Capstone Project - The Battle of Neighbourhoods

Paul Berry

June 2020

Introduction/ Business Problem

Famous Original Ray's Pizza has five locations in New York City. A business consortium is looking at establishing a new location in Toronto, and want to find an appropriate location.

They know that the existing locations in New York are appropriate, to different extents, but don't know anything about Toronto neighbourhoods. The goal of this project is to identify neighbourhoods in Toronto which are most similar to those in New York City with successful existing locations.

Identifying locations that have similar characteristics will give the new store the best chance of success.

Data

Using Foursquare data, we will examine the areas around the existing stores in New York City, and neighbourhoods in Toronto, and then use K-means clustering to find neighbourhoods in Toronto that match existing locations in New York City.

There are five existing stores, we have codes for the existing stores, their relative financial performance, and coordinates:

Existing Store A, Below Average, 40.7460427,-74.001496

Existing Store B, Above Average, 40.7603983,-73.9841844

Existing Store C, Average, 40.7632311,-73.981568

Existing Store D, Average, 40.7642512,-73.9664773

Existing Store E, Average, 40.7841962,-73.9743688

These locations, as well as locations of Toronto neighbourhoods, as derived in a previous project, will be entered into an API call to Foursquare as follows to obtain information on nearby venues.

https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}

This will provide information of nearby venues, which will be aggregated by category (pizza store, asian market, etc.) for analysis.

Ideally, as well as finding a location that is similar to that of existing stores, we will be able to identify which of these locations most closely matches the best performing store.

Methodology

Using the Foursquare API, queries are run for venues around the 103 Toronto Postcodes, and five existing New York locations. This information is grouped by venue category and summarised at an area level. This enables each location to have the proportion of each category calculated, which is the basis for the clustering. The ten most common categories are also derived, to help with descriptions.

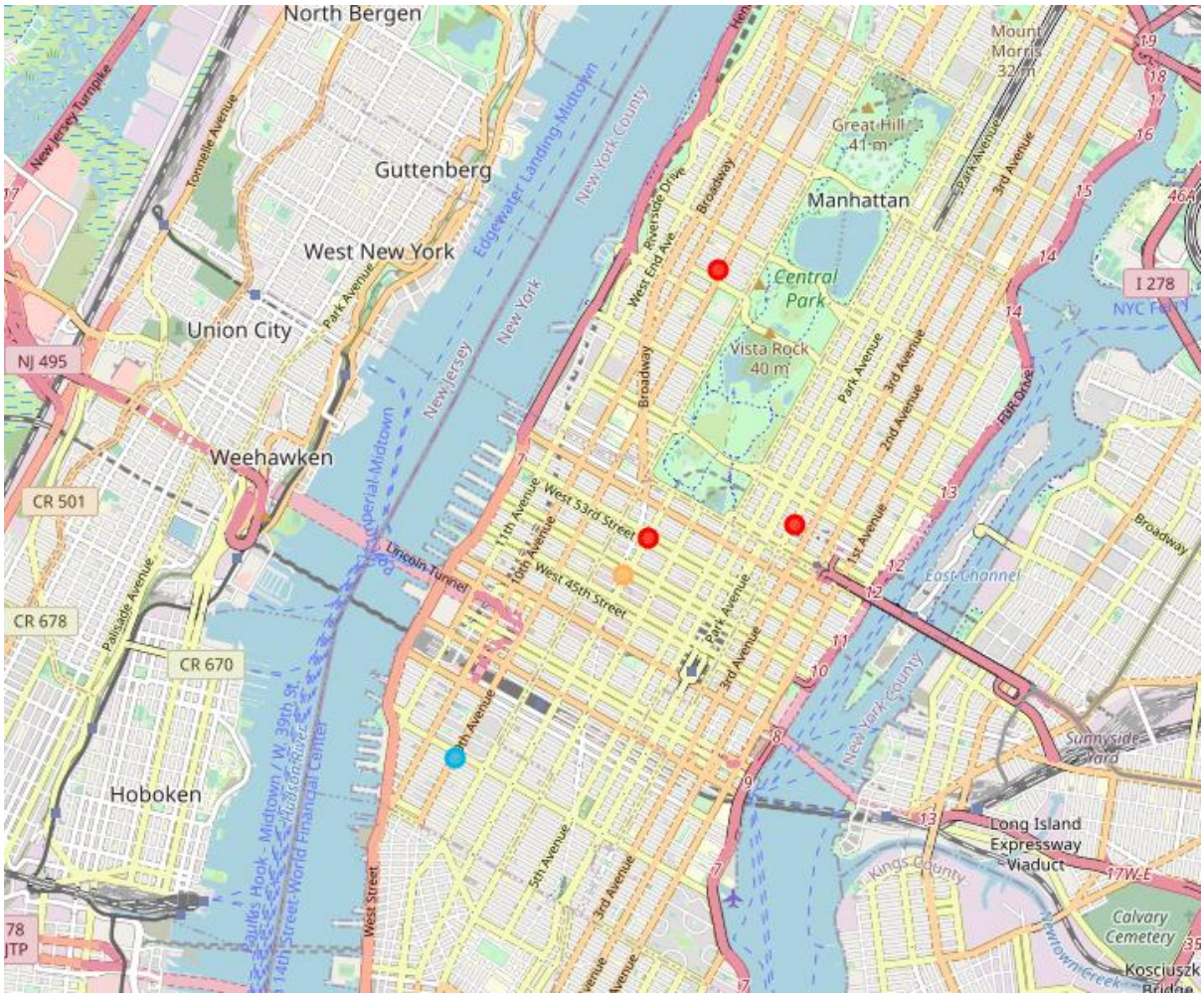
The existing stores are included in the data so that they will form the basis of the clustering.

