Location recommendation for business

Hung NGUYEN - IBM Captone project - 2020
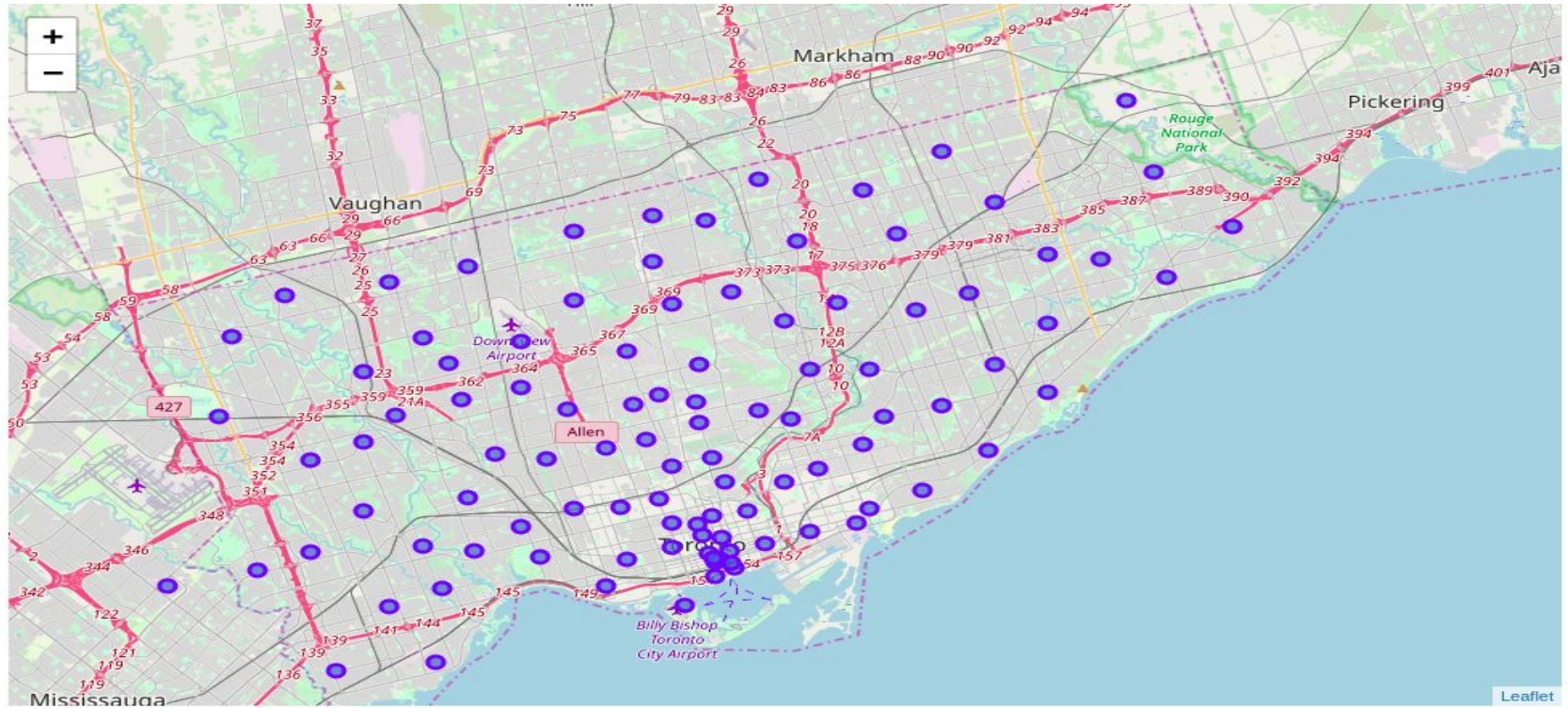
Location recommendation for business

- Location is important factor for business
    - Customer awareness
    - Transportation costs
    - Number of competitors
- Objective: Base on the data of neighborhoods in a given city, give a recommendation for a business owner.
    - City : Toronto, Scarborough
    - Business type: Asian cuisine restaurant
- Interests: Restaurants/ Business owners, bring the most benefit from location.

