Introduction

Opening a new restaurant can be a daunting task. It involves considering some important factors such as; Location, Property prices, Safe areas & availability of the customers as well as Competition. These factors not only have influence on the startup of the restaurant but also on the success of the restaurant.

With the advancements in the field of data science such daunting task could be quantitatively tackled, and useful recommendations can be provided to the potential clients that are planning on opening a new restaurant in a specific city.

In this report we analyze some of the important factors mentioned above utilizing unsupervised machine learning algorithm and visualization tools; hen provide final recommendations for opening a new restaurant in the city of Toronto. Note that this analysis serves as a blueprint and as such can also be carried out with regards to opening a new restaurant in any other city, given availability of similar data.

Data

Complete Project Notebook link

Following data has be utilized in this analysis

1. Wellbeing data provided by City of Toronto[1]. The features utilized from this dataset were:
    a. Neighbourhood
    b. Average Family Income
    c. Lat (Latitude)
    d. Lng (Longitude)

2. Property pricing data obtained from Kaggle[2]. The features utilized from this dataset were
    a. Area Name
    b. Price ($)
    c. Lat (Latitude)
    d. Lng (Longitude)

3. Foursquare Places API[3] for obtaining recommended venues and restaurants in Toronto and their popularity via the number of ‘likes’ from the Foursquare database.
    a. Name (name of the place)
    b. Categories (Which category, for e.g. Italian restaurant, Museum, Art Gallery)
    c. Lat (Latitude)
    d. Lng (Longitude)
    e. Id (Unique identifier of the place)
    f. Likes (Number of ‘likes’ given by people who visited that place)

4. Additionally, following Libraries were utilized in this project.
    a. Numpy, Pandas, Math – Data analysis
    b. Matplotlib – Data visualization
    c. Beautiful Soup – Scraping webpage data
    d. Geopy – Obtaining geospatial coordinates (latitude and longitude)
    e. Folium – Visualization of Toronto map and marking important areas
    f. Sklearn – KMeans unsupervised clustering algorithm
    g. Requests – Sending organic, grass-fed HTTP/1.1 requests

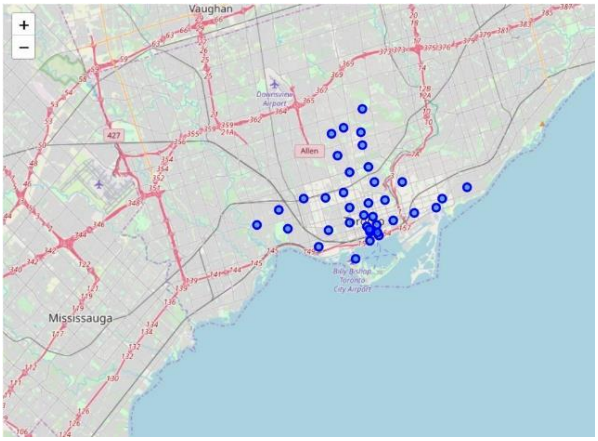
Methodology & Results

As discussed in the introduction, multiple factors need to be considered when opening a new restaurant. Here, we discuss these important factors and motivation in sections,

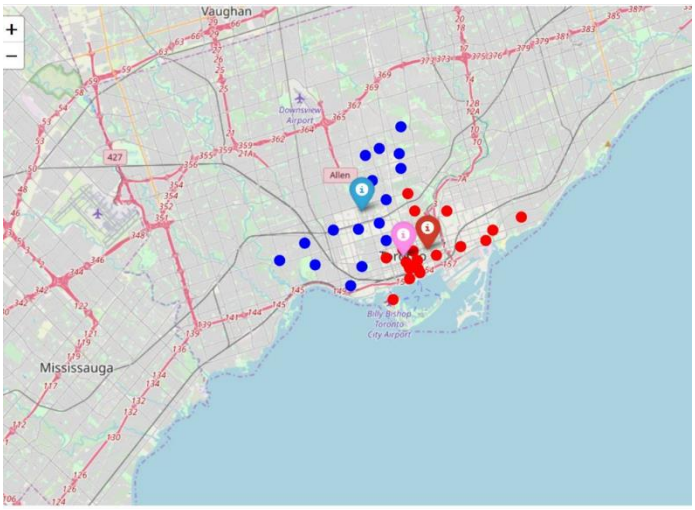
Section – 1: Best Borough and Top Recommended Venues in Toronto

In this section we try to find the best area in Toronto with regards to density of top recommended venues for tourists and then mark these top recommended areas as well top 5 venues and top 3 restaurants on the Toronto map. This would be useful because identification of areas that are in proximity to top recommended venues in Toronto would suggest that opening new restaurant in these areas increases the chances of tourists visiting the restaurant.

