1. Introduction

1.1 Background

An international realtor has identified an opportunity to assist customers in relocating to large cities in North America. The realtor is seeking to a way provide a greater level of detail to their customers by developing information regarding the quantity and variety of venues within prospective cities. This information will allow customers to select the neighborhoods that best suit their interests.   The research identifies the type and frequency of venues which can assist in defining a city's culture and allow customers to align their living preferences with specific neighborhoods.

The realtor's clients most common destinations are Toronto and New York City. To illustrate the available data and provide a comparative analysis, a report describes each city's respective neighborhoods and most common venues. This data is arrayed in illustrative maps that allow clients examine and explore wide array of neighborhoods which comprise each city.

1.2 Business Problem

An international realtor has identified an opportunity to assist customers in relocating to large cities in North America. The realtor is seeking to a way provide a greater level of detail to their customers by developing information regarding the quantity and variety of venues within prospective cities. This information will allow customers to select the neighborhoods that best suit their interests.   The research identifies the type and frequency of venues which can assist in defining a city's culture and allow customers to align their living preferences with specific neighborhoods.

The realtor's clients most common destinations are Toronto and New York City. To illustrate the available data and provide a comparative analysis, a report describes each city's respective neighborhoods and most common venues. This data is arrayed in illustrative maps that allow clients examine and explore wide array of neighborhoods which comprise each city.

1.3 Interest

The interest of this report is applicable to people relocating or considering moving to a new city, tourists and visitors seeking to identify specific areas of interest along with potential investors seeking to understand market dynamics. The most obvious target would be people relocating to Toronto or New York City. They can really begin to explore the city without even putting 'feet on ground'. Additionally, Foreign and Domestic tourists will benefit from this analysis. They can pinpoint an exact neighborhood based on the top amenities in the area. Does it have too much or too little of what they are looking for? In a similar way, a new business owner or investor will be able to thoroughly evaluate the right neighborhood for their newest ventures.

2. Data acquisition and cleaning

2.1 Data Sources.

The websites listed below provided data in support of the research:

    -Toronto Neighborhoods -
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
     -Toronto Latitude and Longitude - http://cocl.us/Geospatial_data
     -New York City neighborhoods - https://geo.nyu.edu/catalog/nyu_2451_34572
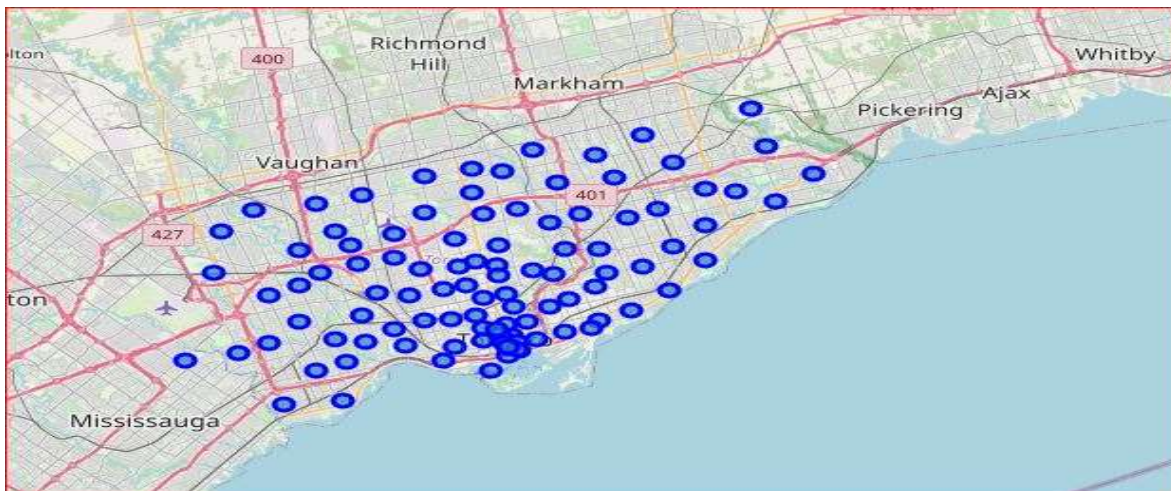   -New York City Latitude and Longitude = Python Geolibrar**y**

2.2 Data Cleaning

1. Obtain Postal Code, Borough, and Neighborhood information from Wikipedia

2. Obtain Latitude and Longitude (lat/long) data from http://cocl.us/Geospatial data. Once collected, combine (lat/long) with previous data obtained from Wikipedia.

