Introduction: Business Problem

Background/Scenario

Moving to a new city can be scary for some, they have grown accustom to their surroundings, whether it be the amenities, venues and/or access transportation. As time progresses individual's appreciation for their neighborhood grows, as a result people tend to look for similar types of neighborhoods when moving. This report will attempt to find an optimal apartment for an individual currently residing in the Uptown Residences in Toronto, Canada, who is looking to move to New York City, New York. Along side the criteria of the specification of the new apartment, this report will attempt to find a home for the individual in a neighborhood similar to the neighborhood they currently reside in (Yorkville).

Throughout this project we will assume that the individual's decision to move to New York City is based around a new job in the Financial District. The Financial District is located in the south end of Manhattan, as a result we will attempt to find an apartment to rent within the surrounding neighborhoods of Tribeca, Soho, Bowery and the Lower East Side.

As previously mentioned, the individual is living in the Uptown Residences, he/she currently rents a 1-bedroom, 1-bathroom apartment. The size of the apartment is 570 sq. ft. and the current rent is $2400 CAD. The distance to the nearest subway station is 207m, giving the individual easy access to transportation throughout the city. The Yorkville neighborhood is one of Canada's most exclusive shopping districts, along side the numerous shopping options are bars, restaurants, art galleries, office spaces and parks.

Problem

In order to create a similar living environment for the individual in New York City, we must be able to find an apartment located in a similar neighborhood that offers similar characteristics to their current situation. In order to live close in proximity to the Financial District we must find the individual a rental unit with the following conditions:

- An apartment with 1-bedroom, and 1-bathroom.
- The unit must not exceed an increased monthly rent of $2300 USD.
- The apartment must be in a neighborhood near the Financial District.
- Must be in proximity to the subway station.
- The neighborhood must have similar attributes to Yorkville to make the transition to New York as seamless as possible.

Interest

This project is desirable for people looking to move and would like to move to a similar neighborhood.

Data

Required Data

Based on the problem at hand, the factors that would affect the decision are:

- The venues in Yorkville which surrounds the Uptown Residences.
- The venues in the different Manhattan neighborhoods.
- The location of the subway stations in Manhattan.
- The available apartments that meet the desired criteria.

Data Sources

The geocoders from the geoPy library will be used to generate approximate longitudinal and latitudinal coordinates for the residences and the two cities. The New York City neighborhoods and their corresponding longitudinal and latitudinal coordinates are retrieved from https://geo.nyu.edu/catalog/nyu_2451_34572. The list of Toronto and New York subway stations and their coordinates was retrieved from http://scruss.com/blog/2005/12/14/toronto-subway-station-gps-locations/, https://data.cityofnewyork.us/Transportation/Subway-Stations/arq3-7z49 respectively. I manually inputted the data from https://www.apartments.com/lower-east-side-new-york-ny/under-2300/?bb=g2ljy0mvvHlryvqD in order to get the current available rental apartments. The median rent data in Manhattan was retrieved from https://streeteasy.com/blog/data-dashboard/. Finally, the Foursquare API is be used to acquire the desired venue data.

Data Cleaning

To create consistency within all of the neighborhood tables the columns within the Toronto Subway CSV file were re-arranged. I decided to remove the data for subway stations not located in Yorkville, as it did not pertain to problem at hand. Within the New York City Subway CSV there were several columns of data that weren't required, such as subway line or notes, as a result they were removed the file. The longitudinal and latitudinal coordinates corresponding to the subway stations were stored within Java syntax for example POINT(longitude, latitude). I wrote a standalone script to stripped and split the data and stored it in separate columns. In order to access the current median asking rent in New York City, I wrote a script to determine the most relevant asking price for each neighborhood.

Collected Data

There are several steps required to determine a desirable apartment within a similar neighborhood for the individual. The first step is to analyze the individual's current surroundings, we need to determine the types of venues that are close in proximity, their distances to the Uptown Residence and the distance to the nearest subway station. Using the geoPy geocoder, lonlat_to_xy function (from the IBM/Coursera Capstone project example) and Foursquare API, the longitudinal and latitudinal coordinates, the cartesian coordinates, and the surrounding venues were determined for the Uptown Residence.

