Capstone Project Report
The Battle of Neighborhoods

1. Introduction

A group of friends decide to travel to Toronto for their vacation and want to have only Italian or Chinese cuisine when they visit, so they would like to stay in a neighborhood with a high density of Italian or Chinese restaurants with great ratings. Toronto is one of the busiest cities and it is the leading travel destination of Canada with approximately 30 million visitors every year. Looking for places to eat while travelling is a tedious procedure and becomes harder when the visitors have preferences for a certain kind of cuisine.

i) Business Problem

The problem the project aims to solve is to list and visualize all major parts of Toronto that have amazing Italian or Chinese restaurants. The neighborhood recommended should have a high average rating of restaurants to ensure a better experience for the visitors. Effectively the objective is to build a restaurant recommender for each neighborhood.

ii) Target Audience

This would be a relevant challenge with valid questions for anyone looking for a specific kind of cuisine when visiting Toronto or any other city. This case is also applicable for anyone interested in exploring, starting or locating a restaurant in any city. The analysis will be helpful for people to discover hidden trends and draw insights about the restaurant demographic of a given city.

2. Data

- A list of neighborhoods of Toronto
  - Source : https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
  - Description : This contains a list of all neighborhoods of Toronto with the Borough and Postal Code information.
- Latitude and Longitude data of the neighborhoods
  - Source : https://cocl.us/Geospatial_data.
  - Description : This contains the map coordinates of all neighborhoods of Toronto with the Borough names.
- A list of venues of Italian and Chinese restaurants in Toronto
  - Source : FoursquareAPI.
  - Description : This should contain the restaurants which serve italian or chinese cuisine.

3. Methodology

Techniques such as web scraping are used to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then the geographical coordinates of the neighbourhoods are obtained using the Python Geocoder package which will give the latitude and longitude coordinates of the neighbourhoods.The example data collected are shown below

	Borough	Neighborhood	Latitude	Longitude
0	North York	Parkwoods	43.753259	-79.329656
1	North York	Victoria Village	43.725882	-79.315572
2	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

Fig 1. A list of neighborhoods of Toronto

	Postal Code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

Fig 2. Latitude and Longitude of the neighborhoods

The number of neighborhoods per borough in Toronto is plotted and shown is Fig 3. From the plot, it is observed that the borough with the most number of neighborhoods is North York with 24 neighborhoods and the borough with the least number of neighborhoods is Mississauga with 1 neighborhood.

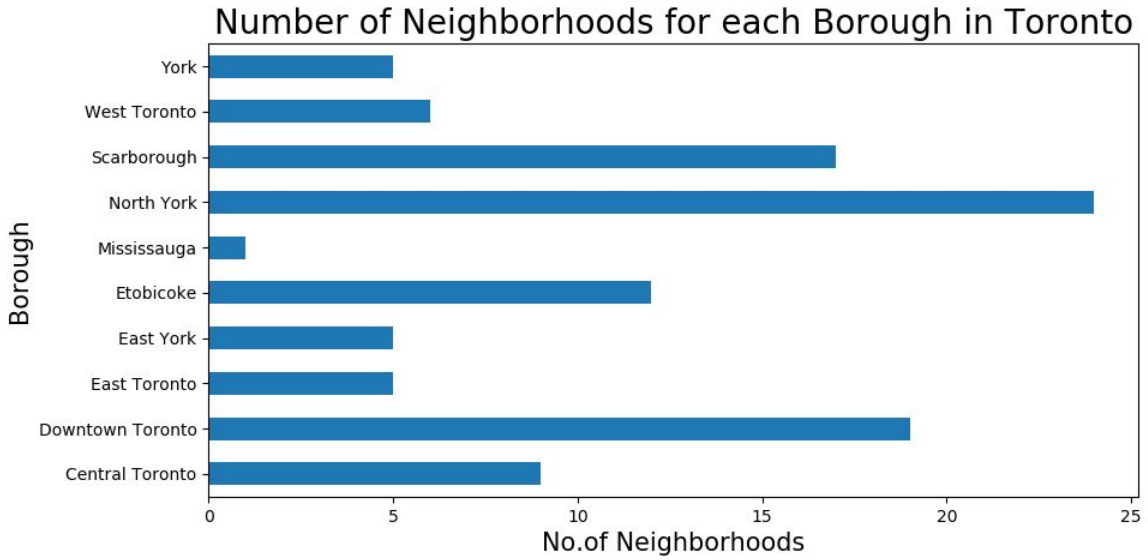

Fig 3. Number of Neighborhoods per Borough

Using the Foursquare API, the top 100 venues that are within a 1 kilometer(1000 meters) radius can be derived. A Foursquare Developer Account is required in order to obtain the Foursquare ID and Foursquare secret key. API calls are made to the Foursquare server by passing in the geographical coordinates of the neighbourhoods in a Python loop. The results obtained from the REST API call will contain the venue data in JSON format. Similarly a REST call is made to obtain the venue details as well. Since the calls are made often for different venues, functions are defined and used. These functions will extract the required and additional data from the JSON. With the data, we can check how many venues were returned for each neighbourhood along with their names, category and ID. The venue details like rating, tips and like count are derived only for Italian and Chinese restaurants. A filter is applied before making the call to the FourSquare server to get the details.

