

# **Coursera IBM Data Science Capstone Project**

## **Open a New Coffee Shop in Toronto**

Congru Yin

### **Introduction:**

For this project, I'm creating a hypothetical scenario for a concept person who wants to open a new coffee shop in Toronto. I am going to expand into to this area and explore neighborhoods in Toronto. My idea is: where there is not many coffee shops might present a great opportunity for him. So I am looking for neighborhoods without many preexisted coffee shops.

### **Data:**

To solve this problem, I will use these data:

1. List of neighborhoods in Toronto, Canada. (from Wikipedia)
2. Latitude and longitude of neighborhoods in Toronto. (csv file provided by Coursera)
3. Venues in Toronto area. (from Foursquare API)

I will use python BeautifulSoup package to scrape Wikipedia page to get neighborhoods information, and assign them coordinates using csv file provided by coursera. Then I will discover venues with Foursquare API to find out which area has lowest number of competitors, and is the most suitable for that man to open a coffee shop.

### **Methodology:**

First, I got the list of neighborhood from [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M), used beautifulsoup package to scrape information. Then I gathered coordinated provided by coursera, to the list I scraped from Wikipedia. Next I used Foursquare API to get top 100 venues within 500 meters radius. With all these prepared, I can analyze each neighborhood by grouping rows by neighborhood.

Here I performed clustering by using k-means clustering. This method aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, and it is one of the most popular unsupervised machine learning methods. In this project, I classified neighborhoods in Toronto to 3 clusters based on frequency of occurrence of coffee shops.

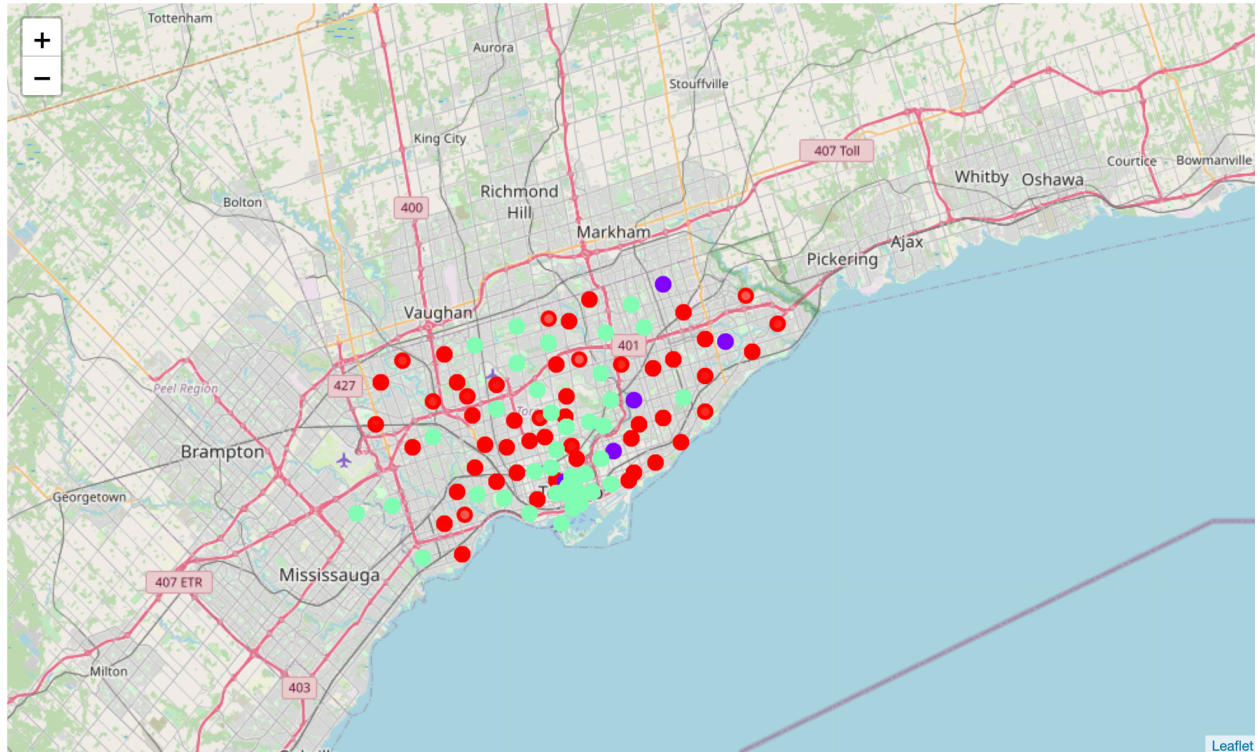
## **Result:**

The clustering result shows that we can classify Toronto neighborhoods into 3 clusters shown as below:

**Cluster 1:** neighborhoods with few or no coffee shops. (as shown in **RED**)  
High Park, The Beaches West, India Bazaar, Trinity, Lawrence Park

**Cluster 2:** neighborhoods with highest number of coffee shops. (as shown in **PURPLE**)  
Woburn, Queen's park, East Toronto, Victoria Village, Milliken

**Cluster 3:** neighborhoods with a few coffee shops. (as shown in **GREEN**)  
Berczy Park, St. James Town, Swansea, Runnymede



## Recommendation:

If the person wants to open a median size coffee shop, and if he wants to avoid competition, cluster 1 is the best choice. There are only few or no coffee shop so he does not need to worry about customer.

If the person wants to open a popular coffee shop, like Blue Bottle, Arabica, something like that, coffee shop that famous enough. For this, I think cluster 2 which is neighborhoods with many coffee shops might be the best choice. Like Woburn, Queen's park, East Toronto, in these places, there are often high traffic, many residences, many office workers, and many tourists. Even if there are many coffee shops there, the new famous one would be the choice of people.