

Best places to stay in North York, Toronto for an Immigrant to Canada

APPLIED DATA SCIENCE CAPSTONE

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

October 2019



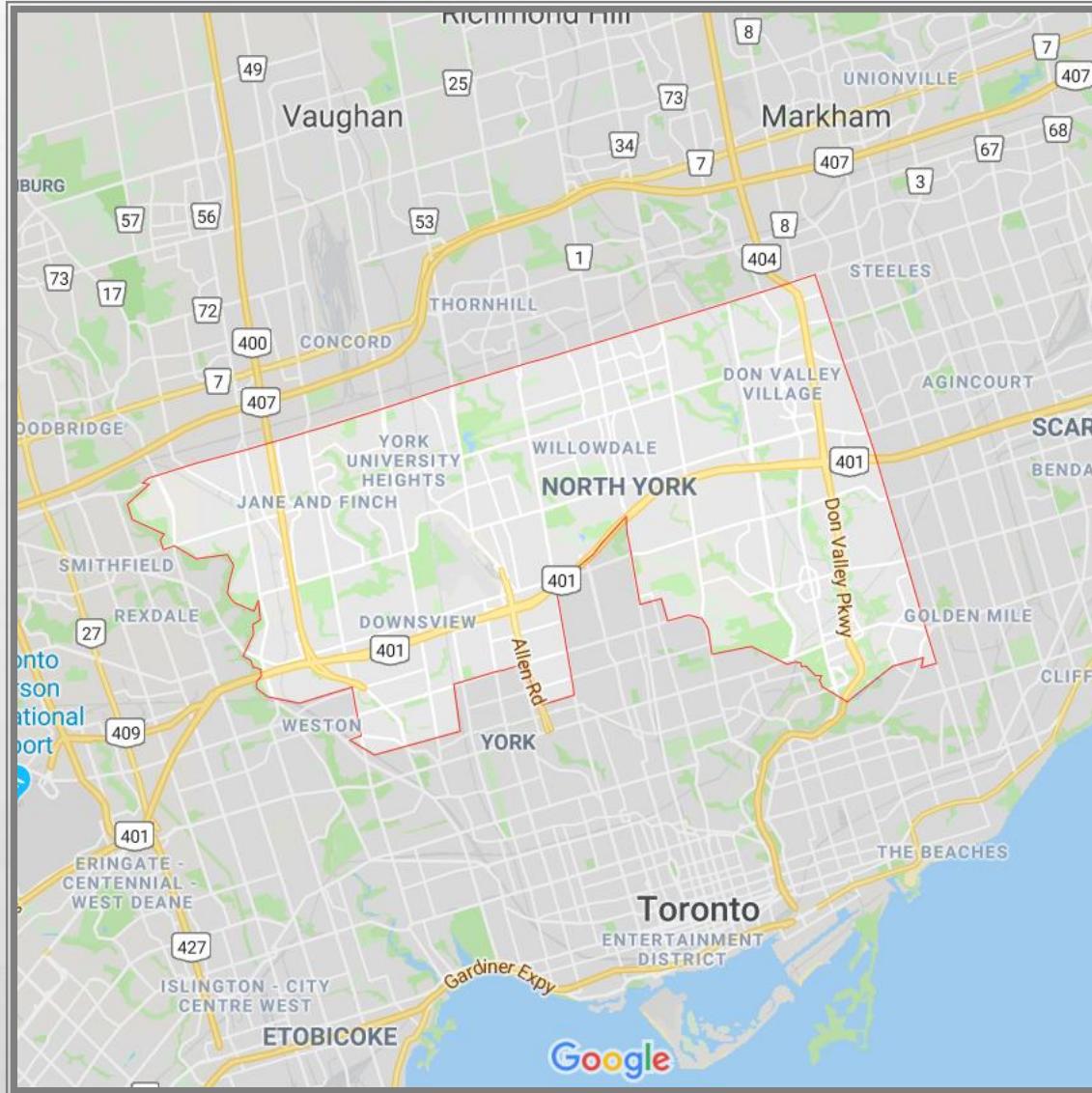
Saunak Bhattacharyya

A WORRYING IMMIGRANT FAMILY

- Toronto is a large city, where do I stay economically?
- North York looks to be a decent borough with all amenities
- But....
- Still there are so many neighbourhoods – North York is so big
- Can we know & compare the amenities of the North York neighbourhoods?
- Which neighbourhood offers quality schooling for my children?
- Lastly, which is the can I get information on the low priced housing neighbourhoods

Where is my NEW HOME?