Folium is used to visualize the neighbourhoods with the highest density or italian and chinese restaurants which have very good ratings. Ranking of neighborhoods is done based on the average ratings of italian and chinese restaurants in that area. Using the ranking, recommendations can be given to the group of friends who are visiting Toronto and want to have only italian or chinese food. For this project the minimum average rating of the neighborhood is set to 8.0 and used to segregate better neighborhoods to ensure a great eating experience.

4. Results

The ratings, like count and tips of all the italian or chinese restaurants in a neighborhood are found. An example of the resulting data frame is shown in Fig 4.

	Borough	ID	Likes	Name	Neighborhood	Rating	Tips
0	Downtown Toronto	4ad776eef964a520e20a21e3	77	Mangia and Bevi Resto-Bar	Regent Park, Harbourfront	8.4	40
1	Downtown Toronto	4cbdc6784495721ea262617a	57	Fusaro's	Regent Park, Harbourfront	8.7	22
2	Downtown Toronto	56d8dff7498eb4e5e661e78d	15	Ardo	Regent Park, Harbourfront	7.8	8
3	Downtown Toronto	4a8355bff964a520d3fa1fe3	56	Mercatto	Queen's Park, Ontario Provincial Government	8.1	26
4	Downtown Toronto	52f6816f11d24a43115dc834	171	Scaddabush Italian Kitchen & Bar	Queen's Park, Ontario Provincial Government	7.9	70

Fig 4. Venue details

The density of restaurants in each borough is found and plotted as a horizontal bar plot. The borough with the highest density of italian and chinese restaurants is Central Toronto with 35 restaurants. The borough with the least density of italian and chinese restaurants is York with 2 restaurants.

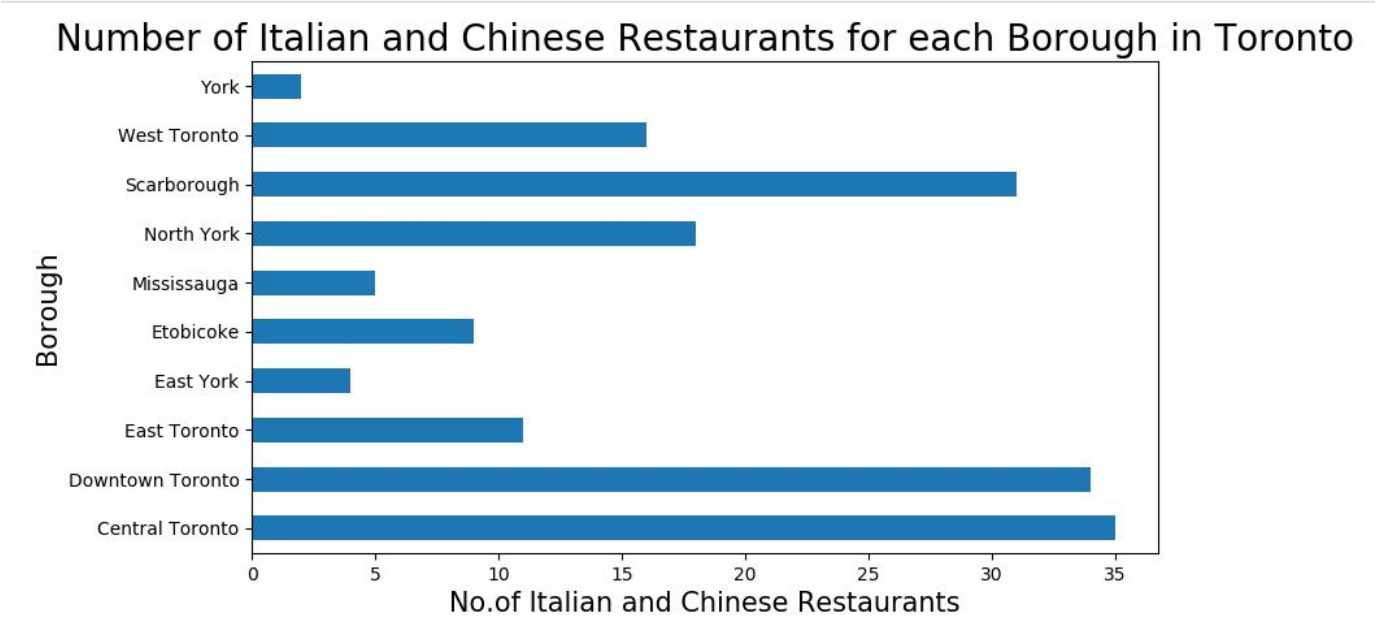

Fig 5. Number of Italian and Chinese Restaurants per Borough

The average ratings for each neighborhood is found after ranking and arranging them from highest average ratings to least. The data frame shown in Fig 6 is part of the result set.