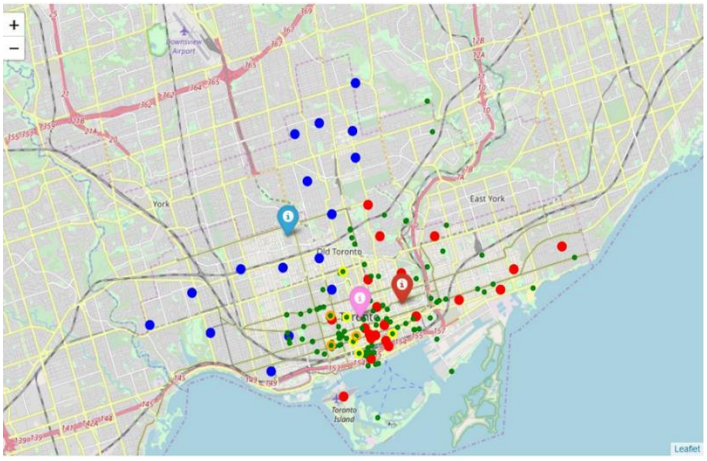
We first scrap the Wikipedia page[4] and obtain information about the Postal codes, boroughs and neighborhoods in Toronto. We plot this codes on the Toronto map (right figure) where blue points depict geographical coordinates of the postal codes on the Toronto map.



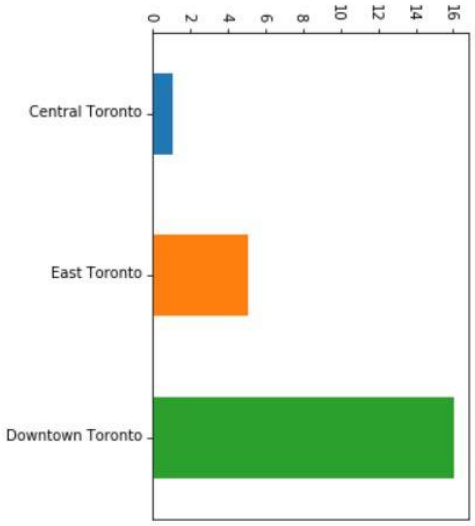
Then we cluster the postal codes into 2 clusters via KMeans clustering (unsupervised machine learning algorithm) and color the postal codes on the Toronto map to indicate the cluster they belong to. In the right figure, red and blue points indicate postal codes belonging to Cluster 0 and Cluster 1, respectively. Also, pink, red and blue pinpoint marker represent the Toronto center, Cluster 0 and Cluster 1 center coordinates, respectively.



Finally, we extract the top recommended venues based on number of 'likes'. For this task Foursquare API[3] was utilized to obtain information about these venues. In the right figure, the green circles depict top recommended venues and yellow hollow rings show top 5 recommended venues. Whereas, orange hollow rings show the top 3 restaurants in Toronto. Interesting observation here is that top recommended venues are largely grouped in Cluster 0 (red circle markers) and few are in Cluster 1 (blue circle markers)



We also create bar plot on the right that depicts the distribution of top recommended venues in the boroughs that reside in Cluster 0 (red circle markers). To summarize the information obtained in this section; Postal codes in Cluster 0 (red circle markers) contain the most top recommended venues and restaurants in Toronto. Further in Cluster 0, the Downtown Toronto Borough has the highest density of top recommended venues for the visiting tourists. Hence, it can be concluded that most influx of tourists is in Cluster 0 and more specifically in the Downtown Toronto surroundings. Thus, it seems like it would be advisable to open restaurant that is close to or in Downtown Toronto.

