Proper location for a new Coffee in Toronto

Gabriel Lopez Rey

February 4, 2021

1. INTRODUCTION

1.1 Background

Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 2,731,571 in 2016 (last Census). It is the most populous city in Canada and the fourth most populous city in North America. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario, while the Greater Toronto Area (GTA) proper had a 2016 population of 6,417,516. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. Its economy is highly diversified with strengths in technology, design, financial services, life sciences, education, arts, fashion, aerospace, environmental innovation, food services, and tourism.

1.2 Business Problem

The idea of this study is to help people planning to open a new restaurant in Toronto to choose the right location. First of all, we need to collect data from all the coffee shops in Toronto, including their name, ID, location (address, latitude, longitude) and then find the "hot" neighbor where most of the place is located. For active data we use FourSquare and apply folium to visualize a particular neighbor in which we will observe customer "traffic" and predict the proper location of a new coffee shop in the city. In this case, you will find your temporary name on the folium map, “New Coffee!”

2. Data acquisition and cleaning

2.1 Data source

Once the objective of geolocation of cafes in Toronto has been established, we will use it as a Foursquare search tool through its API for developers. The geographical location of the coffees can be found (here). Foursquare is a US tech company from New York focusing on location data. Their technology and data powers apps such as Apple's Maps, Uber, Twitter and many other household names.

2.2 Data Cleaning

The data was downloaded into a table, in which there is a lot of information that we do not need for our purposes. According to our plan, we decided to select by latitude (lat), longitude (lng) and City.

	categories	lat	lng	city
0	Coffee Shop	43.650364	-79.388669	Toronto
1	Coffee Shop	43.654053	-79.388090	Toronto
2	Coffee Shop	43.653436	-79.382314	Toronto
3	Coffee Shop	43.652135	-79.381172	Toronto
4	Coffee Shop	43.654270	-79.387448	Toronto

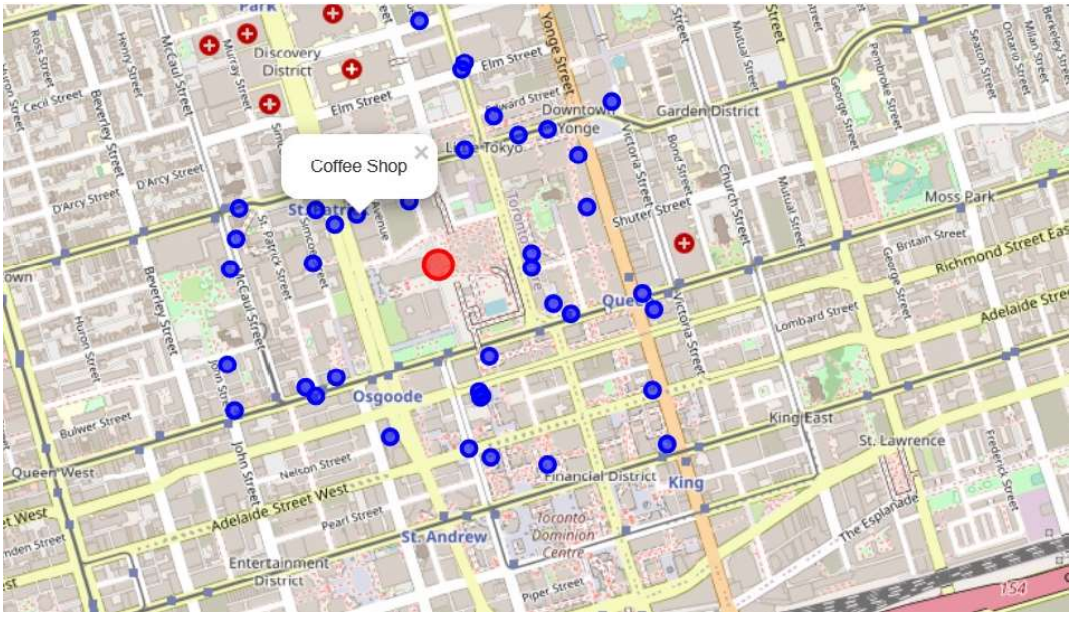
In the table we can visualize the first five data. Starting from this table, we seek to locate a central point between the located coffee houses, which in our opinion is strategic, since it would be located in the neuralgic center of the main coffee shops. For this purpose, we transform the table, through programming, into statistical info.

