Strategic Location for Purchasing a Residential Property for Short-term Rent Out (Airbnb)

Amin Sehati

January 1, 2021

1. Introduction

1.1 Background

Airbnbs are becoming more popular. They offer cheap short-term stay while providing the convinience of a complete property. However, they need to compete with the hotels in their vicinity. Hotels often host a restaurant which could be an important factor for those who care about having their meals close to their temporary location. Hence, for an Airbnb to be attractive enough for these kind of tourists, the number of restaurants in the neighborhood is of high importance.

1.2 Problem

A client is looking for the best neighbourhood to purchase a residential property to rent out as an Airbnb. Where could they purchase a rather cheap property which could attract a reasonable amount of tenants. Our reasoning would be based on number of hotels in the neighbourhood (which indicates level of competition), number of restaurants in the area (as an highly influential factor when choosing an Airbnb), and price of the residential properties in the area (as a decision factor for the client).

In this analysis, the influence of other Airbnbs in the area is neglected. Furthermore, other factors to choose an Airbnb such as number of point-of-interests, museums, etc. are not included in this research. We also neglect the effect of costly restaurants and treat all restaurants as equally influential on the choise of Airbnb location.

1.3 Interest

Anyone with interest in short-term accommodation industry in Toronto could potentially benefit from this analysis. The main target audience, however, is those who seek cheap housing to purchase and rent out as an Airbnb.

2. Data acquisition and cleaning

2.1 Data sources

Foursquare API will be used to fetch data on venues in the neighbourhoods of Toronto. Neighbourhood names will be retrieved and scraped from Wikipedia (https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=945633050) based on their borough and postal codes.

To find out the price of residential properties in Toronto we use the data file provided in the following address which gives us the price, neighbourhood, and geo-coordinates of many residentials properties in Toronto: https://www.kaggle.com/mnabaee/ontarioproperties

Although the data is from 2016, we can still use it for today's comparison since the goal is to compare different neighbourhoods' price level. We assume the residential properties are representitive and sufficient for our conclusion.

2.2 Data cleaning

Once the data on various neighbourhoods in Toronto is scrapped from Wikipedia. We add the geographical coordinates of each neighbourhood to the table.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.75245	-79.32991
1	M4A	North York	Victoria Village	43.73057	-79.31306
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.65512	-79.36264
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.72327	-79.45042
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.66253	-79.39188

Figure 1 - Example of Neighbourhoods in Toronto with their corresponding Postal Code and Borough and Geographical Coordinates

From the CSV file on housing price in Toronto, we have the address, area name (neighbourhood), Price, and latitude and longitude of the property.
We decide to remove invalid data which we assume is the residential properties with prices less than $50k. Although we can fill the missing data using average method, it is decided to remove them since price is the only feature useful for us in this dataset and adding an average value in a neighbourhood will not be advantages.

	Address	AreaName	Price ($)	lat	lng
0	86 Waterford Dr Toronto, ON	Richview	999888	43.679882	-79.544266
4	#1409 - 230 King St Toronto, ON	Downtown	362000	43.651478	-79.368118
5	254A Monarch Park Ave Toronto, ON	Old East York	1488000	43.686375	-79.328918
12	3 Bracebridge Ave Toronto, ON	Old East York	599900	43.697842	-79.317368
15	#710 - 1080 Bay St Toronto, ON	Downtown	805900	43.666794	-79.388756

Figure 2 - Example of Housing Price in Toronto

In order to merge these two datasets, we would need to find the corresponding neighbourhood for each Area Name mentioned in the housing dataset and replace the area name with the name of the neighbourhood from the scraped dataset. Using the mean value of the residential properties in each area, we find the closest neighbourhood. The result is a table which shows the price of a property and its corresponding neighbourhood.

