**Find a Neighborhood for New Hotel**

Chang, 2020

1. **Introduction**

**1.1 Background**

To find a new possible place of hotel, one must consider both competition and surrounding environment. In the city of Toronto, there are 200 neighborhoods. Some of these neighborhoods already have hotels. Especially for neighborhoods in the center part of the city, where competition between hotels are drastic enough. Therefore, choosing a neighborhood with less hotels would decrease the level of competition. However, competition is not the only thing that should be considered. Imagine staying in a hotel but there are nothing around the hotel to explore or enjoy, it would be boring. In this case, the abundant categories of nearby venues are also crucial.

**1.2 Problem**

Data that might contribute to determining where the new hotel should be allocated in the city of Toronto as in neighborhood.

**1.3 Interest**

For the ones that are interested in opening a new hotel in Toronto or explore the density of hotel competitions.

**2. Data acquisition and cleaning**

**2.1 Data sources**

Data that are used in this project are mainly from Foursquare. To segment neighborhoods by borough, postcodes, information listed on Wikipedia are also used as well. As in the figure below shows, there are some information related to each neighborhood.A screenshot of a cell phone

Description automatically generated

**2.2 Data cleaning**

Data downloaded or scraped from multiple sources were combined into one table. First, I deleted the missing data from the original data source. Dropping not assigned information makes the data source clear.

Second, I retrieved data from Wikipedia. However, in the original source (Wikitable), neighborhoods are concatenated by the postal code. I explode the concatenated neighborhood to different rows to see each neighborhood separately.

Third, from website <https://cocl.us/Geospatial_data>, I joined Longitude and Latitude with cleaned information from Wikipedia based on the column postal code

**2.3 Feature selection**

After data cleaning, there were 200 more samples in 5 arrays. According to the Longitude and Latitude recorded in the data set, I retrieved top 5 hotels within neighborhoods. Most of neighborhoods reached the limit of 5 hotels, while some of the neighborhood does not contain more than 5 hotels. These neighborhoods are the possibility neighborhoods to build a new hotel, as the competition within these neighborhoods are not as intense as other neighborhoods.

A screenshot of a cell phone

Description automatically generated

Then I sort out the neighborhoods that have less than 4 hotels:

A screenshot of a cell phone

Description automatically generated

**3. Exploratory Data Analysis**

**3.1 Other information to be consider**

Knowing competitions in the neighborhood is subjective. Other criteria such as nearby interesting places should also be considered. To determine whether a neighborhood is convenient to stay, the variety of existing venues categories should also be considered.

A screenshot of a cell phone

Description automatically generated

**3.2 Cluster analysis of potential neighborhood**

I also applied cluster analysis of these potential neighborhood choices to figure out which neighborhoods are similar to each other:

A screenshot of a cell phone

Description automatically generated

There are in total 5 clusters, most of the neighborhoods are listed as cluster 1: Harbourfront, Davisville North and Regent Park.

**4. Results**

In this study, I analyzed the possible neighborhood choice for starting a new hotel. The most suitable neighborhood should be: 1) situated in the less competitive neighborhoods, 2) also supported by nearby various venues to increase the convenient of staying.

A screenshot of a cell phone

Description automatically generated

Combining with cluster analysis, neighborhoods within same cluster offers more or less similar nearby venues.

**5. Future directions**

In this project, related marketing information are provided by Foursquare. From Foursquare we know which neighborhood is less competitive in hotel industry and are more suitable for opening a new hotel due to less drastic competition. However, even though from 200 neighborhoods I minimized targets to 10 neighborhoods. Then I run Foursquare calls again to get the venue categories of that neighborhood to see if these neighborhood contains various services to offer to hotel guests, and then clustered these neighborhoods into 5 clusters to figure out which neighborhoods are more similar to others.

These methods are only providing supplementary information, further on study should also consider the star rating, room price and facilities provided by the future hotel and also, is applicable, heading out to each neighborhoods to see if the neighborhood has suitable real estates for opening a new hotel.