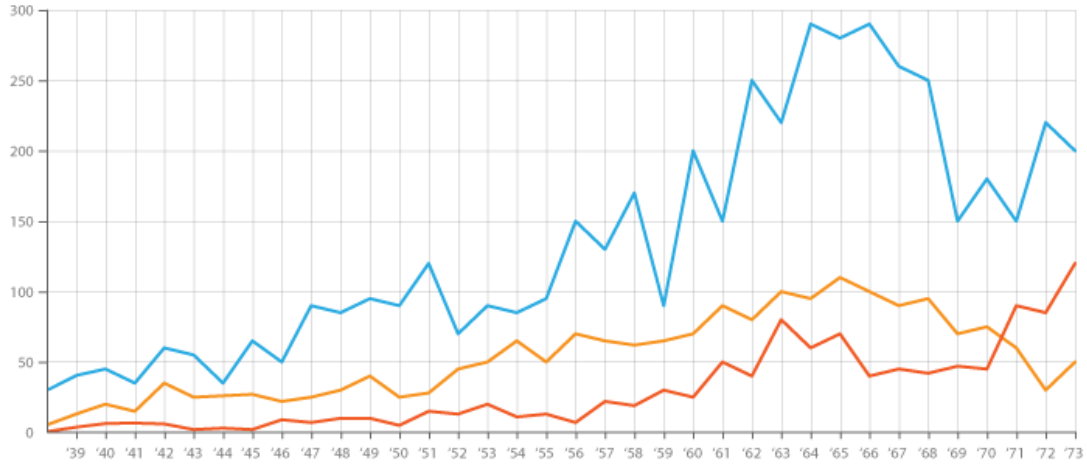
Identify Niche Area for a New Business Venture

Capstone Project

Background

- Investor/Entrepreneur looking for a new venture/business to start/invest in.
- Key questions to answer:
  - Location
  - Type of business



Problem

Key questions:

- Location
  - This selection can be based on his domicile, i.e., where he/she is currently residing or where they want to live long term. In general, people have specific preference on which broad area they want to live in and therefore, we will limit our search space within that area.
  - For this project, **New York City and Toronto are selected** as the potential cities in which we wish to start a business.

- Type of business
  - Identify a niche business sector within these cities that can be exploited.

Data Acquisition and Preprocessing

- Details of neighborhoods and boroughs
  - Dataset for NYC provided by Coursera at https://cocl.us/new_york_dataset.
  - Dataset for Toronto was scraped from the Wikipedia page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
    - The BeautifulSoup (bs4) package was used to scrape the webpage.
- Foursquare data to explore the venues in all the neighborhoods.
  - Up to a 100 venues in the vicinity of each neighborhood was pulled using the Foursquare API
- Use of RideShare data to validate choice of location/neighborhood as a popular/promising location.
  - BikeShare data (https://open.toronto.ca/dataset/bike-share-toronto-ridership-data/) from the Open Data repository from the City of Toronto
  - NYC data available via Kaggle at https://www.kaggle.com/akkithetechie/new-york-city-bike-share-dataset

Methodology

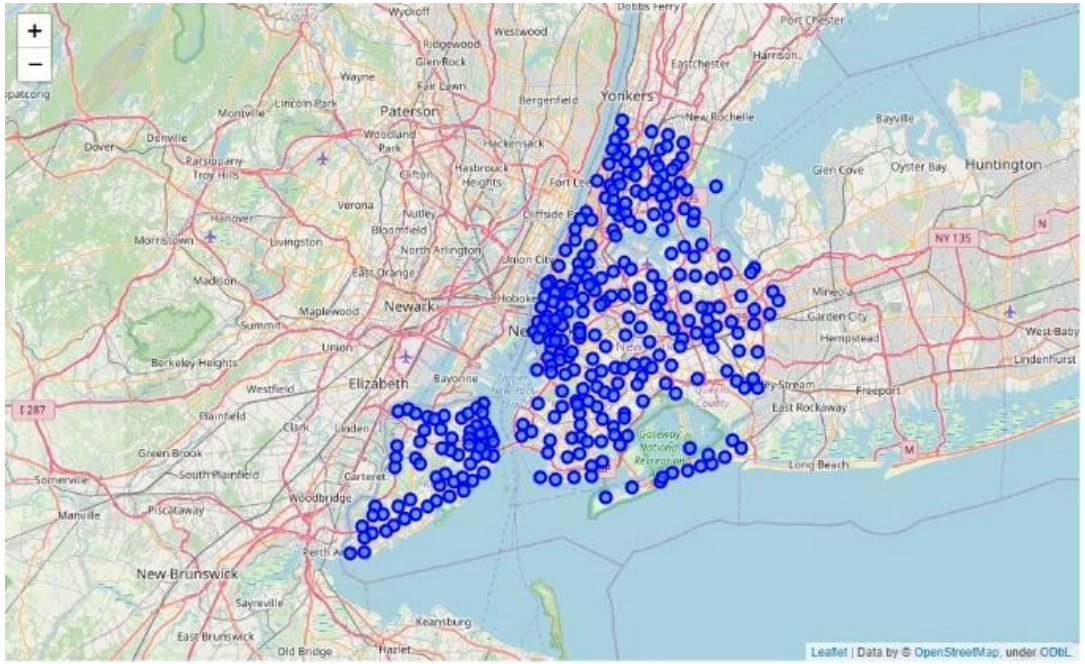
- Exploratory Analysis of NYC and Toronto neighborhood data