Data acquisitions and wrangling

- Data sources:
  - Wikipage about the postal code for neighborhoods of Toronto
  - Population per postal code on the official website of Toronto
  - Foursquare API
- Rows have missing data about borough are removed.
- Different neighborhoods share the same postal code are grouped.
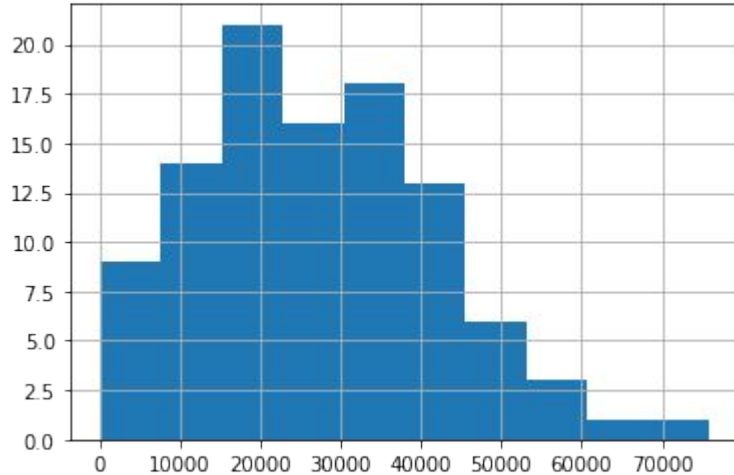- In total: 10 boroughs, 103 super neighborhoods.

EDA (1)

EDA (2)

- Toronto : a sea city, good for seafood supply
- Density of neighborhoods is not homogeneous
- Data about location, population and from Foursquare API are merged together based on the postal code.
- This project considers only Scarborough.
- Foursquare's data suggests 59 categories of venues in this borough.
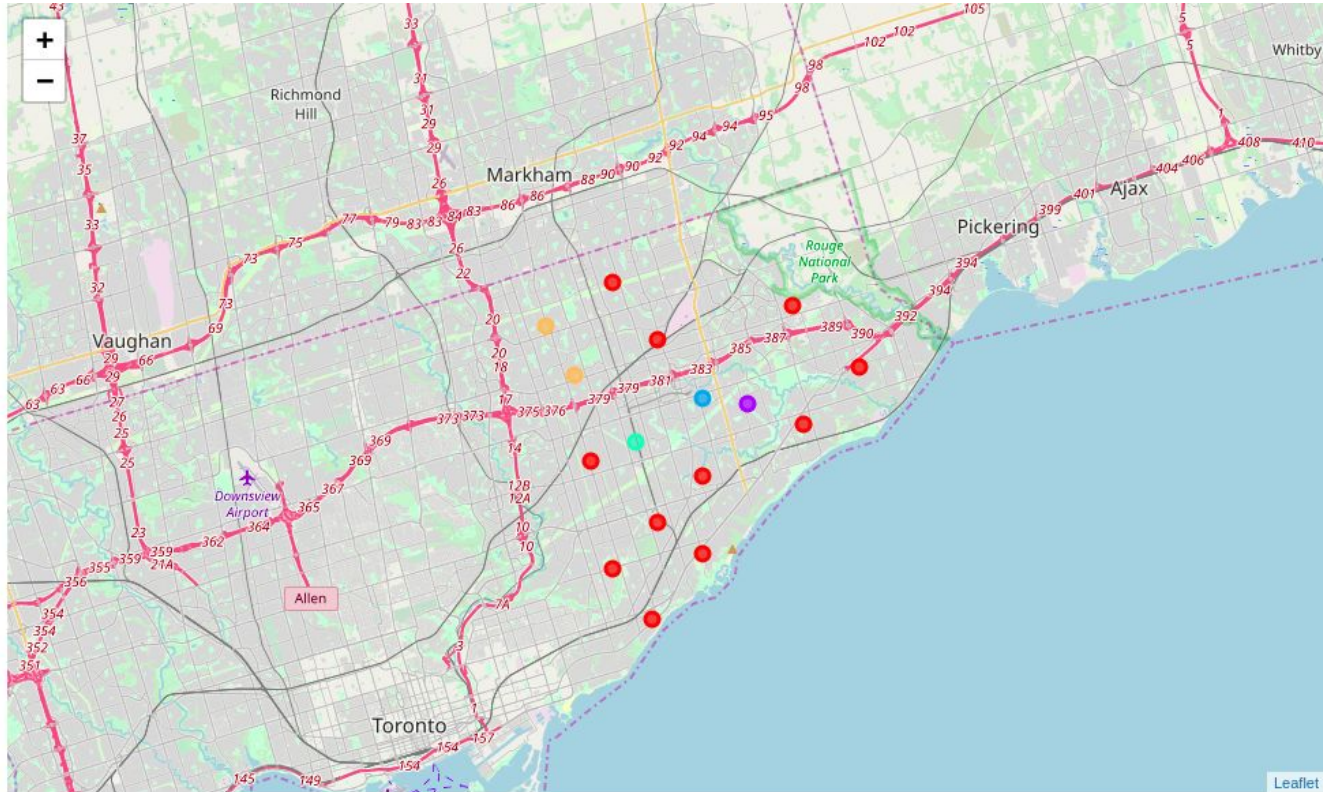
Clustering step - segmentation for super neighborhood