	Neighborhood	Neighborhood Longitude	Neighborhood Latitude	Venue	Venue Longitude	Venue Latitude	Venue X	Venue Y	Distance To Home	Venue Category
0	Yorkville	-79.387415	43.669493	Indigo	-79.389057	43.669065	-5.307983e+06	1.050786e+07	202.748299	Bookstore
1	Yorkville	-79.387415	43.669493	Pi Co.	-79.389852	43.670107	-5.307807e+06	1.050793e+07	299.723712	Pizza Place
2	Yorkville	-79.387415	43.669493	Wish	-79.385694	43.668759	-5.308074e+06	1.050747e+07	231.820606	Restaurant
3	Yorkville	-79.387415	43.669493	Sabai Sabai	-79.384857	43.670647	-5.307785e+06	1.050734e+07	349.898095	Thai Restaurant
4	Yorkville	-79.387415	43.669493	Paramount Fine Foods	-79.389865	43.670677	-5.307716e+06	1.050792e+07	341.933559	Middle Eastern Restaurant

Table 1: The first 5 venues located within a radius of 500m of the Uptown Residences.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Yorkville	Café	Coffee Shop	Sushi Restaurant	Italian Restaurant	Japanese Restaurant	French Restaurant	Hotel	Indian Restaurant	Salad Place	Restaurant

Table 2: Top 10 most common venue categories in which surrounds the Uptown Residences.

From the getNearbyVenue and return_most_common_venues functions from the Segmenting and Clustering Neighborhoods in New York City lab we can get the top 10 most common venues which surrounds the Uptown Residences seen in Table 2. When using the Foursquare API there is a limitation to the number of venues that can be obtained, for that reason, data of 100 surrounding venues was collected. Using the Folium package, I was able to display the venues and the Uptown residences on a map, as seen in Figure 1.

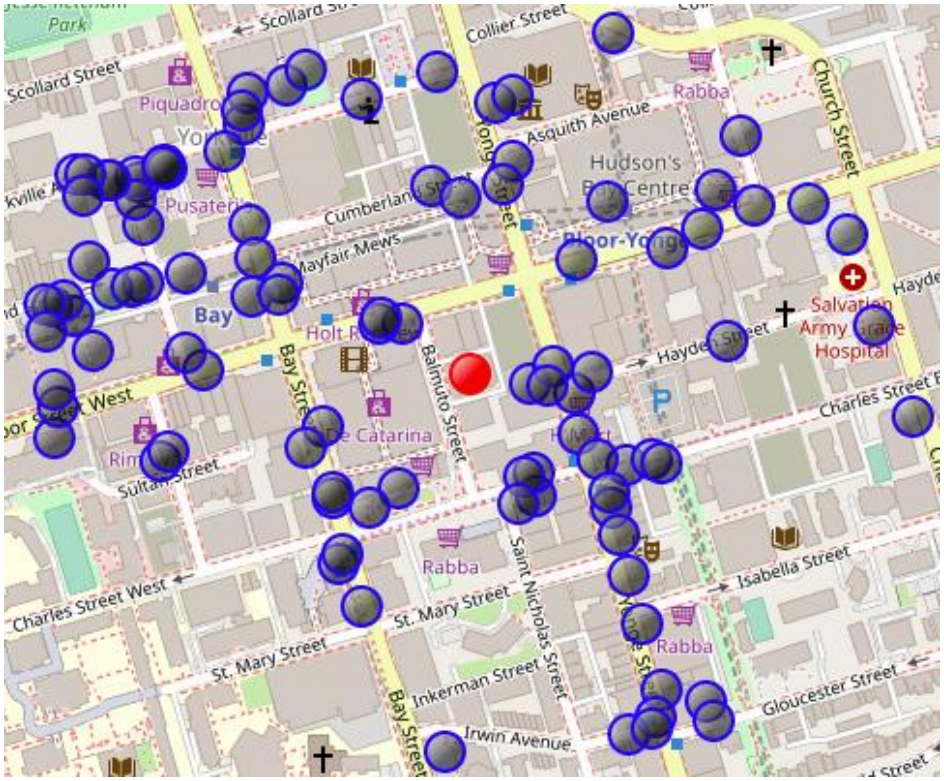

Figure 1: Map of the Uptown Residences (red dot) and the surrounding venues (blue dots).

The individual living in Yorkville, a neighborhood located in downtown Toronto, relies on the subway and public transit as a method of transportation. Table 3 lists all of the subway stations in or near Yorkville and the distance (in meters) to the Uptown Residences. The data illustrates that

subway stations in close proximity to the new apartment is crucial as it will be the main method of transportation.

