Moving to a similar neighborhood in Mexico City

Octavio Ivan Hernandez Salinas
honter1997@gmail.com

April 2020

Introduction

Many people in Mexico City are moving to new houses every day, which is great but a lot of people have problems getting familiarized with the new neighborhood because they miss their old neighborhood and the venues they used to go to . For example: they miss their favorite tacos or their favorite coffee shop.
So, we can prevent people from spending many hours looking for a new neighborhood with the same kind of venues that they have nearby their old house. Instead, we can recommend neighborhoods with almost the same kind of venues as the old neighborhood.

Problem

People spend many hours looking for neighborhoods with similar kind of venues when they are buying a new house

Target audience: People who are looking for a new neighborhood to move in.

Solution

Let's create a basic recommendation model, using clustering and the Foursquare API, to show people the neighborhoods that are similar to theirs.

Data

I am going to use the following dataset, provided by the Mexico City State
https://datos.cdmx.gob.mx/explore/dataset/coloniascdmx/download/?format=csv&timezone=America/Mexico_City&lang=es&use_labels_for_header=true&csv_separator=%2C
The dataset has the following features:

Column name	**Data type**	**Description**
COLONIA	object	Neighborhood's name
ENTIDAD	float64	State's ID
Geo Point	object	Latitude and Longitude separated by comma

Geo Shape	object	JSON with the coordinates of the neighborhood's area shape
CVE_ALC	int64	City's ID
ALCALDIA	object	City's name
CVE_COL	object	Neighborhood's ID
SECC_COM	object	Data used for electoral purposes
SECC_PAR	object	Data used for electoral purposes

In the next sections I will prepare the data in order to use it with the Foursquare API and the K-Means algorithm. The preprocessed dataset will have the following columns:

- neighborhood
- neighborhood_lat
- Neighborhood_lng
- city

Also, I will request, to the Foursquare API, the nearby venues' category for each neighborhood and get the mean value for each category so each row corresponds to each neighborhood and each column to a venue category and this will be the input of the K-Means algorithm.