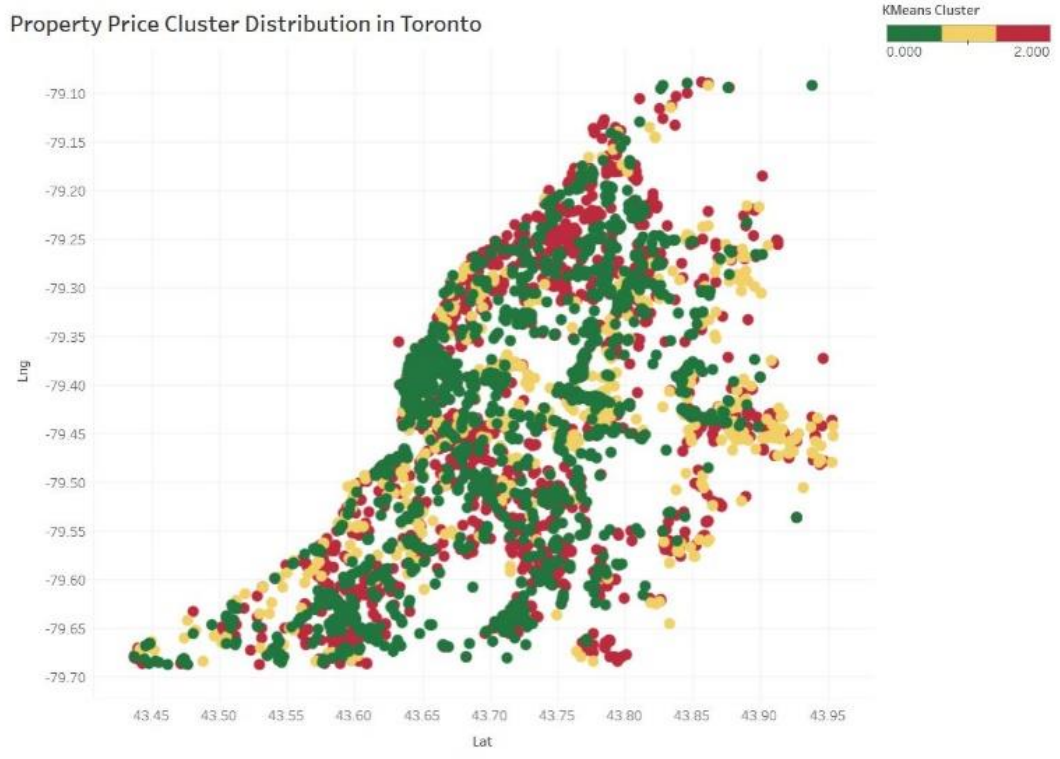


Section – 2: Property Prices in Toronto

It will beneficial to cluster properties in Toronto based on their prices[2] as clustering properties based on their prices can segregate low, medium and high priced properties.

This segregation will provide information about the cost that is to be expected, in the preferable areas, for buying a property with regards to opening a new restaurant. Hence using KMeans algorithm we cluster the property location and prices into 3 groups: Low, Medium and High prices. Note that this is 3D clustering and the dimensions of each point are Latitude, Longitude, Price. The plot below depicts the distribution of the properties on the geographical coordinates along with color coded cluster grouping. The table here depicts the cluster center price values for each of the cluster groups.

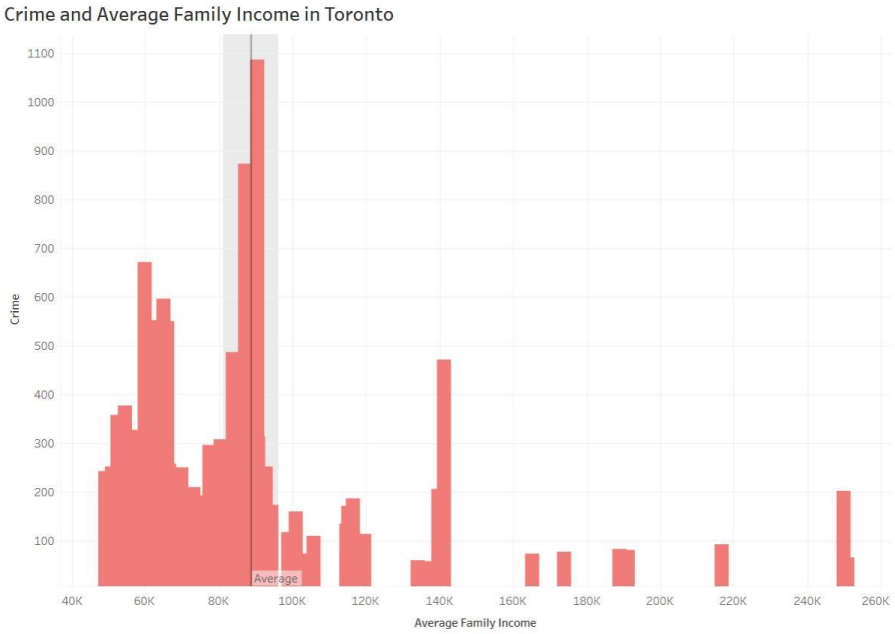
Cluster	Cluster group #	Cluster center value
Low price	0	~ $ 348342
Medium price	2	~ $ 721802
High price	1	~ $ 1309962



