

Capstone Project

Find Neighborhood for A New Supermarket

Zhonghao Sun

Introduction

- Background: Toronto is the capital of Canada. An international supermarket chain company is planning to open its first supermarket in Toronto.
- Aim: The aim of this project is to find a proper neighborhood location for its first store.

Introduction

- Considerations: population, demand, competition
- Specifically, the following need to be explored:
 - Compare population (demand), number of supermarket and grocery stores (competition), number of other venues such as restaurants, bars, hotels, schools that can increase demand,
 - Cluster neighborhoods and find the current development patterns,
 - Identify proper neighborhoods based on different development strategies of the company.

Method—Data

- Neighborhood name and population. The name and population of Toronto neighborhoods can be found in Wikipedia:
https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods.
- Geographical information of neighborhoods. The geographical information is required to request venue data. The latitude and longitude coordinates of neighborhoods can be obtained using Geocoder package or the csv file from the class.
- Venue data. We use Foursquare API to get venue data for neighborhoods.

Method—Data preprocessing

	Neighborhood	Population	Latitude	Longitude
0	Agincourt	44577	43.794200	-79.262029
1	Alderwood	11656	43.602414	-79.543484
2	Bathurst Manor	14945	43.754328	-79.442259
3	Bayview Village	12280	43.786947	-79.385975
4	Bedford Park	13749	43.733283	-79.419750

Neighborhood names, population, and geographical info.

Method—Data preprocessing

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Lounge	Latin American Restaurant	Skating Rink	Breakfast Spot	Population	Gym	Grocery Store	Financial or Legal Service	Health & Beauty Service	Filipino Restaurant
1	Alderwood	Pizza Place	Pharmacy	Skating Rink	Dance Studio	Gym	Pub	Coffee Shop	Athletics & Sports	Sandwich Place	Population
2	Bathurst Manor	Bank	Coffee Shop	Pizza Place	Pharmacy	Bridal Shop	Shopping Mall	Sandwich Place	Restaurant	Deli / Bodega	Sushi Restaurant
3	Bayview Village	Bank	Japanese Restaurant	Café	Chinese Restaurant	Population	Diner	Discount Store	Fish & Chips Shop	Financial or Legal Service	Filipino Restaurant
4	Bedford Park	Sushi Restaurant	Italian Restaurant	Coffee Shop	Sandwich Place	Pizza Place	Greek Restaurant	Indian Restaurant	Juice Bar	Liquor Store	Pharmacy

10 most common venues in each neighborhood.

