CAPSTONE PROJECT

THE CLASH OF NEIGHBORHOODS

1. **INTRODUCTION**
   1.1. Background

   For an individual looking for a new apartment that's close to their new place of work, it can unfortunately be a challenge with the yearly increase in house prices across the province in order to find an affordable place to call home. Yes, one can go on Craigslist or Kijiji and find a sweet deal but with hidden additional expenses in transportation. When considering a new location for settlement, the factors considered are often the proximity of other establishments to a prospective address. Let us take Brian as an example, who is a current grad with an entry-level job position, enjoys the active lifestyle, and occasional visits to restaurants. How can Brian find the best neighborhood that would satisfy all his needs and requirements with his current work situation as an entry-level? A common denominator for Brian would be minimizing the cost of transportation from one place to the other. Brian would need a good gym location, a reputable restaurant with good reviews, of which both establishments are near his location of interest. Brian's search for the best location can be executed effortlessly with the use of location data to find a neighborhood that is close to all the establishments he is going to frequent on a daily or routine basis.

   1.2. Problem

   How can we find the best neighborhood in Toronto based on a given set of criteria? This project aims to provide a recommendation on the best neighborhood that satisfies some given criteria based on the distance from one location to another and ratings (between 0 to 10) for a particular given location (venue).

   1.3. Interest
   1.3.1. **Entrepreneurs**: Individuals looking for the best neighborhood to set up their business to draw in customers from similar establishments operating in that particular neighborhood.
   1.3.2. **Parents**: A family of kids might want to locate the safest neighborhood to raise their children which perhaps has the fewer establishment of businesses, as a result, would offer a quieter lifestyle.
   1.3.3. **Investors**: Real estate investors or corporate entities looking for the most lucrative location to invest to attract a wide range of home renters or clients.

2. **DATA ACQUISITION AND CLEANING**
   2.1. Data Sources

   Two main data sources were utilized in this project. The first one was based on the neighborhood in Toronto along with their postal codes which can be found here on Wikipedia. A sample image of the neighborhoods with their respective postal codes can be found here. The second source of data set used was from Foursquare. Foursquare data were obtained by making a few API requests to obtained information like location coordinates (latitude & longitude), location IDs, category of a location (e.g. Restaurant), name of a location, and ratings of a location (on a scale of 0 to 10). A sample image of a response from a request about venues (or locations) can also be found here.

2.2. Data Cleaning

Data obtained from Wikipedia was put into a DataFrame and contained some "Not Assigned" values which were deleted from the dataset to avoid computational errors down the road. The index of the dataframe was originally based on the postal codes, but this was reset to numeric values. A total of 103 neighborhoods contained valid information out of 180. After dropping the "Not Assigned" values from the dataframe, the corresponding coordinates (latitude and longitude) for each neighborhood was obtained using the "pgeocode" library. Coordinate for one of the postal codes could not be found, and as a result, was also deleted from the dataset. A total of 102 neighborhood coordinates was found out of 103. Dataframe was updated by merging two new columns namely "Latitude" and "Longitude." This dataframe would be used to find the corresponding latitude and longitude for any neighborhood that needs to be analyzed later.

3. EXPLORATORY DATA ANALYSIS
3.1. Calculation of target variables

The target variables for the analysis of the project would be the distance from one location to another and the ratings of a venue. An origin will be considered as a location of interest by a potential user of this application and a destination would be a venue of interest e.g. a restaurant or gym. To calculate the distance between two locations, we make use of their respective coordinate values and calculate the distance using the "geopy" library. This library has a method called "great_circle (origin, destination)" which takes two parameters. The "Great-circle" uses the spherical model of the earth to calculate the distance between two coordinates using 6371.009 km as a radius. The calculations using this radius are estimated to be accurate with an error of up to 0.5%. The arguments passed into the function should be of type tuple i.e. origin = (1,4).

The rating on the other hand for a given location is obtained directly from Foursquare. Every location/venue on Foursquare has its unique id, and a request can be made on this id to obtain additional information like reviews posted, name of venue, latitude, longitude, category, and ratings.

3.2. DataFrames and Analysis
3.2.1. DataFrames

Two neighborhoods; "Christie" and "Richmond, Adelaide, King" were being compared for this project to find the best neighborhood that has a highly rated gym and restaurant within a reasonable distance. Two separate dataframes were populated for each of these two neighborhoods. By obtaining the latitude and longitude for the locations, a request to Foursquare was made with fields of interest to be "ID", "Venue_name", "Latitude" and "Longitude" of the returned venues around such neighborhood of interest. Using the method of distance calculation explained above, the distances of each venue from both neighborhoods coordinate were calculated and updated in the dataframe with a column name "Dist_from_origin(m)" (which is the distance between any two locations in meters). Using the unique id of each venue returned by the previous Foursquare request, a separate request was made for each venue, and the average rating for such venue was extracted from the JSON response and updated in the dataframe with a column name of "Ratings(0-10)." Hence, the final dataframe after data extraction contains 7.