In summary, properties of low to high prices are distributed throughout Toronto and are not localized to particular areas. This is good, because this gives us more options regarding location. However, there are other factors, which we discuss in next sections, that must also be considered to narrow down the location options.

Section – 3: Crime and Average Family Income in Toronto

In this section we explore crime and average family income to select out preferable areas in Toronto for opening a new restaurant.[1] Low crime areas are essential because they provide Safe and relaxing environment for the customer and hence customer is more likely to enjoy the experience offered by the restaurant. On the other hand, areas with high 'Average Family Income' are important because residents of these areas are- more likely to have access to leisure income to spend on eating out at restaurants.



Thus, opening a new restaurant in the areas with low crime incidences and high average family income is a very safe bet! Therefore, we plot here the distribution of crime incidences along with average family income for all the areas in Toronto on the right. This plot suggests that areas with 'Average Family Income' > 95% Confidence Interval (average) = $ 96309, shown in gray color, are best fit for opening restaurant as these areas are not only high income areas (more leisure income to spend on restaurants) but also have less crime incidences. Note that Crime here is combined metric consisting of - Murders, Thefts, Vehicle Thefts, Robberies, Break & Enters, Sexual Assaults, Assaults, Drug Arrests cases. [1]

Hence, we extract areas with 'Average Family Income' > $96309. The table on the right depicts the preferred locations that have low crime incidences and high family income.

	Neighbourhood	Neighbourhood Id	With Bachelor Degree or Higher	Average Family Income	Healthy Food Index	Pop 25-45 years	Crime	Lat	Lng
2	Edenbridge-Humber Valley	9.0	5190.0	119581.0	46.41	3550.0	114.0	43.671478	-79.516731
21	Banbury-Don Mills	42.0	10850.0	115239.0	21.31	6410.0	171.0	43.734804	-79.357243
40	The Beaches	63.0	9235.0	139757.0	31.53	5870.0	207.0	43.671024	-79.296712
59	High Park-Swansea	87.0	10665.0	116465.0	46.53	8085.0	187.0	43.644940	-79.478313
61	Runnymede-Bloor West Village	89.0	4165.0	114798.0	45.64	2940.0	82.0	43.651778	-79.475923
65	Annex	95.0	16590.0	141111.0	38.85	11485.0	471.0	43.670338	-79.407117
67	Yonge-St.Clair	97.0	6775.0	173751.0	32.65	4320.0	78.0	43.688210	-79.394004
69	Yonge-Eglinton	100.0	6255.0	140907.0	31.95	4225.0	144.0	43.706748	-79.398327
71	Lawrence Park South	103.0	7880.0	216754.0	37.53	3390.0	93.0	43.729199	-79.403252
74	Humewood-Cedarvale	106.0	6060.0	105770.0	32.85	5045.0	110.0	43.690248	-79.422097
88	Centennial Scarborough	133.0	4055.0	102278.0	9.35	2935.0	74.0	43.787491	-79.150768
89	Highland Creek	134.0	3055.0	98857.0	0.63	2820.0	118.0	43.790117	-79.173334
95	Rosedale-Moore Park	98.0	11425.0	249884.0	32.20	4895.0	203.0	43.690388	-79.383297
96	Forest Hill South	101.0	5575.0	251035.0	18.49	2745.0	66.0	43.693559	-79.413902

To approximate property prices expected in these preferred areas, it would be useful to approximately assign property pricing cluster, obtained in Section – 2, to each of these preferred areas. We accomplish this task by finding the nearest property from the property pricing data utilized in Section – 2 to each of these preferred areas and then assign the pricing cluster of the nearest property to the respective area. The Toronto map on the right depicts these preferred areas as pinpoint markers, along with the color coding that indicates the pricing cluster these preferred areas belong too.

