The Battle of Neighborhoods (Capstone Project)

Introduction / Business Problem

The city of Toronto is the largest city in Canada by population. This capital city of the province of Ontario is also one of Canada's leading tourism destinations. With its booming tourism industry, the city of Toronto receives a large volume of tourists and visitors every year. In the year 2017 alone, the Toronto area has recorded a whopping total of 43.7 million tourists.

This influx of visitors has propelled the burgeoning demand on hotels and hostels in the region. With the growth in travel accommodations, laundry services would also be increasingly in demand. Investors interested in a business with low investment and high return can look to the business of laundromats in a city with a high tourist volume, such as Toronto. Tourists would often favour a cheaper alternative than hotel laundry services, which explains the mushrooming of laundromats in tourist-rich cities.

Therefore, the problem statement is:

What is the best/most optimal place to set up a laundromat in the city of Toronto that is in close proximity to travel accommodations and preferably at a distance from other competing laundromats?

Data

The data necessary to answer the problem statement includes the locations and the types of venues around the Toronto city area.

First, the locations of all existing laundromats or laundry services in the Toronto city region will be extracted to survey the competition. Ideally, the new laundromat would be set up at a certain distance away from other existing laundromats.

Secondly, the locations of travel accommodations such as hotels, hostels and motels in the region will be obtained. This is to estimate the number of visitors that would frequent the laundromat, based on the assumption that the clientele of the laundromat would be mainly made up of travellers.

A suitable location for the new laundromat would be at the centre of a cluster of hotels and hostels so as to maximise the business potential. A clustering algorithm would be performed on the locations of laundromats and hotels separately to find the concentration of both types of the establishments.

Finally, the ideal location for setting up a new laundromat can be inferred based on the clustering result.

Methodology

In order to search for locations of the existing laundromats and hotels in the city of Toronto, it is necessary to find the coordinates of the neighbourhoods in Toronto for querying the Foursquare API. Hence, the following steps were taken.

1. The Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M_) was scraped using the BeautifulSoup library to build a pandas dataframe listing the boroughs and neighborhoods in Toronto. The dataframe was cleaned and processed appropriately.
2. Next, the geographical coordinates of each neighborhood was obtained from the csv downloaded from http://cocl.us/Geospatial_data. The

3. The postal codes for each of the neighborhoods were matched to the information from the csv file and the original pandas dataframe was updated to list the corresponding geographical coordinate for each neighborhood.
4. To limit the range of the data, only the Central Toronto borough was investigated as this is the centre of Toronto where the tourists are centered. The neighbourhoods from the borough of Central Toronto were extracted and listed in a dataframe.
5. The Foursquare API was then used to query each of the neighbourhood for the keywords {laundromats, laundry} and {hotel, hostel, motel}. The resulting locations for each venue were tabulated and stored in a new dataframe and visualised using Folium.
6. In order to see the aggregation of the laundromats and hotels, hierarchical clustering was used to cluster the laundromats and hotels separately. Hierarchical clustering was used because there was no need for a cluster number to be selected and can be purely based on the distance between the venues. The resulting clusters were plotted on a Folium map.
7. A better visualisation was performed by obtaining the cluster centres for each group of laundromat and hotel. The cluster centres were plotted on a Folium map, with the circle marker weighted corresponding to the number of venues in the specific cluster.

Results

The results are listed as follows.