3. Use Folium to create maps that depict each city's neighborhoods. Once the city data is obtained, identify a more specific location (Borough) to identify neighborhoods.

4. Utilize HTTP requests to the Foursquare API server using (lat/long) of Toronto's and New York City's neighborhoods to pull venue and venue category information.

5. The Foursquare API search enabled the collection venue proximity within specific neighborhoods. Due to http request restrictions the number of places per neighborhood parameter is set to 100 and the radius parameter would be set to 500.

6. Folium- Python visualization library would be used to visualize the neighborhood clusters distribution of Toronto and New York City over an interactive leaflet map.

7. Extensive comparative analysis of two randomly picked neighborhoods (Scarborough and Flushing) is carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.'

8. Unsupervised machine learning algorithm K-mean clustering is applied to form the clusters of different categories of places residing in and around the neighborhoods. These clusters from each of those two chosen neighborhoods are analyzed individually and comparatively to derive the results.

9. In accordance with our problem statement, the data obtained would allow a potential client to review and consider the most frequently occurring venues. By comparing specific neighborhoods and illustrating K-means clustering on a map, clients would be able to select the best neighborhood to fit their needs.

3.0 Exploratory data analysis

To analyze the data, it was necessary to combine the neighborhood data with the latitude and longitude data. A sample of this combination is provided below for illustrative purposes. The complete data set is available in the repository.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.80669	-79.19435
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78454	-79.1605
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.76357	-79.18871
3	M1G	Scarborough	Woburn	43.77099	-79.21692
4	M1H	Scarborough	Cedarbrae	43.77314	-79.23948

 The geographical data and Wikipedia postal code data identified the geographical coordinates of Toronto, Canada are 43.6534817, -79.3839347. Additionally, the data identified that Toronto has 10 boroughs and 103 neighborhoods. The map above illustrates the neighborhoods.

To allow for further analysis, The Toronto borough of Scarborough was selected for a more in depth level of evaluation. Utilizing the combination of neighborhood and geographical data, Foursquare API also provide venue data. The specific analysis of Scarborough identified 86 venues in 17 neighborhoods, 76 distinct venues in 53 categories.

3.1 Sorting venue data by frequency

The data was further refined by identifying the top 10 venues within each district. This refinement is depicted below.

```
----Agincourt----
                       venue  freq
0               Clothing Store  0.25
1   Latin American Restaurant  0.25
2                      Lounge  0.25
3              Breakfast Spot  0.25
4                   Pet Store  0.00
5           Italian Restaurant  0.00
6        Korean BBQ Restaurant  0.00
7              Medical Center  0.00
8               Metro Station  0.00
9          Mexican Restaurant  0.00

----Birch Cliff, Cliffside West----
                       venue  freq
0             College Stadium  0.25
1        General Entertainment  0.25
2                Skating Rink  0.25
3                        Café  0.25
4           Accessories Store  0.00
5          Italian Restaurant  0.00
6        Korean BBQ Restaurant  0.00
7   Latin American Restaurant  0.00
8                      Lounge  0.00
9              Medical Center  0.00

----Cedarbrae----
                     venue  freq
0               Gas Station  0.11
1         Athletics & Sports  0.11
2            Thai Restaurant  0.11
3                     Bakery  0.11
4                       Bank  0.11
5                     Lounge  0.11
6           Hakka Restaurant  0.11
```

```
7     Fried Chicken Joint  0.11
8   Caribbean Restaurant  0.11
9      Accessories Store  0.00
```