3.2.2. Analysis

The analysis is based on the "Category" column because it contains the requirements a potential user would specify. For this analysis, two categorical values were used; "Restaurant" and "Gym." Using these two categories would give an insight into the best gym or restaurant that are around the neighborhoods of interest. The figure below is a snapshot of what a typical dataframe would look like.

Out[342]:

|	ID	Venue_name	Category	Latitude	Longitude	Dist_from_origin(m)	Ratings(0-10)
0	4ad69511f964a520e40721e3	The Keg Steakhouse + Bar - York Street	Restaurant	43.649987	-79.384103	77.626339	8.8
1	4ad4c05df964a52059f620e3	Canoe	Restaurant	43.647452	-79.381320	287.094872	8.6
2	501ae947e4b0d11883b910a7	Equinox Bay Street	Gym	43.648100	-79.379989	314.342569	8.4
3	4ae7199ff964a52068a821e3	Bymark	Restaurant	43.647217	-79.381252	312.013623	8.1
4	4dc842f2e4cd5bc076711d90	GoodLife Fitness Toronto 137 Yonge Street	Gym	43.651242	-79.378068	458.841321	8.3
5	4c1417c182a3c9b676b9fcf8	Portico	Restaurant	43.651856	-79.378055	490.879317	7.9
6	50885719498ea7b5aab3a74c	GoodLife Fitness Toronto Bell Trinity Centre	Gym	43.653436	-79.382314	433.858461	7.5
7	505b8b544c7fd4d07eae01f3	GoodLife Fitness Toronto Street Women's Only	Gym	43.650400	-79.376700	538.418244	7.7

Figure 1. Dataframe of venues around Richmond, Adelaide, King neighborhood (43.6496, -79.3833)