	Neighborhood	Average Rating
52	St. James Town	8.500000
34	Little Portugal, Trinity	8.433333
3	Berczy Park	8.400000
55	Stn A PO Boxes	8.400000
12	Commerce Court, Victoria Hotel	8.400000
39	New Toronto, Mimico South, Humber Bay Shores	8.400000
10	Church and Wellesley	8.400000
48	Richmond, Adelaide, King	8.300000
47	Regent Park, Harbourfront	8.300000
21	First Canadian Place, Underground city	8.300000

Fig 6. Average Rating of Neighborhoods in descending order

Average ratings of the italian and chinese restaurants for each borough is also found using the same method as the neighborhood ranking with 'groupby' clause.

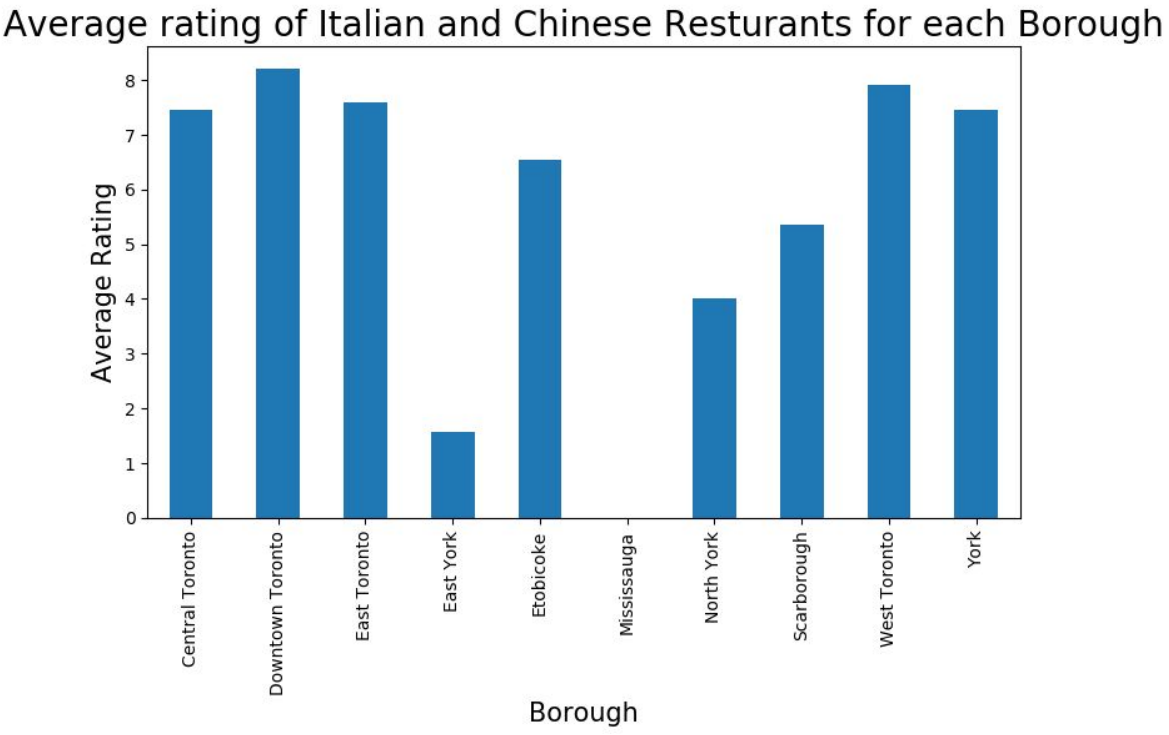

Fig 7. Average Ratings Italian and Chinese Restaurants per Borough

	Neighborhood	Average Rating
3	Berczy Park	8.400000
8	Central Bay Street	8.000000
10	Church and Wellesley	8.400000
12	Commerce Court, Victoria Hotel	8.400000
15	Del Ray, Mount Dennis, Keelsdale and Silverthorn	8.000000
19	Dufferin, Dovercourt Village	8.250000
21	First Canadian Place, Underground city	8.300000
23	Garden District, Ryerson	8.275000
34	Little Portugal, Trinity	8.433333
39	New Toronto, Mimico South, Humber Bay Shores	8.400000
46	Queen's Park, Ontario Provincial Government	8.233333
47	Regent Park, Harbourfront	8.300000
48	Richmond, Adelaide, King	8.300000
52	St. James Town	8.500000
53	St. James Town, Cabbagetown	8.300000
55	Stn A PO Boxes	8.400000
58	The Annex, North Midtown, Yorkville	8.000000
62	Toronto Dominion Centre, Design Exchange	8.200000

