Applied Data Science
Capstone Project - The Battle of Neighbourhoods

RESTAURANT LOCATION SOLUTIONS

Business Problem

- A group of entrepreneurs are interested in opening a new High-end Italian Restaurant in the city of Toronto in the province of Ontario with a focus on the Downtown area.

- The entrepreneurs would like to use data science techniques to analyse Toronto city for possible locations that will help the restaurant to be successful

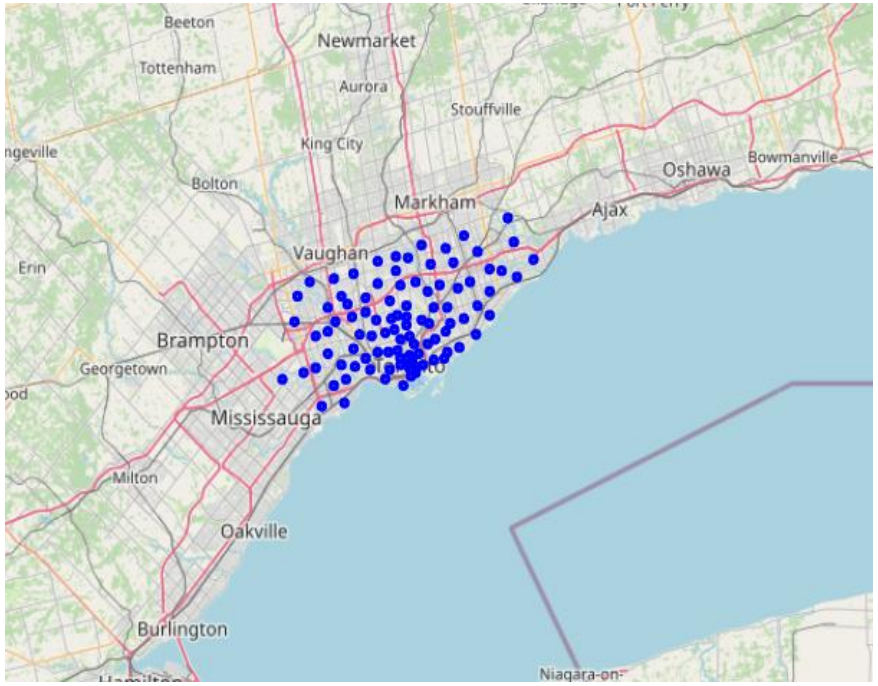
Data Requirements

- Wikipedia site – this will be used to get a list of target neighbourhoods and combined with latitude and longitude location data
- Geopy.geocoders Nominatim python client to locate coordinates of venues throughout Toronto
- Folium will be used to obtain a map of Toronto and matplotlib for plotting locations to help visualize
- Foursquare Data – will be used to explore the target neighbourhoods and segment them into clusters of neighbourhood types. It will also be used to obtain current restaurants in each cluster.
- sklearn.cluster KMeans will be used to help cluster the neighbourhoods

The dataframe has 10 boroughs and 103 neighborhoods.

Approach Overview Toronto City

- Toronto – the name derived from the Huron word for "fishing weir" – is on the northwest shore of Lake Ontario at Latitude 43 39 N, Longitude 79 23 W.
- Utilize Wikipedia to get a table of borough's and neighbourhood names in the city of Toronto.
- Clean dataframe to remove missing data on boroughs & replace missing neighbourhood with borough name. This resulted in a dataframe of 10 boroughs & 103 neighbourhoods.
- Merge dataframe with geospatial latitude and longitude information



The dataframe has 10 boroughs and 103 neighborhoods.