ANALYTICS IN REAL LIFE DECISION MAKING

Selection criteria

For the purposes of this project, the definition of a good neighborhood is one that has an appreciable commercial presence within a given community as well as:

1. Compare mean housing prices
2. Compare school ratings

Tools & Data used to help our immigrant brothers & sisters

Foursquare API

- ❑ Prime data gathering source
- ❑ Provides the ability to perform location search, location sharing and details about a business

Work Flow

- ❑ Foursquare API features of near-by places of the neighborhoods would be mined
- ❑ Extract info on number of places per neighborhood parameter

Clustering Approach

- ❑ Compare and explore similar neighbourhoods in large cities like Toronto
- ❑ Cluster data which is a form of unsupervised machine learning: k-means clustering algorithm

Programming Language Used: **PYTHON**

List of Libraries Used

- ❖ Folium
- ❖ Geocoder
- ❖ Geopy
- ❖ JSON
- ❖ Matplotlib
- ❖ Pandas
- ❖ Requests
- ❖ Scikit Learn
- ❖ XML

Data we used

Longitude and latitude data:

- ❑ We will need geo-locational information about that specific borough and the neighborhoods in that borough. It is "North York" in toronto.
- ❑ Dataset comprising latitude and longitude, zip codes is already available through the previous notebook.
- ❑ The location of North York would be filtered using the same:
 - ✓ https://github.com/SaunakBhattacharyya/Coursera_Capstone/blob/master/Task%202:%20geographical%20coordinates.ipynb



North York / Coordinates

43.7615° N, 79.4111° W

Foursquare API

Connecting to Foursquare and retrieving locational data for each venue in every neighborhood.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood.

For each neighborhood, we have chosen the radius to be 100 meter.

Data Preprocessing

Processing the retrieved data and creating a Data from for all the venues inside North York

When the data is completely gathered, we processed raw data to find our desirable features for each venue.

Our main feature is the category of that venue. After this stage, the column "Venue's Category" will be One-hot encoded and different venues will have different feature-columns.

Creating a Data Frame & Hot Encoding Process

1

Processing the retrieved data and creating a data frame for all the venues inside North York

Out[21]:

	venue.name	venue.categories	venue.location.lat	venue.location.lng
0	The Captain's Boil	[{"id": "4bf58dd8d48988d1ce941735", "name": "S..."}]	43.773255	-79.413805
1	Aroma Espresso Bar	[{"id": "4bf58dd8d48988d16d941735", "name": "C..."}]	43.769449	-79.413081
2	Starbucks	[{"id": "4bf58dd8d48988d1e0931735", "name": "C..."}]	43.768353	-79.413046
3	Konjiki Ramen	[{"id": "55a59bace4b013909087cb24", "name": "R..."}]	43.766998	-79.412222
4	The Keg	[{"id": "4bf58dd8d48988d1cc941735", "name": "S..."}]	43.766579	-79.412131

2

Selection of neighbourhood is dependent on housing price & school ratings, still it is good to know the popular nearby attractions. So, listing the top nearby attraction by categories in North York

Out[23]:

Coffee Shop	6
Pizza Place	5
Ramen Restaurant	5
Bubble Tea Shop	5
Fast Food Restaurant	3
Sushi Restaurant	3
Japanese Restaurant	3
Korean Restaurant	3
Middle Eastern Restaurant	2
Café	2
Name: categories, dtype: int64	

Out[32]:

	Zoo Exhibit	Accessories Store	Afghan Restaurant	African Restaurant	Airport	American Restaurant	Animal Shelter	Antique Shop	Arcade	...
0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0

3

One Hot Encoding of categorical variables like the features to Binary values. One hot encoding is a process by which categorical variables are converted into a form that could be provided to Machine Learning algorithms to do a better job in prediction.

