

# **CAPSTONE PROJECT – FINDING A PLACE TO LIVE IN TORONTO**

BY ANDRES NAVARRE

## 1. Introduction:

The main purpose of this Capstone Project is to serve as a tool for people that is making decisions to move around Toronto looking for a place to live.

Canada is a magnet for immigration, and its biggest city, Toronto; accommodates most of the oncoming migrants. To find a place where facilities and neighbourhoods respond for this requirement is a real challenge. The main goal for this Capstone Project is to display the facilities of neighbourhoods, including schools, super markets, medical clinics, groceries and others in order to provide a decision tool for people to decide where they consider they want to move in.

This Project aims to create an analysis of attributes for a people looking for a place to live, to search the best neighbourhood that responds to their needs. The features that are going to be included in the analysis are: average income, and school ratings, crime rates, recreational facilities(parks, etc), road connectivity, weather conditions, emergency management, water resources.

With this analysis of attributes, people will get a sense of what are they going to find before the move into the neighbourhood, compared with the rest of the city.

## 2. Data Section

Data Link: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)  
and

[https://www.toronto.ca/ext/open\\_data/catalog/data\\_set\\_files/  
2016\\_neighborhood\\_profiles.csv](https://www.toronto.ca/ext/open_data/catalog/data_set_files/2016_neighborhood_profiles.csv)

We will use Toronto's dataset, which we scrapped from wikipedia on Week 3. Dataset includes latitude, longitude and Postal codes.

At the same time we are merging the dataset with the neighbourhood profiles.

Foursquare API Data:

We will need data about different places within the neighbourhoods.

In order to gain that information, we will use "Foursquare" locational information. The data retrieved from Foursquare contained information of place within a specified distance of the longitude and latitude of the postcodes.

The information includes venue names, locations, menus and photos. The foursquare platform will be used as the only data source at the benefit of their API.

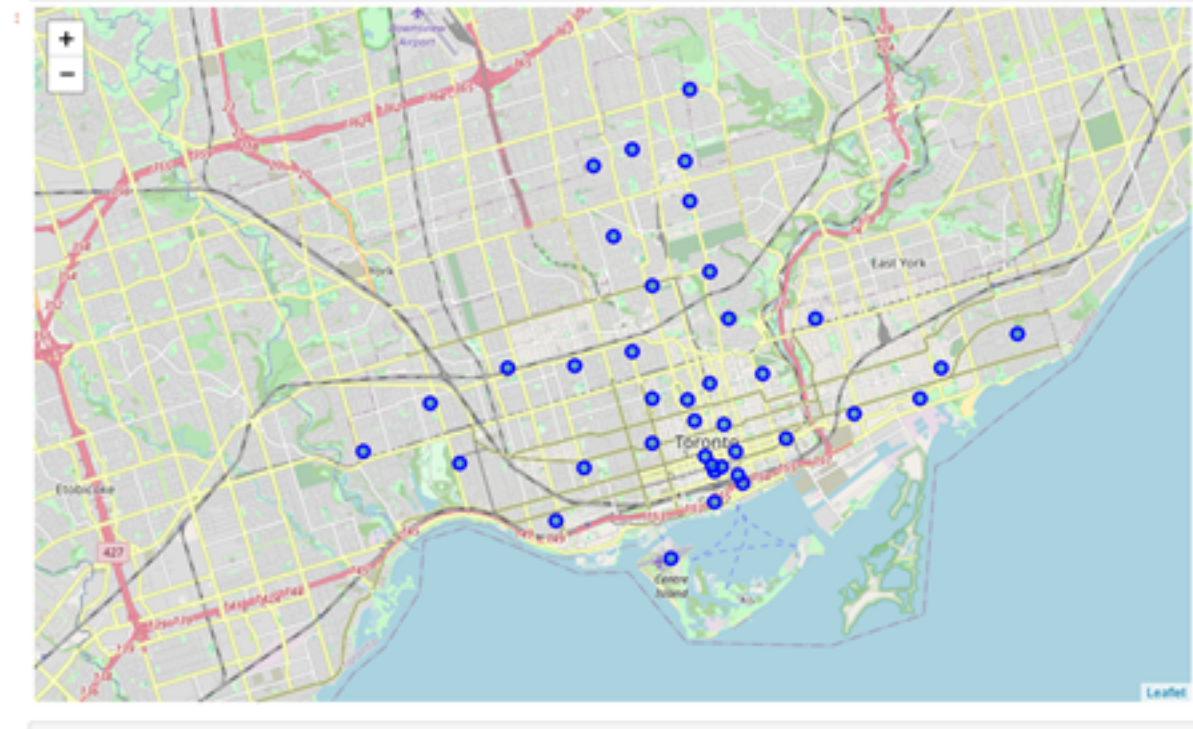
After we find the list of neighbourhoods, we can connect to the Foursquare API to gather the information of the places within each neighbourhood. In order to refine our search, we have chosen a radius of 200 meters.

