# Relocation Neighborhood Recommendations

**Author: Trinath Sahu**

**Version 2.0**

**Last updated date: 3rd November 2019**

## Contents



## Revision History

*Use this table to identify items that are new and/or changed in each release of this document.*

| Revision Number | Revision Date | Author | Summary of Changes | Changes marked |
|---|---|---|---|---|
| 1.0 | 02-Nov-2019 | Trinath Sahu | Initial draft version (added Index, Problems, Interest and data section) | N |
| 2.0 | 03-Nov-2019 | Trinath Sahu | Added remaining section of the report | N |

# 1. Introduction

## 1.1. Background

As per report, between 2012 and 2013, about 36 million Americans moved to a new home. That is about one in every 8 people move to a new location. The United States Census Bureau has a new report with a much bigger dataset about why people move. Top reasons of these movements include new job opportunity, Long term onshore travel, Own home, and job transfer. People who move within the same county are almost twice as likely to do so for housing-related reasons as those who move to a different county. The reasons people move to different counties are pretty evenly distributed among family, jobs and housing, whereas same-county movers do so disproportionately for housing-related reasons.

Advantages of relocation might include:

**Add on to job security.** Sometimes, relocating for work is the only way to keep your job. In many cases, it is strictly mentioned in the offer letter that wherever employee is put he/she has to go to that place and work. If you decide to make the move, your employer may feel you are loyal to the company. This could result in job security if your employer has to make cuts later on.

**It can be an opportunity for career advancement.** It is not uncommon that employees look for job change to get new opportunity in terms of roles, responsibilities, pay etc. In addition to job security, relocating for work can be a career accelerator, leading to promotions. It can give the opportunity to excel, take things to the next level and set the bar for others behind you.

**It can increase your standard of living and quality of life.** Kate Windleton is a relocations manager at Strong Move, a company that helps people relocate both domestically and internationally. She said the possibility of a better standard of living is an important consideration. Take into account the salary offered as well as the cost of living. You may be taking a pay raise or pay cut, but depending on the new cost of living, you can end up with a smaller or larger disposable income.

**It is great for personal development and new experiences.** Relocation is an opportunity to start fresh and reinvent yourself. Try the things you've always wanted to do but haven't made the time for.

**You can make new friends.** Moving is the perfect opportunity to welcome new, positive friendships and leave the toxic ones behind. You can use online apps, rec leagues and local meetups to find people who share your same interests.

**You can move to a better climate.** Whether you are currently living somewhere cold and you've always wanted to live in a tropical area or vice versa, this may be your chance to finally move. Moving to your desired climate may improve your quality of life.

## 1.2. Problem

While there are lot of advantages of relocation, at the same time there are lot of challenges too in the relocation. Some of them include:

**You must find housing in an unfamiliar area.** If you are unable to take advantage of corporate or temporary housing, finding a place to live in a new city can be a major challenge.

**Right schooling and child care can be tough.** If you have children, you will need to find housing that is within the boundaries of the school district you would like your children to attend. You will also have to find trustworthy child care, which can be worrisome if other family members have watched your children, shuttled them to and from school and extracurricular activities, etc. Utilize trusted internet forums and ask around for references.

**Transportation system.** You might not have your own vehicle in the new place, so you will be dependent on public transportation system. So, you will be interested in a place where transportation service is nearby.

And many more…

If you are currently residing in a location where your needs are nearby, you might feel uncomfortable in moving to different location. Sometime, these challenges refrain you from taking up new good career opportunity and hence becomes road blocker for your career growth. So, it is very important as well as a difficult task to find a locality in unfamiliar place which should be similar to the one you are staying currently. The task becomes even more challenging when you relocate to a large city where it is very difficult to decide which factors to consider and which to factor out.

## 1.3. Interest

It will be of interest to all those people who relocates from Newyork to Toronto (which is a big crowd). If some automated system suggests them the recommended places where they can move to and live in, at least it will reduce their effort drastically by limiting scope of their search/exploration. All that you need is a place in the new neighborhood which is like your current neighborhood. If it can say neighborhood x, y and z in Toronto are like your current neighborhood ‘a’ in Newyork then you can imagine how much it will be helpful to them.

