

# Capstone Project - The Battle of Neighbourhoods

## Introduction

Toronto is the provincial capital of Ontario. With a recorded population of 2,731,571 in 2016, it is the most populous city in Canada and the fourth most populous city in North America. The Greater Toronto Area (GTA) as a whole had a 2016 population of 6,417,516. The city covers an area of 630.20 square kilometres (243.32 sq mi) and comprises six districts – East York, Etobicoke, North York, Old Toronto, Scarborough and York – which were amalgamated to form Toronto's present boundaries in 1998. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. More than 50 percent of residents belong to a visible minority population group, and over 200 distinct ethnic origins are represented among its inhabitants. While the majority of Torontonians speak English as their primary language, over 160 languages are spoken in the city. .

## Business Problem

With its excellent business opportunities and diverse cosmopolitan culture, Toronto attract a lot of expats who come to the city for employment. Any new expat to the city would be looking to rent an apartment for living and staying.

A person looking to rent an apartment focusses on few things like rent amount, transportation, stores, entertainment options etc. So, as part of this project, we will list and visualize all major apartments/condos in the city of Toronto that are available for rent. We will cluster the neighbourhoods in order to recommend venues so that a prospective tenant can take an informed decision.

## Assumptions

1. The person is looking for 2-bedroom apartments only.
2. The apartment should be close to public transport.
3. There should be grocery and dining options available near the apartment.

## Data

For this project we need the following data:

- RentCafe data – This is one of the popular websites that lists apartments available for rent.
  - Data source : <https://www.rentcafe.com/apartments-for-rent/ca/on/toronto/?Beds=TwoPlus&PriceCategory=Luxury&OrderBy=RentDesc&page=2>
  - Description : This webpage will provide information regarding the available apartments that fit the selection criteria. We will scrape the webpage to get the required information.
- Toronto City data that lists registered apartments along with site address and various parameters e.g. safety features, amenities, age of building etc.
  - Data source : <https://open.toronto.ca/dataset/apartment-building-registration/>
  - Description : This data set contains the required information about the apartments. We will use this data set to get several important information about the apartments. Some of this data may be available in the RentCafe site, however the data from the 'City of Toronto' will be more authentic as it is a government site.
- Neighbourhood data
  - Data source: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
  - Description – This wiki page will provide us with the neighbourhood corresponding to each postal code where the apartments are located.
- Venues in each neighbourhood of Toronto city where the apartments are located.
  - Data source : Fousquare API
  - Description : By using this api we will get all the venues in each neighborhood.
- GeoSpace data
  - Data source :
    - a. [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)
    - b. geospatial\_data.csv
  - Description : By using either of the above geospatial data we will get the latitude/longitude details of the neighbourhoods.

## Solution Approach

- Collect the below data:
    - a. Toronto city neighbourhood data from the Wikipedia page.
    - b. Apartment data from the RentCafe site and City of Toronto website
- Pre-process the dataCollect

- c. Neighbourhood data from the Wikipedia page.
  - d. Geospatial data.
  - e. Use FourSquare API to find all venues for each neighbourhood.
- Pre-process the above data
  - a. Use BeautifulSoup to scrape the wiki data.
  - b. Format above data to remove headers, NaN etc and populate it into a pandas dataframe.
  - c. Add column names to the dataframe by concatenating the header data.
  - d. Remove boroughs that are 'Not Assigned'
  - e. Use BeautifulSoup to scrape the RentCafe data.
  - f. Get apartment name, street addresses and price.
  - g. Get the City of Toronto data that has details around safety features for each apartment.
  - h. Use the geo packages to get the geospatial data that provides the latitude, longitude details for each postal code. Then merge this data with the neighbourhood, RentCafe and City of Toronto data, so that we have the latitude/longitude for each neighbourhood.
- Use folium and geopy to create the map of Toronto
- Define Foursquare credentials to explore the data for each neighbourhood.
- Find different venues for each neighbourhood and list 10 most common venues.
- Find ratings, tips and like count for each venue using FourSquare API.
- Find ratings of apartment buildings.
- Include apartments that are close to public transport, has at least 2 bedrooms.
- Show the average rent price of each apartment.
- Visualize the different neighbourhoods using folium library(python).
- Use K-Means algorithm (k=5) to cluster the neighbourhoods into 5 clusters.

## **Libraries/Packages**

We will import the required libraries for python.

- pandas and numpy for handling data.
- request module for using FourSquare API.
- geopy to get co-ordinates of City of Toronto.
- BeautifulSoup for Web Scraping.
- folium to visualize the results on a map

## **Benefits of the above solution**

The solution above will provide answers to the below questions that every prospective tenant looks for when searching for an apartment.

- What is the average rental price of different apartments and in different neighbourhoods?
- What are the ratings for each apartment?
- What are the entertainment options available in each apartment neighbourhood?
- What are the public transportation options available in each apartment neighbourhood?
- What are the safety features available for each apartment building?