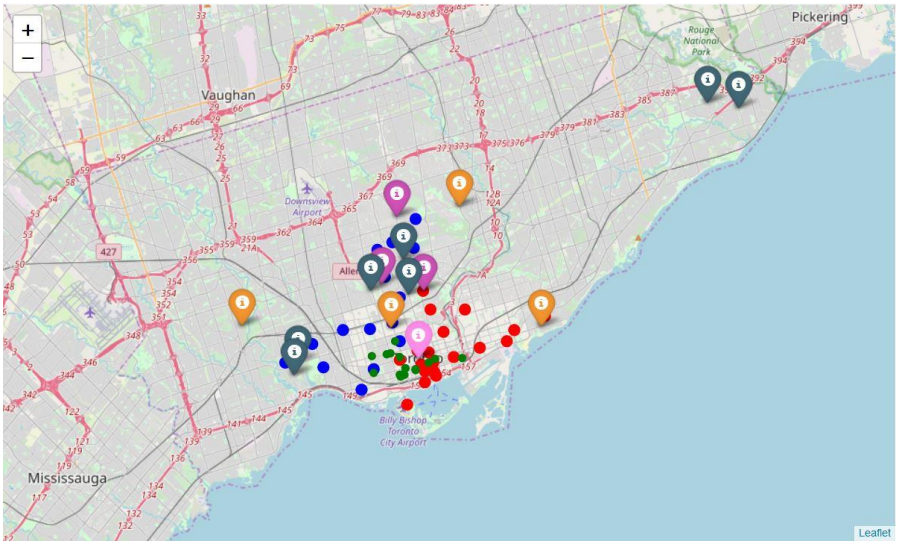
4 areas in Cluster 0 (Low price cluster)
3 areas in Cluster 1 (High price cluster)
7 areas in Cluster 2 (Medium price cluster)

- Low price cluster (~$350K)
- Medium price cluster (~$700K)
- High price cluster (~$1.3M)

In summary, we narrowed down the options to 14 preferred areas in Toronto for opening a new restaurant. All these areas have low crime incidences, high income families.

Section – 4: Comparison with Competing Restaurants in Toronto

In this section we will explore top recommended restaurants in Toronto based on the number of 'likes' given by the customers. These existing recommended restaurants will act as competition and thus their exploration will provide information about the cuisine to offer in the new restaurant and locations to avoid that will be too competitive due to presence of these established restaurants.



Using Foursquare API[3] we extract the information and popularity (in terms of number of 'likes') about the recommended restaurants in Toronto. The green circle markers on the Toronto map on the right depicts these top recommended restaurants. Interestingly these restaurants are largely spread around red circle markers (which depicts the surroundings of Downtown Toronto)

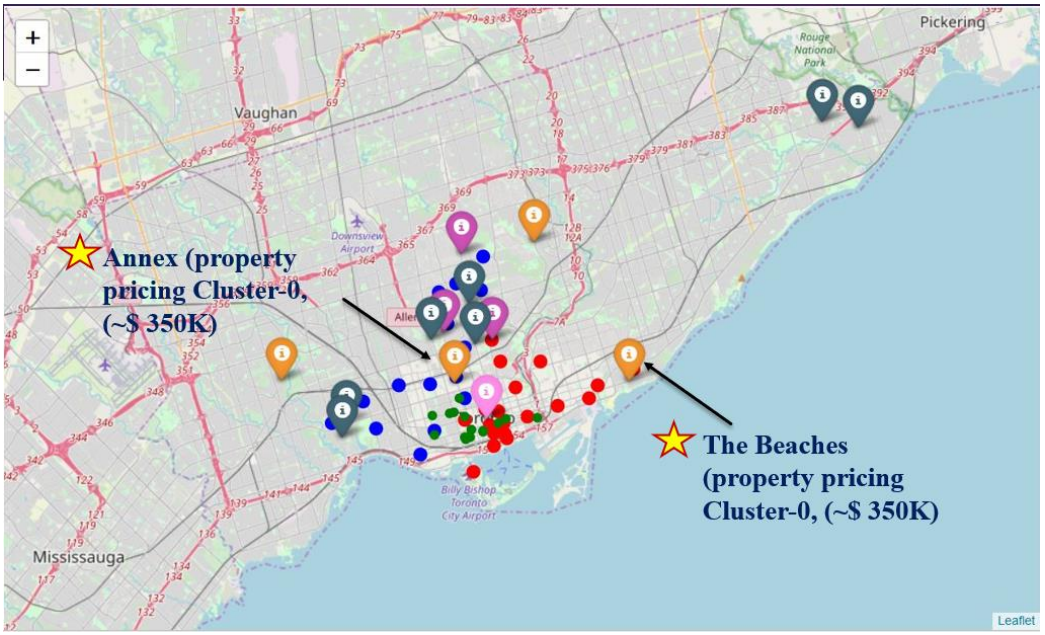
Next, we look at some statistics for these restaurants which is shown in the table on the right. Here, 'count' is the number of restaurants of that category. For e.g. there are 2 top recommended Italian restaurants. 'mean', 'min' and 'max' columns indicates the mean, min and max number of likes, respectively, for the restaurants in each category. The main point to note here is that the Thai and Italian restaurants are the favorites! With max 'likes' of 668 and 407, respectively.