NOTE: Though this project is specifically to help people who is relocating from Newyork to Toronto, the underlying program can easily be modified to make it work for any other countries/locations.

# 2. Data acquisition and cleaning

## 2.1. Data Source

In order to solve the problem, we will need location data (like venues around the Toronto locations), Postal code, Boroughs, neighborhoods in Toronto and New york, and their longitude & latitude.

For location data, we will use Foursquare location data service provider APIs;
For postal code, Boroughs and neighborhoods, we will scrap the Wikipedia page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

We will get latitude and longitudes for Toronto neighborhoods from a csv file called as Geospatial_Coordinates.csv and for Newyork it will be nyu_2451_34572-geojson.json

## 2.2. Data cleansing:

Since there is no file/dataset present which has the Postal Code, Boroughs and neighborhoods and we are using Wikipedia website page for the same, website scrapping feature needs to be used to extract the required data. Luckily, thanks to outstanding Pandas read_html command which can search for any tables present in the page and stores in pandas dataframe directly. But that does not end there. Following tasks need to be done in order to clean up the data:

- Many rows have Borough' column as 'Not assigned'. We will process the cells that have an assigned borough and Ignore cells with a borough that is Not assigned.

- It seems, more than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that M5A is listed twice and has two neighborhoods: Harbourfront and Regent Park. These two rows will be combined into one row with the neighborhoods separated with a comma.

- If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough. So, the value of the Borough and the Neighborhood columns will be Queen's Park.

## 2.3. How data will be used to solve the problem

1. Cleanse the neighborhood data as mentioned above
2. Store the latitudes and longitudes of all neighborhoods in dataframe
3. Call Foursquare APIs to explore the neighborhoods in Toronto city. I.e. get all nearby venues around all the neighborhoods in Toronto. Do the same for current neighborhood in Newyork.
4. Store all the venues around all Toronto neighborhoods.
5. Then add an additional entry with all the venues that are present nearby current neighborhood in Newyork.

6. Group the neighborhoods into Clusters by using clustering algorithm (such as Kmeans). Since we have added an entry for current neighborhood in Newyork, all the Toronto neighborhoods which are similar to the Newyork one will fall into same cluster. That's the trick of adding this additional row.
7. Visualize the data by highlighting different clusters in different colors. So, all the Toronto neighborhoods present in the same cluster as point 5 are similar to current Newyork neighborhood. Suppose cluster# for the current neighborhood is found to be 8 , then all the neighborhoods of Toronto which is assigned cluster no. as 8 will be the ones you are looking for. Those neighborhoods will be highlighted in special color with proper legend. The folio map will clearly show them which will help you to take decision on which locality to move into.

## 2.4. Feature Selection

- Venue categories around neighborhoods of Toronto: We will call Foursquare APIs to get the venues in neighborhoods of Toronto. Then using Onehot coding, we will transform them into features.
- Venue categories around current neighborhood in Newyork. This is required to compare new neighborhoods against the current one.

<u>Some examples of feature columns (which is venue category in this case):</u>

Park, school, College Academic Building, College Arts Building, General College & University, Restaurant, pizza, Bank, Gym, Bus station, bus stop, cafe, dry cleaner, food court, Grocery Store, Health & Beauty Service  Health Food Store, High school, Movie Theater, Kids store, Laundry Service, Medical Center, Hospital, Pharmacy, Metro station, Shopping Mall etc.