	Address	AreaName	Price ($)
0	86 Waterford Dr Toronto, ON	Kingsview Village, St. Phillips, Martin Grove ...	999888
1	#1409 - 230 King St Toronto, ON	Richmond, Adelaide, King	362000
2	254A Monarch Park Ave Toronto, ON	Woodbine Heights	1488000
3	3 Bracebridge Ave Toronto, ON	Woodbine Heights	599900
4	#710 - 1080 Bay St Toronto, ON	Richmond, Adelaide, King	805900

Figure 3 - Example of the resulting merge for the two datasets (neighbourhoods and properties)

For the information on the venues, we retrieve the data using Foursquare API which gives us 2379 entries for 267 different types of venues together with the name of their neighbourhood and geographical location.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.75245	-79.32991	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.75245	-79.32991	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.73057	-79.31306	Wigmore Park	43.731023	-79.310771	Park
3	Victoria Village	43.73057	-79.31306	Memories of Africa	43.726602	-79.312427	Grocery Store
4	Victoria Village	43.73057	-79.31306	Vinnia Meats	43.730465	-79.307520	German Restaurant

Figure 4 - Example of the venues data retrieved from Foursquare API

2.3 Feature Selection

After data cleaning, there were 4901 samples of residential property with its price and neighbourhood. It is obvious that we would need these two features in order to judge on the housing price of each neighbourhood.

From the venue dataset, we would need to extract information on which neighbourhood would be the most attractive to short-time travelers. There are 267 different venue types (as features) to select, each one with the potential to be somehow attractive to travelers. However, based on the business problem that we mentioned, it is decided to select the feature which would give the number of all restaurants in the neighbourhood. Furthermore, to figure out which neighbourhood would have the least amount of competition, we will utilize the number of hotels in the neighbourhood as another feature.

3. Exploratory Data Analysis

3.1 Representative value for price of residential properties

In order to come up with a calculation method to represent the price of residential properties, we need to take a look at our available dataset. As can be seen in the figure below, most of the properties are priced between $100k and $400k. It can also be seen that there are some neighbourhoods which very expensive accommodation.