- Use Foursquare data to identify niche business areas
  - Compare cities to decide between the cities
- Use of Foursquare data to shortlist popular neighborhoods.
  - Select location from among the shortlisted neighborhoods.

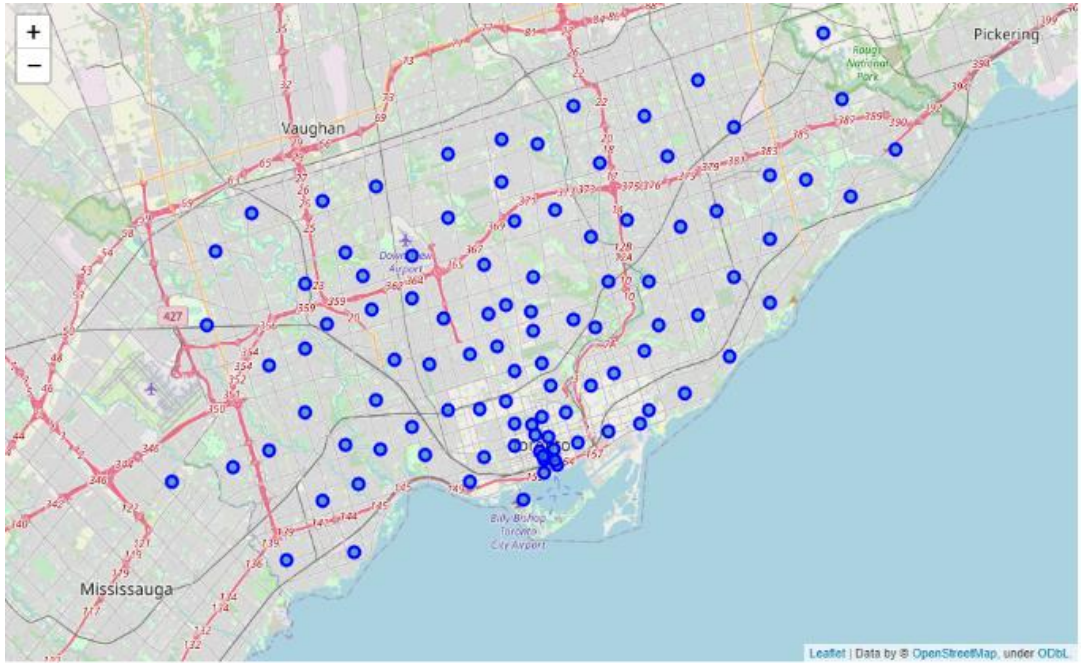
- Use of RideShare data to validate choice of location/neighborhood as a popular/promising location.

