Coursera Capstone Final Project

Teri Nguyen\_Shank

**Battle of the Neighborhoods of Toronto**

**Introduction**

Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 6,197,000 in 2020, it is the most populous city in Canada. It is an international center of business, finance, arts, and culture. It takes a spot in the top ten list of the “World’s Most Economically Powerful Cities” according to Forbes Magazine. Toronto attracts investment due to the size of their economy, their projected future prosperity, and cost of living and quality of life. It also is a lively place when it comes to entertainment, culture, and food. The cities are filled with friendly people. You can find world-class theater, shopping, and restaurants here.

Where to live in Toronto? Depends on your preferences and lifestyle, finding a good place to live can be one of the most daunting tasks when moving to Toronto. Suburbs such as Scarborough, North York, and a few others are favored by families and are among the best places to live in Toronto. While many visitors shirk Scarborough as too suburban, it tends to be a popular location for new immigrants to settle because of its highly affordable part of town with some of the cheapest rents and dining options. North York it is one of the six administrative districts of Toronto. It encompasses most of uptown Toronto with its largely residentials and many micro-neighborhoods of its own.

This project offers a brief analysis of the two different Toronto’s datasets, aimed at giving you an idea about the boroughs, and the neighborhoods. It will then explore two most dense boroughs - North York/Downtown Toronto and Scarborough and their trending venues.

**Data Overview**

The data that will be used for this project are collected from multiple sources as Wikipedia, GeoSpatial Data set that has been prepared as s CSV file, and Foursquare API.

1. List of postal codes of Canada: M from Wikipedia: This is a list of postal codes in Canada that is organized in a form of a table where the first letter is M.

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

1. Geographical coordinates of each postal code from a csv file are ready to be extracted.

GeoSpatial\_Coordinate.CSV

1. Foursquare API: Foursquare location data

[https://api.foursquare.com/v2/venues/search?client\_id={}&client\_secret={}&ll={},{}&oauth\_token={}&v={}&query={}&radius={}&limit](https://api.foursquare.com/v2/venues/search?client_id=%7b%7d&client_secret=%7b%7d&ll=%7b%7d,%7b%7d&oauth_token=%7b%7d&v=%7b%7d&query=%7b%7d&radius=%7b%7d&limit)

1. Toronto information

<https://en.wikipedia.org/wiki/Toronto>

**Methodology**

1. Explore and segment the boroughs in the city of Toronto based on the postal code and borough information.

* Download and Explore Dataset
* Web scrapping: scraping the Wikipedia page using BeautifulSoup package and wrangle the data, clean it, and then read it into a structural format pandas data frame with 3 columns as bellow.

Table

Description automatically generated

* Download csv file using wget command and access data to create the following 3 columns data frame.

Table

Description automatically generated

* Merge the two above data frames together into one with 5 columns.

Table

Description automatically generated

* Once the data is in a structured format, we analyze it.
* Toronto has a total of 15 boroughs and 103 neighborhoods. To segment the neighborhoods and explore them, we create a dataset that contains the 15 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. Then we use the visualization libraries like Folium and Matplot to visualize the results.
* Using geopy library to get the latitude and longitude values of Toronto and then create map of Toronto with Folium using those values.

**Map

Description automatically generated**

Blue dots represent boroughs and neighborhoods.

* Using matplot library to plot the boroughs with its neighborhoods on a bar graph.

**Chart

Description automatically generated**

* According to the graph above the analysis is then narrowed down to two 2 boroughs -Scarborough and North York - due to the highest count of neighborhoods.

1. Define Foursquare API criteria and create get venues functions

The idea is to leverage the Foursquare location data to explore different neighborhoods within two boroughs to come up with different sets of venues within a radius of 500 meters.

* Define limit and radius

LIMIT = 100 # limit of number of venues returned by Foursquare API

radius = 500 # define radius

* Define Foursquare Credentials and Version
* Create 2 functions to repeat the same process to all the boroughs and neighborhoods in Toronto.

get\_category\_type(): extracts the category of the venues.

getNearbyVenues(): extracts the nearby venues.

1. Explore Scarborough.

* Create a new data frame of the Scarborough data.
* Get the latitude and longitude coordinates of Scarborough using geopy library.
* Map Scarborough with 17 neighborhoods

Map

Description automatically generated

Blue dots represent neighborhoods.

* Define the corresponding URL, define limit and radius, and get request to Foursquare for venues categories.