- Only select venues that have impacts on the Asian restaurants which is other Asian restaurants: 'Chinese Restaurant','Hakka Restaurant', 'Vietnamese Restaurant', 'Korean Restaurant', 'Thai Restaurant'
- Incorporate the population to the feature of representation to reflect the market size.

Population in Toronto

- Mostly at least population of 20000 per neighborhoods
- Good market size for business.
- Some of super neighborhood could have large population which is even better for business.

Cluster results - same color, same cluster

Cluster 1 - No Asian restaurant

	Neighborhood	Chinese Restaurant	Hakka Restaurant	Vietnamese Restaurant	Korean Restaurant	Thai Restaurant	Cluster Labels	Population
0	Agincourt	0.0	0.0	0.0	0.0	0.0	0	37769.0
3	Cliffside West,Birch Cliff	0.0	0.0	0.0	0.0	0.0	0	22136.0
5	East Birchmount Park,Ionview,Kennedy Park	0.0	0.0	0.0	0.0	0.0	0	48434.0
6	Golden Mile,Oakridge,Clairlea	0.0	0.0	0.0	0.0	0.0	0	35081.0
7	L'Amoreaux East,Milliken,Agincourt North,Steel...	0.0	0.0	0.0	0.0	0.0	0	54680.0
9	Maryvale,Wexford	0.0	0.0	0.0	0.0	0.0	0	29858.0
10	Rouge Hill,Highland Creek,Port Union	0.0	0.0	0.0	0.0	0.0	0	35626.0
11	Rouge,Malvern	0.0	0.0	0.0	0.0	0.0	0	66108.0
12	Scarborough Village	0.0	0.0	0.0	0.0	0.0	0	36699.0
13	Scarborough Village West,Cliffcrest,Cliffside	0.0	0.0	0.0	0.0	0.0	0	22913.0
14	West Hill,Morningside,Guildwood	0.0	0.0	0.0	0.0	0.0	0	46943.0

Number on this table is number of venues per 100k

Cluster 2 - only Korean restaurant 1.12 rest. /100K

	Neighborhood	Chinese Restaurant	Hakka Restaurant	Vietnamese Restaurant	Korean Restaurant	Thai Restaurant	Cluster Labels	Population
15	Woburn	0.0	0.0	0.0	1.122712	0.0	1	29690.0

Number on this table is number of venues per 100k

Cluster 3 - Only Hakka and Thai restaurants

	Neighborhood	Chinese Restaurant	Hakka Restaurant	Vietnamese Restaurant	Korean Restaurant	Thai Restaurant	Cluster Labels	Population
1	Cedarbrae	0.0	0.455691	0.0	0.0	0.455691	2	24383.0

Number on this table is number of venues per 100k

Cluster 4 - only Chinese and Vietnamese restaurants

	Neighborhood	Chinese Restaurant	Hakka Restaurant	Vietnamese Restaurant	Korean Restaurant	Thai Restaurant	Cluster Labels	Population
4	Dorset Park,Wexford Heights,Scarborough Town C...	0.313483	0.0	0.313483	0.0	0.0	3	45571.0

Number on this table is number of venues per 100k

Cluster 5 - Chinese (and Thai) restaurants

	Neighborhood	Chinese Restaurant	Hakka Restaurant	Vietnamese Restaurant	Korean Restaurant	Thai Restaurant	Cluster Labels	Population
2	Clarks Corners,Sullivan,Tam O'Shanter	0.206513	0.0	0.0	0.0	0.206513	4	34588.0
8	L'Amoreaux West	0.229232	0.0	0.0	0.0	0.000000	4	48471.0

Number on this table is number of venues per 100k

Conclusion and future directions

- Asian restaurants are not popular in Scarborough, more chance for Asian business owners.
- The best option for super neighborhood is Rouge and Malvern for their large population and lacking of Asian restaurants.
- Good clusters showed that number of venues per 100k population is good feature.
- Real project will require more features for better recommendations, current model is simple for realistic recommendation.
- This capstone project is a base for future projects.