Results

- Exploratory Analysis of NYC and Toronto neighborhood data



Neighborhoods in NYC

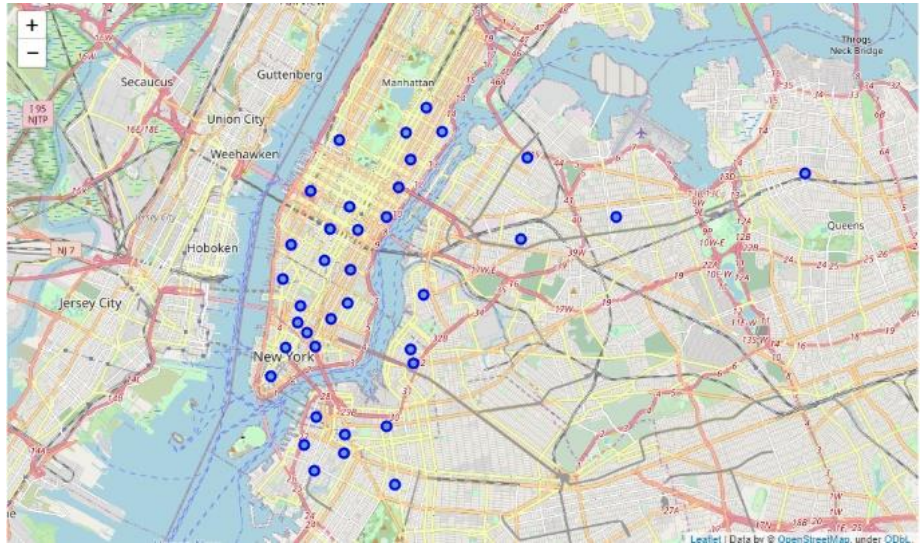


Neighborhoods in Toronto

- **The neighborhoods in NYC are found to be spread with uniform (high) density.**
- **The neighborhoods in Toronto become sparser as we move further from the Downtown. It is concentrated in the Downtown Toronto area.**

Results

- Identify popular neighborhoods (> 80 venues in close proximity based on Foursquare data)



Popular neighborhoods in NYC

Popular neighborhoods in Toronto

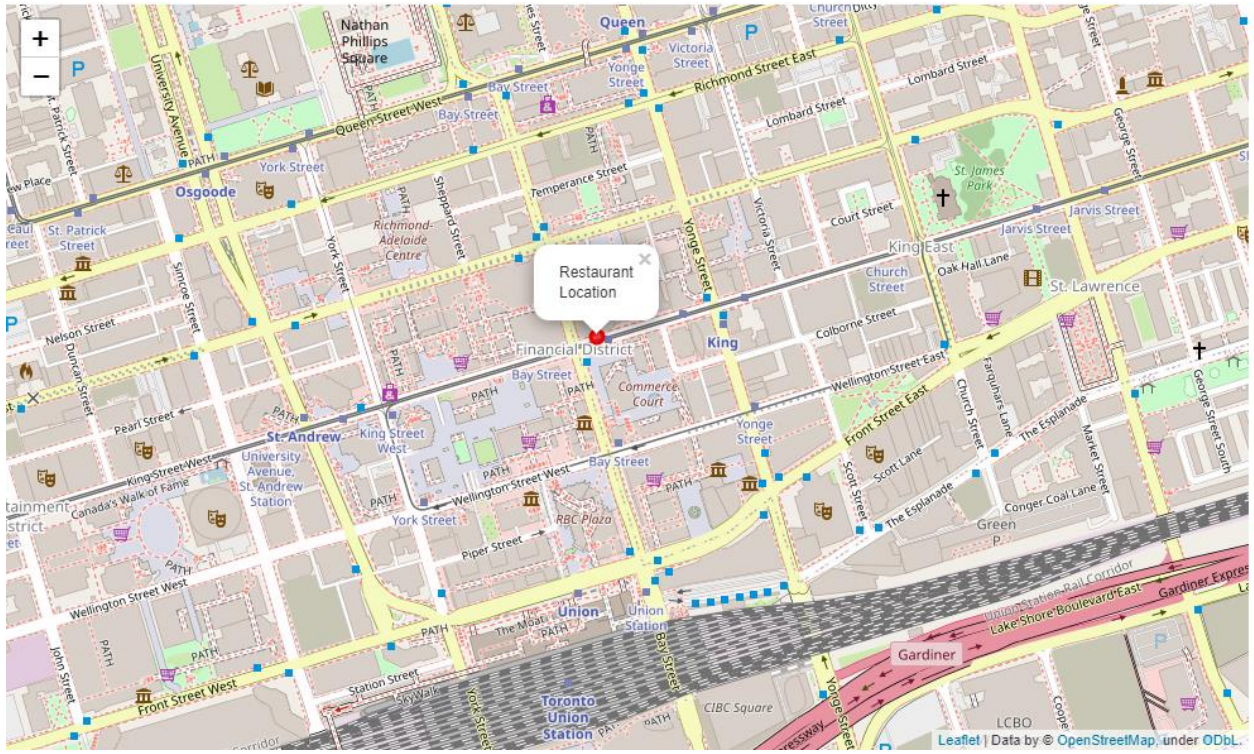
- **The key observation is that all the popular locations in Toronto are in the Downtown Toronto area, whereas its spread over 4 of the 5 boroughs in NYC, while it is localized to one borough in Toronto.**

