Capstone Project

for IBM Professional Certification in Data science

The Battle of the Neighborhoods

Analysis of preferred neighborhood in Toronto to visit

Prepared by: Parvaneh Pouladzadeh

February 2010

Problem Statement

- Toronto, is the capital of the province of Ontario, Canada, is a major Canadian city along Lake Ontario's northwestern shore with a population of 2,531,571 in 2016.
- The demographics of Toronto is important (because Toronto also has established ethnic neighborhoods such as Chinatown, Corso Italia, Little Italy, Little India, Greektown, Koreatown, Little Jamaica, Little Portugal and Roncesvalles, which celebrate the city's multiculturalism).
- **Based on the mission of this project is to use Foursquare location data and regional clustering of venue information to determine which neighborhood in Toronto is the best place to visit.**

Data

```
1 url = "https://lvdesign.com.fr/coursera_data/toronto_base.csv"
2 df_geoToronto = pd.read_csv("https://lvdesign.com.fr/coursera_data/toronto_base.csv")
3 df_geoToronto.head()
4
```

	Unnamed: 0	PostalCode	Borough	Neighborhood
0	0	M3A	North York	Parkwoods
1	1	M4A	North York	Victoria Village
2	2	M5A	Downtown Toronto	Regent Park , Harbourfront
3	3	M6A	North York	Lawrence Manor , Lawrence Heights
4	4	M7A	Queen's Park	Queen's Park

- In this project we considered Toronto for analysis and we will be using the datasets below for analyzing Toronto. In order to obtain the data that is in the table of postal codes to figure out the locations and neighborhood of Toronto to realize which place could be most effective for travelers to visit. https://lvdesign.com.fr/coursera_data/toronto_base.csv
- Using the Foursquare API:

✓ One of the important features of this API is to provide a list of venues within a specific location

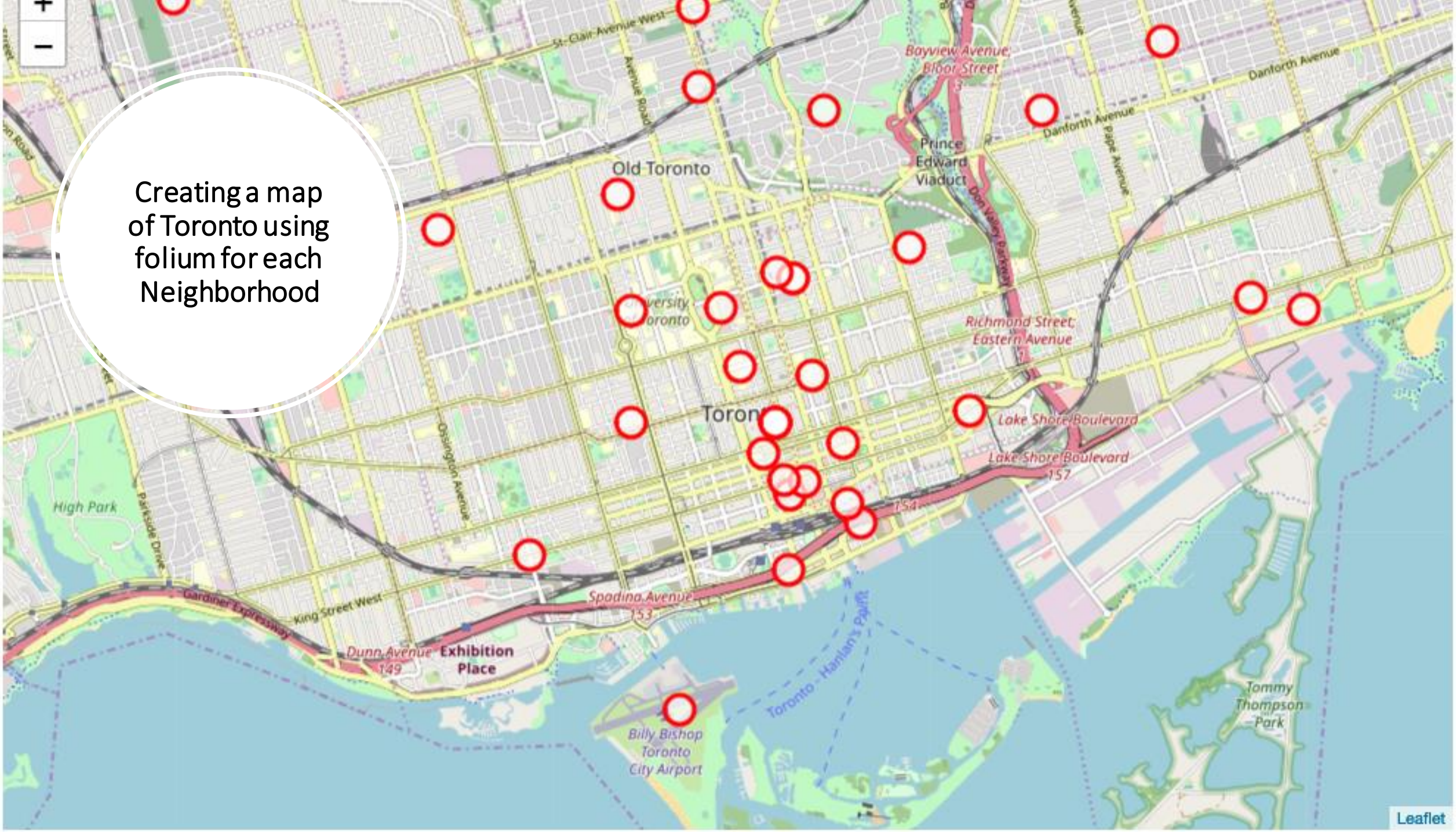
✓ The Foursquare API was used to find the venues on each postal code zone using a radius based on the area cover by each postcode.