1. The neighbourhoods falling under Central Toronto amount to 9 neighbourhoods in total. The laundromats and hotels around these neighbourhoods were explored.
2. A total of 106 laundromats and 87 hotels were found after performing the query on Foursquare API around these neighbourhoods.
3. Hierarchical clustering on the laundromats and hotels have produced a total of 5 clusters of laundromats and 4 clusters of hotels, as seen below (red: laundromats, blue: hotels).
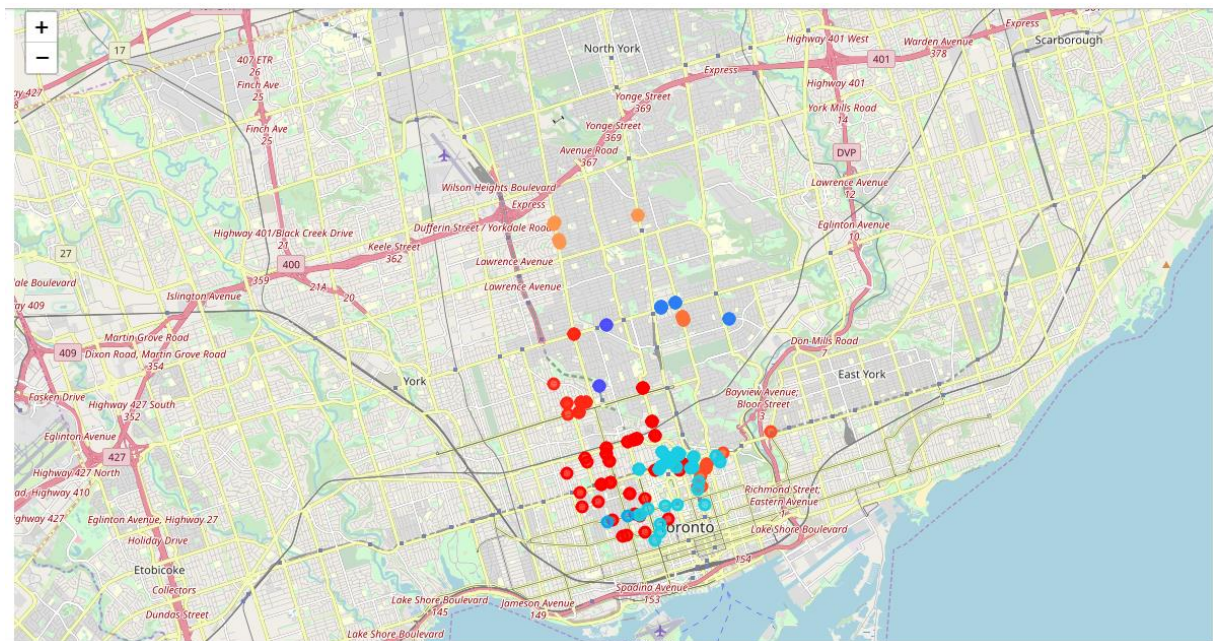
Discussions

1. It can be observed that the laundromat and hotel clusters are generally overlapping or close to each other. This observation makes sense as the laundromats depend on the tourist population for supporting its business.

2. A large concentration of laundromats and hotels alike can be found surrounding The Annex neighbourhood. This can be seen in the large clusters shown in the Folium map.

3. Conversely, there is a relatively smaller hotel concentration at the Forest Hill neighbourhood with only a very sparse number of laundromats serving it. The closest cluster of laundromats is located further away in the neighbourhood of Humewood.

