Pridicting best neighbourhoods to open a Pizza place in Toronto

Introduction:

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Current to 2016,

the Toronto census metropolitan area (CMA),

of which the majority is within the Greater Toronto Area (GTA), held a population of 5,928,040,

making it Canada's most populous CMA. 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.

Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

Brief history about Toronto: People have travelled through and inhabited the Toronto area, located on a broad sloping plateau interspersed with rivers,

deep ravines, and urban forest,
for more than 10,000 years. After the broadly disputed Toronto Purchase,

when the Mississauga surrendered the area to the British Crown, the British established the town of York in 1793 and later designated it as the capital of Upper Canad

During the War of 1812,

the town was the site of the Battle of York and suffered heavy damage by United States troops.

York was renamed and incorporated in 1834, as the city of Toronto.
It was designated as the capital of the province of Ontario in 1867 during Canadian Confederation.

Problem Description:

Say you want to open a new pizza place where would you open it ,

thats what we will solve in this project .

Objective :

The aim of this report is to study and analyze the neighborhoods of Toronto city

and group them into similar clusters and,

to analyze those clusters to gather meaningful information.

That information can be used to find out neighborhoods that can you open pizza place in it .

Data :

To solve this problem, I will need below data :
- List of neighbourhoods in Toronto, Canada
- Latitude and longitude of these neighbourhoods
- Venue data related to Pizza places

Methodology

First, I need to get the list of neighbouthoods in Toronto, Canada.

This is possible by extracting the list of neighbourhoods from Wikipedia page

("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M")

I did the web scraping by utilizing pandas html table scraping method as it is easier and more

conventient to pull tabular data directly from a web page into dataframe.

However, it is only a list of neighbourhood names and postal codes.

I will need to get their coordinates to utilize Foursquare to pull the list of venues near these neighbourhoods.

To get the coordinates, I used the csv file provided by IBM team to match the coordinates of Toronto neighbourhoods.

After gathering all these coordinates, I 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, categoris, 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 neighbourhood by grouping the rows by neighbourhood and taking the mean

in the frequency of occurence of each venue category.

This is to prepare clustering to be done later.

Here, I made a justification to specifically look for ‘ Pizza Place ’.

Lastly, I performed the clustering method by using k-means clustering.

k-means clustering algorithm identifies k number of centeroids,

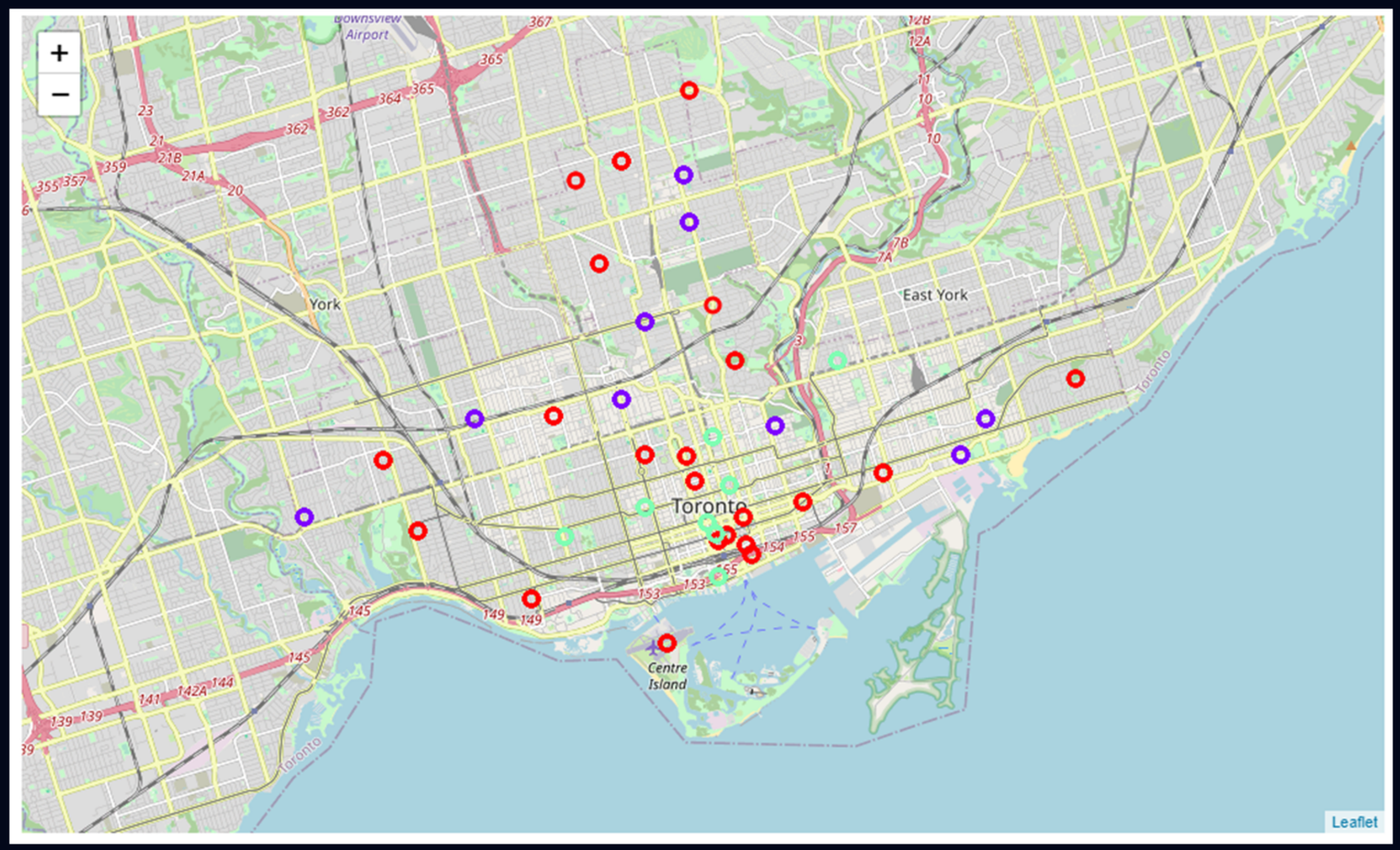
and then allocates every data point to the nearest cluster,

while keeping the centroids as small as possible.

I have clustered the neighbourhoods in Toronto into 3 clusters based on their frequency of occurence for ‘ Pizza place ’.
Based on the results (the concentration of clusters), I will be able to recommend the ideal location to open the restaurant.

Results

Clusters



The results from k-means clutering show that we can categorize Toronto neighbourhoods into 3 clusters based in how many Pizza place in each neighbourhood:

- cluster 0 : neighbouthoods with no Pizza place
- cluster 1 : neighbourhoods with high number of Pizza places
- cluster 2 : neighbourhoods with little or no Pizza restaurants

The results are visualized in the above map with Cluster 0 in red color, Cluster 1 in purple, Cluster 2 in light green color.

Recommendations

Most of Pizza places are in Cluster 1 which is around St. James Town, Cabbagetown areas and lowest (close to zero) in Cluster 0 areas

which are Berczy Park and St. James Town areas. Also, there are good opportunities to open near India Bazaar,

The Beaches Wes, Runnymede, Swansea as the competition seems to be low. Looking at nearby venues,

it seems Cluster 2 might be a good location as there are not a lot of Pizza places in these areas.

Therefore, this project recommends the entrepreneur to open an pizza place in these locations with little to no competition.

Nonetheless, if the food is authentic, affordable and good taste, I am confident that it will have great following everywhere