	lat	lng
count	39.000000	39.000000
mean	43.652904	-79.384648
std	0.002680	0.003842
min	43.648894	-79.391256
25%	43.650475	-79.388068
50%	43.653084	-79.383864
75%	43.654491	-79.381845
max	43.658421	-79.378339

Here, we can visualize the mean value of latitude (lat) and longitude (lng), which places us in the central neuralgic point we are looking for.

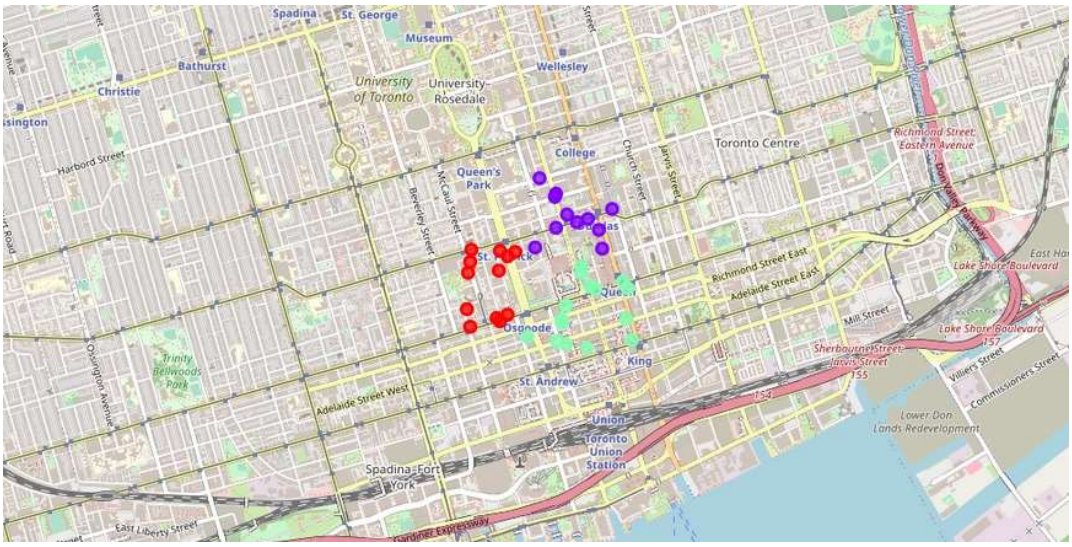
3. Methodology

To visualize it on the map, we use the Python library called Folium



Looking at the image of the map, we can visualize in blue bubbles, the different coffee shops, and in a red bubble in the center of the image, our central point found to make the investment of a new coffee. It will be located just meters from it, at the crossroads between Armory Street and Chestnut Street.

Likewise, we carry out the analysis of separation of zones in clusters, to visualize them. There are many models for clustering. We use the model that is considered the one of the simplest model among them. Despite its simplicity, k-means is vastly used for clustering in many data science applications.



From the image above, we can clearly see the three well-defined clusters.

4. Conclusion

This report can be useful for someone who plans to open a cafe in Toronto, comparing the geographic location of the competition, to achieve a good location in the investment of the new coffee shop, which consisted of locating it in the neuralgic center of all the most important coffee shops, using different Data Science tools.

5. Future directions

A complete business plan could be carried out, with the help of all the Data Science tooling, such as the following:

- Who are potential clients?
- Are they interested in the product or service you offer?
- Who is the most likely to purchase it?
- What are they willing to pay for it?
- What are the current and future market trends?
- Will the demand be sufficient to sustain the business?

The market research process includes:

- Primary Market Research – using data collected through surveys, observation, opinion polls.
- Secondary Market Research – using existing data.