

1. Introduction and Business Problem

Introduction:

People seeking better career opportunities migrate from one city to another city. It is often difficult for them to find a neighborhood which is nearby to their new work location and suitable for their family in the new city. Most of the cases people love their current neighborhood because of its amenities, venues and tourist spots, so they look for neighborhood which is more similar to their current neighborhood.

Business Problem:

Today migration from one place to another place for better career opportunities is a most common phenomenon. Such a drastic change in the neighborhood from their currently habituated neighborhood is usually difficult for an individual and his family to adjust. On a positive note every city has its own unique and diverse neighborhoods which usually people like to explore. We can find aspects in which both the cities are similar based on the information like venues, tourist spots, restaurants etc. The goal is to analyze neighborhoods in cities and find similar neighborhoods to help people migrating to new cities to live comfortably as they are doing in their current city. In my case study, A person from Toronto wants to migrate to New York seeking a better career opportunity. By, Analyzing the Current Neighborhood of that person in Toronto we should be able to suggest a Similar Neighborhood for that person in New York.

2. Data Acquisition and Data Cleaning

Data Acquisition:

1. The data on Neighborhoods and Boroughs in New York City is obtained from an open source spatial data repository. Link - https://geo.nyu.edu/catalog/nyu_2451_34572.

- Borough – Name of the Borough
- Neighborhood – Name of the Neighborhood in Borough
- Latitude – Latitude of the Neighborhood
- Longitude – Longitude of the Neighborhood

2. The data on Neighborhoods and Borough in Toronto the information is taken from the Wikipedia list. Link - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- Borough – Name of the Borough
- Neighborhood – Name of the Neighborhood in Borough

- Postal Code – Postal Code of the Neighborhood in Borough
3. The Geo Spatial Coordinates CSV of the Toronto Neighborhoods and Borough is obtained from different dataset of Latitude and Longitude of different Postal Codes.
- Postal Code – Postal Codes of the Neighborhoods
 - Latitude – Latitude of the Neighborhoods
 - Longitude – Longitude of the Neighborhoods
4. The information such as venues, restaurants, parks, hotels etc. is collected using the Foursquare API for the neighborhoods in Toronto and New York. Foursquare API provides location-based experiences with information like venues, check-ins. Foursquare API provides the venue information in the JSON format.
- name – Name of the Venue in the Neighborhood
 - categories – Type of the Venue in the Neighborhood
 - lat – Latitude of the Venue in the Neighborhood
 - lng – Longitude of the Venue in the Neighborhood

Data Cleaning:

1. The data on Neighborhoods and Boroughs in New York City is obtained in the JSON Format (See Fig 2.1). so pandas.io.json is used to transform this JSON data into Pandas Dataframe (See Fig 2.2).

```
{'type': 'FeatureCollection',
  'totalFeatures': 306,
  'features': [{'type': 'Feature',
    'id': 'nyu_2451_34572.1',
    'geometry': {'type': 'Point',
      'coordinates': [-73.84720052054902, 40.89470517661]},
    'geometry_name': 'geom',
    'properties': {'name': 'Wakefield',
      'stacked': 1,
      'annoline1': 'Wakefield',
      'annoline2': None,
      'annoline3': None,
      'annoangle': 0.0,
      'borough': 'Bronx',
      'bbox': [-73.84720052054902,
        40.89470517661,
        -73.84720052054902,
        40.89470517661]}},
  ]}
```

Figure 2.1 New York City JSON format data from Spatial Data Repository

	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

Figure 2.2 Panda DataFrame of Neighborhoods and Borough of New York

2. The data of Neighborhoods of Toronto is obtained by Web Scrapping the List of Postal Code of Canada : M using Beautiful Soap python package. Beautiful Soap is a popular python package for web scraping and parsing HTML files. The Resultant Pandas DataFrame of Neighborhood of Toronto is created from the Wikipedia lists web page. (See Fig 2.3).

	Postal code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

Figure 2.3 Pandas DataFrame of Neighborhoods and Borough of Toronto

3. The Geo Spatial Coordinates CSV is loaded into Pandas Dataframe to obtain the latitude and longitude information of Postal Code of Borough and Neighborhoods of Toronto (See Fig 2.4).

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Figure 2.4 Geo Spatial Coordinates CSV of Toronto converted to Pandas Dataframe

The Pandas Dataframe on Borough and Neighborhoods of Toronto is merged with the Geo Spatial Coordinates CSV of Toronto to obtain a single Dataframe consisting Neighborhoods and Borough of Toronto and their respective Latitude and Longitude information (See Fig 2.5).

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...	43.757410	-79.273304
11	M1R	Scarborough	Wexford, Maryvale	43.750072	-79.295849

Figure 2.5 Dataframe of Neighborhoods of Toronto with corresponding Latitude and Longitude

4. The FourSquare API venue information is obtained through API call using requests python package through FourSquare API credentials. The obtained result is JSON using pandas.io.json the JSON is transformed into a pandas Dataframe (See Fig 2.6 and Fig 2.7).

	name	categories	lat	lng
0	Rosedale Park	Playground	43.682328	-79.378934
1	Whitney Park	Park	43.682036	-79.373788
2	Alex Murray Parkette	Park	43.678300	-79.382773
3	Milkman's Lane	Trail	43.676352	-79.373842

Figure 2.7 Venues in Neighborhood and their Categories DataFame Toronto

	name	categories	lat	Ing
0	Arturo's	Pizza Place	40.874412	-73.910271
1	Bikram Yoga	Yoga Studio	40.876844	-73.906204
2	Tibbett Diner	Diner	40.880404	-73.908937
3	Starbucks	Coffee Shop	40.877531	-73.905582
4	Dunkin'	Donut Shop	40.877136	-73.906666

Figure 2.7 Venues in Neighborhood and their Categories DataFrame New York

5. The Final Dataframe after combining the FourSquare API venue data with the Dataframe of Neighborhoods and Borough of Toronto and New York Cities Separately and renaming the Columns of the resulting Dataframe { name – Venue , lat – Venue Latitude , Ing – Venue Longitude and categories – Venue Category }. The Dataframe obtained for Newyork after keeping only Manhattan Borough (See Fig. 2.8) and The Dataframe obtained for Toronto after keeping only Downtown Toronto Borough (See Fig. 2.9)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

Figure 2.8 Manhattan Borough Venues DataFrame

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rosedale	43.679563	-79.377529	Rosedale Park	43.682328	-79.378934	Playground
1	Rosedale	43.679563	-79.377529	Whitney Park	43.682036	-79.373788	Park
2	Rosedale	43.679563	-79.377529	Alex Murray Parkette	43.678300	-79.382773	Park
3	Rosedale	43.679563	-79.377529	Milkman's Lane	43.676352	-79.373842	Trail
4	St. James Town, Cabbagetown	43.667967	-79.367675	Cranberries	43.667843	-79.369407	Diner

Figure 2.9 DownTown Toronto Borough Venues DataFrame