Applied Data Science Capstone: Final Project

Business Problem

The objective of this project will be to discover the best location to open a Spanish food restaurant in Toronto. To do this we will have the information related to other restaurants of this type that are already open in the city.

Data

To solve this problem we will use the information from the different neighbourhoods in Toronto and the Foursquare API to find out which Spanish restaurants there are.

The process for extracting this information will be:
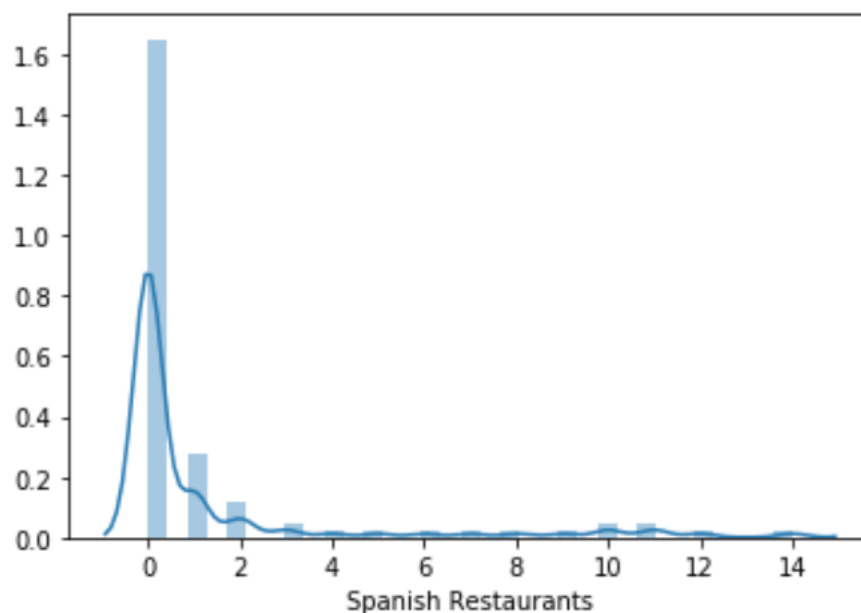Scrapping of Toronto neighborhoods via Wikipedia.
Getting Latitude and Longitude data of these neighborhoods via Geocoder package.
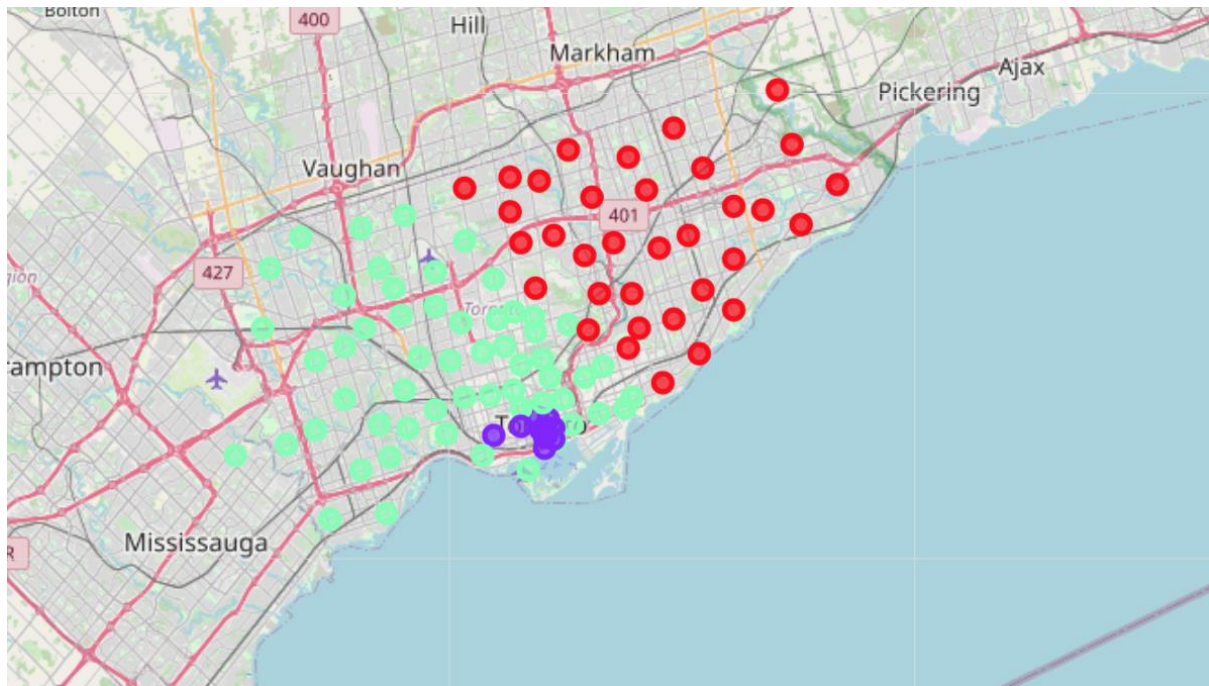Using Foursquare API to get venue data related to these neighborhoods .

Methodology

To be able to know the neighborhoods with the best opportunities to open a Spanish restaurant we must first know the areas where there are already neighborhoods of this type. For these we have made requests to Foursquare's API, using the endpoint search and the category of Spanish restaurant.

The following figure shows the distribution of the number of restaurants per neighborhood.

After that, we will use the KMeans algorithm to group the neighborhoods according to the competition in Spanish restaurants. In the following map the three groups of neighborhoods are presented.

- Red Cluster: Low Offer.
- Green Cluster: Medium Offer.
- Purple Cluster: High Offer.

We have been able to group the neighborhoods according to their latitude, longitude and number of Spanish residents.