Results and Discussion

- Analyzed the diversity of venues in these popular neighborhoods.
  - New York City has 423 different types of venues whereas Toronto has 263 different types.
  - NY has more venues offering the same facilities/experience compared to Toronto.
    - Ex: NY has 77 yoga studios as opposed to 13 in Toronto. Similar trend in a lot of facilities.
  - NY offers more multicultural experiences than Toronto.
    - Examples of multicultural venues in NY not in Toronto: Argentinian Restaurant, Australian Restaurant, Austrian Restaurant, Arepa restaurant, etc.
- Select city and business
  - Advantageous to open a multicultural restaurant in Downtown Toronto.
  - Rationale:
    - Lack of such restaurants in Toronto. For example, Downtown Toronto lacks a Kebab restaurant.
    - Downtown is a very popular place with many neighborhoods with more than 80 venues. This makes it an ideal place to start a business as the neighborhood is frequently visited by people in general.
  - Business idea: **Open a Kebab restraint in Toronto**

Results and Discussion

- Find a location in Toronto to open the Kebab restaurant
  - **Key idea:** Find a popular location.
  - Only a handful of popular places – no point in clustering, therefore select centroid location.



Results and Discussion

- Use Ride share data for validating choice of location

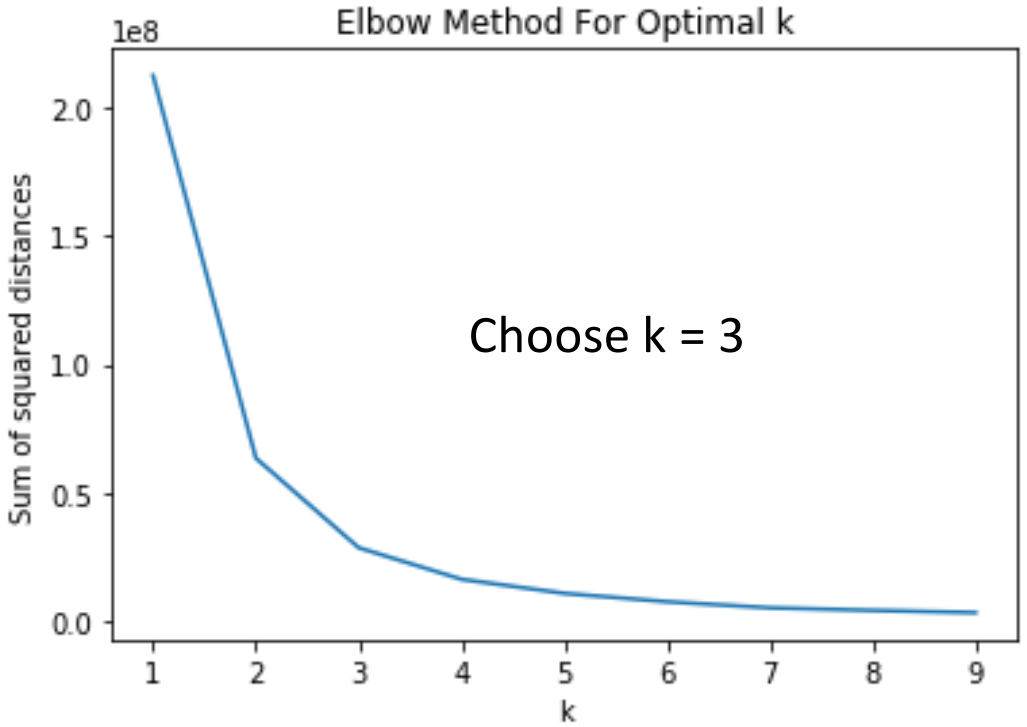
	trip_id	trip_duration_seconds	from_station_id	trip_start_time	from_station_name	trip_stop_time	to_station_id	to_station_name	user_type
0	4158592	749	7061	10/1/2018 0:01	Dalton Rd / Bloor St W	10/1/2018 0:14	7042	Sherbourne St / Wellesley St E	Annual Member
1	4158593	433	7003	10/1/2018 0:06	Madison Ave / Bloor St W	10/1/2018 0:13	7280	Charles St E / Jarvis St - SMART	Annual Member
2	4158594	285	7024	10/1/2018 0:14	Dundonald St / Church St	10/1/2018 0:19	7028	Gould St / Mutual St	Annual Member
3	4158595	150	7190	10/1/2018 0:16	St. George St / Hoskin Ave	10/1/2018 0:18	7161	Beverley St / College St	Annual Member
4	4158596	744	7265	10/1/2018 0:21	Wallace Ave / Symington Ave - SMART	10/1/2018 0:33	7136	Queen St W / Close Ave	Annual Member

