The Battle Of Neighborhoods

Neighborhood recommendation for XYZ Grocery's first venture in Toronto.

INTRODUCTION

- Toronto city overview for XYZ Grocery
- XYZ Grocery Strategic Plan and Overview
- XYZ Grocery Business Problem
  - Choice of the first neighborhood to offer services.
  - The neighborhood should enable easy replication.
  - The neighborhood should have high demand.
  - The neighborhood should have lower competition.
- Success Criteria:
  - Good recommendation of neighborhood that meets above points.
  - It should have capability to enable latest information whenever required.

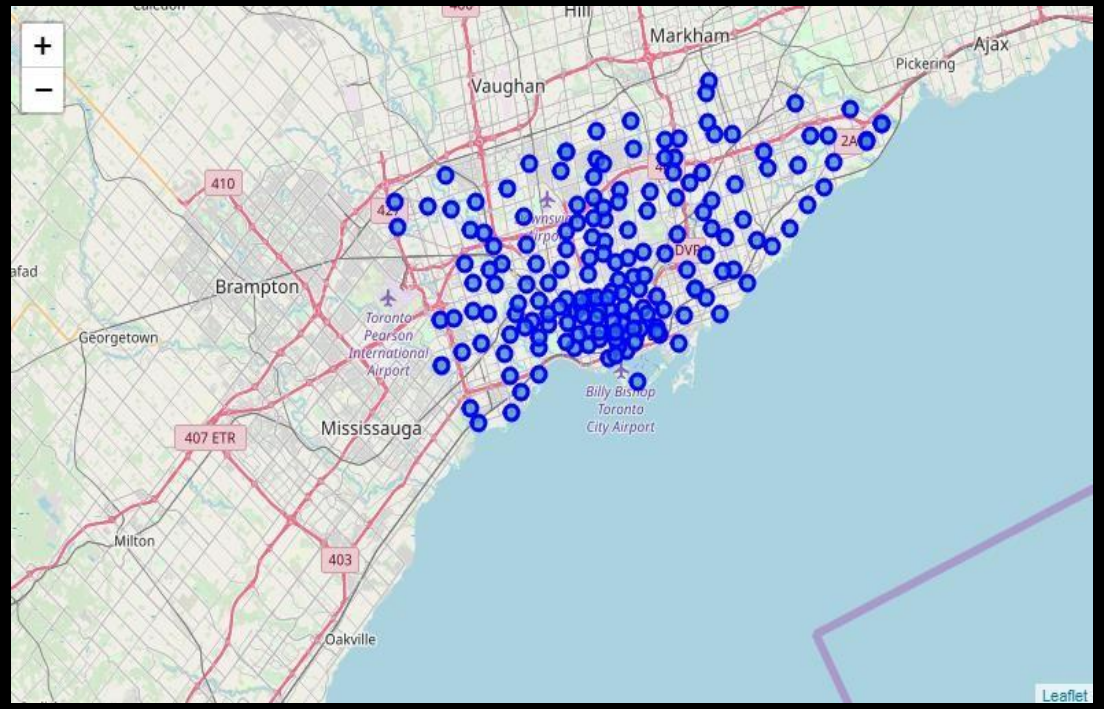
DATA DESCRIPTION

- Data Requirements for this project:
  - Neighborhood Information (i.e. name, coordinates, populations).
  - Venue information (i.e. name, category, coordinates)
- Data Sources for this project:
  - Wikipedia Pages
  - Foursquare.com
- Data Processing for this project:
  - Data cleaning is required
  - Data needs to be in a structured format

