CAPSTONE PROJECT

TOPIC: Opening a hotel in Toronto, Canada

By Jayanti Bhattacharya

Date – 5th January 2021

Introduction:

Toronto is one of Canada's leading tourism destinations. In 2017, the Toronto-area received 43.7 million tourists, of which 10.4 million were domestic visitors and 2.97 million were from the United States, spending a total of $8.84 billion. Toronto has an array of tourist attractions and a rich cultural life.

Hotels are definitely one of the fastest-growing sectors in the tourism sector and it is truly justified as accommodation is the key part in the development of any country or region's tourism. Tourism and Hotel Industry always go hand in hand and the presence of enough hotels also adds value and quite a lot of factors and punches it within the region's economy. The Existence of a Hotel isn't enough to single-handedly boost a region's tourism but they also give out a symptom of health tourism.

Hence it is safe to say hotels are very important to bloom tourism in a country and hence a very profitable business. Toronto is a megacity in Canada where opening a hotel is highly profitable and useful. Opening a hotel is a source of consistent income and determining the location of it is a business problem which is very important to determine success and failure of the business of the hotel.

Business Problem:

The objective of this project is to analyse and select the best locations to build a hotel in Toronto, Canada. Using data methodology techniques, data visualisation techniques like Folium and machine learning techniques like clustering, this project is aimed at answering the following business problem:

If a property developer is planning to build a hotel in Toronto where is a great location to build it and where are the recommended locations?

Target Audience:

The target audience are property developers who want to build a hotel in Toronto.
This project is aimed at analysing locations which can be great locations for creating hotels avoiding extreme competition and oversupply and hence leading to boost the tourism in Canada.

Data:

For this project we require the following:

- *Lists of neighbourhoods in Toronto, to determine the specific places we require*
- *Coordinates of the neighbourhood, i.e., Latitudes and Longitudes of the neighbourhood to specify the position.*
- *The data of venues in the neighbourhoods to later do clustering and find out the required positions.*

Sources for retrieving and gathering this data:

- *For finding all the neighbourhoods in Toronto, the Wikipedia link https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M was used to derive all the required information with help of python library pandas and put into a database to use it later.*
- *For finding the coordinates, i.e., latitude and longitude for all neighbourhoods the excel spreadsheet https://cocl.us/Geospatial_data and then combined them and put them in a csv file named Toronto.csv.*

- *Used Foursquare API, which provides location based experiences with diverse information about venues, users, photos, and check-ins and supports real time access to places, Snap-to-Place that assigns users to specific locations, and Geo-tag and used by 125,000 developers are building location-aware experiences with Foursquare technology and data*, to get data on the venue of the place.

Methodology:

First, we get the information about neighbourhoods by web scraping and the coordinates by python geocoder package and put into a dataframe. Now, explore all the neighbourhoods in Toronto. Now let's visualize all the neighbourhoods using Folium.

Now, after defining Foursquare API credentials, get tNow, let's get the top 100 venues that are in Toronto within a radius of 2000 meters using the GET request. Now, group data by neighbourhood and taking mean of frequency of each venue. Next, check if 'Hotel' category is present in the dataframe or not. If yes, filter the venue data with the category 'Hotel'.

Let's see number of hotels present

```
[15]: len(t_grouped[t_grouped["Hotel"] > 0])
```

```
[15]: 47
```