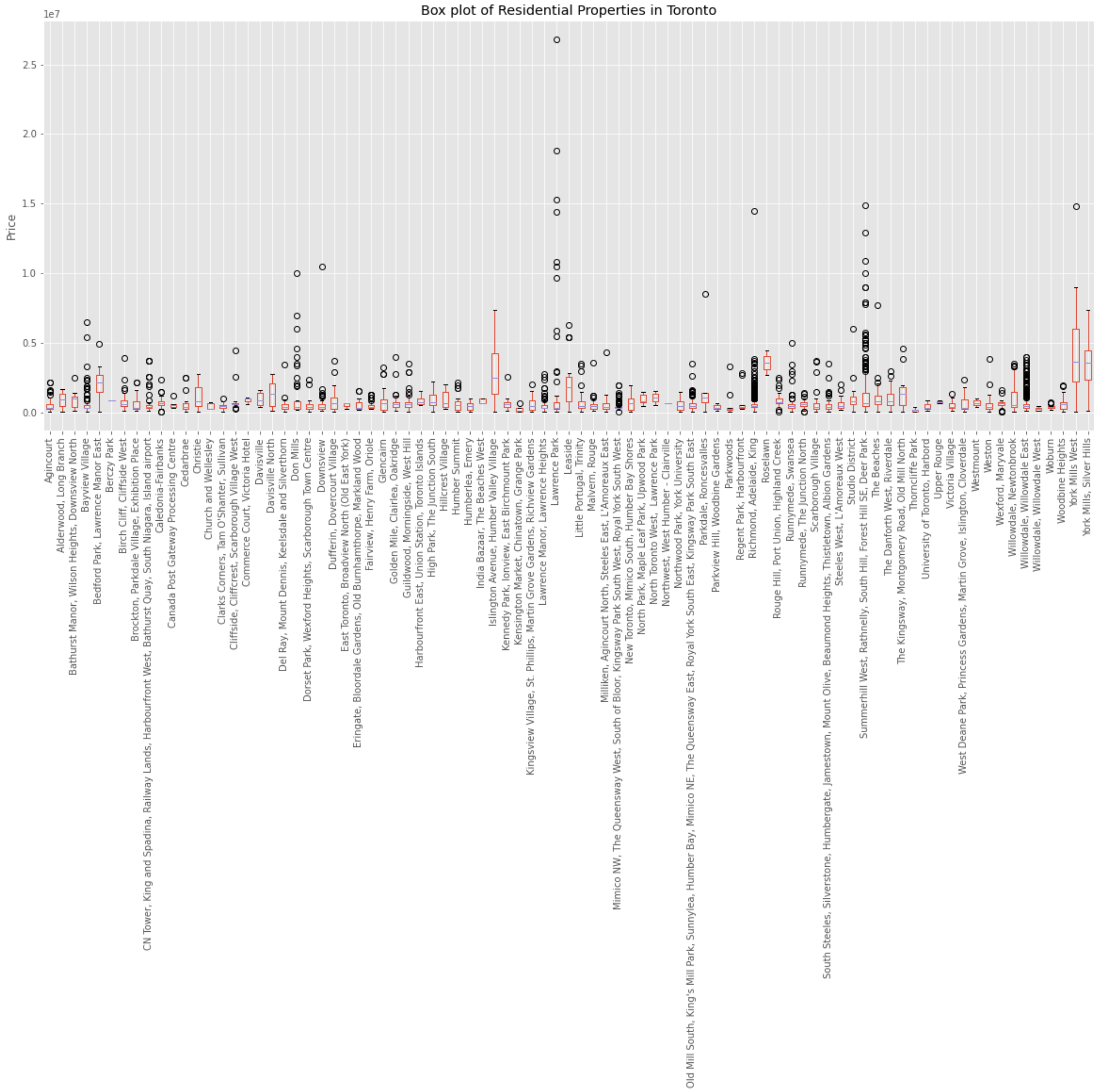

Figure 5 - Box plot of prices for residential properties

Based on the distribution of prices over the neighbourhoods, it seems that taking the mean value of the properties in each neighbourhood would be an acceptable representative of how much on average would cost to purchase a property in each neighbourhood with the assumption that the data here represents different sizes, built year, and characteristics of the buildings in that area. Hence, we calculate the mean value of the prices accordingly.

	Neighbourhood	Price ($)	Latitude	Longitude
0	Agincourt	4.858069e+05	43.79452	-79.26708
1	Alderwood, Long Branch	9.331101e+05	43.60124	-79.53879
2	Bathurst Manor, Wilson Heights, Downsview North	8.759328e+05	43.75788	-79.44847
3	Bayview Village	6.717563e+05	43.78112	-79.38060
4	Bedford Park, Lawrence Manor East	2.037382e+06	43.73545	-79.41916

Figure 6 - Mean value of prices in each neighbourhood

3.2 Representative value for hotels and restaurants in each neighbourhood

Using the dataset on the venues, we can simply sum up the number of all types of hotels in each neighbourhood calculate its mean frequency. This would show how frequent hotels exists in each neighbourhood and thus the amount of competition.