The information per venue is as follows:

- 1 Neighbourhood
- 2 Neighbourhood latitude
- 3 Neighbourhood longitude
- 4 Venue
- 5 Name of the Venue

- 6 Venue Latitude
- 7 Venue Longitude
- 8 Venue Category

## Map of Toronto



## 3. Methodology Section

### Defining Clustering:

We are going to use a data cluster methodology which is a unsupervised machine k-means cluster algorithm. With the methodology we can compare cities and segment them into groups by its similarities.

By using Foursquare API features, near-by places of the neighbourhoods would be mined. Due to connectivity of request limitations, the number of places per neighbourhood parameter would reasonably be set to 100 and the radius parameter would be set to 200.

### Location:

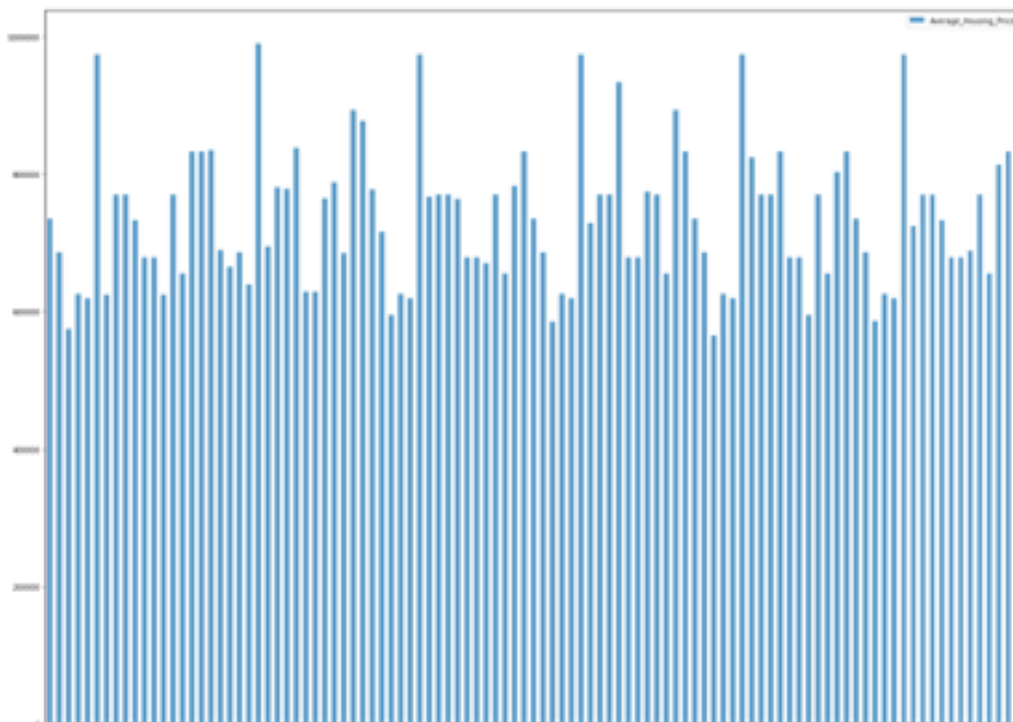
Toronto is a popular destination for new immigrants in Canada to reside. As a result, it is one of the most diverse and multicultural areas of Canada. Immigration trend continue to be on the rise so more people is arriving into the country.