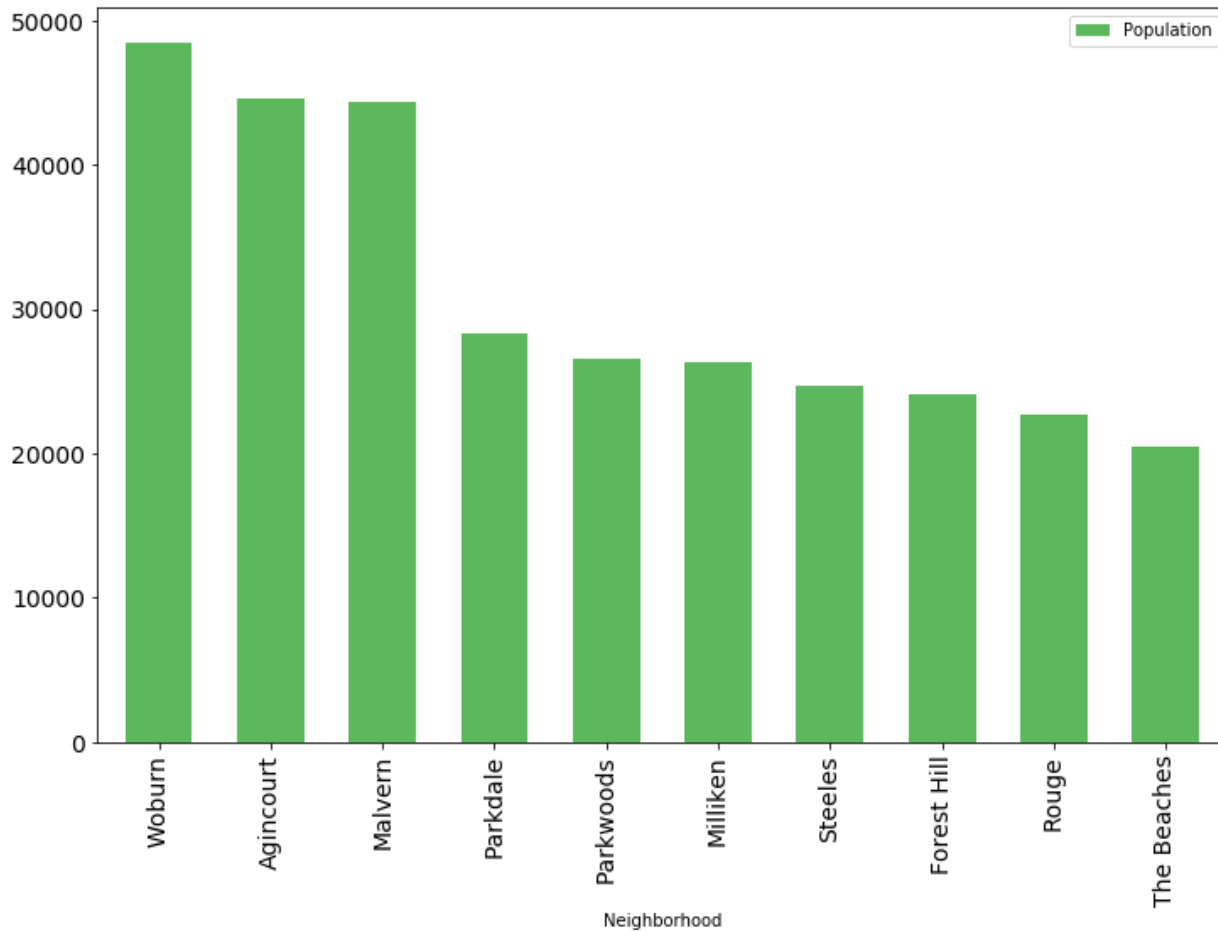
Method—Data preprocessing

Restaurants will attract people and increase the demand.