# 3. Methodology

## 3.1. Exploratory Data Analysis

The Latitudes and Longitudes of Toronto Neighborhoods were stored in the json file nyu_2451_34572-geojson.json. I used the open feature of Python to read the file into a dictionary variable and then extracted Boroughs, Neighborhoods, and their Latitudes & Longitudes from feature section of the json file (which is now in dictionary) and then stored in Pandas so that data can be accessed easily by using excellent features of Pandas. Once that is done, used Pandas head function to display top 5 rows as below:

| | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|
| 0 | Bronx | Wakefield | 40.894705 | -73.847201 |
| 1 | Bronx | Co-op City | 40.874294 | -73.829939 |
| 2 | Bronx | Eastchester | 40.887556 | -73.827806 |
| 3 | Bronx | Fieldston | 40.895437 | -73.905643 |
| 4 | Bronx | Riverdale | 40.890834 | -73.912585 |

Since the project is to find similar neighborhoods of current Newyork neighborhood, I read the row corresponding to this neighborhood and stored it in a separate dataframe. Then called Foursquare API function 'Explore venue' with latitude/longitude of inputted neighborhood to get all the venues within reach of 500 meter, which is then stored in a dataframe. Please note that this is a regular call to Foursquare, so I was not worried at all to try multiple times for many neighborhoods just for exploration purpose.

I displayed the data to see if the API call was successful and all data were retrieved in proper format. In this case, I was using the Newyork neighborhood 'Wakefield' and it returned 8 venues nearby it.

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Wakefield | 40.894705 | -73.847201 | Lollipops Gelato | 40.894123 | -73.845892 | Dessert Shop |
| 1 | Wakefield | 40.894705 | -73.847201 | Rite Aid | 40.896649 | -73.844846 | Pharmacy |
| 2 | Wakefield | 40.894705 | -73.847201 | Carvel Ice Cream | 40.890487 | -73.848568 | Ice Cream Shop |
| 3 | Wakefield | 40.894705 | -73.847201 | Shell | 40.894187 | -73.845862 | Gas Station |
| 4 | Wakefield | 40.894705 | -73.847201 | Dunkin' | 40.890459 | -73.849089 | Donut Shop |
| 5 | Wakefield | 40.894705 | -73.847201 | SUBWAY | 40.890656 | -73.849192 | Sandwich Place |
| 6 | Wakefield | 40.894705 | -73.847201 | Cooler Runnings Jamaican Restaurant Inc | 40.898276 | -73.850381 | Caribbean Restaurant |
| 7 | Wakefield | 40.894705 | -73.847201 | Koss Quick Wash | 40.891281 | -73.849904 | Laundromat |

Then was the time to process Toronto neighborhoods. The data is not present in any file. Rather, it is stored in a Wikipedia page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
So, website scrapping features were required to extract the required data from the Wikipedia page. Luckily, since the needed data is present in the website in tabular format, and we have excellent Pandas read_html command to look for tables in the website and extract the same, I could achieve

this easily. I used head function of Pandas to display the top 5 rows of Toronto data to ensure data was extracted and stored in Pandas dataframe in proper format.

| | 0 | 1 | 2 |
|---|---|---|---|
| 1 | M1A | Not assigned | Not assigned |
| 2 | M2A | Not assigned | Not assigned |
| 3 | M3A | North York | Parkwoods |
| 4 | M4A | North York | Victoria Village |
| 5 | M5A | Downtown Toronto | Harbourfront |

As mentioned in the data cleansing section, Rows with Borough 'Not assigned' were removed, Neighborhood with value 'Not assigned' were made it same as respective Borough name and merged multiple Neighborhoods for same Borough into single row by using Pandas groupby clause. Then again used head function of Pandas to display top 5 rows and see if everything was done properly or not. Shape function was used to see count of number of neighborhoods in Toronto after data cleansing. There were 103 neighborhoods. Pandas dtype attribute was used to see datatypes of various columns. On the other hand, I used Geospatial_Coordinates.csv file to get the Latitudes/Longitudes of Toronto neighborhoods. Data were merged to store Boroughs, neighborhoods, Postal Code, Latitude and Longitudes in single dataframe as below:

Out[42]:

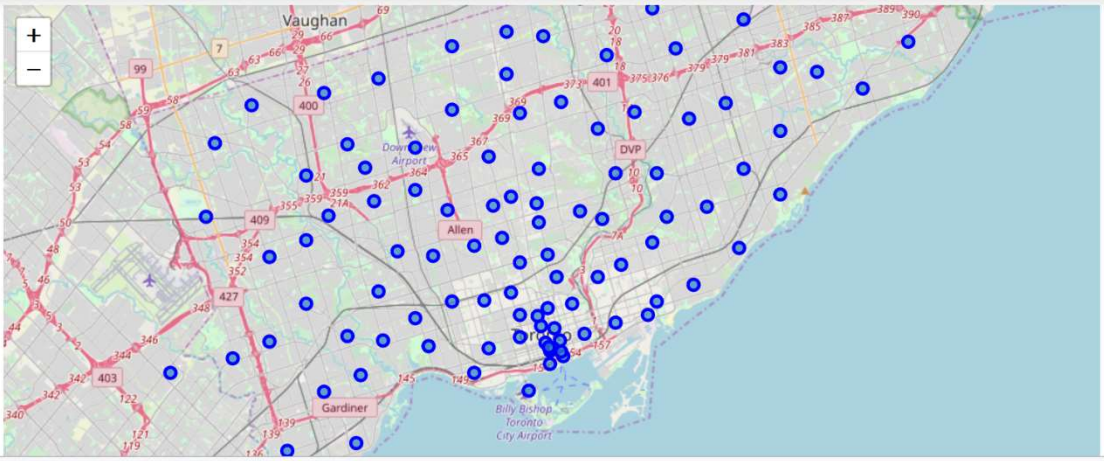
| | PostalCode | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | M1B | Scarborough | Rouge,Malvern | 43.806686 | -79.194353 |
| 1 | M1C | Scarborough | Highland Creek,Rouge Hill,Port Union | 43.784535 | -79.160497 |
| 2 | M1E | Scarborough | Guildwood,Morningside,West Hill | 43.763573 | -79.188711 |
| 3 | M1G | Scarborough | Woburn | 43.770992 | -79.216917 |
| 4 | M1H | Scarborough | Cedarbrae | 43.773136 | -79.239476 |

```
In [43]: downt_data.shape
```

```
Out[43]: (103, 5)
```

I displayed neighborhoods of Toronto using Folium map. Added popup label for each neighborhood so that when neighbor circle is clicked it pops up a text having name of the neighborhood. Zoom level was set properly so that we can view each neighborhood clearly.

Out[46]:




Once that was done, I called Foursquare API 'Explore Venue' for all the neighborhoods of Toronto and then stored in the dataframe. Sometime, this API function does not work properly, therefore I displayed the neighborhood after each call to API to ensure it was successful. Got to know that there were 2252 venues returned for all the neighborhoods of Toronto.

In [50]:
```
print(downt_venues.shape)
downt_venues.head()
```

```
(2252, 7)
```

Out[50]:

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Rouge,Malvern | 43.806686 | -79.194353 | Wendy's | 43.807448 | -79.199056 | Fast Food Restaurant |
| 1 | Highland Creek,Rouge Hill,Port Union | 43.784535 | -79.160497 | Royal Canadian Legion | 43.782533 | -79.163085 | Bar |
| 2 | Guildwood,Morningside,West Hill | 43.763573 | -79.188711 | Swiss Chalet Rotisserie & Grill | 43.767697 | -79.189914 | Pizza Place |
| 3 | Guildwood,Morningside,West Hill | 43.763573 | -79.188711 | G & G Electronics | 43.765309 | -79.191537 | Electronics Store |
| 4 | Guildwood,Morningside,West Hill | 43.763573 | -79.188711 | Big Bite Burrito | 43.766299 | -79.190720 | Mexican Restaurant |

By now, I had all the venues of Toronto neighborhoods and 1 neighborhood of Newyork(which is the current residence neighborhood in Newyork). I merged both the data into 1 dataframe. Please note that this is the main trick which will give us the result we are looking for, i.e get similar neighborhoods in the Toronto compared to current Newyork neighborhood. Used 'unique' function of Pandas to get the total number of unique venues. I found that there were 274 unique venues.

Some examples are as follows:

Park, school, College Academic Building, College Arts Building, General College & University, Restaurant, pizza, Bank, Gym, Bus station, bus stop, cafe, dry cleaner, food court, Grocery Store, Health & Beauty Service  Health Food Store, High school, Movie Theater, Kids store, Laundry Service, Medical Center, Hospital, Pharmacy, Metro station, Shopping Mall etc.