### Foursquare API:

This project have used Foursquare API as its first data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

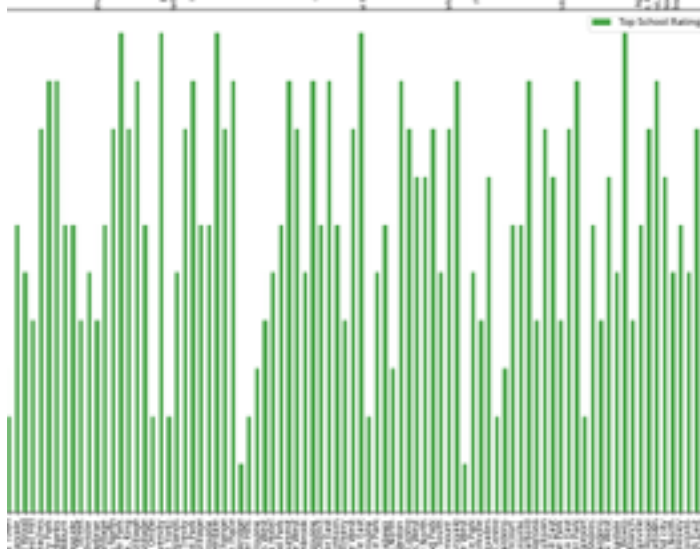
## 6. Estimated results

In this project, we used different methodologies to bring data analysis to people that wants to move to Toronto.



The main problem to solve on this project, is to suggest a neighbourhood to live in that respond to the needs of people in a city like Toronto. Houses affordability and School rankings are elements that could be listed and analyzed.

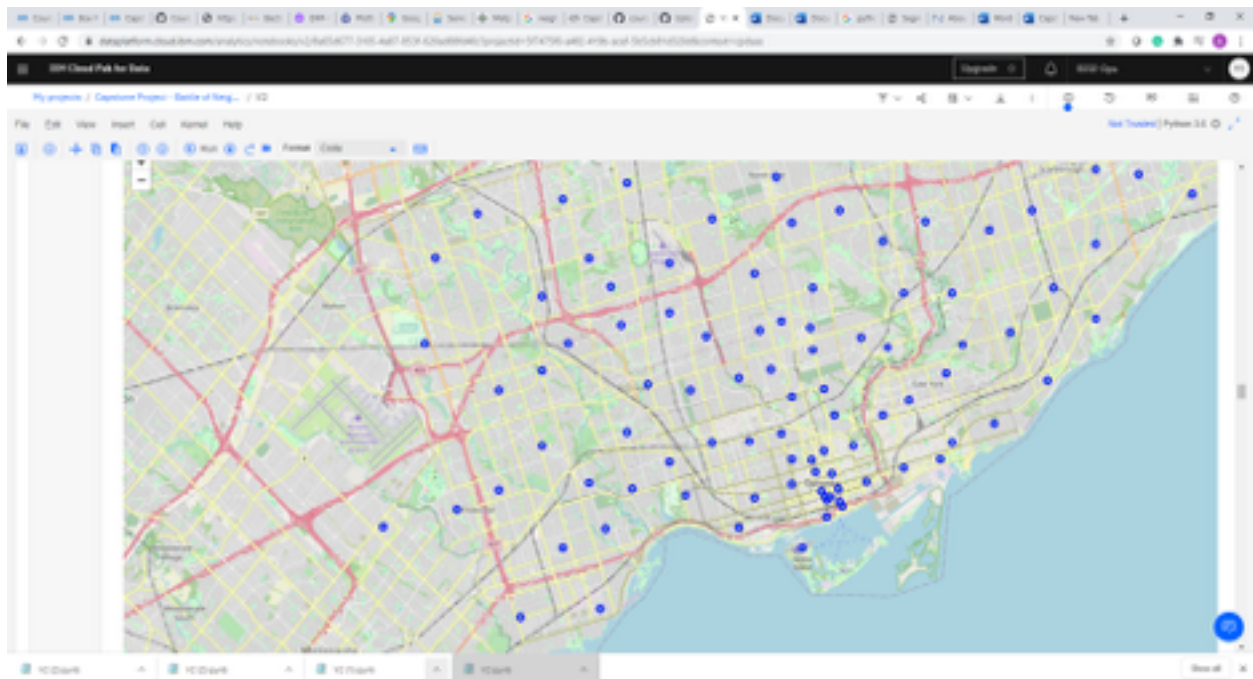
In terms of schools, there is an interest of new migrants in register their children in the best schools possible. In Canada, School depends in the location of the house, so, it seems to be a relative correlation between house prices and ranking of schools.