Grocery stores, other supermarkets, and shopping malls may increase competition

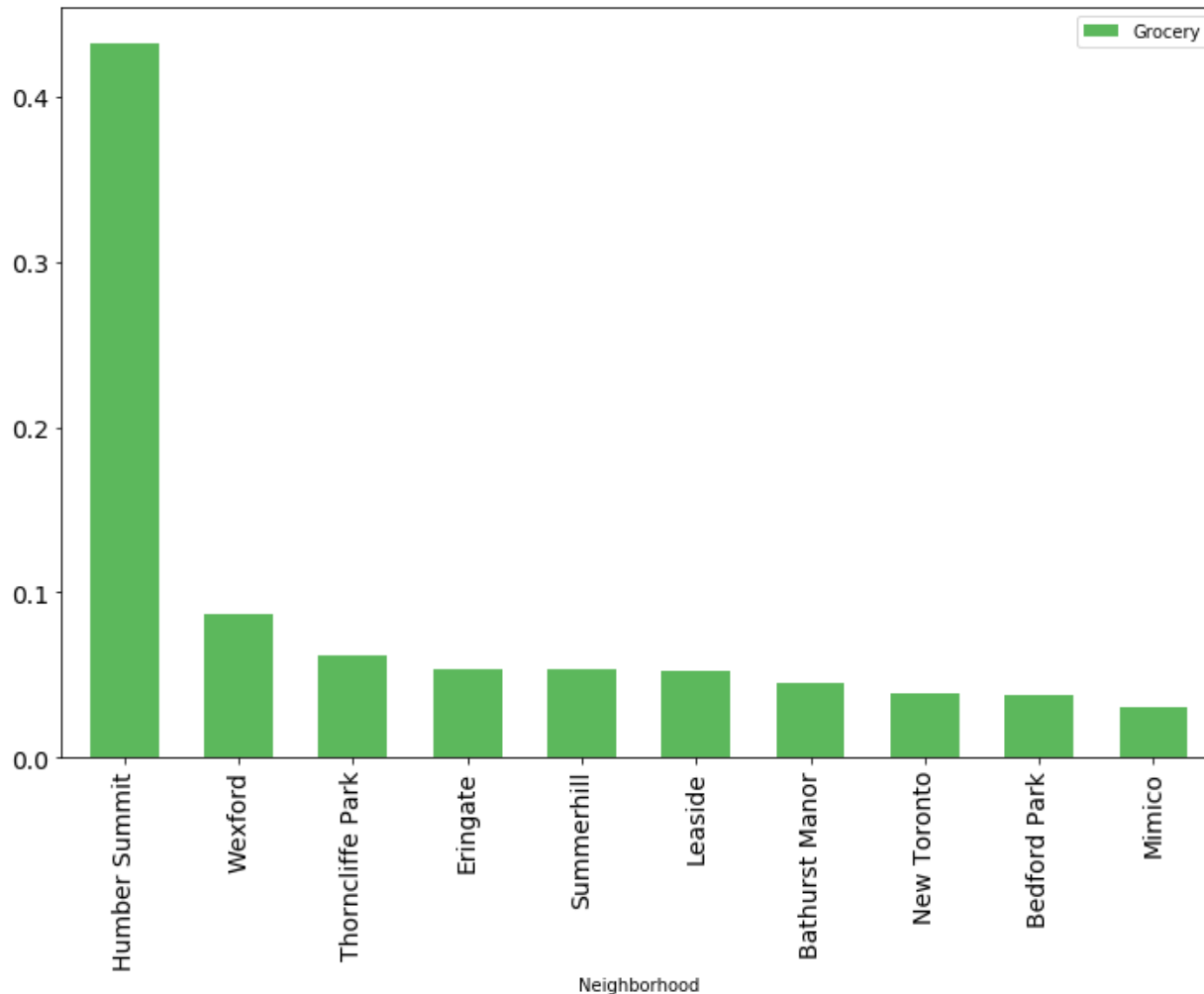
	Neighborhood	Restaurant	Grocery	Population
46	Woburn	0.084351	0.000000	0.067109
43	Westmount	0.070292	0.000000	0.008103
4	Bedford Park	0.058679	0.037562	0.019021
18	Kingsview Village	0.056234	0.000000	0.022487
3	Bayview Village	0.056234	0.000000	0.016989
38	The Danforth	0.050878	0.020570	0.010859
9	Dorset Park	0.048200	0.000000	0.019630
34	Steeles	0.048200	0.030854	0.034166
1	Alderwood	0.044987	0.000000	0.016126