This dataframe was then split into two separate dataframes based on these categorical values; "Gym" and "Restaurant." Below is a map showing clusters of venues around two neighborhoods and some bar plots.
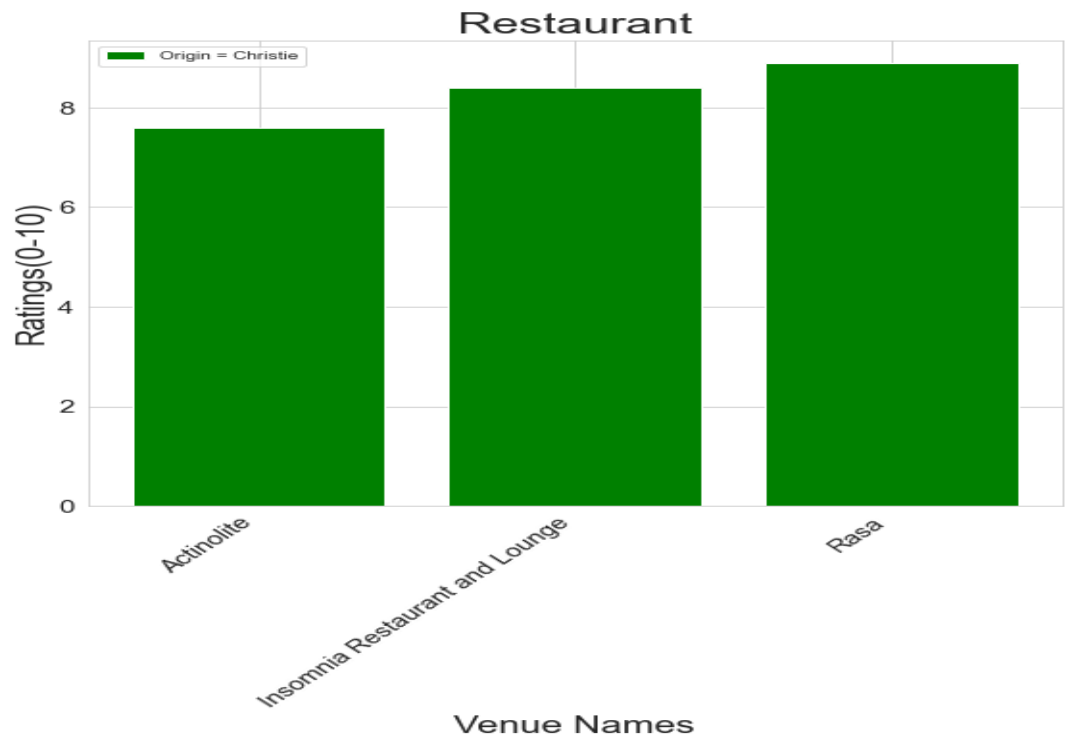
Out[357]:
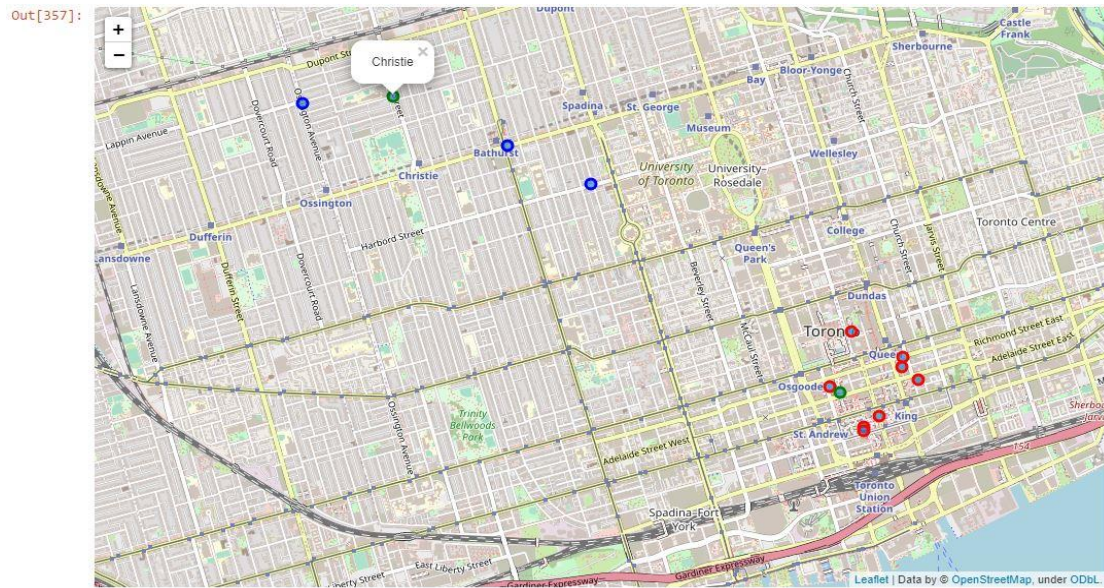