----Clarks Corners, Tam O'Shanter, Sullivan----
```
                   venue   freq
0                Pharmacy   0.14
1              Pizza Place  0.14
2    Fast Food Restaurant  0.14
3     Fried Chicken Joint  0.07
4             Gas Station  0.07
5       Convenience Store  0.07
6       Italian Restaurant 0.07
7       Chinese Restaurant 0.07
8             Noodle House 0.07
9                    Bank  0.07
```

----Cliffside, Cliffcrest, Scarborough Village West----
```
                       venue  freq
0        American Restaurant   0.5
1                      Motel   0.5
2                Noodle House  0.0
3           Indian Restaurant 0.0
4                Intersection 0.0
5           Italian Restaurant 0.0
6       Korean BBQ Restaurant 0.0
7   Latin American Restaurant 0.0
8                      Lounge 0.0
9              Medical Center 0.0
```

----Dorset Park, Wexford Heights, Scarborough Town Centre----
```
                       venue  freq
0           Indian Restaurant  0.4
1          Chinese Restaurant  0.2
2       Vietnamese Restaurant  0.2
3                   Pet Store  0.2
4                Soccer Field  0.0
5                Intersection  0.0
6          Italian Restaurant 0.0
7       Korean BBQ Restaurant 0.0
8   Latin American Restaurant 0.0
9                      Lounge 0.0
```

----Golden Mile, Clairlea, Oakridge----
```
                       venue  freq
0                      Bakery 0.25
1                    Bus Line 0.25
2              Ice Cream Shop 0.12
3                 Bus Station 0.12
4               Metro Station 0.12
5                Intersection 0.12
6                        Park 0.00
7          Italian Restaurant 0.00
8       Korean BBQ Restaurant 0.00
9   Latin American Restaurant 0.00
```

----Guildwood, Morningside, West Hill----
```
                venue  freq
0        Intersection  0.12
1                Bank  0.12
2       Breakfast Spot 0.12
3       Medical Center 0.12
```

```
4              Restaurant  0.12
5  Rental Car Location  0.12
6    Electronics Store  0.12
7   Mexican Restaurant  0.12
8          Noodle House  0.00
9    Italian Restaurant  0.00


----Kennedy Park, Ionview, East Birchmount Park----
                      venue  freq
0              Hobby Shop  0.25
1        Department Store  0.25
2          Discount Store  0.25
3             Coffee Shop  0.25
4            Soccer Field  0.00
5            Skating Rink  0.00
6       Italian Restaurant  0.00
7     Korean BBQ Restaurant  0.00
8  Latin American Restaurant  0.00
9                   Lounge  0.00


----Malvern, Rouge----
                      venue  freq
0       Fast Food Restaurant   1.0
1         Accessories Store   0.0
2              Noodle House   0.0
3              Intersection   0.0
4        Italian Restaurant   0.0
5     Korean BBQ Restaurant   0.0
6  Latin American Restaurant   0.0
7                   Lounge   0.0
8            Medical Center   0.0
9             Metro Station   0.0


----Milliken, Agincourt North, Steeles East, L'Amoreaux East----
                      venue  freq
0                      Park  0.33
1              Intersection  0.33
2                Playground  0.33
3              Noodle House  0.00
4        Italian Restaurant  0.00
5     Korean BBQ Restaurant  0.00
6  Latin American Restaurant  0.00
7                   Lounge  0.00
8            Medical Center  0.00
9             Metro Station  0.00


----Rouge Hill, Port Union, Highland Creek----
                      venue  freq
0                       Bar   1.0
1         Accessories Store   0.0
2                      Park   0.0
3              Intersection   0.0
4        Italian Restaurant   0.0
5     Korean BBQ Restaurant   0.0
6  Latin American Restaurant   0.0
7                   Lounge   0.0
8            Medical Center   0.0
9             Metro Station   0.0


----Scarborough Village----
                      venue  freq
0                Playground   0.5
```

```
1              Cosmetics Shop  0.5
2             Accessories Store  0.0
3                Noodle House  0.0
4                Intersection  0.0
5           Italian Restaurant  0.0
6         Korean BBQ Restaurant  0.0
7    Latin American Restaurant  0.0
8                     Lounge  0.0
9              Medical Center  0.0


----Steeles West, L'Amoreaux West----
                   venue  freq
0   Fast Food Restaurant  0.18
1            Coffee Shop  0.09
2          Breakfast Spot  0.09
3               Pharmacy  0.09
4             Pizza Place  0.09
5     Chinese Restaurant  0.09
6          Sandwich Place  0.09
7      Indian Restaurant  0.09
8                   Bank  0.09
9            Supermarket  0.09


----Wexford, Maryvale----
                      venue  freq
0           Accessories Store  0.17
1                    Bakery  0.17
2             Sandwich Place  0.17
3   Middle Eastern Restaurant  0.17
4       Vietnamese Restaurant  0.17
5             Auto Garage  0.17
6                    Bank  0.00
7                     Bar  0.00
8       Korean BBQ Restaurant  0.00
9   Latin American Restaurant  0.00

----Woburn----
                      venue  freq
0              Coffee Shop  0.50
1              Soccer Field  0.25
2       Korean BBQ Restaurant  0.25
3           Accessories Store  0.00
4                Intersection  0.00
5           Italian Restaurant  0.00
6   Latin American Restaurant  0.00
7                     Lounge  0.00
8              Medical Center  0.00
9              Metro Station  0.00
```

