl_merged['Cluster Labels'] == 2]
```

Out[60]:

	Neighborhood	Restaurant	Cluster Labels	Latitude	Longitude
48	La Merced (neighborhood)	0.2	2	18.8625	-99.625

## Discussion

Most of the restaurants are concentrated in the central area of Mexico City, with the highest number in cluster 0 and moderate number in cluster 1.

On the other hand, cluster 2 has very low number to totally no restaurants in the neighborhoods. This represents a great opportunity and high potential areas to open new restaurants as there is very little to no competition from existing restaurants. Meanwhile, Restaurants in cluster 0 and 3 are likely suffering from intense competition due to oversupply and high concentration of other business.

From another perspective, this also shows that the oversupply of restaurants mostly happened in the central area of the city, with the suburb area still have very few restaurants. Therefore, this

project recommends property developers to capitalize on these findings to open new restaurants in neighborhoods in cluster 2 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new restaurants in neighborhoods in cluster 1 with moderate competition. Lastly, property developers are advised to avoid neighborhoods in cluster 0 and 3 which already have high concentration of restaurants and suffering from intense competition.

## Conclusion

In conclusion we can say that the best idea is to open a restaurant in the neighborhood of the cluster 2, because there is no competition and this give a good opportunity to open new chain restaurants, also is ok to consider cluster 1 for expansions.