

Applied Data Science Capstone
Joshua Triana
April 2020

Introduction

In this project, we want to explore a strategic location for a new coffee shop in the Toronto Area. We want to convince a client looking to invest in the Toronto Area. Coffee shops are a desired business all around the world, so it's a "safe business" to invest in. Nevertheless, we want to analyze the best possible location in the Toronto Area in order to increase the odds.

BUSINESS PROBLEM

The main objective is to find the best location to open a new coffee shop in Toronto, Canada. We will be aiming to apply all the knowledge gained in this specialization in order to answer the business question: In Toronto, if new clients want to open a new coffee shop, where would be the best location;

TARGET AUDIENCE

A client looking to open a new coffee shop in Toronto Area

DATA

To solve this problem, we will need below data:

- List of neighborhoods in Toronto, Canada
- Latitude and Longitude of these neighborhoods
- Venue data related to Coffee shops.

EXTRACTING THE DATA

- Scrap of Toronto neighborhoods in Wikipedia
- Using Foursquare API to get venue data
- Latitude and Longitude by using Geocoder package

METHODOLOGY

We need to extract data from:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Then we are going to scrap the table data to a dataframe and then merge it with the corresponding coordinates.

The map's visualization will be key as a first approach to understand the data.

Then, we are going to use the Foursquare API to get the venues data.

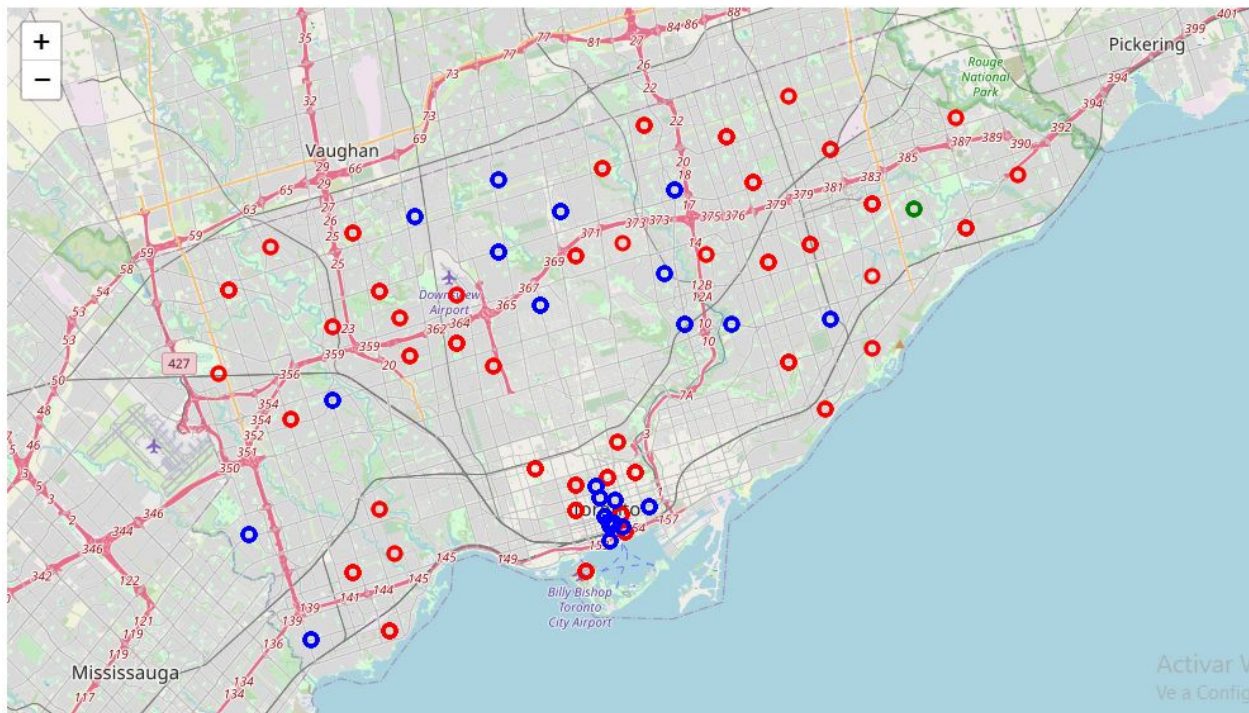
We will take the top four boroughs to work with. The idea is to put the coffee shop in key and popular locations.

Borough	
Central Toronto	9
Downtown Toronto	19
East Toronto	5
East York	5
Etobicoke	12
Mississauga	1
North York	24
Scarborough	17
West Toronto	6
York	5

The analysis of frequencies by groups will be necessary in order to apply k-means. Kmeans is a popular and very used machine learning algorithm and is appropriate to this problem. We get two clusters for coffee shops and finally analyze the data in order to get the best cluster to open the new coffee shop

RESULT

K-means clustering show that we can categorize Toronto neighborhoods into 2 clusters



- Cluster 0: Neighborhoods with the less number of coffee shops(24)
- Cluster 1: Neighborhoods with more coffee shops(118)

This results recommends the client to open the coffee shop around cluster 0, with less competition