Fig 9. Neighborhoods with ratings equal to or above 8.0

The neighborhood with the highest average rating is St. James Town

Finally, we filter out the neighborhoods with an average rating equal to or above 8.0. The higher the rating the better the food experience and convenience. We plot these neighborhoods using the folium package. The folium package helps add markers with popups and hence makes it easier to analyse the data.

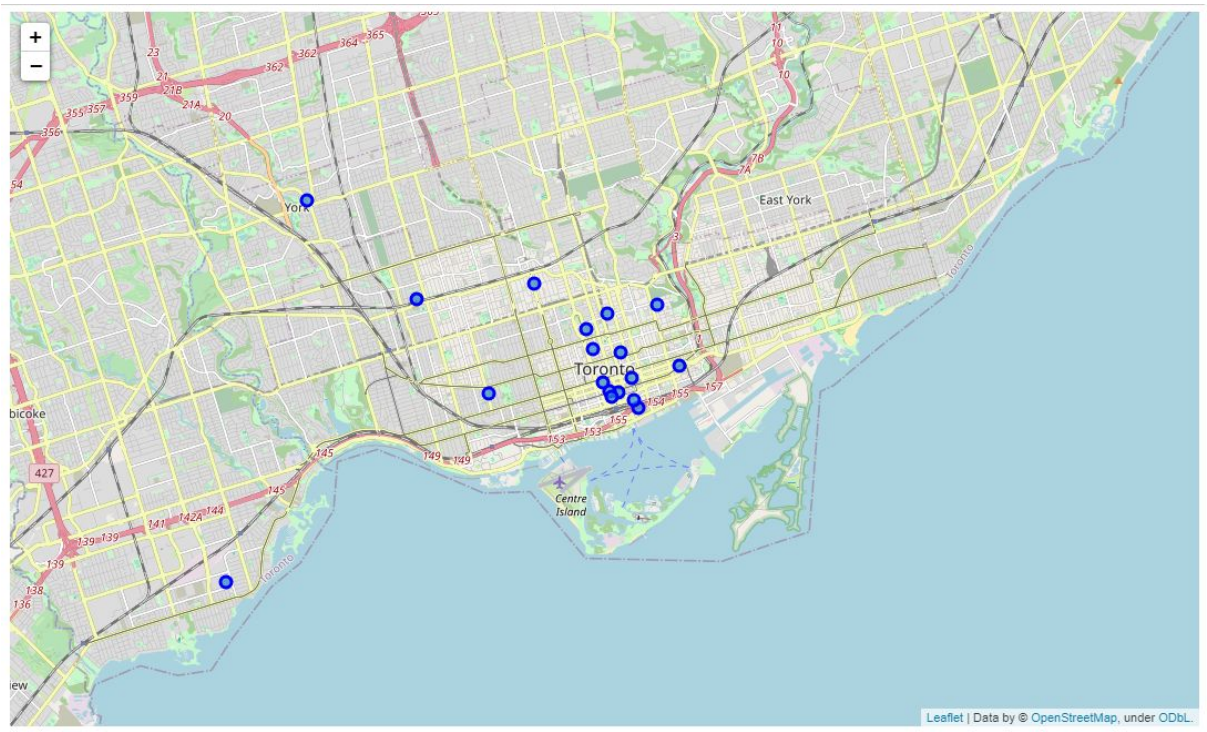

Fig 10. Neighborhoods with ratings above or equal to 8.0

5. Discussion

The built restaurant finder is able to aptly recognize the best neighborhood to stay at during a visit to Toronto. The data sources can be further optimized and increased in quality and quantity to get better results. In this case the data is solely dependent on the FourSquare API for the ranking algorithm hence it does not rank the venues without any ratings.

The'No rating' scenario is observed in the case of Mississauga borough.To avoid such overdependence, the data source can be enriched by using a kaggle dataset in combination.

The future scope of the project would be to include a set of relevant features, instead of just one deciding feature for ranking, to better the rating algorithm and the recommendations.

6. Conclusion

For the group of friends visiting Toronto, St. James Town would be the best place to stay. Downtown Toronto has the maximum Average rating of Italian and Chinese Restaurants which suggests a good potential market for Italian and Chinese Restaurants.

The model built was able to determine a good set of recommendations to propose to the visitors. The model can be used, with a few minor changes, for finding different cuisines density and ratings in any city.