	Neighbourhood	Population	Land Area	Density	Population %	Income	Commuting	2nd Language	2nd Language %	Latitude	Longitude
1	Agincourt	44577	12.45	3580	4.6	25,750	11.1	Cantonese (19.3%)	19.3% Cantonese	43.788	-79.2839
2	Alderwood	11656	4.94	2360	-4.0	35,239	8.8	Polish (6.2%)	06.2% Polish	43.6035	-79.5464
3	Alexandra Park	4355	0.32	13,609	0.0	19,687	13.8	Cantonese (17.9%)	17.9% Cantonese	43.6498	-79.4015
4	Allenby	2513	0.58	4333	-1.0	245,592	5.2	Russian (1.4%)	01.4% Russian	43.7077	-79.4127
5	Amesbury	17318	3.51	4,934	1.1	27,546	16.4	Spanish (6.1%)	06.1% Spanish	43.7011	-79.481
6	Armour Heights	4384	2.29	1914	2.0	116,651	10.8	Russian (9.4%)	09.4% Russian	43.7454	-79.4226
7	Banbury	6641	2.72	2442	5.0	92,319	6.1	Unspecified Chinese (5.1%)	05.1% Unspecified Chinese	43.7491	-79.3664
8	Bathurst Manor	14945	4.69	3187	12.3	34,169	13.4	Russian (9.5%)	09.5% Russian	43.7627	-79.4563