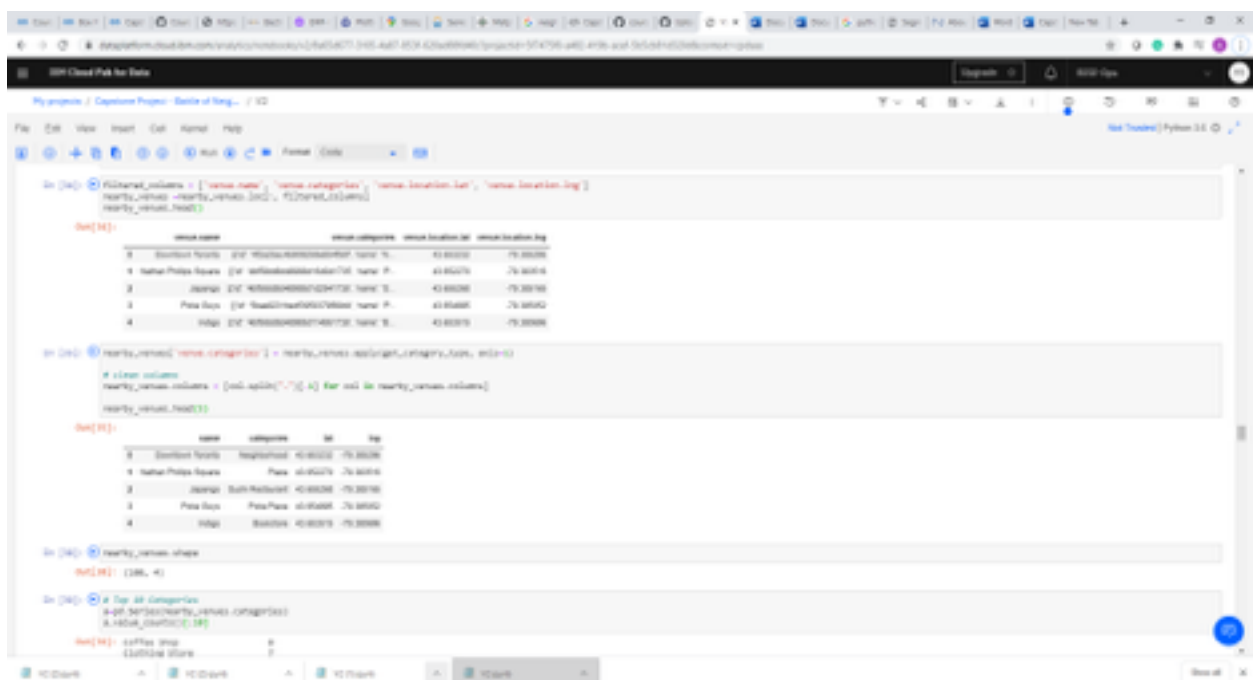
The theory suggest that there is a correlation between house prices and school rankings. We weren't able to find that correlation in our analysis and this can be

explained in our database, which seems to be from different years. so it is recommended to bring more updated data especially in house prices as they seem to be a more accurate response to the reality of the market.

The result of clusters is as follows. First we plotted the neighbourhoods and use a non supervised model in order to include a clustering algorithm.

