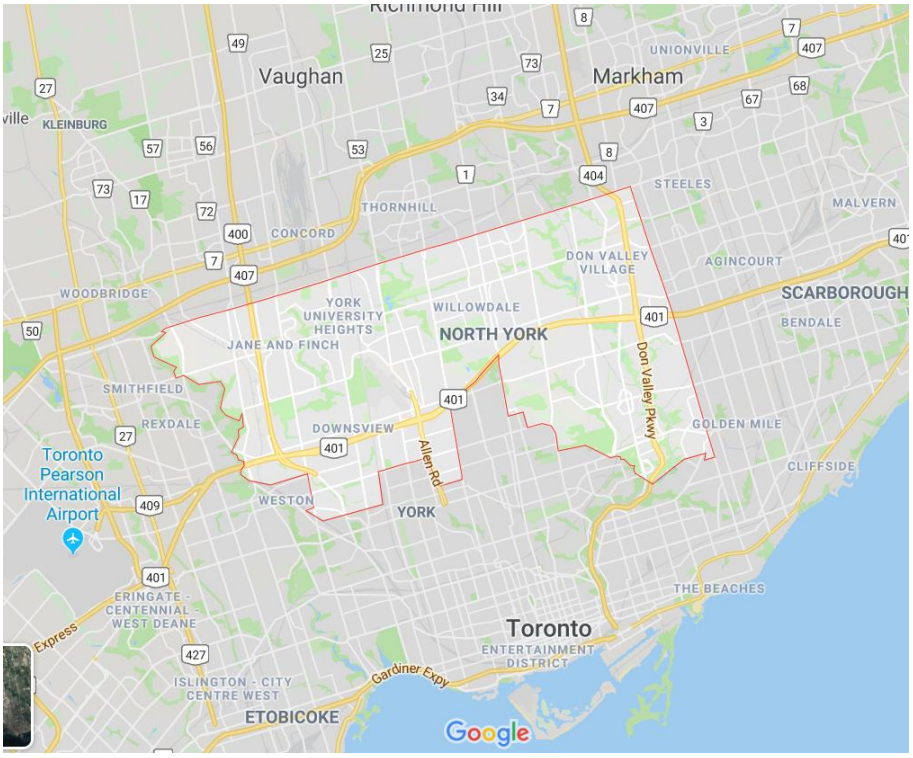
Capstone Project - Battle of Neighborhoods

North York, Toronto - Popular Neighbourhood Guidance to Immigrants

Data sets & Methodology

The Location:

North York is a popular destination for new immigrants in Canada to reside. As a result, it is one of the most diverse and multicultural areas in the Greater Toronto Area, being home to various religious groups and places of worship. Although immigration has become a hot topic over the past few years with more governments seeking more restrictions on immigrants and refugees, the general trend of immigration into Canada has been one of on the rise. Again referring to Wikipedia over the late 20th century and early 21st century, North York City Centre have emerged as secondary business districts outside Downtown Toronto. High-rise development in these areas has given the former municipalities distinguishable skylines of their own with high-density transit corridors serving them.
Thus, this projects aim to create an analysis of features for **North York neighborhood**. The features include like median house price, school ratings, population rate, crime rates, recreational facilities, etc.



We will discuss the data used in this problem in three buckets. Firstly, the geo location data, then the local search API data namely Foursquare and lastly the web research data that was fed in the Python analytical model.

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. This project will require knowledge of the different neighborhoods in Toronto, school ratings and median house prices. As such the neighborhood data required will be:

1. Neighborhood location in terms of latitude and longitude

2. School Ratings

3. Average Housing Prices

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

We get the list of neighborhoods, boroughs and postal codes from the Wikipedia website. https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto

Then for further deep dive and study of the demographics of each of the neighbourhoods, we use Wikipedia information again.

https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods

In order to establish the targeted neighborhood(s), we will explore the demographics of the neighborhoods in the city of Toronto by segmenting the data and conducting descriptive analysis using Panda. Additional data will be gleaned by web scraping and API will be used to generate data.

Foursquare API Data:

We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

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.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

Secondary Research Data:

North York as a borough occours in multiple post codes (M2, M3, M6, etc.) it brings in the chart neighbourhoods from other borough also. So we create a simple excel file with the 102 neighbourhoods names listed above and two colums containing the average housing price and school ratings. Using ranking method we arrive at the probable best neighbourhoods in North York, Toronto.

- Now we do some extensive secondary research over websites to find the average housing price for these over 100 neighbourhoods. As there is no sigle website that provides the average housing price in one place or one table. We collect information from 3 websites namely:
    - Numbeo https://www.numbeo.com/cost-ofliving/in/Toronto
    - Point2homes https://www.point2homes.com/CA/Real-Estate-Listings/ON/Toronto.html
    - Squarespace Report - https://static1.squarespace.com/static/546bbd2ae4b077803c592197/t/5c5c92ae15fcc0cc392edc40/15495707760 and collate the information on an excel sheet for further data processing and ranking
- We face the same problem with the school ratings of the neighbourhoods. There is no single website or a report that provides this information in one table or place. So we use various filters and arrive at the school ratings for the neighbourhood using Compare School Rankings website
    - http://ontario.compareschoolrankings.org/elementary/SchoolsByRankLocationName.aspx?schooltype=elementary

    Note we use only elemantary school rankings here for simplicity in decision making.