Also, displayed top 10 venues nearby each neighborhood. In order to achieve this, groupby clause of Pandas dataframe was used.

| | 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 | Adelaide,King,Richmond | Coffee Shop | Café | Bar | Sushi Restaurant | Hotel | Steakhouse | Asian Restaurant | Thai Restaurant | American Restaurant | Bakery |
| 1 | Agincourt | Sandwich Place | Lounge | Skating Rink | Breakfast Spot | Print Shop | Women's Store | Dim Sum Restaurant | Diner | Discount Store | Dog Run |
| 2 | Agincourt North,L'Amoreaux East,Milliken,Steel... | Park | Playground | Donut Shop | Dessert Shop | Dim Sum Restaurant | Diner | Discount Store | Dog Run | Doner Restaurant | Drugstore |
| 3 | Albion Gardens,Beaumond Heights,Humbergate,Jam... | Grocery Store | Fried Chicken Joint | Pharmacy | Pizza Place | Fast Food Restaurant | Beer Store | Sandwich Place | Dog Run | Dessert Shop | Dim Sum Restaurant |

## 3.2. Model development (Machine learning)

Since this type of problem is not to predict something, and it is to find similarities between the items, clustering algorithm is suitable to solve this kind of problems. It is a type of unsupervised algorithm. A cluster is a group of data points or objects in a dataset that are similar to other objects in the group, and dissimilar to datapoints in other clusters. In clustering dataset are un-labelled.

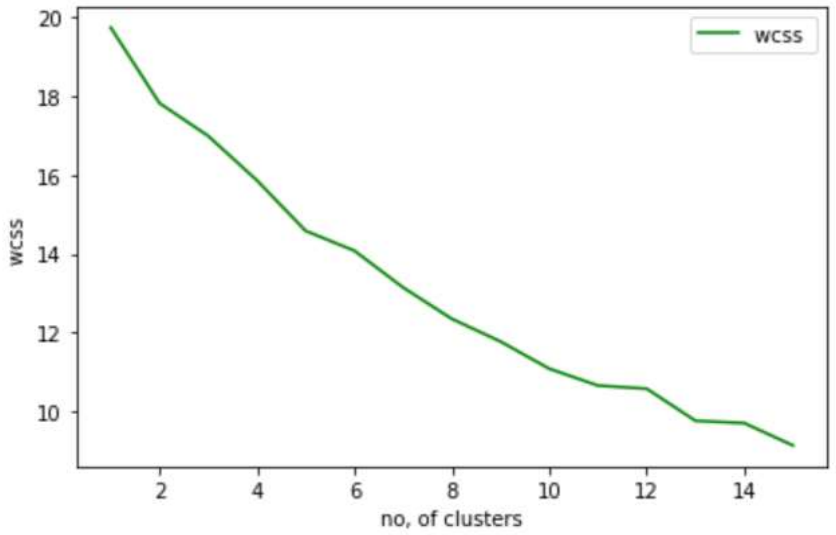
There are several types of Clustering algorithm such as K-means, Hierarchical and Density based (DBSCAN). DBSCAN is used for spatial clustering of applications with noise. Hierarchical clustering is not efficient model for such problems. Hence, I used K-means clustering algorithm to solve this problem.

Used Onehot coding feature to convert the venue category values into separate columns with values 0's and 1's. There were multiple rows for each neighborhood. So, grouped by neighborhood with count option to merge into 1 row per neighborhood with average no. of venues nearby the neighborhood.

| | Neighborhood | Yoga Studio | Accessories Store | Afghan Restaurant | Airport | Airport Food Court | Airport Lounge | Airport Service | Airport Terminal | American Restaurant | Antique Shop | Aquarium | Art Gallery | Ar Museun |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Adelaide,King,Richmond | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.030000 | 0.000000 | 0.00 | 0.010000 | 0.01000 |
| 1 | Agincourt | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.00000 |
| 2 | Agincourt North,L'Amoreaux East,Milliken,Steel... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.00000 |
| 3 | Albion Gardens,Beaumond Heights,Humbergate,Jam... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.00000 |
| 4 | Alderwood,Long Branch | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.00000 |
| 5 | Bathurst Manor,Downsview North,Wilson Heights | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.00000 |
| 6 | Bayview Village | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.00000 |
| 7 | Bedford Park,Lawrence Manor | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.040000 | 0.000000 | 0.00 | 0.000000 | 0.00000 |