	Station Name	Neighborhood	Longitude	Latitude	X	Y	Distance To Home
0	Bay Station	Yokville	-79.3907	43.6701	-5.307797e+06	1.050803e+07	393.829541
1	Museum Station	Yokville	-79.3935	43.6671	-5.308238e+06	1.050840e+07	803.984971
2	Bloor-Yonge Station	Yokville	-79.3865	43.6706	-5.307771e+06	1.050753e+07	206.607575
3	Rosedale Station	Yokville	-79.3889	43.6768	-5.306754e+06	1.050770e+07	1181.986019
4	Sherbourne Station	Yokville	-79.3764	43.6722	-5.307646e+06	1.050634e+07	1350.817727

Table 3: The coordinates of the subway stations within Yorkville.

Using the newyork_data JSON file from the Segmenting and Clustering Neighborhoods in New York City lab along with geoPy's geocoder I was able to collect the coordinates for the four neighborhoods, seen in Table 4.

	Borough	Neighborhood	Longitude	Latitude	X	Y
0	Manhattan	Soho	-74.000657	40.722184	-5.820668e+06	9.868956e+06
1	Manhattan	Tribeca	-74.010683	40.721522	-5.820815e+06	9.870247e+06
2	Manhattan	Bowery	-73.994390	40.719457	-5.821109e+06	9.868135e+06
3	Manhattan	Lower East Side	-73.980890	40.717807	-5.821342e+06	9.866385e+06

Table 4: The coordinates of the four neighborhoods of interest.

In order to mimic the future analysis of the Yorkville data, we require the location of all the subway stations with Tribeca, Bowery, Soho and the Lower East Side in order to determine the distance from the stations to the potential apartments.

	Station Name	Neighborhood	Longitude	Latitude
0	Delancey St - Essex St	Lower East Side	-73.987409	40.718306
1	East Broadway	Lower East Side	-73.990177	40.713855
2	Lower East Side - 2nd Ave	Lower East Side	-73.989938	40.723402
3	Spring St	Soho	-74.003739	40.726227
4	Prince St	Soho	-73.997702	40.724329
5	World Trade Center	Tribeca	-74.009745	40.712564
6	Chambers St	Tribeca	-74.003407	40.713234
7	Canal St	Soho	-73.999826	40.718174
8	Broadway - Lafayette St	Lower East Side	-73.996204	40.725297
9	Bowery	Bowery	-73.993807	40.720247
10	Franklin St	Tribeca	-74.006886	40.719318
11	Spring St	Soho	-73.997141	40.722301

Table 5: The subway stations within each neighborhood.

Similar to Table 2, Table 6 displays the top 10 most common venues in Soho, Tribeca, Bowery and the Lower East Side.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Soho	Italian Restaurant	French Restaurant	Clothing Store	Optical Shop	Coffee Shop	Women's Store	Yoga Studio	Pizza Place	Boutique	Salon / Barbershop
1	Tribeca	Park	American Restaurant	Spa	Hotel	Coffee Shop	Men's Store	Wine Bar	Café	French Restaurant	Bakery
2	Bowery	Cocktail Bar	Café	Ice Cream Shop	American Restaurant	Wine Bar	Bakery	Pizza Place	Coffee Shop	Sandwich Place	Optical Shop
3	Lower East Side	Italian Restaurant	Ice Cream Shop	Sandwich Place	Coffee Shop	Cocktail Bar	Café	Japanese Restaurant	Boutique	Wine Shop	Wine Bar

Table 6: Top 10 venues per neighborhood.

Table 7 displays the available apartments for rent in the Lower East Side that fit the individual's specifications.

	Address	Longitude	Latitude	X	Y	Price	Size	Rooms
0	166 Suffolk St	-73.985019	40.720717	-5.820863e+06	9.866931e+06	2250	500-600	1
1	25 Clinton St	-73.984442	40.720542	-5.820891e+06	9.866856e+06	2050	500-600	1
2	43 Clinton St	-73.984887	40.719949	-5.820993e+06	9.866910e+06	2253	650	1
3	93 Clinton St	-73.985535	40.718592	-5.821225e+06	9.866988e+06	1950	500-600	1
4	57 Pitt St	-73.983062	40.717764	-5.821357e+06	9.866665e+06	2250	500	1
5	191 Henry St	-73.987951	40.713705	-5.822063e+06	9.867277e+06	2300	NaN	2
6	164 Henry St	-73.989205	40.713268	-5.822141e+06	9.867437e+06	1995	NaN	1
7	167 Ludlow St	-73.987762	40.721634	-5.820717e+06	9.867289e+06	2195	NaN	1
8	176-178 Ludlow St	-73.987392	40.721742	-5.820697e+06	9.867242e+06	1850	500-600	1
9	120 Orchard St	-73.989321	40.719365	-5.821107e+06	9.867480e+06	2249	400	1
10	72 Orchard St	-73.990289	40.717546	-5.821419e+06	9.867597e+06	2300	500-600	1

Table 7: Available apartment listing in the Lower East Side.

