

Applied Data Science

Capstone Project - The Battle of Neighbourhoods

Restaurant Location Solutions

Introduction/Business Problem

As a Data scientist I have been requested to help a group of entrepreneurs who are interested in opening a new High-end Italian Restaurant in the city of Toronto in the province of Ontario.

They have requested analyses of the city to identify the best location that will help the restaurant to be successful.

The entrepreneurs are particularly interested in the Downtown Toronto area.

Data Requirements

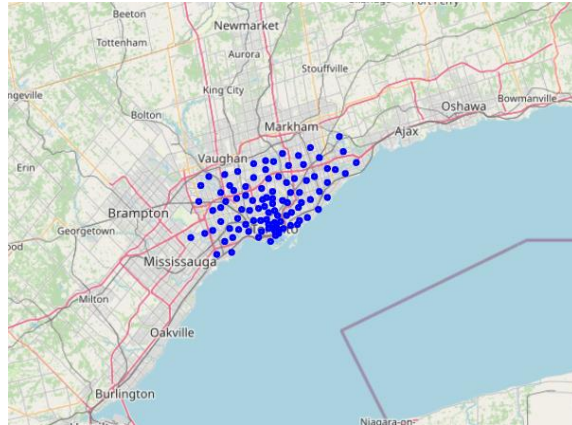
To help the stakeholders identify the best location to open an Italian restaurant I will be using the following data elements;

- Foursquare Location Data – this 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.
- 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
- sklearn.cluster KMeans will be used to help cluster the neighbourhoods

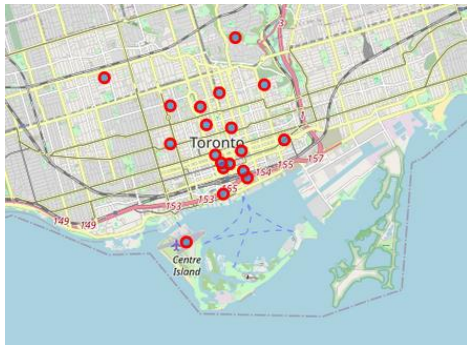
Methodology / Approach taken

1. Utilize Wikipedia to get a table of borough's and neighbourhood names in the city of Toronto. I then used python BeautifulSoup package to webscrab the html table and converted the json into a dataframe
2. Clean dataframe to remove missing data on boroughs & replace missing neighbourhood with borough name. This resulted in a dataframe of 10 boroughs & 103 neighbourhoods.
3. Merge dataframe with geospatial latitude and longitude information
4. Visualise the result
 - Import geopy.geocoders import Nominatim which I then used to get the geo location, latitude and longitude of Toronto City
 - Import folium to get map of Toronto and matplotlib for plotting locations on map