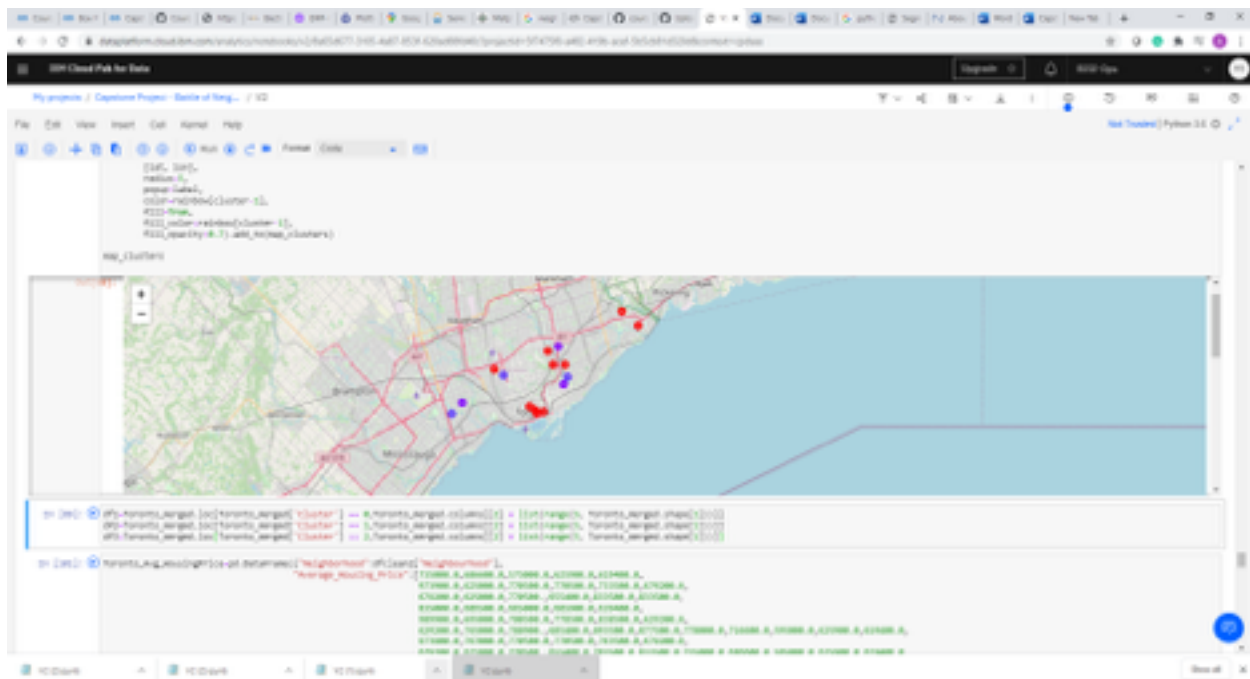


We included the k-means cluster algorithm, separating neighbourhood into different clusters (initially estimated in 15) and for 103 different latitude and longitude from dataset. Schools ratings and House prices were included.



[illegible][illegible]

Then, we use colours to indicate the clusters ( in terms of basket of services) that different neighbourhoods have and provide to households. We plotted a graph like this one.



So, there are different attributes in which new immigrants can solve their needs and at the end of the day, it is going to be, their decision.

6.

## Discussion Section

The information of the House Prices were a little old as I wasn't able to find more updated data. Nevertheless, as prices in Toronto continue to raise, it is assumed that using average prices will keep a homogeneous price structure in between neighbourhoods.

Rentals were excluded of the analysis. The main reason is that rent has a different evolutionary trajectory in time, and I wasn't able to exclude rentals that were provided by social housing as the data base didn't include that information.

## Problem to Solve:

The main purpose of this project, is to suggest a neighbourhood to live in that respond to the needs of people in a city like Toronto. Elements to be considered could be Airport accessibility, Public Transportation, Downtown access, market access, and others. Houses affordability and School rankings are elements that could be listed and analyzed.

Another aspect is that, even that it is possible to find differences between neighbourhoods, there are neighbourhoods that could strong variations within the neighbourhood and some others than don't have it, and present a more homogeneous structure.

#### 7. Recommendations:

The analysis excludes a very important piece of the puzzle, which is working places. It is extremely relevant where people works and which mean of transportation is using to move. Living a 1 hour drive of the work place is could be challenging for a new migrant and not a having a vehicle even more. These are elements that should be incorporated in the model.

Another element to include is Social Services and Libraries. I didn't include it in the analysis but it could also be needed as new migrants need a lot of support in their first years of living in a city.

For some migrants with children of certain age, finding college or higher education is also relevant. This was not included in the analysis and could be of value.

#### Libraries used in the Project:

Pandas: For creating and manipulating data frames.

Folium: Python visualization library would be used to visualize the neighbourhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.