Applying a ML techniques (K-Means Clustering)

Finding Most Common Venues

Out[35]:

	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 M Comm Ve
0	Adelaide, King, Richmond	Coffee Shop	Café	Hotel	Steakhouse	Gastropub	Burger Joint	Break
1	Agincourt	Chinese Restaurant	Shopping Mall	Malay Restaurant	Shanghai Restaurant	Sushi Restaurant	Supermarket	Break
2	Agincourt North, L'Amoreaux East, Milliken, St...	Pharmacy	Sandwich Place	Yoga Studio	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electro S
3	Albion Gardens, Beaumont Heights, Humbergate, ...	Grocery Store	Coffee Shop	Fast Food Restaurant	Beer Store	Liquor Store	Pizza Place	Food
4	Alderwood, Long Branch	Coffee Shop	Gym	Convenience Store	Gas Station	Pub	Pizza Place	Sandv Pl

Clustering of the neighbourhood by k-means is done. At first, we need to set number of clusters.

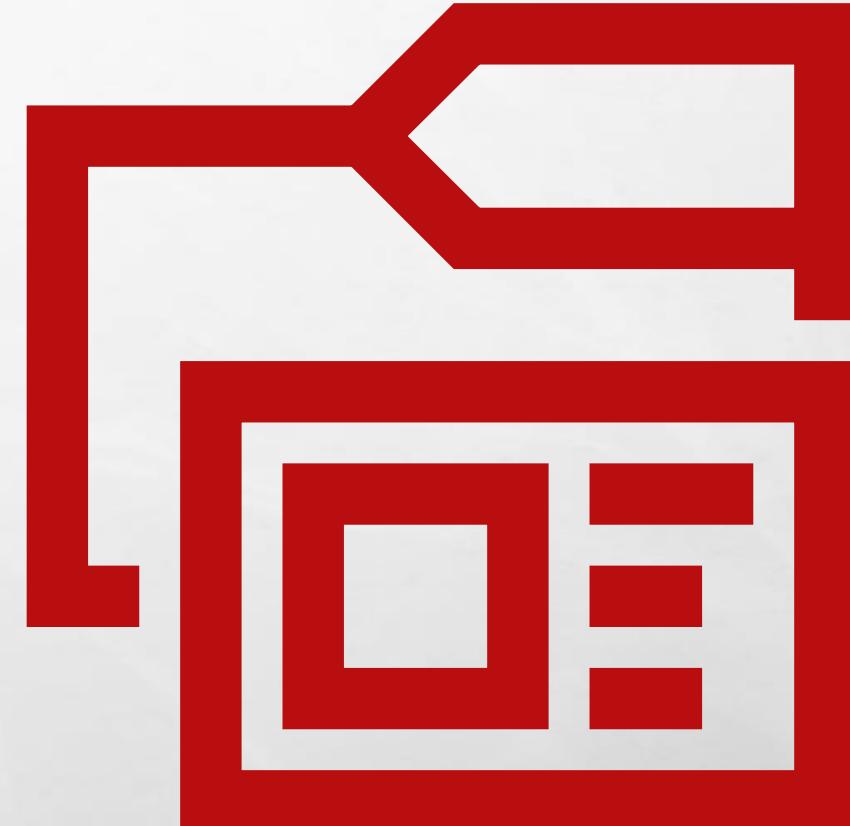
For using k-means technique, let us assign the neighborhood into 3 clusters. Then, k-means clustering is run. We check cluster labels generated for each row in the data frame.

Out[37]:

	Postalcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venu
0	M1B	Scarborough	Rouge, Malvern	43.811525	-79.195517	0	Zoo Exhibit	Business Servic
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.785730	-79.158750	0	Bar	Fish Chip Shc
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.765690	-79.175256	0	Construction & Landscaping	Athletics Spor
3	M1G	Scarborough	Woburn	43.768359	-79.217590	0	Coffee Shop	Pai
4	M1H	Scarborough	Cedarbrae	43.769688	-79.239440	0	Bakery	India Restaurai