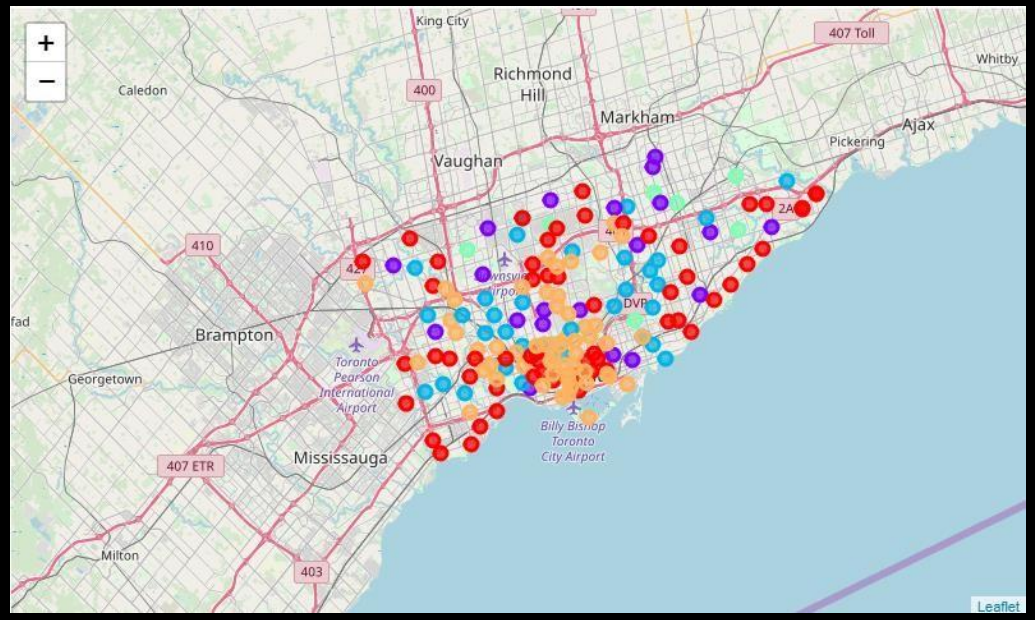
METHODOLOGY

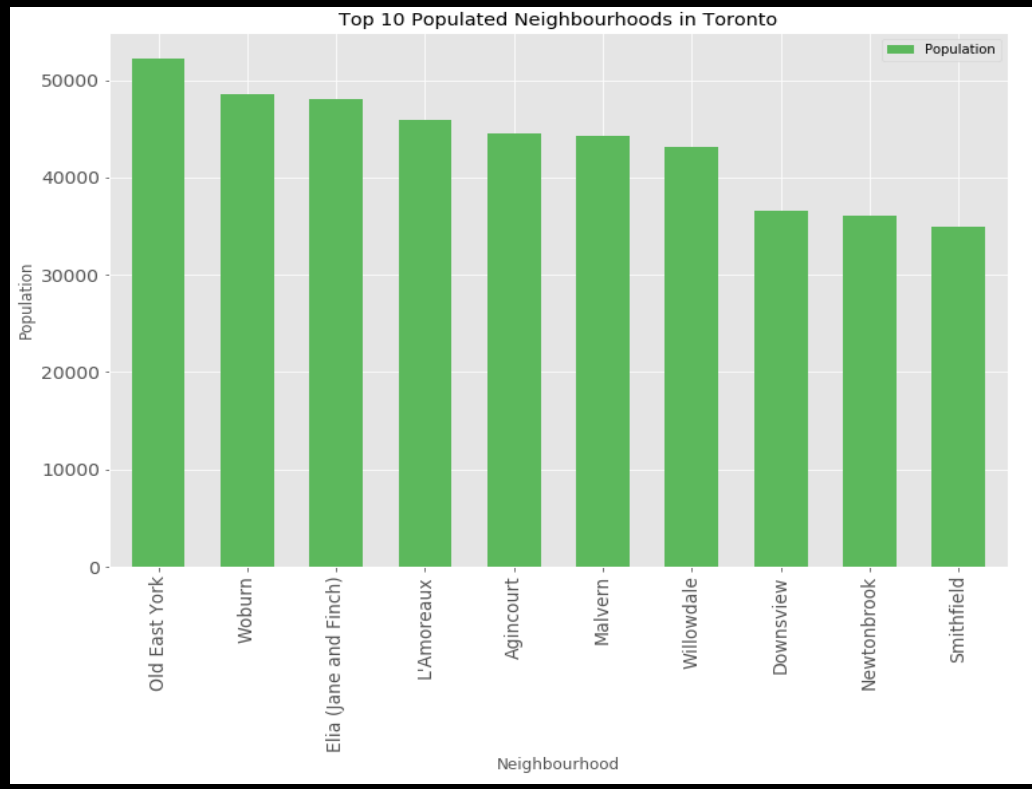
- Data Scrapping Technique
  - To scrap the neighborhood data from Wikipedia
  - To make the data into a structured format
- Foursquare.com
  - Retrieving venue information from above API.
- K-Means Machine Learning
  - Segment and cluster similar neighborhoods
  - Group them based on most common venues
- Scoring Weightage
  - Population count has 50% weightage
  - Venue (fewer grocery) has another 50% weightage



RESULT

- Clustering Similar Neighborhoods
  - Grouping based on common venues
- Neighborhood with high population
  - Ranking the neighborhoods based on population count
- Topmost common venue
  - Ranking the neighborhoods based on common venue w/o Grocery





DISCUSSION

- Result Review:
  - Fifth cluster looks to offer the best business expansion opportunity.
  - Downsview neighborhood in the first cluster looks to meet the key business criteria whereby it will offer the highest demand (i.e. high population) and lowest competition (i.e. lower supply).
- Additional Observation:
  - High number of Italian people.
  - Worthwhile to consider Italian food or related supplies offering in this neighborhood.

	Neighbourhood	Population	Income	Commuting	2nd Language	2nd Language %	Latitude	Longitude	Population Score	Venue Score	Total Score	Cluster Labels	1st Most Common Venue
7	Downsview	36613	26,751	14.4	Italian (11.7%)	11.7% Italian	43.7323	-79.4934	1.515581	0.0	0.757790	4	Spa
8	Newtonbrook	36046	33,428	16.6	Russian (8.8%)	08.8% Russian	43.7901	-79.4197	1.492110	0.0	0.746055	4	Korean Restaurant
9	Smithfield	34996	24,387	12.8	Punjabi (11.8%)	11.8% Punjabi	43.7394	-79.5884	1.448645	0.0	0.724323	4	Grocery Store
10	Fairbank	34121	28,403	21.6	Portuguese (11.3%)	11.3% Portuguese	43.6964	-79.4563	1.412425	0.0	0.706213	4	Furniture / Home Store
11	Riverdale	31007	40,139	20.0	Cantonese (6.7%)	06.7% Cantonese	43.6697	-79.3532	1.283522	0.0	0.641761	4	Chinese Restaurant

CONCLUSION

- Recommendation Summary:
  - Region: Fifth Cluster.
  - Neighborhood: Downsview
  - Additional Offering: Italian food or related supplies.
- Additional Note:
  - Re-run of the program is encouraged to get latest information.
  - Updated information is critical to consider as part of business expansion plan.
- Thank You:
  - Appreciation Notes from Data Science Team for this project.