Methodology

Using the sklearn package, k means clustering was used to analyze the similarities between the four neighborhoods in Manhattan and Yorkville. Given that there are five neighborhoods we do not require a lot of clusters. We will examine the clustering with 2 or more clusters initialized.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Yorkville	Café	Coffee Shop	Italian Restaurant	Japanese Restaurant	Sushi Restaurant	Restaurant	Indian Restaurant	Hotel	Sandwich Place	French Restaurant
1	Soho	Italian Restaurant	Clothing Store	French Restaurant	Coffee Shop	Gym	Women's Store	Bakery	Boutique	Art Gallery	Dessert Shop
3	Bowery	Hotel	Pizza Place	Shoe Store	Cocktail Bar	Sandwich Place	Rock Club	Pilates Studio	Paper / Office Supplies Store	Men's Store	Mediterranean Restaurant
4	Lower East Side	Italian Restaurant	Ice Cream Shop	Asian Restaurant	Coffee Shop	Sandwich Place	Japanese Restaurant	Pizza Place	Boutique	Mexican Restaurant	Wine Bar

a) Cluster 1

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Tribeca	Park	American Restaurant	Hotel	Spa	Coffee Shop	Italian Restaurant	Bakery	Men's Store	Sushi Restaurant	Café

b) Cluster 2

Figure 3: Cluster when k=2.

When we run the k means clustering algorithm with 2 clusters defined, we get an interesting result seen in Figure 3. Figure 3a) displays the neighborhoods in the first cluster: Yorkville, Soho, Bowery and the Lower East Side, this indicates that these 4 neighborhoods share similarities. Looking at the second cluster Figure 3b) we see that Tribeca is by itself which suggests that this neighborhood has the least similarities to the others.

Running the k means clustering algorithm with 3 defined clusters will allow us to refine our search in similar neighborhoods.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Soho	Italian Restaurant	Clothing Store	French Restaurant	Coffee Shop	Gym	Women's Store	Bakery	Boutique	Art Gallery	Dessert Shop
3	Bowery	Hotel	Pizza Place	Shoe Store	Cocktail Bar	Sandwich Place	Rock Club	Pilates Studio	Paper / Office Supplies Store	Men's Store	Mediterranean Restaurant
4	Lower East Side	Italian Restaurant	Ice Cream Shop	Asian Restaurant	Coffee Shop	Sandwich Place	Japanese Restaurant	Pizza Place	Boutique	Mexican Restaurant	Wine Bar

a) Cluster 1

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Yorkville	Café	Coffee Shop	Italian Restaurant	Japanese Restaurant	Sushi Restaurant	Restaurant	Indian Restaurant	Hotel	Sandwich Place	French Restaurant

b) Cluster 2

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Tribeca	Park	American Restaurant	Hotel	Spa	Coffee Shop	Italian Restaurant	Bakery	Men's Store	Sushi Restaurant	Café

c) Cluster 3

Figure 4: Cluster when k=3.

Similar to Figure 3, we get an expected result in which Tribeca is in a cluster by itself. Different to Figure 3, we see that the three remaining Manhattan neighborhoods are in the first cluster seen in Figure 4a) and that Yorkville is now in a separate cluster. This indicates that Yorkville has enough similarities to Soho, Bowery and the Lower East Side that it can be clustered together when there are 2 clusters defined. When the third cluster is added Yorkville is placed in a separate cluster indicating that the similarities between the 3 Manhattan neighborhoods is a lot stronger compared to their similarities with Yorkville. Since Yorkville was clustered by itself, we do not need to run the algorithm with 4 clusters. From Figure 3 we can determine that of the 3 Manhattan neighborhoods, the Lower East Side is the most similar to Yorkville. The venues that these two neighborhoods share are: Italian Restaurants, Coffee Shops, Sandwich Places, and Japanese Restaurants.

Distance To Home	35 Balmuto	166 Suffolk St	25 Clinton St	43 Clinton St	93 Clinton St	57 Pitt St	191 Henry St	164 Henry St	167 Ludlow St	176-178 Ludlow St	120 Orchard St	72 Orchard St
<100	2	5	3	3	2	0	2	2	15	12	5	12
100 - 200	15	3	8	10	19	2	5	8	12	13	18	10
200 - 300	21	18	11	9	7	4	7	6	17	16	21	28
300 - 400	16	19	10	13	26	16	7	12	17	19	19	31
400 - 500	25	34	34	24	17	13	18	19	23	24	27	18
500>	21	19	23	32	29	34	19	17	16	16	10	1

Table 8: The distribution of venues within a 500m radius relative to their distance to the corresponding apartment.