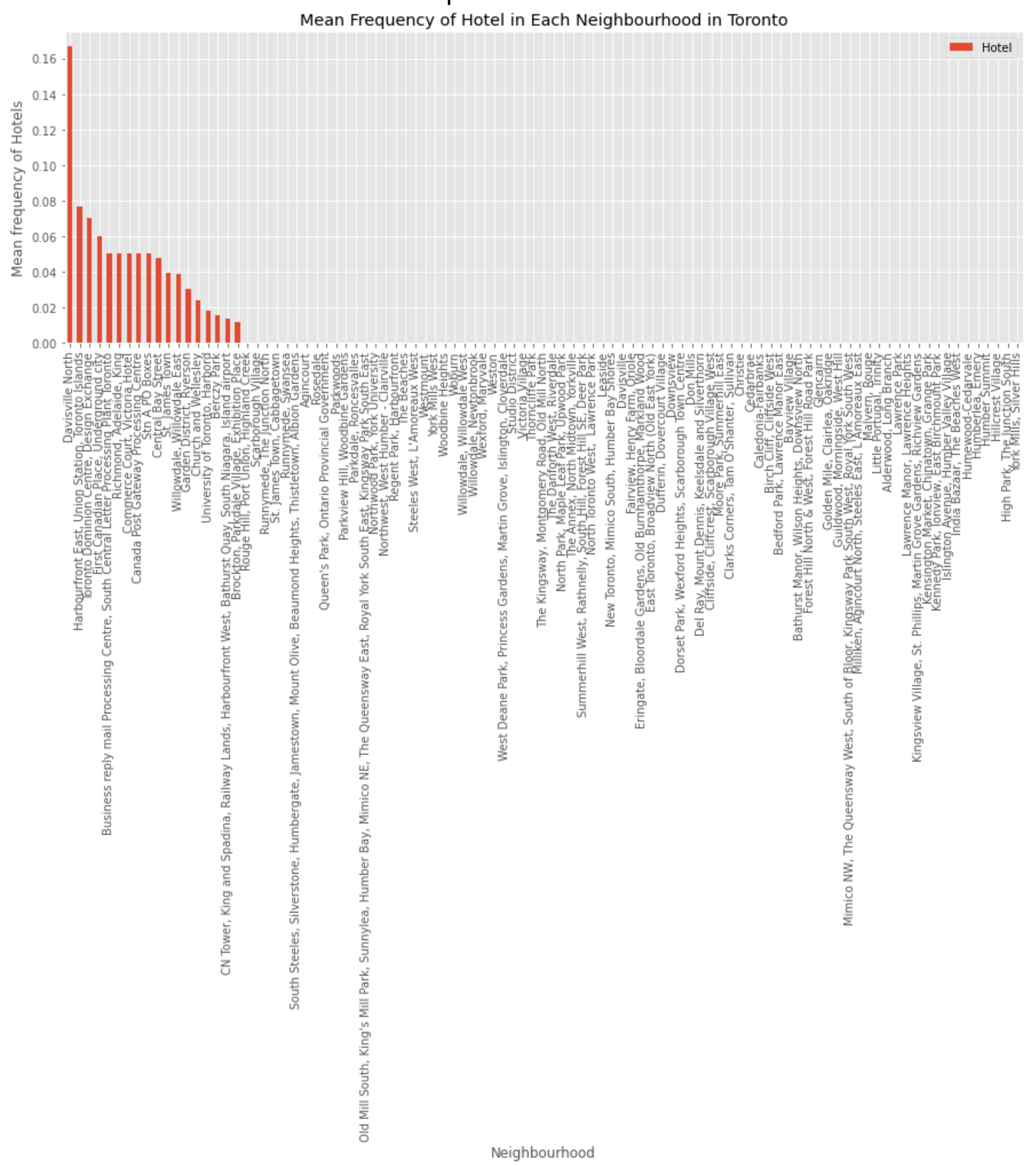

Figure 7 - Mean Frequency of Hotel in Each Neighbourhood in Toronto

Using the same method, we calculate the mean frequency of restaurants in each neighbourhood in Toronto which would result in the figure below.

