Exploring and clustering Restaurants in Mexico City, Mexico.

Capstone Project Report

This is an advance of the final report and only contains the sections: ***Introduction/Business Problem*** *and* ***Data*****.**

Student: Daniel Chong Sánchez

10-22-2019

Introduction/Business Problem

Mexico City (CDMX), is the capital of Mexico and the most populous city in North America. It is one of the most important cultural and financial centers in the Americas. (According to the most recent definition agreed upon by the federal and state governments, the population of Mexico City is 21.3 million, which makes it the largest metropolitan area of the Western Hemisphere. It is a multicultural city that in the last decades have been also an important touristic destination in Mexico.

This city includes thousands of restaurants in **Mexico City Metropolitan Area, (CDMX)** is also a favorite destination for food lovers. But this city has presented problems of insecurity since some years ago, with a tendency that has increased and that affects the businesses of the city.

Explaining The Problem.

In this project regional restaurants in the Mexico City Metropolitan Area (CDMX), will be analyzed and with the data obtained from Mexican government provided by the official web site, regional restaurants will be clustered according to the perceived insecurity by National Survey on Insecurity ENVIPE-2018. National Survey of Victimization and Public Security Perception (ENVIPE-2018). This survey established as 43.1 the score for CDMX as one of the most higher in the country, therefore this region is to considered as an insecure for the people who live in there. Also the National Urban Public Security Survey (ENSU) in his web site shows the numbers to develop this project. The purpose of this project is helping people who wish to go out to different places, like Coffee Shop for lunch or dinner, to be aware about the level of insecurity of the chosen restaurant for this purpose and beside find similar options in less insecure places.

The problem of insecurity in CDMX is a current and very relevant issue, as the figures show in the web site http://nl.semaforo.com.mx/ (the site showing lots statistics about crime following different indicators in Mexican states), also in the web site of the National Institute of Statistics and Geography (INEGI), https://www.inegi.org.mx/programas/ensu/default.html#Datos_abiertos, which is the official site, exist lots of mexican statistics in particular about social issues.

This project pursuit to find an optimal location for a restaurant in CDMX. In particular, this report is aimed to stakeholders interested in opening an **italian restaurant** in Mexico City.

Taking in consideration that there are lots of restaurants in Mexico City, we will try to detect **areas or locations that are not already crowded with restaurants**. We are also particularly interested in **areas with no Italian restaurants in vicinity**. We would also prefer locations **as close to the city center as possible**, assuming that first two conditions are met.

We will use our data science powers to generate a few most promissing neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Who will be the target people.
Target people will be visitor or people living in CDMX that want to go out for have a meal, and are concerned about the insecurity in this city.

About the data.
For implementing the proposed analysis following data is needed:

- CMDX municipalities longitude and latitude (this data will be downloading from the national web site of Mexican Postal Service).
- CDMX data about regional Restaurants grouped by municipalities (this data will be obtained with the help of foursquare API).
- CDMX municipalities perceived insecurity in the last 10 years for calculating the mean perceived insecurity (this data will be obtained from the web site of the National Urban Public Security Survey (ENSU))

What is next?
In the next section (2nd week) we will cover the methodology for solving the problem and also the implementation of a python notebook that implements the methodology for making it reproducible and reused for analyzing similar problems.