✓ Using K-Means clustering algorithm, the postal codes were grouped based on the venues density (venues/area) and the result was showed on a map of Toronto.

Methodology

- We prepared the data to be used for the machine learning process.

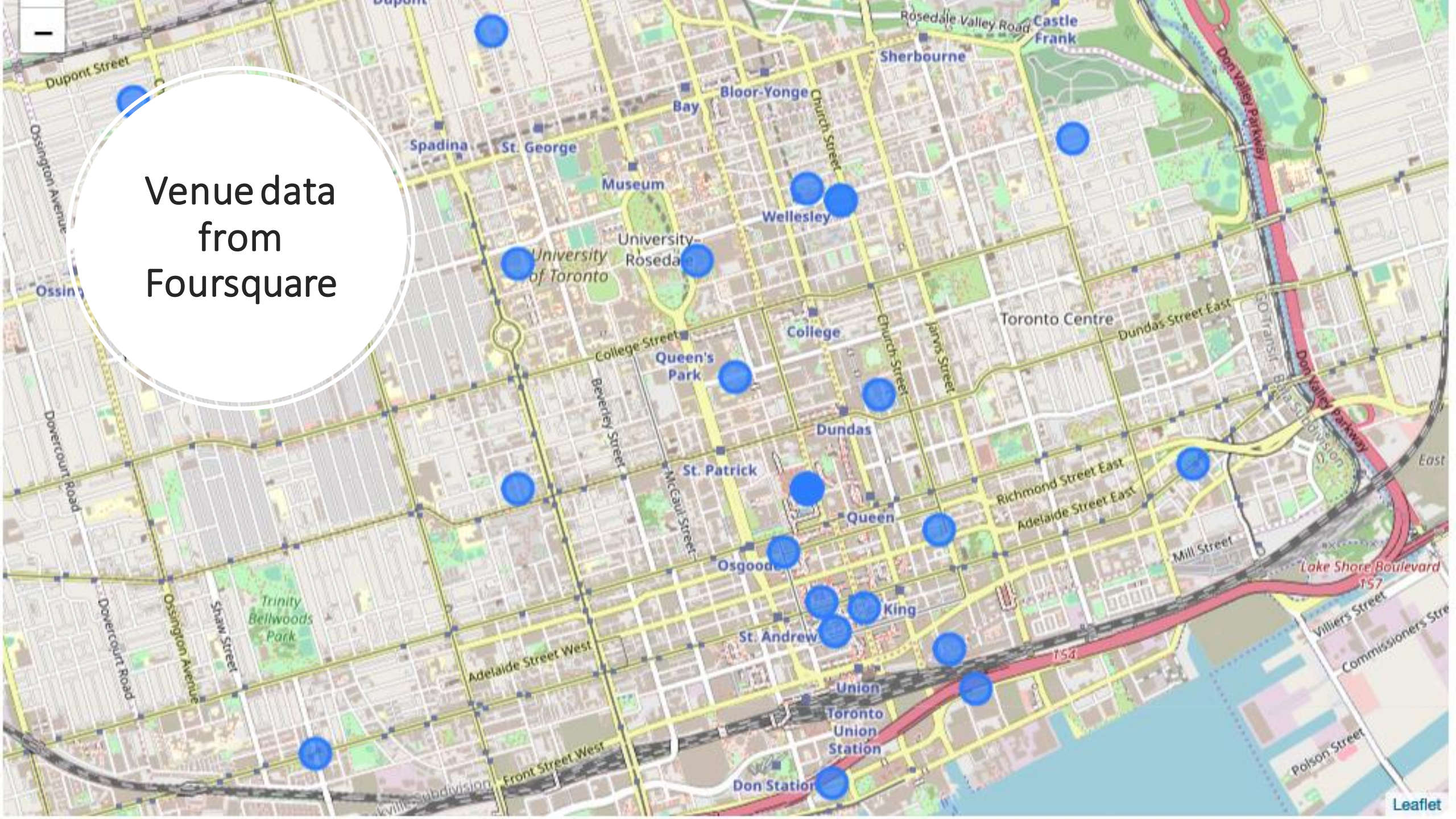
1. Based on coordinate data, we extract data of Toronto neighborhoods from Wikipedia.
2. We need to explore and clean neighborhoods dataset.
3. Get venues in order to get the venues in the perimeter of each Postal code.
4. Analyze venues dataset in order to get a better sense of the best way of clustering the postal codes.
5. Cluster Postcodes, given that each postcode has a different radius passed to the venues request.
6. Examine each Clusters based on the centroids of each cluster.