Once the data is combined at a postcode/store level, K-means clustering is run on this dataset, with five clusters, as the data is not wide enough to support more without introducing an unacceptable level of noise.

Results

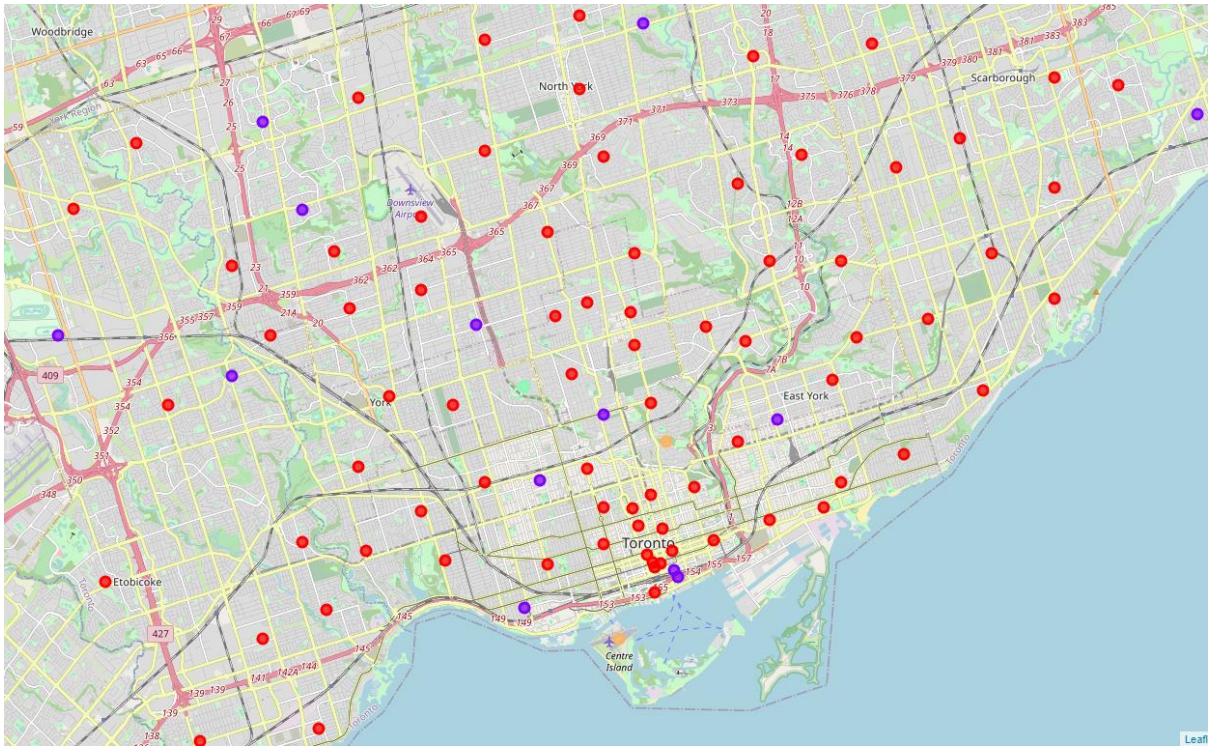
Borough	**Location**	**Cluster**	**Financial Performance**
New York	Existing Store A	2	Below Average

New York	Existing Store B	4	Above Average
New York	Existing Store C	0	Average
New York	Existing Store D	0	Average
New York	Existing Store E	0	Average



Clustering of the existing stores aligned well with the prior known financial performance metric.

Most of the regions in Toronto are likewise clustered into group 0, several in group 4 (the highest performing store), and other clusters that don't have an analogue in New York.



The Toronto neighbourhoods that are in the desirable cluster 4 are as follows (shown in orange on the map:

Cluster Labels	Postal Code	Borough	Neighbourhood
4	NYB	New York	Existing Store B
4	M2H	North York	Hillcrest Village
4	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport
4	M4W	Downtown Toronto	Rosedale

Discussion

The clustering aligns well with the known financial performance of the stores, which gives good confidence that the projection onto Toronto will be representative of the existing situation in New York.

However, we must also keep in mind that the types of venues in a neighbourhood is not the only variable when comparing environs. There are many social differences between the United States and Canada in general, and New York City and Toronto in particular that must be considered. Thus, while this analysis is a good first step in finding an appropriate location, it should not be considered sufficient and complete for these purposes.

The analysis yields three neighbourhoods that are superficially similar to the highest performing existing location in New York City. Hillcrest Village, Rosedale, and the downtown area comprising King and Spadina, Harbourfront West, Bathurst Quay, etc. These areas indicate further study, either through further analysis of existing data, or commissioned market research, such as pop-up stores in these areas. Further analysis regarding socioeconomic indicators may also yield useful results.

Conclusion

This study has found three areas in Toronto which are good matches for existing successful locations. Hilcrest Village, Rosedale, and Harbourfront West/Bathurst Quay. These locations closely match the best performing location in New York City, and are good candidates for expansion, or further analysis.