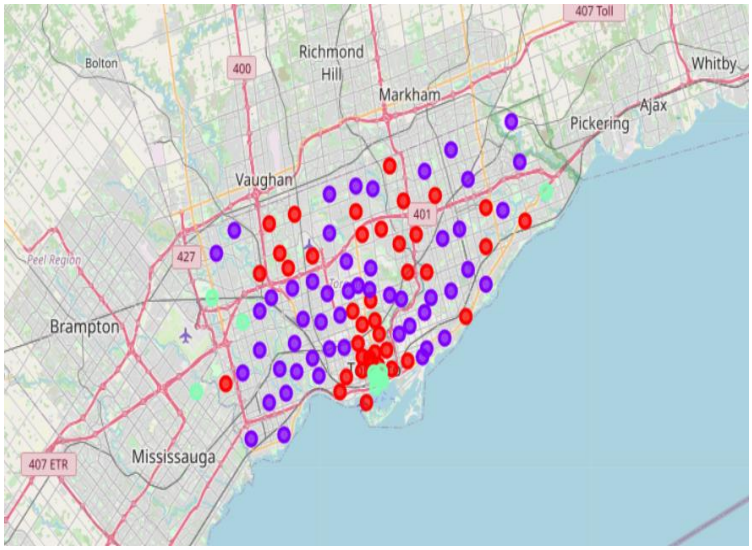
Let's see the number of hotels in all Neighbourhood

```
[16]: t_sh = t_grouped[["Neighbourhoods","Hotel"]]
      t_sh
```

[16]:

	Neighbourhoods	Hotel
0	Agincourt	0.000000
1	Alderwood, Long Branch	0.000000
2	Bathurst Manor, Wilson Heights, Downsview North	0.000000
3	Bayview Village	0.000000
4	Bedford Park, Lawrence Manor East	0.000000

Now, perform clustering on the data by the machine learning technique of K-clustering and cluster them into 3 clusters represented by 3 distinct colours. Then visualize the data using folium.



Finally, examine the clusters one by one to get the desired result and output and examine the clusters and derive results from it better.

Examine Clusters

CLUSTER 1

In [68]:
```
t_merged.loc[t_merged['Cluster Labels'] == 0]
```

Out[68]:

	Neighbourhood	Hotel	Cluster Labels	Latitude	Longitude
68	Rouge Hill, Port Union, Highland Creek	0.025641	0	43.784535	-79.160497
15	Church and Wellesley	0.020000	0	43.665860	-79.383160
78	Summerhill West, Rathnelly, South Hill, Forest...	0.030000	0	43.686412	-79.400049
13	Central Bay Street	0.030000	0	43.657952	-79.387383
31	Garden District, Ryerson	0.030000	0	43.657162	-79.378937
39	Humberlea, Emery	0.021277	0	43.724766	-79.532242
88	West Deane Park, Princess Gardens, Martin Grov...	0.020833	0	43.650943	-79.554724
9	CN Tower, King and Spadina, Railway Lands, Har...	0.030000	0	43.628947	-79.394420
64	Regent Park, Harbourfront	0.020000	0	43.654260	-79.360636
93	Willowdale, Willowdale East	0.024390	0	43.770120	-79.408493
44	Kensington Market, Chinatown, Grange Park	0.020000	0	43.653206	-79.400049
45	Kingsview Village, St. Phillips, Martin Grove ...	0.044776	0	43.688905	-79.554724
27	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.031250	0	43.643515	-79.577201
74	St. James Town, Cabbagetown	0.020000	0	43.667967	-79.367675
34	Guildwood, Morningside, West Hill	0.027778	0	43.763573	-79.188711

CLUSTER 2

```
In [69]: t_merged.loc[t_merged['Cluster Labels'] == 1]
```

Out[69]:

	Neighbourhood	Hotel	Cluster Labels	Latitude	Longitude
63	Queen's Park, Ontario Provincial Government	0.010000	1	43.662301	-79.389494
61	Parkview Hill, Woodbine Gardens	0.000000	1	43.706397	-79.309937
66	Rosedale	0.010000	1	43.679563	-79.377529
62	Parkwoods	0.010989	1	43.753259	-79.329656
0	Agincourt	0.000000	1	43.794200	-79.262029
59	Old Mill South, King's Mill Park, Sunnylea, Hu...	0.000000	1	43.636258	-79.498509
67	Roselawn	0.000000	1	43.711695	-79.416936
56	North Toronto West, Lawrence Park	0.000000	1	43.715383	-79.405678
55	North Park, Maple Leaf Park, Upwood Park	0.000000	1	43.713756	-79.490074
54	New Toronto, Mimico South, Humber Bay Shores	0.000000	1	43.605647	-79.501321
53	Moore Park, Summerhill East	0.010000	1	43.689574	-79.383160
52	Mimico NW, The Queensway West, South of Bloor,...	0.000000	1	43.628841	-79.520999
51	Milliken, Agincourt North, Steeles East, L'Amo...	0.000000	1	43.815252	-79.284577
60	Parkdale, Roncesvalles	0.000000	1	43.648960	-79.456325
58	Northwood Park, York University	0.012048	1	43.767980	-79.487262
71	Scarborough Village	0.013514	1	43.744734	-79.239476
70	Runnymede, The Junction North	0.000000	1	43.673185	-79.487262
96	Woodbine Heights	0.000000	1	43.695344	-79.318389
95	Woburn	0.000000	1	43.770992	-79.216917

CLUSTER 3

```
In [70]: t_merged.loc[t_merged['Cluster Labels'] == 2]
```

Out[70]:

	Neighbourhood	Hotel	Cluster Labels	Latitude	Longitude
84	Toronto Dominion Centre, Design Exchange	0.060000	2	43.647177	-79.381576
18	Commerce Court, Victoria Hotel	0.060000	2	43.648198	-79.379817
11	Canada Post Gateway Processing Centre	0.080000	2	43.636966	-79.615819
29	First Canadian Place, Underground city	0.060000	2	43.648429	-79.382280
57	Northwest, West Humber - Clairville	0.105263	2	43.706748	-79.594054
65	Richmond, Adelaide, King	0.060000	2	43.650571	-79.384568
5	Berczy Park	0.060000	2	43.644771	-79.373306
76	Stn A PO Boxes	0.060000	2	43.646435	-79.374846
73	St. James Town	0.060000	2	43.651494	-79.375418
35	Harbourfront East, Union Station, Toronto Islands	0.050000	2	43.640816	-79.381752

Result:

From all the above methodology used in this project, the neighbourhoods of Toronto are categorised into 3 parts, namely, Cluster 1, Cluster 2 , Cluster 3. Cluster 1 as seen from the examination of the 3 clusters has a moderate number of hotels and hence moderate level of competition among hotels in those neighbourhoods. Cluster 2, on the other hand, has low number of hotels in that cluster corresponding to low level of competition among hotels in those neighbourhoods. Last but not the least, Cluster 3, has a large number of hotels in that cluster corresponding to high competition among hotels in that neighbourhood.

Discussion:

It is clear from the conclusion that Most of the hotels are concentrated in the center part of the city which is cluster 3. Highest number of hotels are in cluster 3. Lowest number of hotels are in cluster 2 and moderate number in cluster 1. Hence, there is high competition rate in Cluster 3, while there are less in Cluster 1 and minimum in Cluster 2. It seems that Cluster 3 has good competition among hotels and also slightly overcrowded and Cluster 1 has some competition and Cluster 2 with very less competition.

Recommendations:

This project tells us that It is better to avoid high competition places like Cluster 3 and it is preferrable to not build a hotel there to maximize profits as that cluster has oversupply, high competition and hence not a very good option. Cluster 1 or 2, on the other hand can be a suitable option for building a hotel. Cluster 1 with moderate competition which might be a profitable area if the place stands out from the others. Cluster 2 is a safe position will almost no competition and a highly profitable area to maximize profits.

Conclusion:

Hence, the answer to our business problem, 'if a property developer is planning to build a hotel in Toronto where is a great location to build it and where are the recommended locations?', is that neighbourhoods in Cluster 2 are most preferred to open a hotel in followed by Cluster 1 for maximum profits and less competition. Cluster 3 is overcrowded and has high competition and thus should not be preferred over the others due to reasons like oversupply and high competition.

The findings of this project might help project developers and future hotel owners to maximise their profit by selecting a better location.

THAT'S THE END OF THIS REPORT

Thank you