# Open a Mexican Food (Oaxaca) Restaurant

## Table of contents

## 1.1 Introduction

The approach of my project is based on the fact that I have one Mexican Restaurant (Oaxacan Food) in Mexico City and I like to open a second one in another neighbourhood of Mexico City.
I want to use the techniques learned in the course to find neighbourhoods in Mexico City which can be good candidates to place the new restaurant.
I would like to explore only 4 of 16 districts of Mexico City. These districts are:

- Benito Juarez
- Coyoacan
- Cuauhtemoc
- Miguel Hidalgo

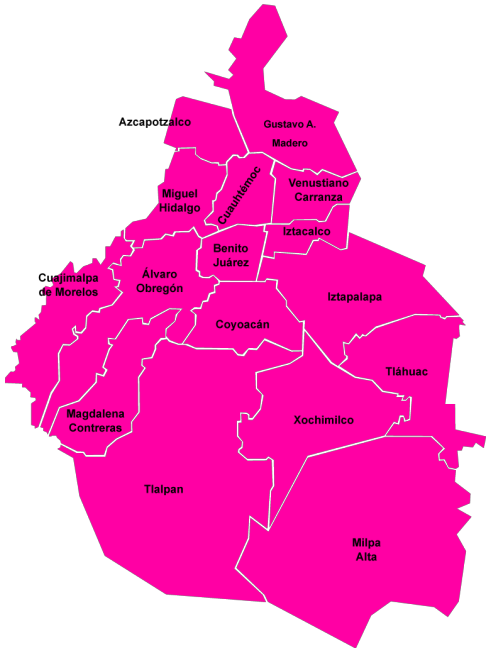



Figure 1. Mexico City Districts

The restaurant will be of Oaxacan cuisine, which is a very tasty type of Mexican food.
Oaxaca is one of the main gastronomic centers of Mexico whose cuisine is internationally known. Oaxacan cuisine is a regional cuisine of Mexico, centered in the city of Oaxaca, capital of the state of the same name located in southern Mexico. Like the rest of Mexican cuisine, Oaxacan food is based on basic foods such as corn, beans, and chilies, but there is a wide variety of other ingredients and food preparations due to the influence of the state's varied geography and indigenous cultures. In the well-known features of the kitchen include ingredients such as chocolate (often drunk in a hot preparation with spices and other condiments), Oaxacan cheese, mezcal and grasshoppers with dishes like tlayudas, tamalesoaxaqueños and seven notable varieties of mole.

Figure 2. Oaxaca State

I will collect information of the neighbourhoods in Mexico City which has restaurants of different types of food not only Mexican Food.

I will collect the places of interest of my current neighborhood using the Foursquare API, and then I will explore the venues for each of the neighbourhoods in other districts of Mexico City.
Finally, I will use KMEANS to find neighbourhoods similar to my current restaurant neighborhood in terms of the places of interest found on Foursquare.


## 1.2 Data

### 1.2.1 For my current restaurant

My current restaurant is located in a neighbourhood in Mexico City called "Credito Constructor". It has the following coordinates: 19.3658436 -99.1815044
I used the Foursquare API with the explore endpoint which returns a list of recommended venues near the current location. I limit the search to a radius of 1000 meters and 100 venues at most.

**GET https://api.foursquare.com/v2/venues/explore**

The result were 100 venues returned with 59 unique categories.

### 1.2.2 For new restaurant

To get data from Mexico City neighbourhoods I will use data from https://download.geonames.org/export/zip. Specifically I downloaded a MX.zip file.
This file contains information of all the estates in the country of Mexico, but we only need the list of 16 Districts and Neighborhoods in Mexico City.

| | Delegacion | Colonia |
|---|---|---|
| 0 | Coyoacan | Villa Coyoacan |
| 1 | Coyoacan | Delegacion Politica Coyoacan |
| 2 | Coyoacan | Barrio Santa Catarina |
| 3 | Coyoacan | Barrio La Concepcion |
| 4 | Coyoacan | Barrio San Lucas |

To get geographical coordinates I used Nominatim from geopy.geocoders:

| | Delegacion | Colonia | Latitud | Longitud |
|---|---|---|---|---|
| 0 | Coyoacan | Villa Coyoacan | 19.350175 | -99.162254 |
| 1 | Coyoacan | Barrio La Concepcion | 19.344669 | -99.157403 |
| 2 | Coyoacan | Barrio San Lucas | 19.315626 | -99.132968 |
| 3 | Coyoacan | Parque San Andres | 19.345394 | -99.148279 |
| 4 | Coyoacan | Del Carmen | 19.354290 | -99.161797 |

The resulting list includes 257 neighbourhoods in districts: Benito Juarez, Coyoacan, Cuauhtemoc and Miguel Hidalgo.
Then, I used the Foursquare API with the explore endpoint to get the recommended venues for each neighbourhood (using radius of 1000 meters and limit of 100 venues per neighbourhood).