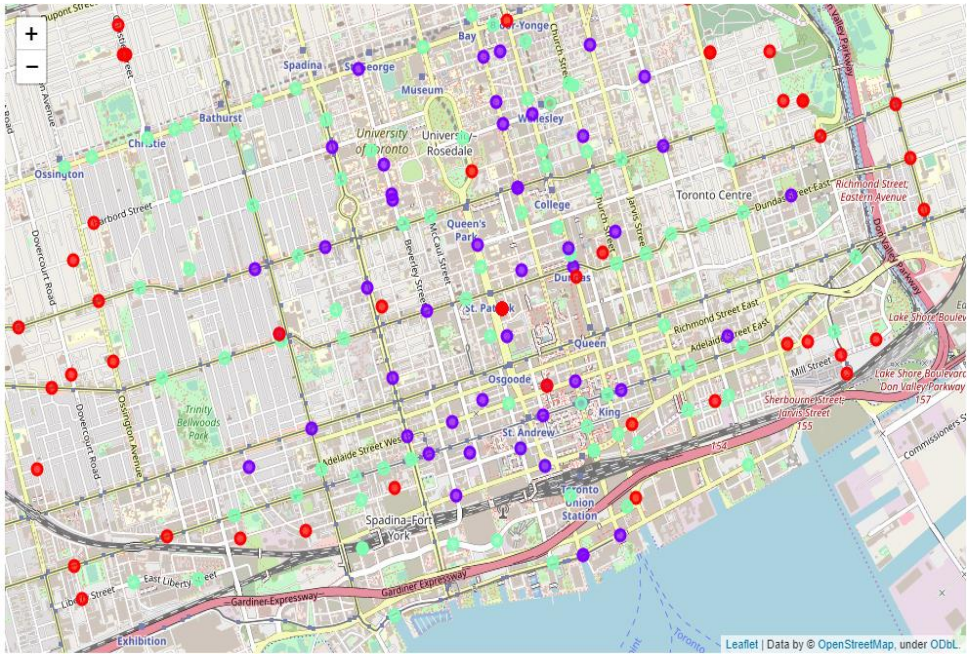
- Process data, add location information. Aggregate trip information to get total trips from each bike station.

	Street	Latitude	Longitude	Zip	Count
0	Fort York Blvd / Capreol Ct	43.640400	-79.399500	M5V	2354
1	Lower Jarvis St / The Esplanade	43.648220	-79.370922	M5E	1106
2	St. George St / Bloor St W	43.667510	-79.399821	M5R	2744
3	Madison Ave / Bloor St W	43.686100	-79.402500	M4V	1662
4	University Ave / Elm St	43.656322	-79.389114	M5G	1271
5	King St W / York St	43.647914	-79.383565	M5H	1918
6	Bay St / College St (East Side)	43.660809	-79.385849	M5S	3447

Results and Discussion

- Use elbow method to determine number of clusters for bike stations. Cluster using k-means algorithm.





We can see that the area we selected to start the new Kebab restaurant does indeed lie in a very popular area. The area has a high concentration of cluster 1 (purple points) and cluster 2 (green points) than Cluster 0 (red) points. The cluster 1 and cluster 2 points have a higher average number of rides than cluster 0.

Conclusion

- Cities of Toronto and New York were studied to identify a good location and potential ideas for a new business venture.
- Datasets on neighborhoods of New York City and Toronto was used in conjunction with FourSquare data to determine choice of city as Toronto.
  - location and business idea was decided based on these datasets.
- Choice of location in Toronto was compared with the BikeShare data to get more confidence in the analysis.

Thank you!