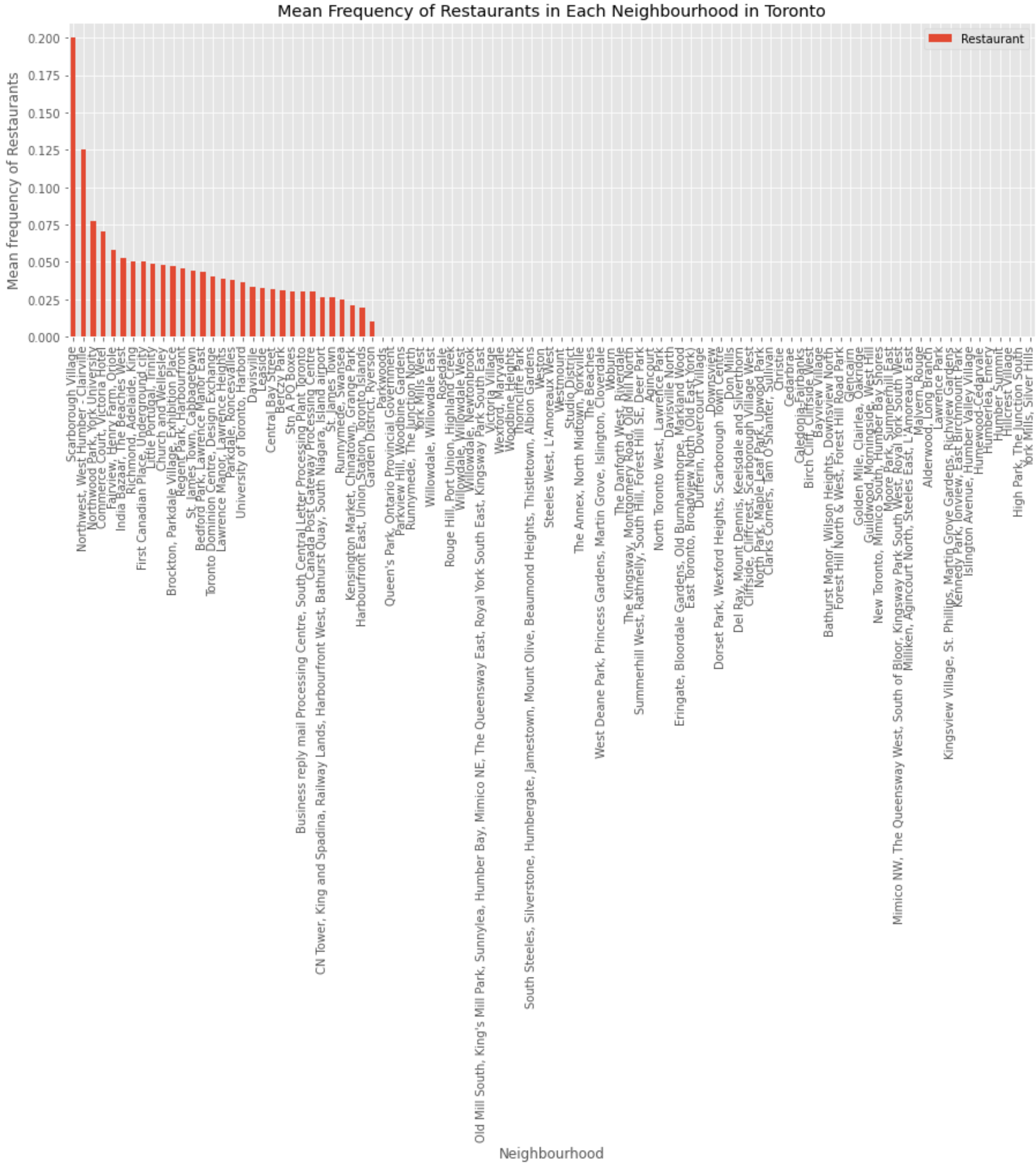

Figure 8 - Mean Frequency of Restaurants in Each Neighbourhood in Toronto

3.3 Normalizing

Once we have the mean price, mean frequency of hotels and restaurants, we can normalize them in order to reach higher quality in our statistical analysis. The following table shows part of this normalized dataset.

	Price	Hotels	Restaurants
0	-0.595279	-0.275696	-0.460351
1	0.046552	-0.275696	-0.460351
2	-0.035492	-0.275696	-0.460351
3	-0.328462	-0.275696	-0.460351
4	1.631059	-0.275696	0.961882

Figure 9 - Normalized dataset

4. Results

Since we would like to group neighbourhood based on our available dataset, we choose to use clustering. With this method we can show our client that for each pricing level, what level of competition and attractiveness does each neighbourhood offer.

4.1 Finding the optimal number of clusters (k)

To find an optimal number of clusters we use the “elbow method”. Although it is hard based in the following figure to judge on the optimal value, we decide to proceed with k=5 as it provides an acceptable squared error while maintaining a rather low number of clusters.