From this data, we can show the top 10 most common venues per neighborhood in Scarborough.

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
Agincourt	Clothing Store	Latin American Restaurant	Breakfast Spot	Lounge	Vietnamese Restaurant	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store	Department Store
Birch Cliff, Cliffside West	General Entertainment	Skating Rink	College Stadium	Café	Clothing Store	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store
Cedarbrae	Hakka Restaurant	Thai Restaurant	Athletics & Sports	Gas Station	Bakery	Bank	Fried Chicken Joint	Caribbean Restaurant	Lounge	College Stadium
Clarks Corners, Tam O'Shanter, Sullivan	Pizza Place	Fast Food Restaurant	Pharmacy	Fried Chicken Joint	Gas Station	Noodle House	Convenience Store	Italian Restaurant	Bank	Thai Restaurant
Cliffside, Cliffcrest, Scarborough Village West	Motel	American Restaurant	Vietnamese Restaurant	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store	Department Store	Cosmetics Shop
Dorset Park, Wexford Heights, Scarborough Town...	Indian Restaurant	Vietnamese Restaurant	Chinese Restaurant	Pet Store	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store	Department Store	Cosmetics Shop
Golden Mile, Clairlea, Oakridge	Bakery	Bus Line	Ice Cream Shop	Intersection	Bus Station	Metro Station	Vietnamese Restaurant	College Stadium	Fried Chicken Joint	Fast Food Restaurant
Guildwood, Morningside, West Hill	Bank	Restaurant	Medical Center	Breakfast Spot	Intersection	Rental Car Location	Mexican Restaurant	Electronics Store	Department Store	Clothing Store
Kennedy Park, Ionview, East Birchmount Park	Hobby Shop	Discount Store	Department Store	Coffee Shop	Hakka Restaurant	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Cosmetics Shop
Malvern, Rouge	Fast Food Restaurant	Vietnamese Restaurant	Hakka Restaurant	Gas Station	Fried Chicken Joint	Electronics Store	Discount Store	Department Store	Cosmetics Shop	Convenience Store
Milliken, Agincourt North, Steeles East, L'Amo...	Intersection	Playground	Park	Vietnamese Restaurant	Chinese Restaurant	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store	Department Store
Rouge Hill, Port Union, Highland Creek	Bar	Vietnamese Restaurant	Clothing Store	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store	Department Store	Cosmetics Shop
Scarborough Village	Playground	Cosmetics Shop	Vietnamese Restaurant	Chinese Restaurant	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store	Department Store
Steeles West, L'Amoreaux West	Fast Food Restaurant	Pizza Place	Indian Restaurant	Coffee Shop	Breakfast Spot	Pharmacy	Bank	Chinese Restaurant	Sandwich Place	Supermarket
Wexford, Maryvale	Vietnamese Restaurant	Sandwich Place	Auto Garage	Bakery	Middle Eastern Restaurant	Accessories Store	Restaurant	Caribbean Restaurant	Electronics Store	Discount Store
Woburn	Coffee Shop	Soccer Field	Korean BBQ Restaurant	Vietnamese Restaurant	Chinese Restaurant	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store	Department Store

After organizing and reviewing the data for Toronto and specifically Scarborough, it was time to analyze the data for New York City.

