

Coursera capstone: the report

Introduction

Toronto is one of the most important cities in Canada. This is true in terms of economy, population density and tourism. This popularity has resulted in the establishments of multiple restaurants and hotels.

Hotels generally charge their customers insane amounts of money for laundry. This inconveniences and discourages customers from wanting to use hotel laundry services. People then prefer to look around the area to find Laundromats and pay a lot less for laundry services.

Hence, this project aims to find the best Laundromats nearest to hotels and farther away from other 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 amount 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 center of a cluster of hotels and hostels so as to maximize 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 neighborhoods 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.

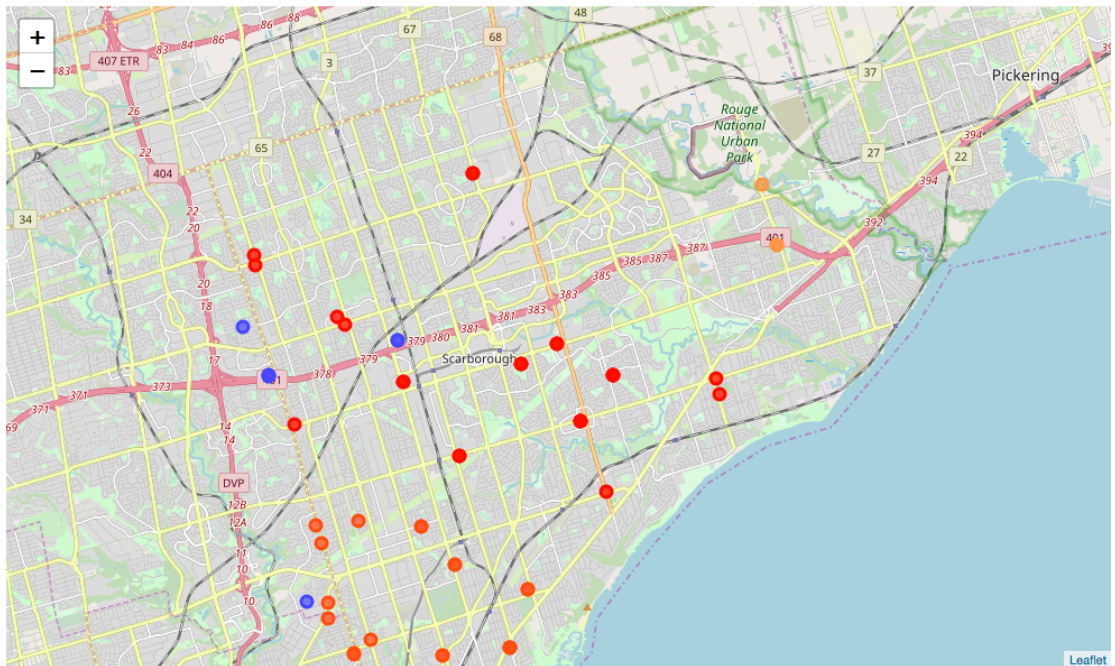
2. Next, the geographical coordinates of each neighborhood was obtained from the csv downloaded from http://cocl.us/Geospatial_data. The
3. The postal codes for each of the neighborhoods were matched to the information from the csv file and the original pandas dataframe was updated to list the corresponding geographical coordinate for each neighborhood.
4. To limit the range of the data, only the Scarborough borough was investigated as this is the center of Toronto where the tourists are centered. The neighborhoods from the borough of Scarborough were extracted and listed in a dataframe.
5. The Foursquare API was then used to query each of the neighborhood for the keywords {Laundromats, laundry} and {hotel, hostel, motel}. The resulting locations for each venue were tabulated and stored in a new dataframe and visualized 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 visualization was performed by obtaining the cluster centers for each group of Laundromat and hotel. The cluster centers 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.

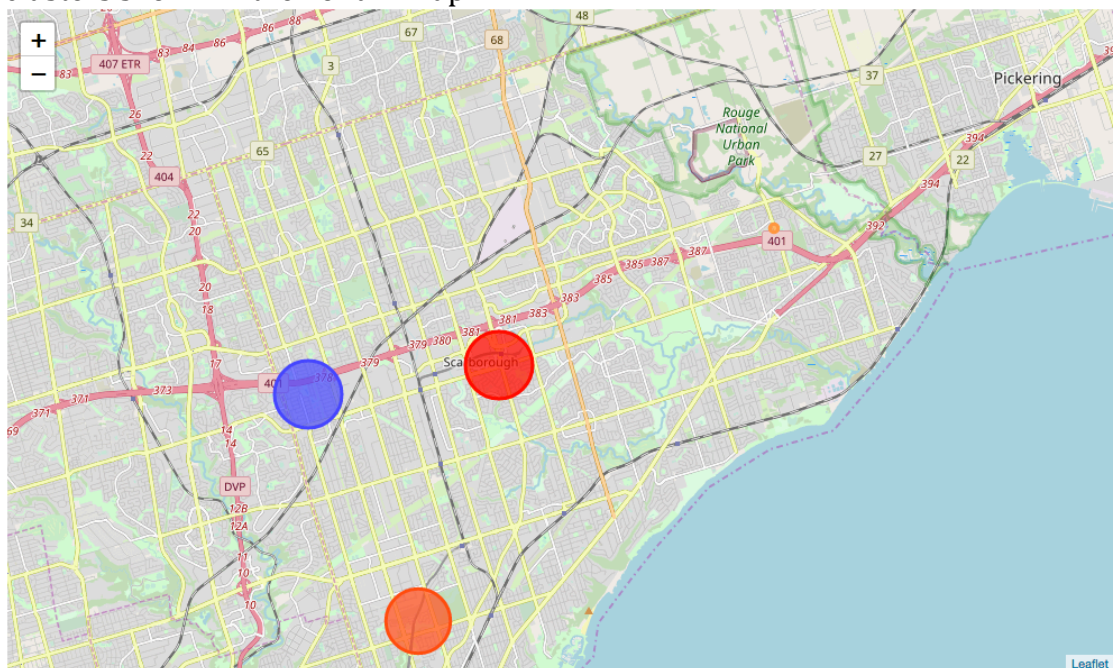
1. The neighborhoods falling under Scarborough amount to 17 neighborhoods in total. The Laundromats and hotels around these neighborhoods were explored.
2. A total of 50 Laundromats and 7 hotels were found after performing the query on Foursquare API around these neighborhoods.

3. Hierarchical clustering on the Laundromats and hotels have produced a total of 2 clusters of Laundromats and 1 clusters of 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 Ellesmere neighborhood. This can be seen in the large clusters shown in the Folium map.



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 neighborhood of Scarborough center in Scarborough. This is due to the lack of Laundromats in its close proximity while still having a decent amount of hotels.