| Neighborhood | Neighbourhood Latitude | Neighbourhood Latitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|
| Villa Coyoacan | 19.350175 | -99.162254 | Coyoacán | 19.350126 | -99.163159 | Neighborhood |
| Villa Coyoacan | 19.350175 | -99.162254 | El Kiosko de Coyoacán | 19.349964 | -99.162764 | Plaza |
| Villa Coyoacan | 19.350175 | -99.162254 | Tierra Garat Coyoacán | 19.349609 | -99.160795 | Coffee Shop |
| Villa Coyoacan | 19.350175 | -99.162254 | Tostadas Amatista | 19.351255 | -99.163793 | Mexican Restaurant |
| Villa Coyoacan | 19.350175 | -99.162254 | Fuente de los Coyotes | 19.349150 | -99.163446 | Historic Site |

The resulting dataframe has 11,231 rows with 353 unique categories.

## 1.3 Methodology

To find similar neighborhoods to my restaurant neighbourhood in Mexico City I used KMEANS to group similar neighbourhoods in cluster. One of these clusters includes my restaurant neighbourhood in Mexico City, so other neighbourhoods in the same cluster will be similar to mine.
First I transformed the data so all attributes are numeric. For this purpose I followed the following steps:

- Put together the location data of CDMX districts and neighbourhoods in Mexico City(variable geo_barrios)
- Collect the data of places of interest of CDMX (variable cdmx_venues)
- Use "onehot encoding" to transpose the categories of the places of interest and convert them to numerical values
- Group the resulting matrix by neighbourhood, using the average value of each category
- Apply KMEANS using different number of clusters (K)
- Measure clustering quality using Silhouette Coefficient and Calinski-Harabaz index
- Select the best K from these results

It turned out that best value for K was 3.

## 1.4 Analysis

For each neighbourhood I looked for what are the 10 most frequent venue categories.
To further limit candidate neighbuorhoods, I used the distance for all points in cluster 0, which is the one that contains my restaurant neighbourhood in Mexico City, to its centroid. Then I picked the closest neighbourhoods in Mexico City to my current restaurant.

- Find the distances of all the points to the cluster where the Mexico City neighbourhood is located.
- Add cluster label to each point.
- Sort rows by cluster and distance.
- Keep only row for the cluster where the Mexico City neighborhood is located.
- Get the index of Mexico City in this last dataframe.
- Select the neighbourhoods closest to the one in Mexico City, according to the distances to the centroid.

## 1.5 Results and Discussion

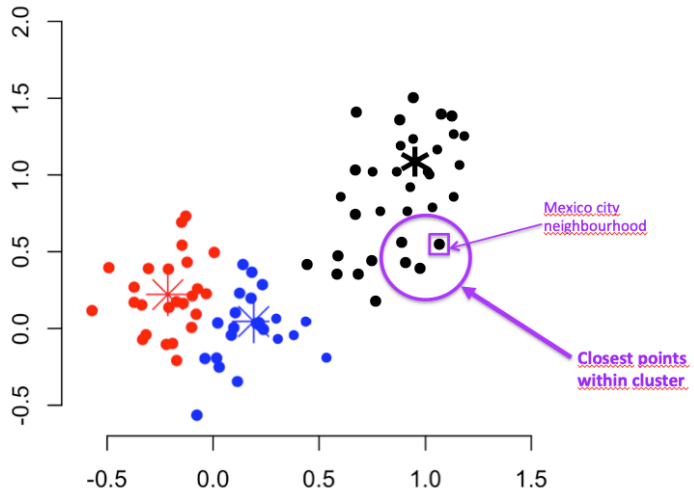
The result of this exercise shows that it is possible to help people who are in a situation similar to the one described in this case, that is, someone who wants to find a neighbourhood similar to the place of the current restaurant, to open a new restaurant using public data available through the Foursquare API.

In KMEANS one of the difficulties is the choice of the value for K. To decide what value to use, I executed the algorithm with different K values and, for each case, I calculated the Silhouette Coefficient and the Calinski-Harabaz Index. These are 2 metrics that allow us to decide if we obtain dense and well separated clusters. As a result both indicators obtained the best value for K equal to 3.

The venues of my neighbourhood in Mexico City were added to the venues of the 117 neighbourhoods and I generated 3 clusters. I searched within the cluster where my restaurant neighborhood in Mexico City was assigned (cluster 0), for those neighborhoods that were closest considering the Euclidean distance of each one of them to the cluster centroid.

The characteristics that distinguish my neighbourhood, according to the results of Foursquare, is the diversity of places to eat, shops and places to exercise. These same characteristics are present in almost all the selected neighbourhoods.

The technique used to select the candidate neighbourhoods is illustrated in the following figure:




## 1.6 Conclusion

Of course, the final decision can not be based solely on the results of this analysis. Rather it should be considered as a tool to narrow the options that must be investigated in greater detail.

For example, one way to enrich the results of the analysis would be by adding demographic and socioeconomic attributes to each of the neighbourhoods. This would result in more similar and homogeneous clusters.