Strategic Clustering for Establishing a Restaurant in Mexico City

1. Introduction

1.1 Background

To have success in a restaurant, we need to consider multiple factors such as type of the restaurant, ratings, prices, locality, etc. But one of the most important is understanding the customers: Which places do they prefer? What type of food is their favorite? What type of restaurant do they prefer? All these questions can be helpful to locate the ideal place for a restaurant. However, finding the best location in a big city such as Mexico City could be very difficult. So, implementing a Data Science analysis and using a Machine Learning algorithm in order to classify some of the neighborhoods from Mexico City, based on their most popular restaurant, it will give some insight about the customers' preferences.

1.2 Problem

A hamburger restaurant is going to be established in Mexico City. The owner wants to locate it in a strategic place, in a neighborhood where the customers like to go to big restaurants and where the restaurant has the opportunity to grow despite the competition of other restaurants.

The objective of this capstone project is to find the optimal neighborhood in Mexico City for the restaurant. The Foursquare API would be used in order to get the most popular restaurants of each neighborhood and then classify them in order to know the preferences of the people around the location.

1.3 Interest

Restaurant analytics are really useful to restaurant owners. It delivers insights that help to understand what drives the profitability of the restaurant , saving time and winning money. The location of the restaurant could be one of the most important factors, as it could tell us a lot about what the customers like and what they are used to.

2. Data Acquisition and cleaning

2.1 Data source

The Mexican Postal Codes data are obtained from the Great Data.com Data Set. Which has the latitude and longitude information of each neighborhood of Mexico City. The sample data set can be freely downloaded in a csv file.

All these locations are used to extract information of venues, specifically of the restaurants, utilizing Foursquare API and its explore function.

2.2 Data Cleaning

The neighborhood data is parsed and stored in a data frame. There are 5 different type of code:

- ❖ BO: Barrio.
- ❖ CAMP: Campamento.
- ❖ COL: Colonia.
- ❖ EQUIP: Equipamiento.
- ❖ PBO: Axotla.
- ❖

	Postal Code	Sequence No.	State Code	Type Code	Place Name	Municipality Name	Major City Name	Latitude	Longitude
0	1000	1	DF	COL	San Angel	Álvaro Obregón	México	19.346257	-99.193169
1	1010	1	DF	COL	Los Alpes	Álvaro Obregón	México	19.360502	-99.193958
2	1020	1	DF	COL	Guadalupe Inn	Álvaro Obregón	México	19.349653	-99.197087
3	1030	1	DF	PBO	Axotla	Álvaro Obregón	México	19.359138	-99.176073
4	1030	2	DF	COL	Florida	Álvaro Obregón	México	19.356070	-99.181523
...	...	...	...	...	...	...	...	...	...
995	10840	1	DF	BO	Las Calles	La Magdalena Contreras	México	19.304080	-99.232064
996	10840	2	DF	BO	Plazuela del Pedregal	La Magdalena Contreras	México	19.304080	-99.232064
997	10900	1	DF	PBO	San Nicolás Totolapan	La Magdalena Contreras	México	19.297665	-99.242156
998	10910	1	DF	PBO	La Magdalena	La Magdalena Contreras	México	19.302101	-99.245791
999	10920	1	DF	COL	Las Huertas	La Magdalena Contreras	México	19.298948	-99.240467

We drop the CAMP and EQUIP from the Type Code column of the Data frame, in order to get only information related to the neighborhoods (or colonias).

Type Code	Postal Code	Sequence No.	State Code	Place Name	Municipality Name	Major City Name	Latitude	Longitude
BO	55	55	55	55	55	55	55	55
CAMP	1	1	1	1	1	1	1	1
COL	890	890	890	890	890	890	890	890
EQUIP	2	2	2	2	2	2	2	2
PBO	52	52	52	52	52	52	52	52

The most important restaurant of 10 Municipality will be obtained. In the table below, the number of neighborhoods of each municipality is shown. And all of them are located in a Mexico CIty map.

Municipality Name	Postal Code	Sequence No.	State Code	Type Code	Place Name	Major City Name	Latitude	Longitude
Azcapotzalco	87	87	87	87	87	87	87	87
Benito Juárez	53	53	53	53	53	53	53	53
Coyoacán	95	95	95	95	95	95	95	95
Cuajimalpa de Morelos	45	45	45	45	45	45	45	45
Cuauhtémoc	41	41	41	41	41	41	41	41
Gustavo A. Madero	177	177	177	177	177	177	177	177
Iztacalco	38	38	38	38	38	38	38	38
Iztapalapa	199	199	199	199	199	199	199	199
La Magdalena Contreras	40	40	40	40	40	40	40	40
Álvaro Obregón	222	222	222	222	222	222	222	222