categories	count	mean	min	max
Restaurant	3.0	64.0	40.0	78.0
Italian Restaurant	2.0	290.5	174.0	407.0
Mediterranean Restaurant	2.0	145.5	49.0	242.0
Mexican Restaurant	2.0	295.0	292.0	298.0
Tapas Restaurant	2.0	277.0	261.0	293.0
Thai Restaurant	2.0	369.5	71.0	668.0
American Restaurant	1.0	249.0	249.0	249.0
French Restaurant	1.0	74.0	74.0	74.0
New American Restaurant	1.0	179.0	179.0	179.0
Ramen Restaurant	1.0	176.0	176.0	176.0
Seafood Restaurant	1.0	177.0	177.0	177.0
Theme Restaurant	1.0	27.0	27.0	27.0

In summary, lot of different cuisines are offered in Toronto, out of which Thai and Italian restaurants are highly liked by the customers. Additionally, multiple recommended restaurant options for certain cuisines are available in Toronto, as indicated from 'count' column in the table. Overall, this exploration of the competing restaurants indicates that there is a tough competition in the restaurant business in Toronto and it would be advisable to offer cuisine which competing restaurants don't offer.

Discussion & Conclusion

Based on the analysis performed in the methodology & results section, the two recommended location for opening a new restaurant are depicted on the Toronto map.



Some noteworthy reasons for these final recommended locations are listed below,

1. Both locations belong to the low property price cluster, hence low initial investment in the property for opening the restaurant. Furthermore, as indicated in Section – 2, lots of property choices in each property price cluster, throughout Toronto, thus multiple options ~$ 350K are likely available at both locations.
2. Both areas have low crime incidences.
3. Both areas have high family income which suggest that people have more leisure income to spend on eating out at the restaurants.
4. Both areas reside close to the red circle markers that has the top recommended venues in Toronto (Section - 1). Hence, plenty of tourist customer influx is also expected.
5. An unexpected insight – competing restaurants are tightly grouped (green circle markers) but recommended areas are further away from these restaurants. Therefore, less competition since the customer wouldn't have multiple choices for restaurants in these recommended areas.

Lastly, we would like to provide some extra suggestions with regards to opening a new restaurant at these locations.

1. New restaurant could offer cuisine not offered by competing restaurants such as Indian or Chinese Cuisine.
2. 'The Beaches' is by the ocean is a beneficial location due to the view of the ocean. Moreover, Seafood cuisine restaurant is a good option to open at this area.
3. French or Theme restaurants are also viable in both recommended areas since there is only 1 restaurant for each and either aren't very popular.
4. It is advised not to open Italian or Thai restaurants because popular Thai and Italian restaurants with high amount of likes already exists in Toronto. Not only that but multiple recommended restaurants for these cuisines are present in the Toronto area.

References

1. City of Toronto, "Neighbourhood Profiles," *City of Toronto*, 05-Dec-2018. [Online]. Available: https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/.

2. M. Nabaee, "House Sales in Ontario," *Kaggle*, 07-Oct-2016. [Online]. Available: https://www.kaggle.com/mnabaee/ontarioproperties.

3. *Foursquare Developer*. [Online]. Available: https://developer.foursquare.com/places.

4. "List of postal codes of Canada: M," *Wikipedia*, 15-Nov-2019. [Online]. Available: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M