- Created map of Toronto City with neighbourhoods superimposed on top

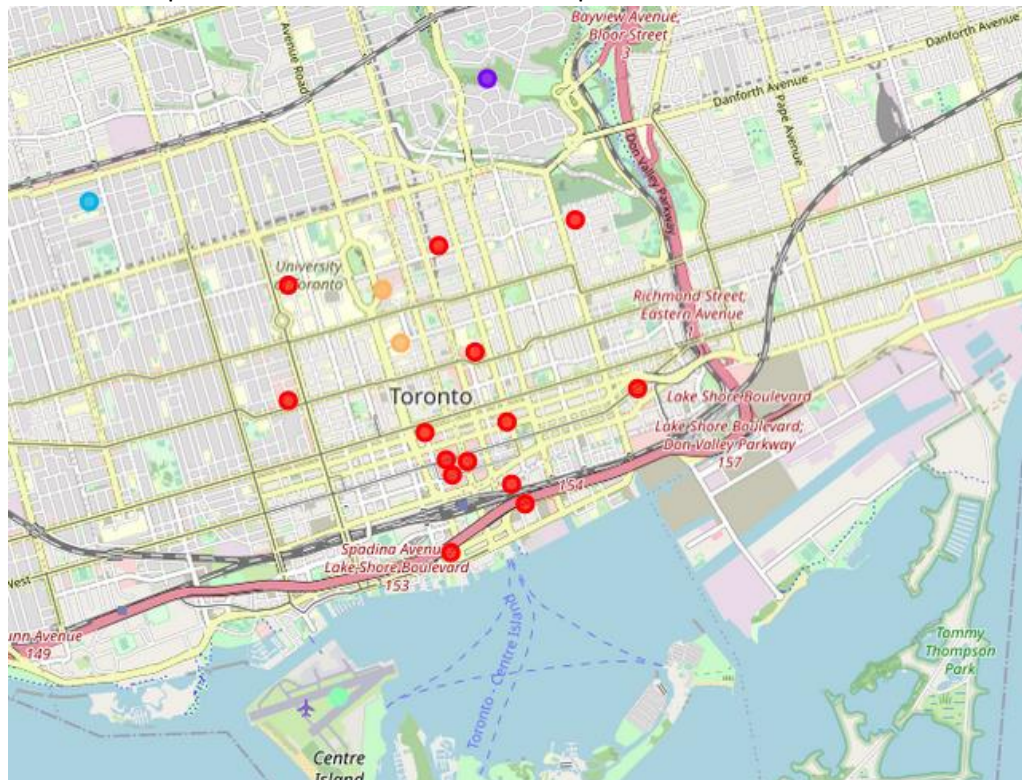


5. Since the entrepreneurs are particularly interested in the Downtown Toronto area I dropped the other borough's from the dataframe to focus in on Downtown. The Downtown area consists of 19 neighbourhoods
6. I created a new map to visualise Downtown neighbourhood locations



7. 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
8. Starting with the first neighbourhood 'Rosedale', I made a call to the Foursquare API to obtain the top 100 venues
9. Converted json result into dataframe and created a function to extract the name & category of each venue. 4 Venues were returned by Foursquare
10. Created a new function to repeat the same process to create a list of the top 100 venues from each of the neighbourhoods in Downtown. Of the 19 neighbourhood 1224 venues were returned
11. Grouped the data to check the number of venues returned for each Neighbourhood & checked the number of unique categories, 208.
12. Used python one-hot to create a list of the categories converted into binary vectors. Grouped the resulting rows by neighbourhood and by taking the mean of the frequency of occurrence of each category
13. Created a new dataframe to rank the top 10 most common venues for each neighbourhood. This allowed me to get an understanding of the most common venues in each neighbourhood.
14. Used K-means to cluster the neighbourhoods into 5 clusters. This allowed me to group similar neighbourhoods. I created a new dataframe that included the top 10 venues for each neighbourhood in each cluster

15. Created a map to visualize the clusters in their specific locations in Downtown Toronto



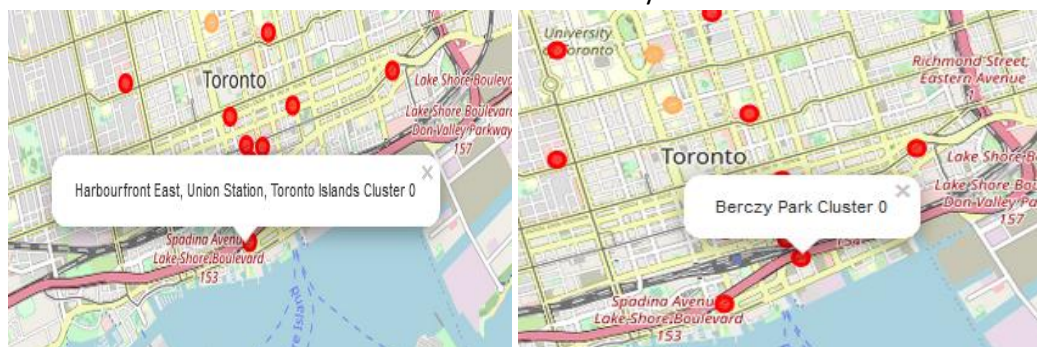
16. I then classified each cluster based on their city location & top 10 venue's as follows;

- 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 has 4 Italian restaurants.

17. Based on the analysis of the clusters I identified the red cluster for further investigation.

The cluster consists of 14 neighbourhoods. Within 2 neighbours 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.



I then analysed the neighbourhoods to I identified the number of Italian restaurants to check for specific competition

	Neighborhood	Italian Restaurant
0	Central Bay Street	4
1	Christie	1
2	Commerce Court, Victoria Hotel	3
3	First Canadian Place, Underground city	1
4	Garden District, Ryerson	2
5	Harbourfront East, Union Station, Toronto Islands	3
6	Queen's Park, Ontario Provincial Government	1
7	St. James Town	2
8	St. James Town, Cabbagetown	2
9	Stn A PO Boxes	3
10	Toronto Dominion Centre, Design Exchange	3
11	University of Toronto, Harbord	2

Harbourfront East had 3 Italian Restaurants & Berczy Park had zero.

18. I used foursquare to request the reviews for each restaurant in Harbourfront East.

Taverna Mercatto - Rating 8.0

Amno Pasta - Rating 7.6

Piazza Manna - Rating 6.6

Results

Using the locations and neighbourhood dynamics I identified 2 locations that the entrepreneurs should consider for their new restaurant location Harbourfront East and Berczy Park.

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.

Harbourfront East already has 3 Italian restaurants with good reviews so competition maybe a concern.

Berczy does not have any Italian restaurants and is the ideal location recommended to the entrepreneurs.

Discussion

The approach and methods used provided good overall research and analysis of Toronto to identify a good location.

Conclusions

I would recommend the entrepreneurs spend more time collecting information on specific buildings and rental costs before making any final decisions.