The data frame produced 5 boroughs and 306 neighborhoods.

Similar to the data for Toronto Canada, the borough of Queens was selected for a more in depth study.

Once again in utilizing geographical data along with Fourquare API, 2097 venue in 81 neighborhoods with identified in Queens, NY. The data was analyzed to identify the top 10 most common venues in Queens. Due to the size of the chart, I pasted only the first five neighborhoods. A full length report is available in the notebook.

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
Arverne	Surf Spot	Metro Station	Playground	BBQ Joint	Bus Stop	Pizza Place	Café	Board Shop	Sandwich Place	Beach
Astoria	Bar	Middle Eastern Restaurant	Greek Restaurant	Indian Restaurant	Seafood Restaurant	Mediterranean Restaurant	Hookah Bar	Café	Food Truck	Deli / Bodega
Astoria Heights	Italian Restaurant	Chinese Restaurant	Playground	Bakery	Bus Station	Plaza	Motel	Supermarket	Bowling Alley	Food
Auburndale	Italian Restaurant	Miscellaneous Shop	Comic Shop	Supermarket	Noodle House	Fast Food Restaurant	Mobile Phone Shop	Bar	Mattress Store	Toy / Game Store
Bay Terrace	Clothing Store	Shoe Store	Cosmetics Shop	Women's Store	Donut Shop	Kids Store	American Restaurant	Mobile Phone Shop	Men's Store	Pharmacy

3.2 . K-Means Clustering

For both the Scarborough and Queens boroughs, I used K-means clustering to group neighborhood. The Scarborough Borough in Toronto, Canada I use k-means to group the neighborhoods in Scarborough into 3 clusters. Cluster_0 has 15 (Red) neighborhoods and the most common venues are, international cuisine restaurants, shops and public parks. Cluster 1 (Purple) has 1 neighborhood , and the most common venues are general entertainment and skating rings.   Cluster 2 (Green) has 1 neighborhood, and the most common venues are skating rinks and Latin restaurants.

Queens Borough in New York City I used k-means to group the Queens borough into 5 clusters. Cluster_0 has 1 neighborhood, with deli/bodega as the most common venue. Cluster 1's most common venue is Caribbean restaurant. Cluster 2 and consist of many international cuisine restaurants and grocery stores. The most common venues are pizza places, deli, and Chinese restaurants. Cluster_1 has 1 neighborhood, and the most common venue is a dance studio. Cluster_2 has 80 neighborhoods, and the most common venue are donut shops and international cuisine restaurants. Cluster_3 has 2 neighborhoods, and the most common venues are the donut shop and Indian restaurant. Cluster_4 has 1 neighborhood, and the most common venues are gyms and fitness centers.

4. Results and Discussion. Both Scarborough and Queens borough consist of neighborhood cluster that contain majority of the neighborhoods, and the remaining cluster had 1-5 neighborhoods. Although Toronto and New York City are the largest metropolitan cities in their respective countries, the depth and breadth of neighborhoods and venues in New York City far outpaces Toronto. It is interesting to identify the similarities between both broughs given their large and diverse metropolitan populations while in some ways reflecting the local cultures of the neighborhoods.

5. Conclusions. Based on analysis of the data, both locations offer a wide array of venues for dining and entertainment. However, if someone was looking for which borough offers the greatest number of venues, Queens offers a much larger array of choices across all categories. It was interesting to evaluate and analyze the data while learning more about each of these locations.

If interested in learning more, the presentation, notebook and report are located in the Capstone-Project-Battle of the Neighborhoods located at
https://github.com/mkess76/Capstone-Project-Battle-of-the-Neighborhoods