IBM APPLIED DATA SCIENCE CAPSTONE

WEEK 5

INTRODUCTION

This is my capstone project for IBM DataScience Professional Certificate. In the last months, people in everywhere is using more than usual their PCs, notebooks, tablets, etc, and it means that hardware or software problems has increased a lot.

I've created an scenario for people who has skill of IT and wants to open a business in Toronto-Canada or for somebody who wants to open a different bussiness.

I've analizated that there are not enough Computer Repair Centers in Toronto area. Therefore, it could be a great opportunity for a entrepeneur who is based in Canada.

BUSINESS PROBLEM

The main objective of this capstone project is to find the most suitable location for the entrepreneur to open a Computer Repair Center in Toronto, Canada. I've used data science methods and tools along with machine learning algorithms such as clustering.

This project aims to provide a solution to answer the business question: In Toronto, if an entrepreneur wants to open an Computer Repair Center, where should they consider opening it?

DATA

I’ve used the following data:
- List of neighborhoods in Toronto, Canada
- Latitude and Longitude of these neighborhoods
- Venue data related to each borough along with the top 6 most common venues. This will help us find neighborhoods that are more suitable to open an Repair Computer Center.

EXTRACTING THE DATA

- The 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.

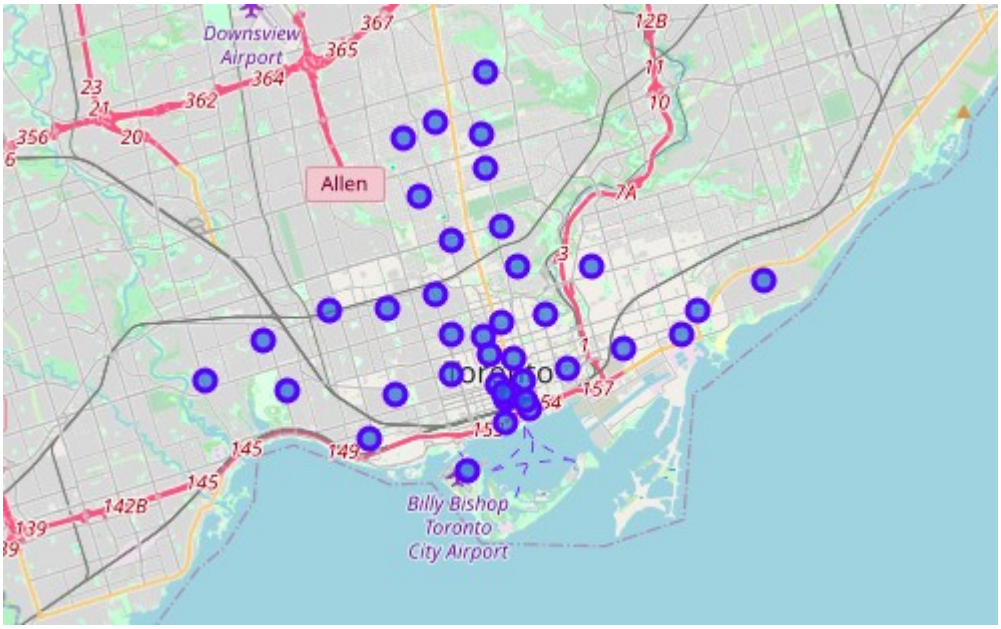
METHODOLOGY

- First, I got the list of neighborhoods in Toronto, Canada. This is possible by extracting the list of neighborhoods from Wikipedia:
“https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M”
- Second, I’ve made an analysis with the top 6 most common venues and the top 10 venues for each borough in Toronto – Canada. I've identified that there are a few business dedicated to repair computers near to top 10 venues.
- Third, I’ve visualized the map of Toronto using Folium package to verify whether these are correct coordinates. Next, I use Foursquare API to pull the list of top 100 venues within 500 meters radius. I have created a Foursquare developer account in order to obtain account ID and API key to pull the data. From Foursquare, I am able to pull the names, categories, latitude, and longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues. Then, I analyze each neighborhood by grouping the

rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later
- Four and final step, I will focus on most promising areas and within those create clusters of locations where an entrepeneur can open a business related to computer repairs. I've presented a map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