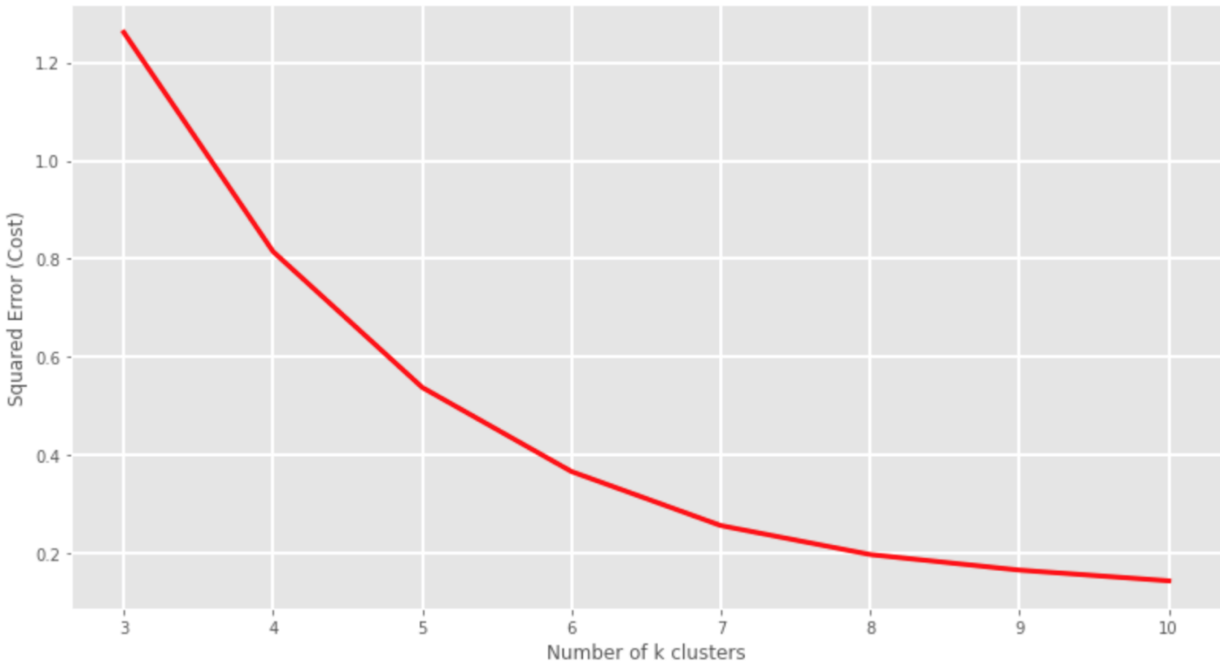

Figure 10 - Finding the optimal number of clusters

4.2 Analysing the clusters

By plotting the cluster of neighbourhoods on the map of Toronto, it can be seen that there is a cluster populated mainly near downtown Toronto (as expected) shown as orange, one big cluster distributed over the city shown as purple, one cluster around green areas (parks) of the city shown as light green. One cluster with one neighbourhood shown as cadet blue, and two neighbourhoods each on one side of the city shown as red.