Creating a map of Toronto using folium for each Neighborhood
Old Toronto
Toron
Prince Edward Viaduct
Bayview Avenue
Bloor Street
Danforth Avenue
Richmond Street
Eastern Avenue
Lake Shore Boulevard
Lake Shore Boulevard
157
154
153
149
Spadina Avenue
King Street West
Gardiner Expressway
Dunn Avenue
Exhibition Place
High Park
Parkside Drive
Ossington Avenue
St. Clair Avenue West
Avenue Road
Billy Bishop Toronto City Airport
Toronto - Hanlan's Point
Tommy Thompson Park
Leaflet

Clustering

	PostalCode	Address	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7 C
1	M4A	M4A, North York	43.725882	-79.315572	1.0	Coffee Shop	Portuguese Restaurant	Intersection	Hockey Arena	Dumpling Restaurant	Drugstore	Dor
2	M5A	M5A, Downtown Toronto	43.654260	-79.360636	1.0	Coffee Shop	Park	Café	Bakery	Pub	Theater	E
3	M6A	M6A, North York	43.718518	-79.464763	1.0	Clothing Store	Furniture / Home Store	Coffee Shop	Event Space	Shoe Store	Boutique	Misce
4	M7A	M7A, Queen's Park	43.662843	-79.389638	1.0	Coffee Shop	Bubble Tea Shop	Japanese Restaurant	Gym	Diner	Sushi Restaurant	Re
7	M3B	M3B, North York	43.745906	-79.352188	1.0	Pool	Gym / Fitness Center	Japanese Restaurant	Baseball Field	Caribbean Restaurant	Café	
							Fast Food	Gym / Fitness	Athletics &	Rock		

- We examined the most common venues by neighborhood listed in Foursquer using clustering techniqes.
- Cluster 1 shows the highest venues for travellers wich is in the North York area that is great to visit.


Venue data from Foursquare
Dupont Street
Ossington Avenue
Spadina
St. George
Bay
Bloor-Yonge
Church Street
Sherbourne
Rosedale Valley Road
Castle Frank
Don Valley Parkway
Museum
Wellesley
University of Toronto
University–Rosedale
College Street
College
Queen's Park
Beverley Street
McCaul Street
Jarvis Street
Toronto Centre
Dundas Street East
Dundas
St. Patrick
Richmond Street East
Adelaide Street East
Queen
Osgoode
King
St. Andrew
Mill Street
Lake Shore Boulevard
TS7
TS4
Villiers Street
Commissioners Street
Union
Toronto Union Station
Front Street West
Adelaide Street West
Dovercourt Road
Shaw Street
Trinity Bellwoods Park
Polson Street
Leaflet

Results

- After clustering the Toronto neighborhoods based on the results from Foursquare API data, we were able to separate our data set into three distinct clusters, and then from our target cluster, we can choose the best choice for our customers to open a new restaurant.
- Restaurants are the most common places of interest in most districts/clusters, followed by coffee shops and shopping malls.
- Cluster 1 contains the North York neighborhood of Toronto, with a larger variety of places of interest than the other clusters, and is therefore recommended to travelers.

Conclusion

- The project has explored the different neighborhoods of Toronto to find top places of interest based on Foursquare location data. The results indicate that the central area of the city encompasses the most districts with various restaurants, cafés, stores and shopping malls. A map of Toronto with different clusters as well as tables of top venues is presented as a reference for travelers and visitors.