4. There is a small cluster of laundromats in the neighbourhood of Caribou park which does not have any neighbouring hotel clusters. Upon closer inspection, this is probably due to the fact that there is are boarding schools and hospitals in that region, which serve as the laundromats' clientele.
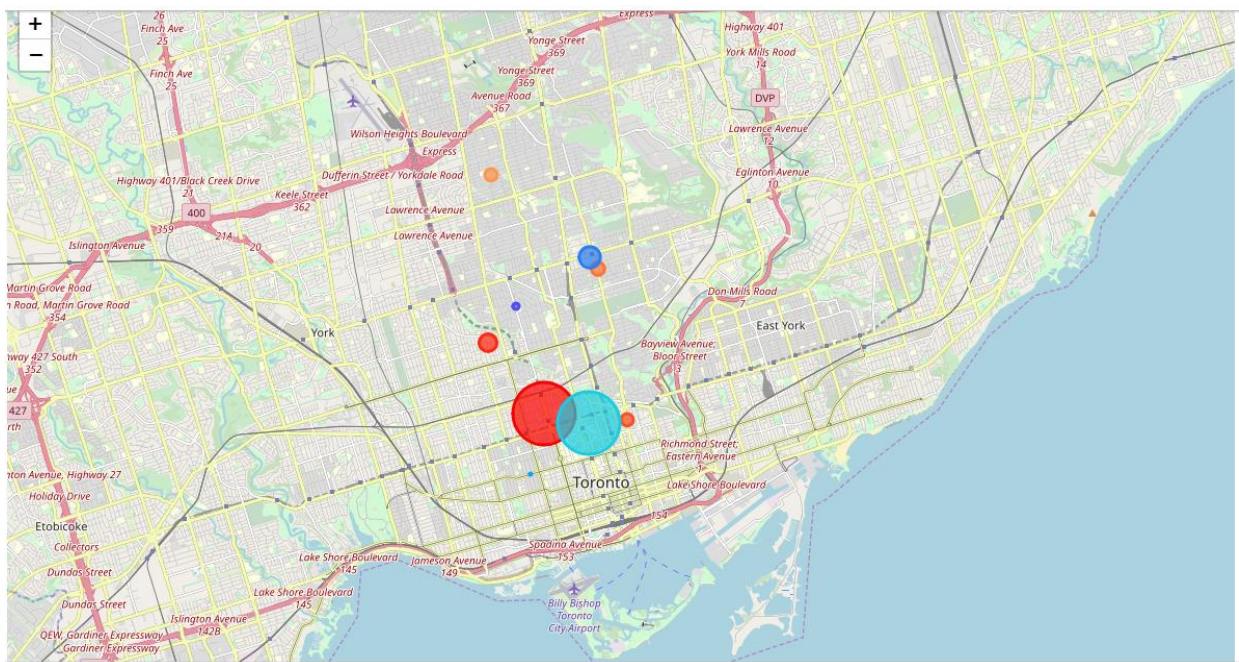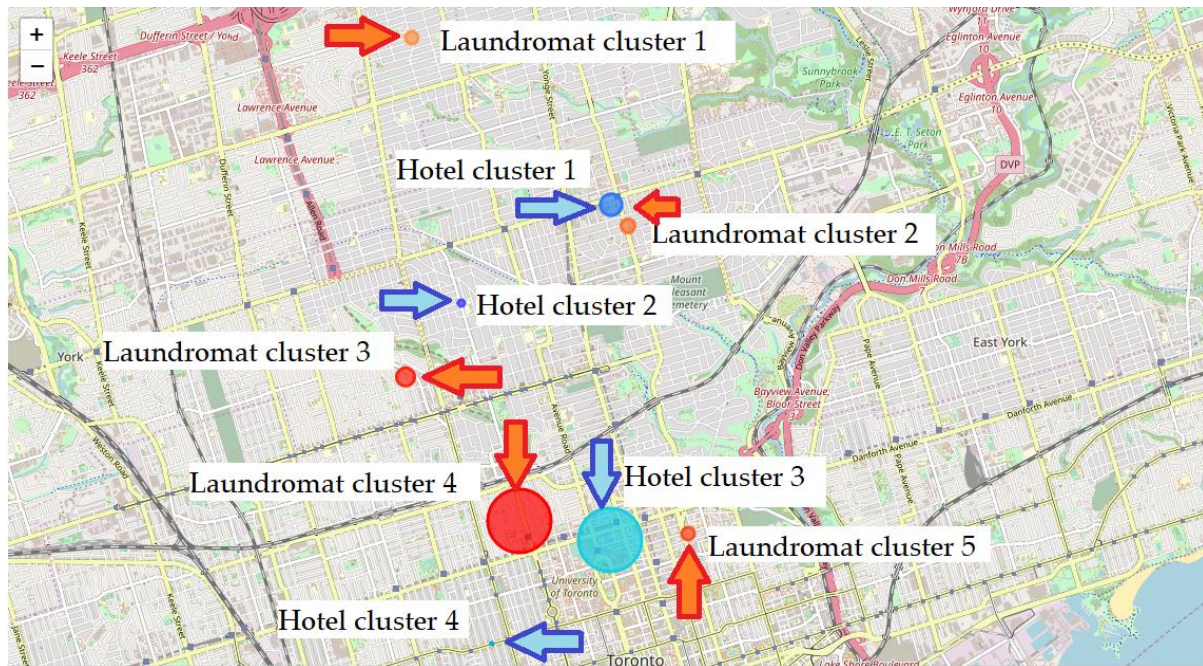
Conclusion

To answer the initial question of 'What is the best/most optimal place to set up a laundromat in the city of Toronto that is in close proximity to travel accommodations and preferably at a distance from other competing laundromats?', the most suitable place for setting up a new laundromat is in the neighbourhood of Forest Hill in Central Toronto. This area can be shown in the above screenshot as the region of Hotel cluster 2. This is due to the lack of laundromats in its close proximity while still having a decent number of hotels.