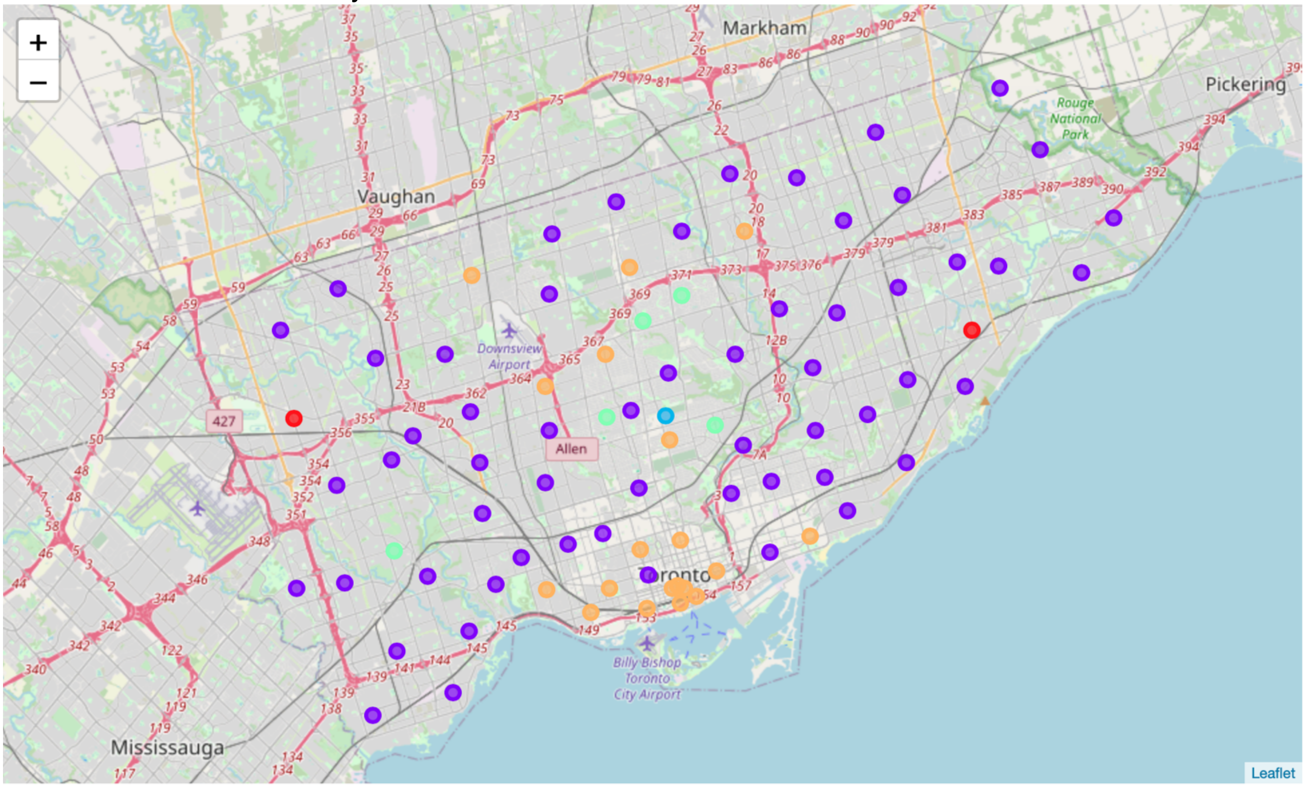

Figure 11 - Clusters of neighbourhoods shown on Toronto's map

5. Discussion

Cluster 0 (red): Low Housing Price, Low Number of Hotels (competition), High amount of restaurants

With two neighbourhoods on each side of the city, it offers low housing price and very high number of restaurants while having rather low amount of competition which makes it very attractive as our target.

	Cluster Label	Neighbourhood	Price ($)	Latitude	Longitude	Hotels	Restaurants
50	0	Northwest, West Humber - Clairville	649999.00	43.71174	-79.57941	-0.275696	3.628570
62	0	Scarborough Village	631907.96	43.74446	-79.23117	-0.275696	6.081922

Cluster 1 (purple): Med Housing Price, Low Number of Hotels (competition), Low number of restaurants

As expected, most of the neighbourhoods in the city have medium housing price with a few hotels and restaurants in their neighborhood.

	Cluster Label	Neighbourhood	Price ($)	Latitude	Longitude	Hotels	Restaurants
0	1	Agincourt	4.858069e+05	43.794520	-79.267080	-0.275696	-0.460351
1	1	Alderwood, Long Branch	9.331101e+05	43.601240	-79.538790	-0.275696	-0.460351
2	1	Bathurst Manor, Wilson Heights, Downsview North	8.759328e+05	43.757880	-79.448470	-0.275696	-0.460351
3	1	Bayview Village	6.717563e+05	43.781120	-79.380600	-0.275696	-0.460351
6	1	Birch Cliff, Cliffside West	8.311808e+05	43.695100	-79.264660	-0.275696	-0.460351
9	1	Caledonia-Fairbanks	7.186862e+05	43.687840	-79.450460	-0.275696	-0.460351
11	1	Cedarbrae	5.316326e+05	43.769440	-79.238920	-0.275696	-0.460351
12	1	Christie	1.229167e+06	43.668690	-79.420710	-0.275696	-0.460351
14	1	Clarks Corners, Tam O'Shanter, Sullivan	4.691393e+05	43.784910	-79.297220	-0.275696	-0.460351
15	1	Cliffside, Cliffcrest, Scarborough Village West	1.035046e+06	43.723600	-79.234960	-0.275696	-0.460351
19	1	Del Ray, Mount Dennis, Keelsdale and Silverthorn	5.250651e+05	43.695170	-79.483970	-0.275696	-0.460351
20	1	Don Mills	1.257380e+06	43.735455	-79.352690	-0.275696	-0.460351

Cluster 2 (cadet blue): High Housing Price, Very high number of Hotels, Low number of Restuarants
This cluster which contains only one neighbourhood, shows very high housing price and hotels with very low number of restaurants in the vicinity. It would be least attractive choice of all clusters for us.