### 3.3. Model Evaluation and refinement

In K-means algorithm, value of k (which is number of clusters) need to be pre-specified. Since there are around 100 neighborhoods, somewhere around 15 will be a good k-value. This value should not be too high where we might end of having only 1 or 2 neighborhoods in some clusters or should not be too low in which case too many neighborhoods might come in one cluster. Intra cluster distance (wcss) is one of the way to calculate accuracy of K-means algorithm. To ensure this is the right k-value for problem in concern, I developed the K-means algorithm for various values of k (ranging from 1 to 15) and gathered their respective error value (kmeans.inertia_). Then I plotted a line graph between k-value and error value to determine a right k-value. I found that error value is less for k=15.




## 4. Results

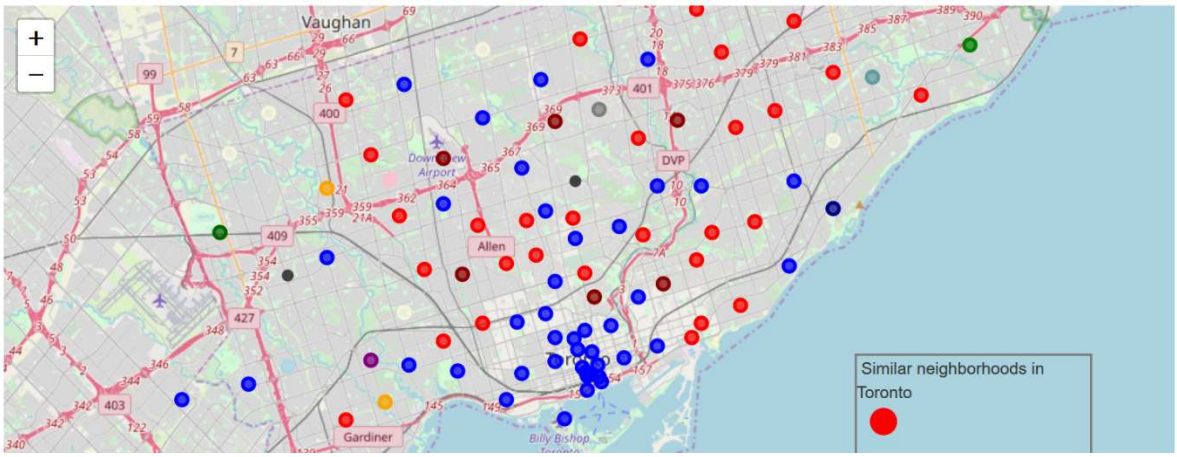
Now is the time to present the empirical findings of your data research. The clustering algorithm assigned a cluster to each neighborhood based on similarities, as shown below.

| | PostalCode | Borough | Neighborhood | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | M1B | Scarborough | Rouge,Malvern | 43.806686 | -79.194353 | 3.0 | Fast Food Restaurant | Drugstore | Dim Sum Restaurant | Diner | Discount Store |
| 1 | M1C | Scarborough | Highland Creek,Rouge Hill,Port Union | 43.784535 | -79.160497 | 6.0 | Bar | Women's Store | Drugstore | Diner | Discount Store |
| 2 | M1E | Scarborough | Guildwood,Morningside,West Hill | 43.763573 | -79.188711 | 8.0 | Electronics Store | Mexican Restaurant | Moving Target | Rental Car Location | Pizza Place |
| 3 | M1G | Scarborough | Woburn | 43.770992 | -79.216917 | 14.0 | Coffee Shop | Indian Restaurant | Korean Restaurant | Women's Store | Drugstore |
| 4 | M1H | Scarborough | Cedarbrae | 43.773136 | -79.239476 | 8.0 | Hakka Restaurant | Bakery | Bank | Athletics & Sports | Caribbean Restaurant |