Toronto City – Downtown Area

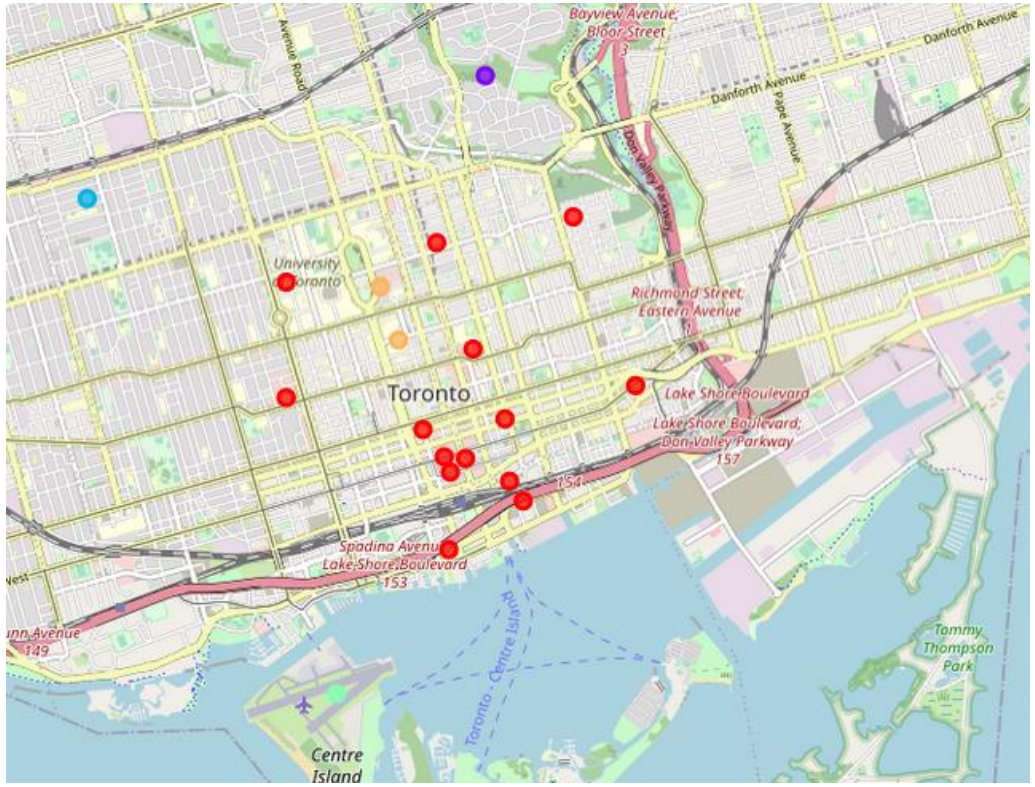
- Since the entrepreneurs are particularly interested in the Downtown Toronto area I dropped the other borough's from the dataframe to focus in on Downtown.

- The Downtown area consists of 19 neighbourhoods

The dataframe has 10 boroughs and 103 neighborhoods.

Toronto City – Downtown Area Analysis

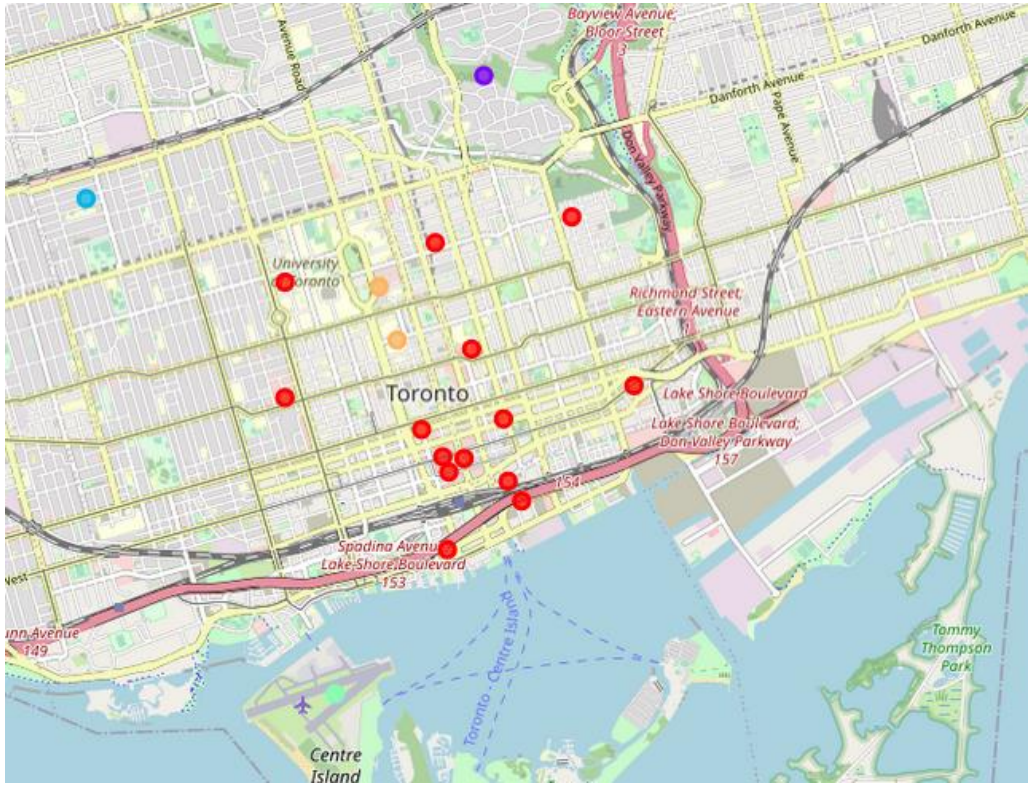
- Utilize Foursquare API to explore neighbourhoods to understand the types and dynamics of each to be able classify and segment them for consideration of new restaurant locations
- Of the 19 neighbourhood 1224 venues were returned with 208 Unique Categories ranging from parks, café's , gyms, airports and shopping centres.
- Focus in on the top ten locations in each neighbourhood to find the most popular venues
- Using K-means clustering I grouped the neighbourhoods that had similar venues



The dataframe has 10 boroughs and 103 neighborhoods.