Figure 2. "Christie" at the top left corner and "Adelaide, Richmond, King" at the bottom right conner.
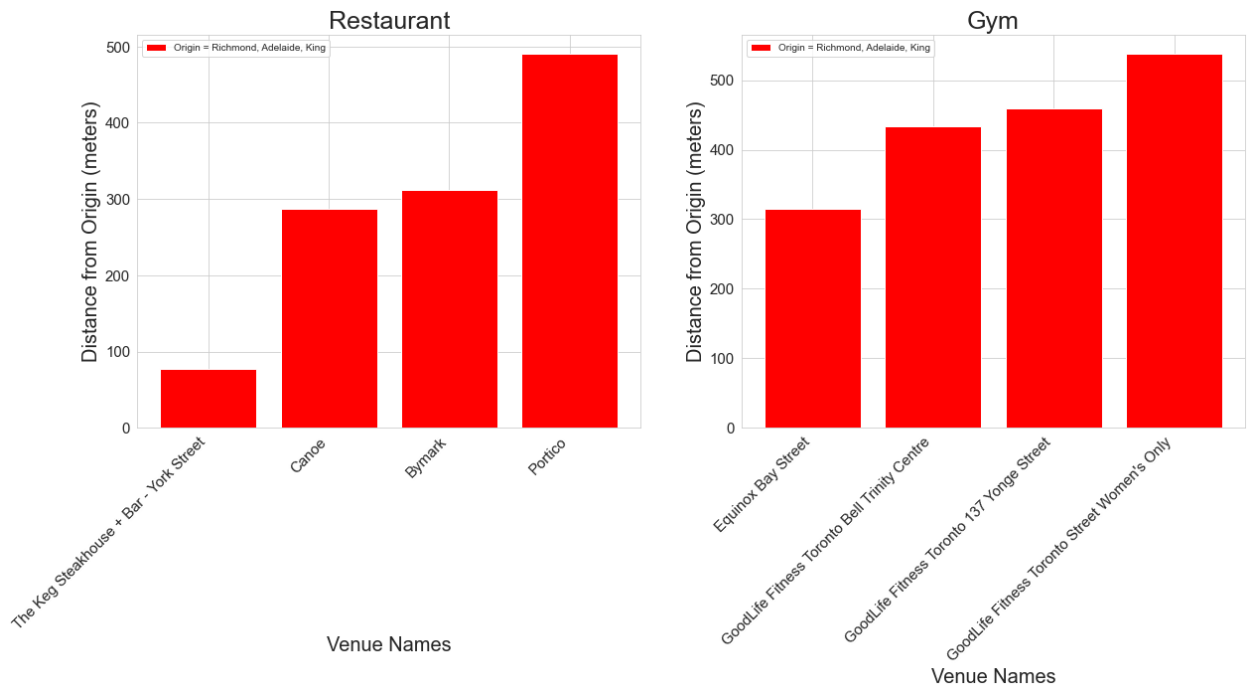
Bar Plots for Christie Neighborhood
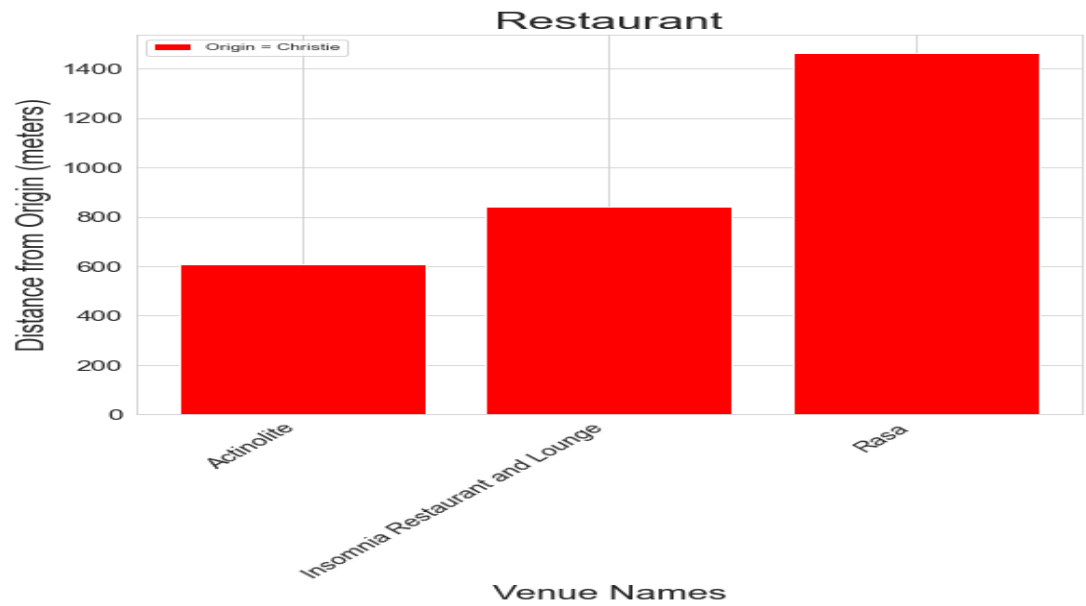
Figure 3. Actinolite with shortest distance from "Christie" coordinate
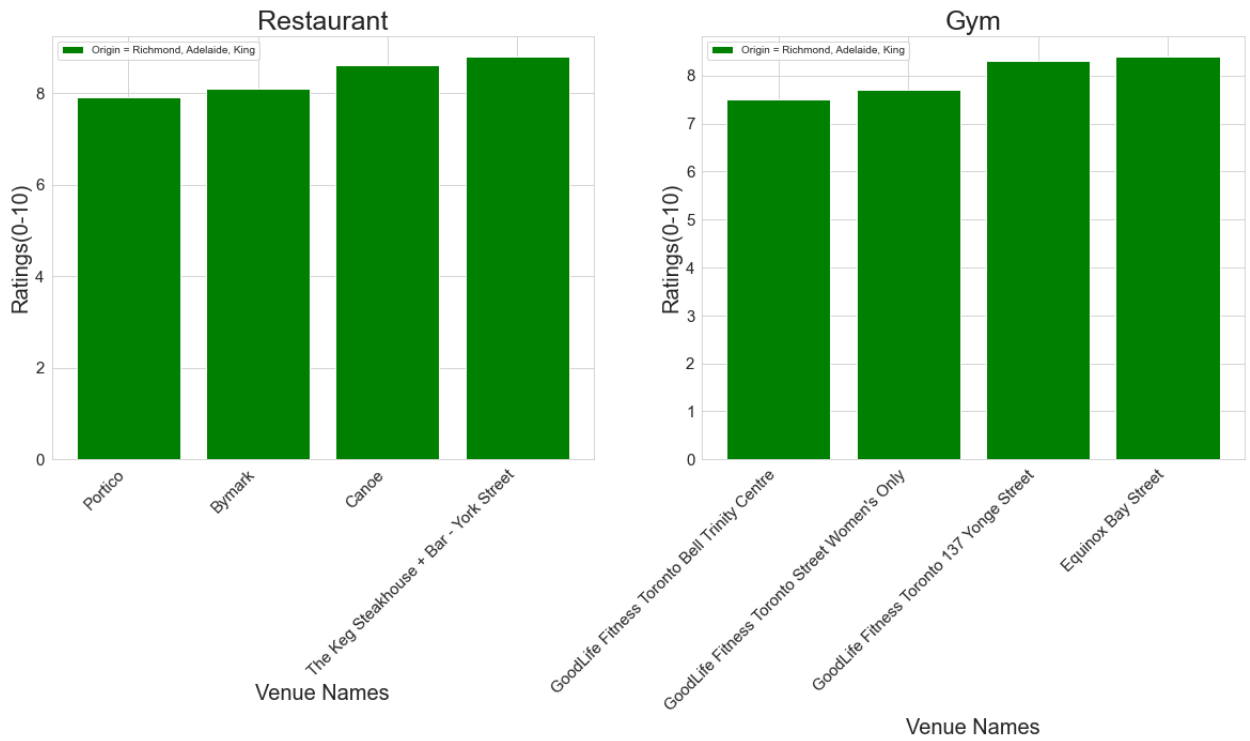