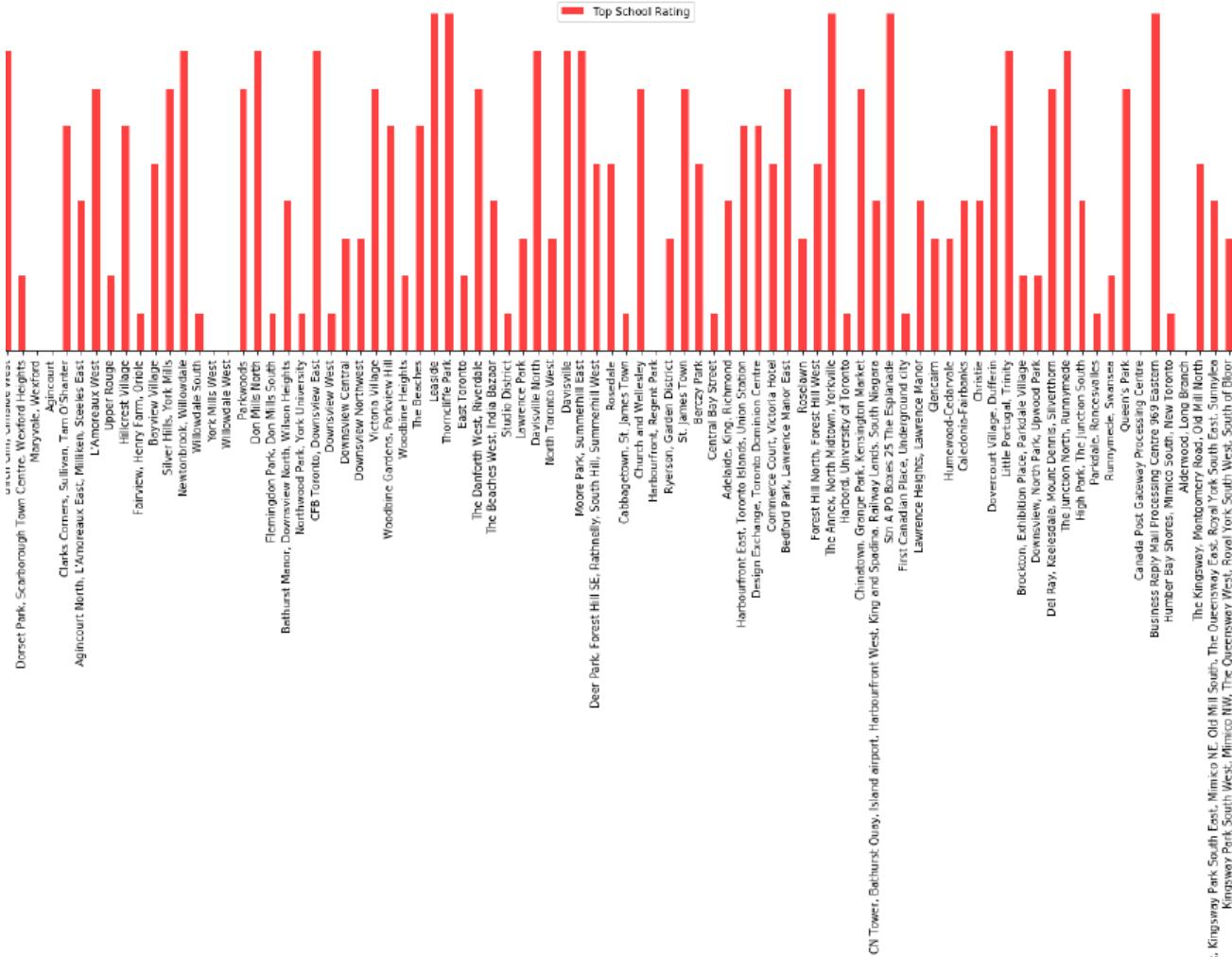
Visual analytics & Secondary research (data collection)

- Bar chart of average housing price for each of the neighbourhood were charted along with bar chart of average school ratings for each of the neighbourhood.
- Secondary web research was done to find out the average housing prices and school ratings in the neighbourhoods of Toronto.
- As mentioned in the ‘data section – as this information was not to be found in a singular website in a tabular format – various websites were referred.