Results and Discussion



High population neighborhoods

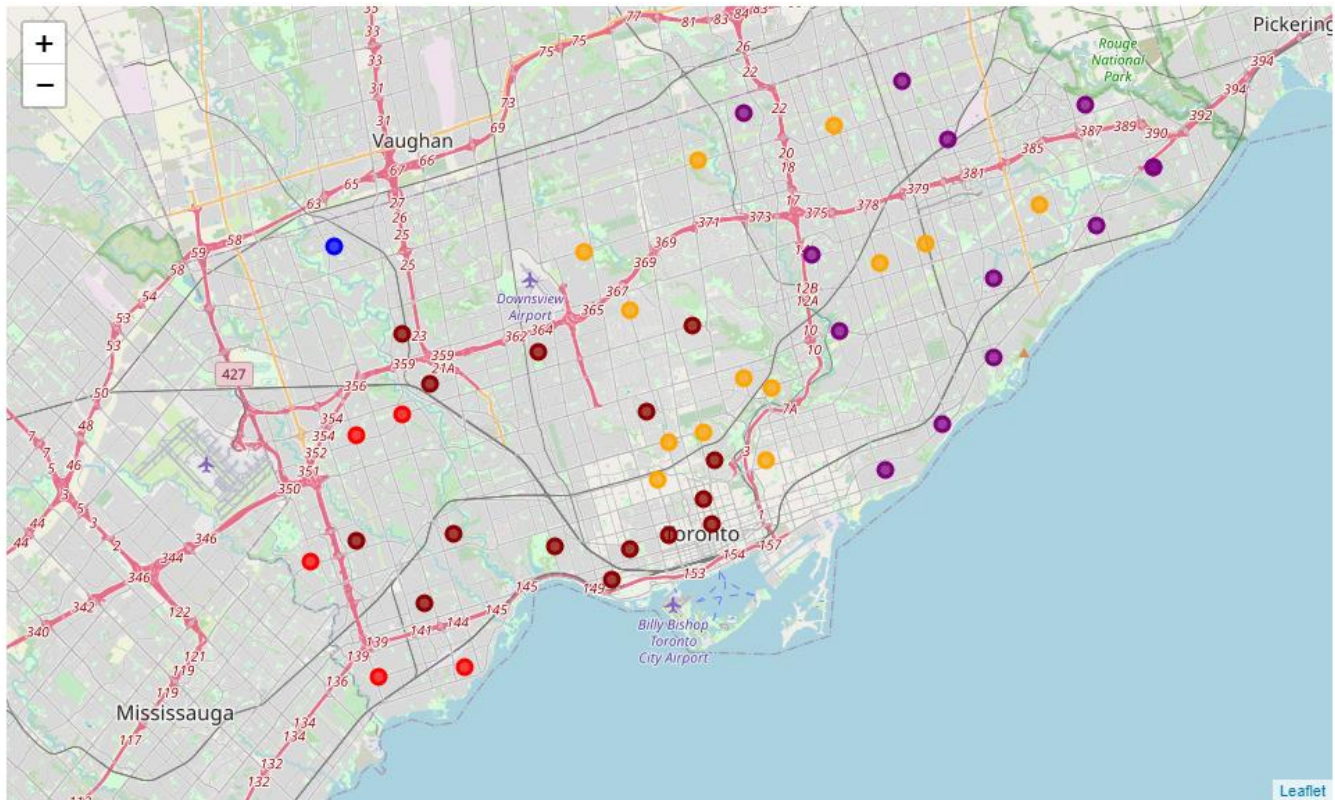
Results and Discussion



High competition neighborhoods

Results and Discussion

Clustering:



Results and Discussion

We identify 5 clusters:

- (1) Cluster #1 (darkred) has a relative low score of restaurant and grocery, and a relative low score of population.
- (2) Cluster #2 (red) has a relative high score of restaurants, and a relative low score of population.
- (3) Cluster #3 (blue) has a very high score of restaurants, and a very low score of population. There is only one neighborhood in this cluster: Humber Summit.
- (4) Cluster #4 (purple) has a very low score of restaurant/grocery, and a relative high score of population, such as Bayview Village, Bedford Park, Dorset Park, and Woburn.
- (5) Cluster #5 (orange) has a relative high score of restaurants.

Results and Discussion

Strategies	Neighborhoods
Highest competition	Cluster #3: Humber Summit
Around restaurants, high competition	Cluster #2: Eringate, New Toronto Cluster #5: Bathurst Manor, Bedford Park
Around restaurants, low competition	Cluster #2: Alderwood, Kingsview Village Cluster #5: Bayview Village, Dorset Park, Woburn
High population, low competition	Cluster #4: Agincourt, Malvern, Parkwoods

Conclusions

- We analyzed the population and nearby venue data of Toronto neighborhoods.
- We find neighborhoods with high population, or a large number of supermarkets, or a large number of restaurants.
- We identify 5 clusters of neighborhoods based on the level of demand and competition.
- The choose of the location depends on the development strategy.

THANK YOU!