	Cluster Label	Neighbourhood	Price ($)	Latitude	Longitude	Hotels	Restaurants
18	2	Davisville North	1.305829e+06	43.71276	-79.38851	7.311636	-0.460351

Cluster 3 (light green): Very high Housing Price, Low Number of Hotels (competition), Low number of restaurants
Although the amount of competition is low in this cluster, but also the property prices are very high and there are not many restaurants nearby. Probably these neighbourhoods are luxury and quiet parts of the city and not an interesting choice for us.

	Cluster Label	Neighbourhood	Price ($)	Latitude	Longitude	Hotels	Restaurants
36	3	Islington Avenue, Humber Valley Village	2.682881e+06	43.66263	-79.52831	-0.275696	-0.460351
42	3	Leaside	2.167378e+06	43.70902	-79.36349	-0.275696	0.594855
58	3	Roselawn	3.589950e+06	43.71208	-79.41848	-0.275696	-0.460351
83	3	York Mills West	4.315027e+06	43.74778	-79.40033	-0.275696	-0.460351
84	3	York Mills, Silver Hills	3.548245e+06	43.75698	-79.38060	-0.275696	-0.460351

Cluster 4 (orange): Med Housing Price, Med number of Hotels, Med number of Restaurants
As mentioned before, the orange neighbourhoods are mostly located near downtown of Toronto, where there are all kinds of properties, hotels, and restaurants but with moderate amount and price. This could be also a good candidate although demanding further analysis of each neighbourhood in its cluster.

	Cluster Label	Neighbourhood	Price ($)	Latitude	Longitude	Hotels	Restaurants
4	4	Bedford Park, Lawrence Manor East	2.037382e+06	43.73545	-79.41916	-0.275696	0.961882
5	4	Berczy Park	8.999000e+05	43.64536	-79.37306	0.435616	0.561879
7	4	Brockton, Parkdale Village, Exhibition Place	5.702839e+05	43.63941	-79.42676	0.259880	1.079008
8	4	CN Tower, King and Spadina, Railway Lands, Har...	6.235956e+05	43.64082	-79.39818	0.331290	0.411952
10	4	Canada Post Gateway Processing Centre	5.393529e+05	43.64869	-79.38544	2.000503	0.520990
13	4	Church and Wellesley	4.274600e+05	43.66659	-79.38133	0.808208	1.097333
16	4	Commerce Court, Victoria Hotel	9.029500e+05	43.64840	-79.37914	2.000503	1.829445
17	4	Davisville	9.634667e+05	43.70340	-79.38659	-0.275696	0.630028
26	4	Fairview, Henry Farm, Oriole	4.867577e+05	43.78097	-79.34781	-0.275696	1.426843
30	4	Harbourfront East, Union Station, Toronto Islands	9.042000e+05	43.64285	-79.38076	3.226149	0.168714

6. Conclusion

In this study, the aim was to recommend the best neighbourhood for a client to invest in by purchasing a residential property to rent out as an Airbnb. We assumed that the main decision criterion would be high number of restaurants in the vicinity of the property. Based on this analysis, we found that there are two neighbourhoods Northwest and Scarborough Village are the first choices to consider. Further analysis should be done on the neighbourhoods to see what other attraction points they offer, and which one would be more preferable in matter of tourist attraction.

7. Future directions

Since we used the housing price data from 2016, one could find a more recent data with more samples in order to find a more reliable data on the property values. Also, our assumption of features was restricted to restaurants while it could be expanded to other categories such as park, museums, and coffee shops.

As a future research, one could use scraping to extract the data of location and pricing from Airbnb in order to analyse the competition in more detail.