Toronto City – Downtown Area Analysis cont.

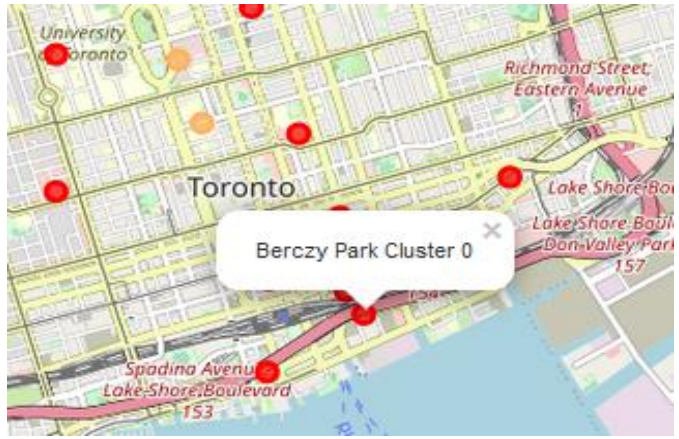
- The K-means clustering resulted in 5 clusters of neighbourhoods
- Analysis of the neighbourhoods allowed me to label them and start to identify suitable locations for a new restaurant.
  - Cluster Red - Food and Services areas. The clusters top ten venues across each of the neighbourhoods contain numerous cafes, restaurants, pubs and coffee shops.
  - Cluster Purple – Parks and residential. The cluster main venues are parks and playgrounds. It would mainly be residential and not a good location for a top end restaurant
  - Cluster Blue - Parks and residential. This cluster includes grocery stores, parks and cafés. It is residential and not a good location
  - Cluster Aqua – Airport and travel. This cluster is close to the airport and consists of airport services, car rental etc. Its not a good location for a restaurant
  - Cluster Orange – Food and Services area. The cluster has numerous restaurants and coffee shops. It's a popular food location but already has 4 Italian restaurants.



The dataframe has 10 boroughs and 103 neighborhoods.

Toronto City – Downtown Area Analysis cont.

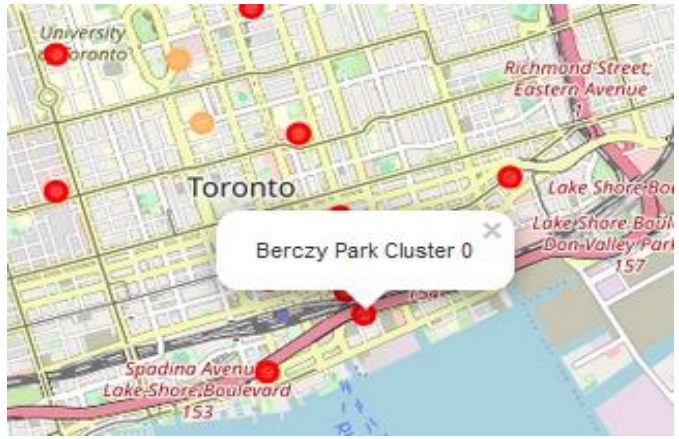
- Based on the analysis of the clusters I identified the red cluster for further investigation.
- The cluster consists of 14 neighbourhoods. Two of the neighbourhoods stand out as ideal locations for a top end restaurant due to their location close to Lake Ontario which provides beautiful water views and their proximity to transports with a train station near bye.
- The 2 ideal locations are Harbourfront East and Berczy Park.



The dataframe has 10 boroughs and 103 neighborhoods.

Toronto City – Downtown Area Analysis cont.

- Further analysis of the 2 neighbourhoods identified that Harbourfront East already has 3 Italian restaurants and Berczy Park had 0.
- Rather then rule out Habourfront straight away I cross checked the reviews of the restaurants
- Taverna Mercatto - Rating 8.0
- Amno Pasta - Rating 7.6
- Piazza Manna - Rating 6.6



The dataframe has 10 boroughs and 103 neighborhoods.

Toronto City – Downtown Area Conclusion

- Both locations are recognised for food services providing good visibility & foot traffic that would attract customers. They are close to Lake Ontario which provides beautiful scenery with water views and their proximity to transports with a train station nearby.
- With Harbourfront East already having 3 Italian restaurants with good reviews competition maybe a concern.
- Berczy does not have any Italian restaurants and is the ideal location recommended to the entrepreneurs.