Address	Neighborhood	Mean Venue Distance
35 Balmuto St	Yorkville	363.166296
166 Suffolk St	Lower East Side	392.270042
25 Clinton St	Lower East Side	399.482213
43 Clinton St	Lower East Side	407.259253
93 Clinton St	Lower East Side	376.170508
57 Pitt St	Lower East Side	459.192298
191 Henry St	Lower East Side	414.020100
164 Henry St	Lower East Side	394.193779
167 Ludlow St	Lower East Side	321.753972
176-178 Ludlow St	Lower East Side	329.725865
120 Orchard St	Lower East Side	325.686626
72 Orchard St	Lower East Side	280.811958

Table 9: The mean venue distance for each apartment.

Using the distribution of venues from Table 8 with the mean venue distances in Table 9, we are able to narrow down the selection of possible apartments.

Address	Neighborhood	Mean Venue Distance
167 Ludlow St	Lower East Side	321.753972
176-178 Ludlow St	Lower East Side	329.725865
120 Orchard St	Lower East Side	325.686626
72 Orchard St	Lower East Side	280.811958

Table 10: The remaining apartments that had the same or smaller mean venue distance to 35 Balmuto Street.

From the data collected in Table 9, we were able to determine the apartments that had the same or a smaller mean venue distance compared to the Uptown Residence, seen in Table 10. Along with the mean venue distance, we can compare the four remaining neighborhoods distribution of venues in Table 8 to the venue distribution of 35 Balmuto Street.

Station Name	Neighborhood	Longitude	Latitude	167 Ludlow St	176-178 Ludlow St	120 Orchard St	72 Orchard St
Delancey St - Essex St	Lower East Side	-73.987409	40.718306	512.410049	539.185898	428.765528	246.824828
East Broadway	Lower East Side	-73.990177	40.713855	1341.372938	1354.954398	1239.391987	1002.734509
Lower East Side - 2nd Ave	Lower East Side	-73.989938	40.723402	781.542993	859.573095	876.572959	994.506363
Broadway - Lafayette St	Lower East Side	-73.996204	40.725297	1639.734414	1719.407901	1719.797384	1786.375893

Table 11: Distance (in meters) from the corresponding subway stations to the remaining apartment buildings.

The final step in the analysis process is to examine the distance the apartments are from the subway stations. As previously mentioned, the individual currently relies on the subway as their main vehicle of transportation and intends to use the subway system as their primary method of transportation in New York as well. From Table 11 we see that 120 and 72 Orchard Street are the closest to a subway stop.

Discussion and Results

Using the k means clustering algorithm we are able to examine the characteristics of the 5 neighborhoods. We saw that Tribeca had the least in common to Yorkville and the other 3 Manhattan neighborhoods. Based on the frequency of the venues in each neighborhood it was determined that the Lower East Side was the most similar to Yorkville.

Using the mean venue distance and the distribution of the venues, the number of available apartments for rent in the Lower East Side was reduced to:

- 167 Ludlow St
- 176-178 Ludlow St
- 120 Orchard St
- 72 Orchard St

The apartment located at 72 Orchard St is the most desirable for the individual in this capstone project. It is a 1-bedroom, 1-bathroom apartment priced at $2300 USD located in the Lower East Side neighborhood. The selection was based on the individuals specified criteria and the analysis previously conducted, those being:

- The apartment is located in the neighborhood most similar to Yorkville.
- It is the closest in proximity to a subway station.

Conclusion

The purpose of this project was to determine the best apartment for an individual to rent in either Tribeca, Soho, Bowery or the Lower East Side (neighborhood's which surrounds the financial district in Manhattan), based on the neighborhood's similarity to Yorkville (the individual's current neighborhood), proximity to venues, distance to the subway stations and the price of the apartment.

Using the k means clustering algorithm, the four Manhattan neighborhoods along with Yorkville are grouped together based on the top 10 frequent venues. The use of the algorithm was to determine the neighborhood most similar to Yorkville. The algorithm was run with varying number of clusters, k = 2, 3…, N until Yorkville was either grouped with 1 other neighborhood or was in a cluster alone.

After the neighborhood is determined, analysis of potential rental apartments is conducted. The distribution of the venues based on proximity to the apartment, the mean venue distance and the distance to the surrounding subway stations is determined.

The apartment on 72 Orchard St is the best fit for the individual based off of statistical analysis. In reality there are sometimes other factors that can affect an individual's selection, such as the appearance of the apartment, or the landlord, in that case this program provides alternative options.