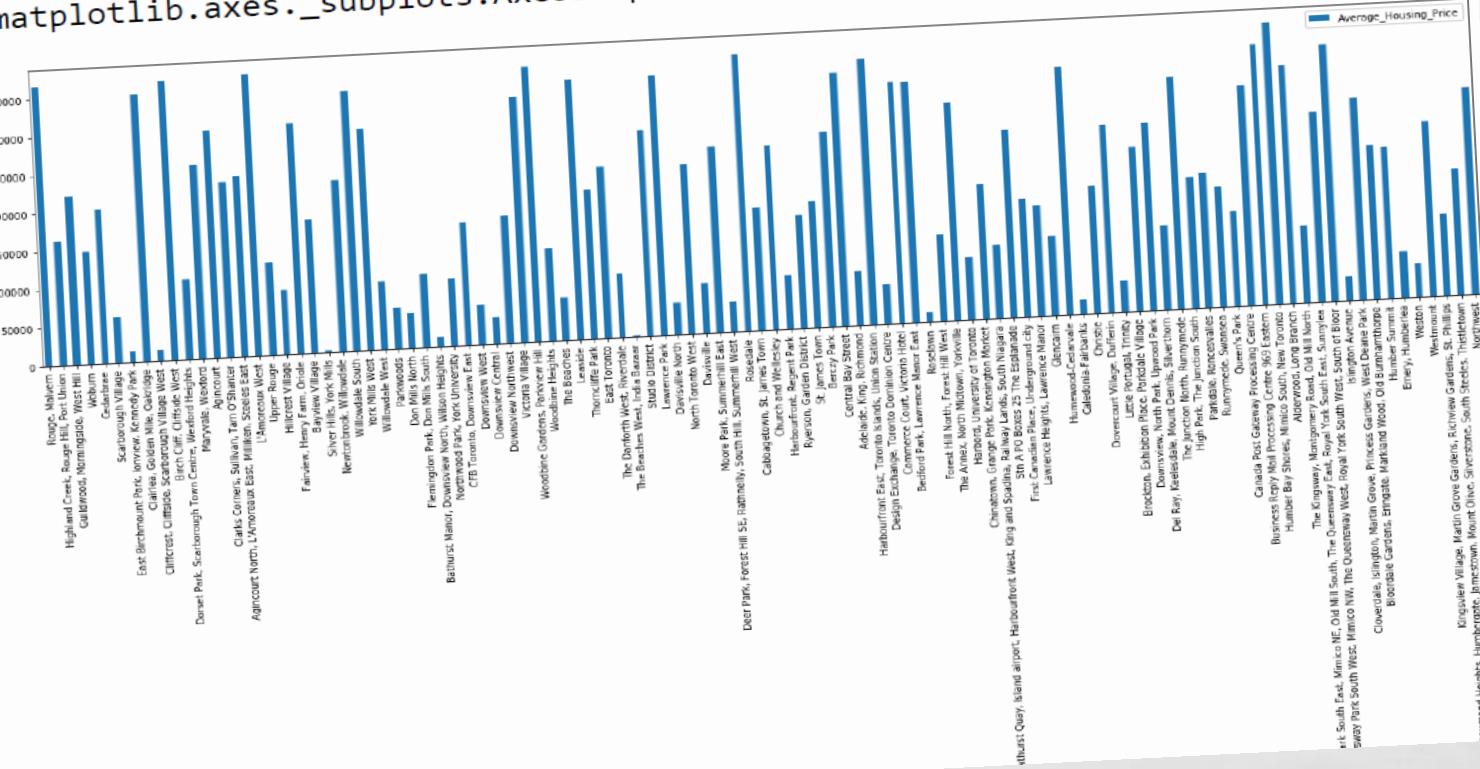
Neighborhood Average Housing Prices

lib.axes._subplots.AxesSubplot at 0x1cc0e090828>



Neighborhood Average School Rating

Out[42]: <matplotlib.axes._subplots.AxesSubplot at 0x1cc0e178400>



So where are the

probable HOME of an immigrant?