If you remember, we had intentionally added a row for current neighborhoods in Newyork in the feature dataset. I found that cluster number assigned to this was 8. So. I extracted all the Toronto neighborhoods which were assigned cluster 8 and displayed them as below:

| | PostalCode | Borough | Neighborhood | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th I Com Ve |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | M1E | Scarborough | Guildwood,Morningside,West Hill | 43.763573 | -79.188711 | 8.0 | Electronics Store | Mexican Restaurant | Moving Target | Rental Car Location | Pizza Place | Brea |
| 4 | M1H | Scarborough | Cedarbrae | 43.773136 | -79.239476 | 8.0 | Hakka Restaurant | Bakery | Bank | Athletics & Sports | Caribbean Restaurant | Restau |
| 7 | M1L | Scarborough | Clairlea,Golden Mile,Oakridge | 43.711112 | -79.284577 | 8.0 | Bakery | Bus Line | Soccer Field | Bus Station | Fast Food Restaurant | Interse |
| 10 | M1P | Scarborough | Dorset Park,Scarborough Town Centre,Wexford He... | 43.757410 | -79.273304 | 8.0 | Indian Restaurant | Pet Store | Chinese Restaurant | Latin American Restaurant | Vietnamese Restaurant | Ga |
| 11 | M1R | Scarborough | Maryvale,Wexford | 43.750072 | -79.295849 | 8.0 | Auto Garage | Vietnamese Restaurant | Sandwich Place | Shopping Mall | Middle Eastern Restaurant | Brea |
| 12 | M1S | Scarborough | Agincourt | 43.794200 | -79.262029 | 8.0 | Sandwich Place | Lounge | Skating Rink | Breakfast Spot | Print Shop | Won |
| 13 | M1T | Scarborough | Clarks Corners,Sullivan,Tam O'Shanter | 43.781638 | -79.304302 | 8.0 | Pizza Place | Bank | Italian Restaurant | Rental Car Location | Noodle House | Pharr |

Best way to present the output is via map. So, I used folium map to visualize the clustering and segmentation of the neighborhoods in 'Toronto". Used different color for different cluster so that it is easy to distinguish; added popup label for each neighborhood. Added a legend to highlight which colored neighborhoods are similar to the inputted Newyork neighborhood (in this example it is red ones). So, the person can choose one among the highlighted neighborhoods to move into in Toronto.



## 5. Discussion

As said in problem statement, 1 in every 8 people relocate, so this project is going to help all of them. At least it will reduce their effort drastically by limiting scope of their search/exploration, which otherwise would have needed lot of manual effort. Roughly, around 70% of manual effort will be saved. This solution is going to help employees in choosing their career path without thinking much about the relocation from housing location perspective. All that you need to do is input a Newyork neighborhood name and the get the similar neighborhoods in Toronto in tabular as well as in map format. Isn't wonderful?

## 6. Conclusion

In this project, I picked up a real-life problem which is relocating to a unfamiliar place and highlighted challenges employees face. I limited scope of this project to Newyork to Toronto relocation. Crowd who will be interested in this project is almost 1 out of every 8 people. Determined how data can solve this problem. If the system recommends list of neighborhoods in Toronto which is similar to one's current neighborhood in Newyork then that reduces lot of manual effort (around 70%). I collected data from various sources like Foursquare, Wikipedia website and various files. Chose a best fit machine learning algorithm (K-means clustering) to solve this kind of problem. Before fitting the algorithm, I cleansed the data and did exploratory analysis. Once data was fit, I showed the desired output (list of similar neighborhoods) in tabular format as well as via folium map highlighting necessary features.

Though this project is specifically to help people who is relocating from Newyork to Toronto, the underlying program can easily be modified to make it work for any other countries/locations. That will be my next goal to achieve.

## 7. References


1. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2. Foursquare APIs
3. Geospatial_Coordinates.csv
4. nyu_2451_34572-geojson.json


**Thank you!**