RESULT

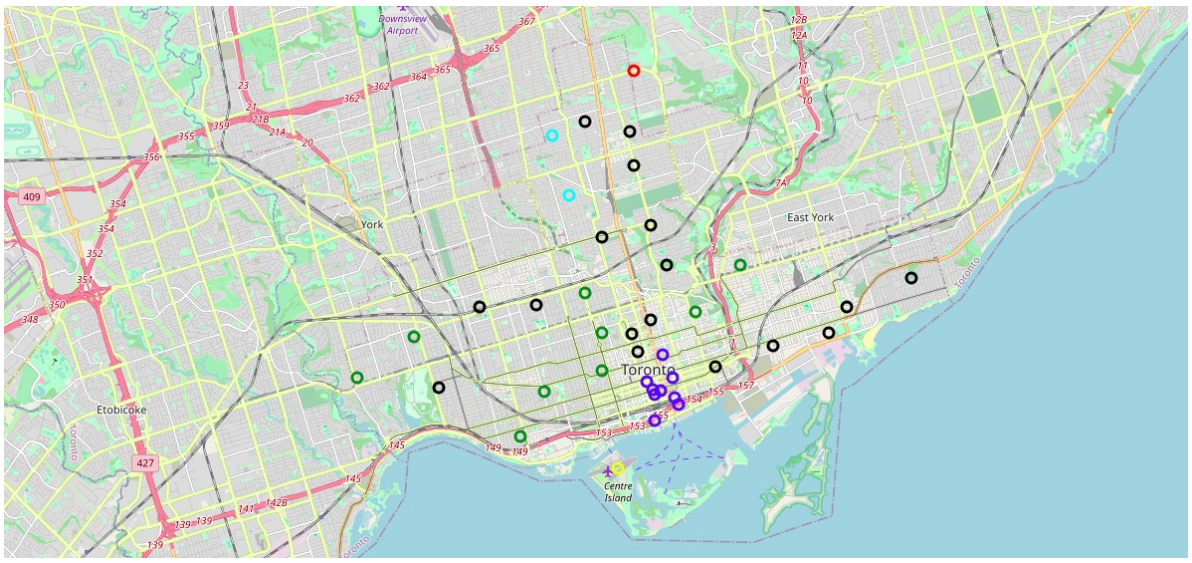
Map of locations in Toronto – Canada.
I've create a map to visualize neighborhoods for 'Toronto'. I've use the first entries latitude and longitude values.



Clusters
The results from k-means clustering show that we can categorize Toronto neighborhoods into 6 clusters based on the frequency of occurrence of each venue category

Theser cluster are the most promising areas to open a business related to repair computers.
Cluster [0] = color 'red'.
Cluster [1] = color 'blue'.
Cluster [5] = color 'black'.

And the followings are the least promising areas to open a business related to repair computers.
Cluster [2] = color 'green'
Cluster [3] = color 'yellow'
Cluster [4] = color 'cyan'



RECOMMENDATIONS

The most promising areas to open a business related to repair computers are in:

Cluster 0: which is around Regent Park, Harbourfront / Queen's Park, Ontario Provincial Government / The Beaches / India Bazaar, The Beaches West /Studio District / Roselawn / Davisville North / Forest Hill North & West, Forest Hill Road Park / North Toronto West, Lawrence Park / Davisville /Runnymede, Swansea / Moore Park, Summerhill East / Summerhill West, Rathnelly, South Hill, Forest / Church and Wellesley / Business reply mail Processing Centre, South C.

Cluster1: which is around Christie / Dufferin, Dovercourt Village / Little Portugal, Trinity / The Danforth West, Riverdale /Brockton, Parkdale Village, Exhibition Place / High Park, The Junction South / The Annex, North Midtown, Yorkville / Parkdale, Roncesvalles / University of Toronto, Harbord / Kensington Market, Chinatown, Grange Park.

Cluster5: which is around Garden District, Ryerson / St. James Town / Berczy Park / Central Bay Street / Richmond, Adelaide, King / Harbourfront East, Union Station, Toronto Islands / Toronto Dominion Centre, Design Exchange / Commerce Court, Victoria Hotel / Stn A PO Boxes / St. James Town, Cabbagetown / First Canadian Place, Underground city.

The others clusters (Cluster 2, 3 y 4) can’t be considering because they have the lowest frequency of occurrence.