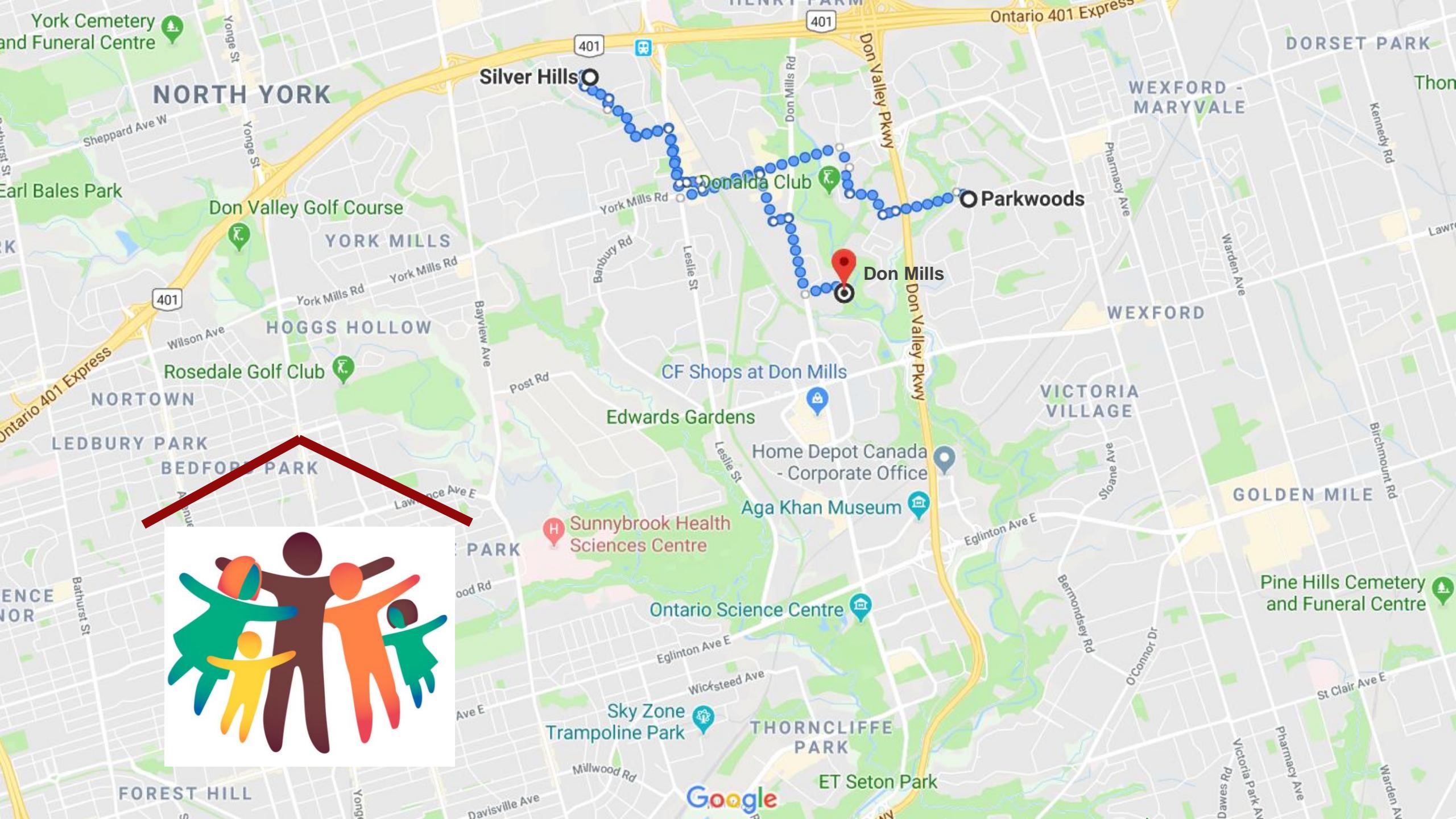
As North York as a borough occurs in multiple post codes (M2, M3, M6, etc.) it brings in the chart neighborhoods from other borough also.

So, we create a simple excel file with the 102 neighborhoods names listed above and two columns containing the average housing price and school ratings.

Using ranking method, we arrive at the probable best neighborhoods in North York, Toronto.

Out[52]:

	Neighbourhood Names	School Ratings	Housing Price Ratings	School Ratings_Rank	Ranking_avg
5	Scarborough Village	9	19.0	4.0	11.50
77	Little Portugal, Trinity	8	11.0	13.5	12.25
45	Davisville North	8	12.0	13.5	12.75
20	Silver Hills, York Mills	7	2.0	25.5	13.75
26	Don Mills North	8	14.0	13.5	13.75
47	Davisville	8	21.0	13.5	17.25
100	Kingsview Village, Martin Grove Gardens, Richv...	9	34.0	4.0	19.00
25	Parkwoods	7	17.0	25.5	21.25
41	The Danforth West, Riverdale	7	25.0	25.5	25.25
60	Design Exchange, Toronto Dominion Centre	6	16.0	35.0	25.50
37	The Beaches	6	18.0	35.0	26.50
49	Deer Park, Forest Hill SE, Rathnelly, South Hi...	5	10.0	43.5	26.75
42	The Beaches West, India Bazaar	4	1.0	55.0	28.00



THANK YOU

**COURSERA
& THE FACULTY**