Figure 4. Actinolite with the least good rating around "Christie"

Bar Plots for Adelaide, Richmond, and King Neighborhood

Figure 5. "The Keg Steakhouse + Bar" and Equinox with the shortest distances

Figure 6. "The Keg Steakhouse + Bar" and Equinox with the best ratings

Based on the plots provided above, it was observed that no result for the "Gym" category was found for the neighborhood of Christie, and 3 results were found for the "Restaurant" category. However, a total of 8 results was obtained for the "Adelaide, Richmond, King" neighborhood" with an even split between the "Gym" and "Restaurant" category. The number of results obtained from both neighborhoods could be a potential deciding factor as the "Christie" neighborhood offers fewer options based on the number of restaurant and gym venues available.

Furthermore, the bar plots are displayed in order of increasing distance from the starting location and average rating. For the "Christie" location the best restaurant around is "Rasa" with an 8.9 rating, and it is also the farthest away from the starting location with 1.5 km. Conversely, the closest restaurant is "Actinolite" with a shorter distance of 609 meters, but with a less impressive rating of 7.6. However, for the "Adelaide, Richmond, and King" location with a handful of results for both the "Gym" and "Restaurant" category, this location is worth looking into. The best "Restaurant" around is "The Keg Steakhouse + Bar" with a rating of 8.8 and 78 meters from the starting location. The best "Gym" around is called "Equinox" with a rating of 8.4 and 315 meters away from the location of origin. These two locations are considered the best for this neighborhood because they have the highest rating and the closest distances in their respective categories.

Neighborhood	Venue Name	Category	Distance from Origin	Ratings (0-10)
Christie	Rasa		1464 meters	8.9
Adelaide, Richmond, and King	The Keg Steakhouse + Bar	Restaurant	78 meters	8.8
	Canoe		287 meters	8.4
	Bymark		312 meters	8.1
	Portico		490 meters	7.9
	Equinox Bay Street	Gym	314 meters	8.4
	GoodLife Fitness Toronto 137 Yonge Street		459 meters	8.3

Figure 7. Summary of "Christie" Vs "Adelaide, Richmond, and King" Neighborhood

4. CONCLUSIONS

In this study, I analyzed the search results on two neighborhoods; "Christie" and "Adelaide, Richmond, and King" based on the "Gym" and "Restaurant" categories. I calculated the individual distances between each result from the location the search was made and calculated the average rating for every location (venue) returned. A cluster of the results around each neighborhood was built based on proximity, and bar plots were generated in order of increasing distance and ratings. It was observed that "Adelaide, Richmond, and King" neighborhood had the best results both under the "Gym" and "Restaurant" category. It was also noted that all the restaurants result for "Adelaide, Richmond, and King" are within a radius of 500 meters which are closer in distance than the best restaurant in "Christie" neighborhood with a distance of 610 meters. Below is a summary of the results from both locations. Based on the analysis and observation made, I conclude that "Adelaide, Richmond, and King" is a better neighborhood for someone looking to relocate to a new location with interests in good gyms and restaurant locations.