LIMIT = 100 # limit of number of venues returned by Foursquare API

radius = 500 # define radius

* Using get\_category\_type() to get venues and their categories in Scarborough and then clean and filter venues into a data frame with 4 columns: Postal Code, Boroughs, Neighborhoods, Latitudes, and Longitudes

Table

Description automatically generated

Foursquare returned 39 venues

* Using getNearbyVenues() to extract number of venues in Scarborough then use groupby function to filter and group venues into a list of unique venues returned by Foursquare.

89 venues were returned resulting in 56 unique categories.

* Visualize Scarborough’s venues on the map using Folium

A picture containing text, map

Description automatically generated

* Analyze these neighborhoods to get the top common venues.

A picture containing text, receipt

Description automatically generated

Analyze the neighborhoods to get the top common 10 venues using a machine learning algorithm called One hot coding.

* Explore the first neighborhood in Scarborough and its venues – Malvern/Rouge.

Graphical user interface, text, application

Description automatically generated

* Send a request to the API to search for a particular venue – Vietnamese restaurant - within 500 meters radius of Scarborough.



* Get the venue's overall rating. Make a call to the Foursquare API with venue ID listed above to get the ratings.

Pho Vietnam has a rating of 5.9 while Banh Mi Metro has none.

1. Explore North York

We do the same process and data analysis as the above borough.

* Data frame of North York along with Postal code, Neighborhood, Latitude, Longitude
* Map North York with 24 neighborhoods according to its latitude and longitude
* Define the corresponding URL and get request.
* Repeat the process of calling functions get\_category\_type() and getNearbyVenues(): to extract categories and venues then filter and group venues into a list of unique venues returned by Foursquare.

Table

Description automatically generated

Map

Description automatically generated

Blue dots represent neighborhoods.

Table

Description automatically generated

The above result of 244 venues group into a list of 101 unique categories

Text, table

Description automatically generated

Analyze the neighborhoods to get the top common 10 venues using a machine learning algorithm called One hot coding.

* Explore the first neighborhood in North York and its venues – Fairview / Henry Farm / Oriole
* Send a request to the API to search for a specific type of venues – Vietnamese restaurant - in North York within 500 meters radius.



* Get the venue's overall rating. Make a call to the Foursquare API with venue ID listed above to get their rating.

Pho 88 has a rating of 6.1 and Pho Lac Vien has a rating of 7.6 while Pho Lien Huong has none.

1. Explore Downtown Toronto

We do the same process and data analysis as the above boroughs.

* Data frame of Downtown Toronto along with Postal code, Neighborhood, Latitude, and Longitude
* Map Downtown Toronto with its according latitude and longitude
* Define the corresponding URL and get request.
* Repeat the process of calling functions get\_category\_type() and getNearbyVenues(): to extract categories and venues then filter and group venues into a list of unique venues returned by Foursquare.

Table

Description automatically generated

Map

Description automatically generated

Blue dots represent the neighborhoods.



* Foursquare finally returned 202 unique categories

A picture containing text, receipt, screenshot

Description automatically generated

Analyze the neighborhoods to get the top common 10 venues using a machine learning algorithm called One hot coding.

* Send a request to the API to search for a specific type of venues – Vietnamese restaurant - in Downtown Toronto within 500 meters radius. Banh Mi Nguyen Hung has the top rating of 8.9 followed by Asian Bowl: 6.8 and Pho Tien: 6.1

Graphical user interface, application

Description automatically generated

**Results**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Scarborough | North York | Downtown Toronto |
| Neighborhoods | 17 | 24 | 17 |
| Total Venues | 89 | 244 | 1102 |
| Unique Categories | 56 | 101 | 202 |
| Vietnamese restaurants | 1 | 3 | 4 |
| Rating | 5.9 | 7.6, 6.1, and 0 | 8.9, 6.8, and 6.1 |

According to my data analysis and the above table, Downtown Toronto has the most venues and options for all types of cuisine, but since North York surrounds most of uptown Toronto, it is the best borough between the 2.

**Discussion/Conclusion**

Toronto is one of the world’s most multicultural cities and famous for a diverse choice of food. Smaller suburbs surround the large cities like Scarborough’s neighborhoods offering options for people who prefer to step back from big city life while uptown neighborhoods in North York offers a wider selection of necessities. The two boroughs are diverse and culturally unique, varying from a bohemian hangout to a foodie paradise. The objective of this analysis is to help newcomers compare the 2 boroughs, within the boroughs segmented into each neighborhood to discover the top trending venues, and finally narrow it down to one cuisine.