Location selection for a bubble tea shop

Why site selection is important?

- Less competition
- More cooperation
- Target the right customers
- Brand recognition

Data acquisition and cleaning

- The postal code, borough and neighborhood of Toronto are found in Wikipedia. (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

- The latitude and longitude of each neighborhood: CSV file. (http://cocl.us/Geospatial_data)

- After the cleaning, there are 24 neighbourhoods in North York with 248 venues and there are 106 unique venue categories

Methodology

- Two factors to be considered:
  - Competition
    - If the category contains the keywords of shopping, Thai, Chinese, etc. , the cooperation number will return 1; otherwise, it will return 0.
  - Cooperation
    - If the keyword such as coffee, café, tea showed in the category, the competition number will return 1; otherwise, it will return 0.

    →extract all the neighborhood that only contains 0 competition number and 1 cooperation number.

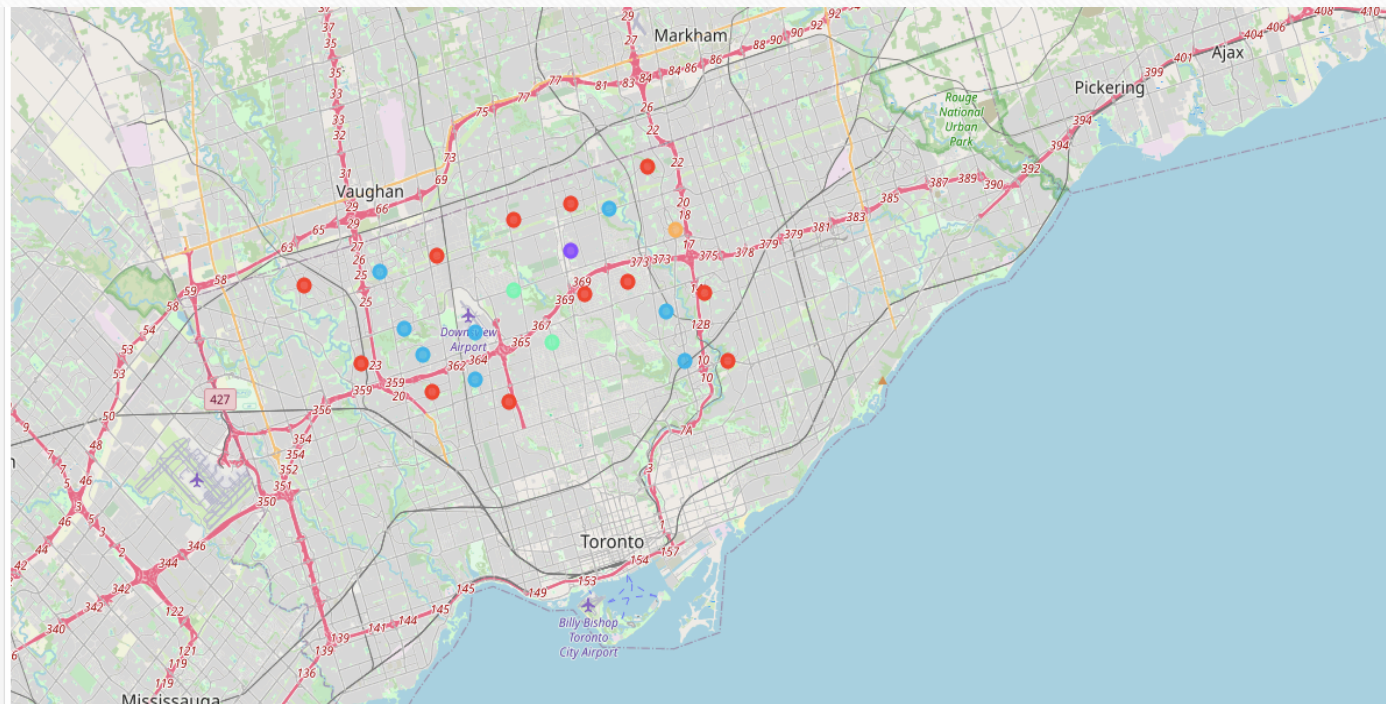   →summed up both the cooperation numbers and competition number for all the neighbourhoods in North York

	Cluster Labels	Postal Code	Borough	Neighbourhood	Latitude	Longitude	CompetitionNumber	CooperationNumber
0	0	M3A	North York	Parkwoods	43.753259	-79.329656	0	0
1	0	M4A	North York	Victoria Village	43.725882	-79.315572	0	0
2	2	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	0	1
3	2	M3B	North York	Don Mills	43.745906	-79.352188	0	1
4	0	M6B	North York	Glencairn	43.709577	-79.445073	0	0
5	2	M3C	North York	Don Mills	43.725900	-79.340923	0	1
6	0	M2H	North York	Hillcrest Village	43.803762	-79.363452	0	0
7	3	M3H	North York	Bathurst Manor, Wilson Heights, Downsview North	43.754328	-79.442259	0	2
8	3	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	0	2
9	0	M3J	North York	Northwood Park, York University	43.767980	-79.487262	0	0

Table 5: Summed up competitors and cooperation business in each neighborhood

K-Means Clusters of 5

Map of clustered neighbourhoods

Cluster 1 and 2

	Postal Code	Latitude	Longitude	CompetitionNumber	CooperationNumber
0	M3A	43.753259	-79.329656	0	0
1	M4A	43.725882	-79.315572	0	0
4	M6B	43.709577	-79.445073	0	0
6	M2H	43.803762	-79.363452	0	0
9	M3J	43.767980	-79.487262	0	0
12	M2L	43.757490	-79.374714	0	0
14	M6L	43.713756	-79.490074	0	0
15	M9L	43.756303	-79.565963	0	0
16	M2M	43.789053	-79.408493	0	0
19	M9M	43.724766	-79.532242	0	0
22	M2P	43.752758	-79.400049	0	0
23	M2R	43.782736	-79.442259	0	0

	Postal Code	Latitude	Longitude	CompetitionNumber	CooperationNumber
20	M2N	43.77012	-79.408493	0	5

Cluster 3 and 4

	Postal Code	Latitude	Longitude	CompetitionNumber	CooperationNumber
2	M6A	43.718518	-79.464763	0	1
3	M3B	43.745906	-79.352188	0	1
5	M3C	43.725900	-79.340923	0	1
10	M2K	43.786947	-79.385975	0	1
11	M3K	43.737473	-79.464763	0	1
13	M3L	43.739015	-79.506944	0	1
17	M3M	43.728496	-79.495697	0	1
21	M3N	43.761631	-79.520999	0	1

	Postal Code	Latitude	Longitude	CompetitionNumber	CooperationNumber
7	M3H	43.754328	-79.442259	0	2
18	M5M	43.733283	-79.419750	0	2

Cluster 5

	Postal Code	Latitude	Longitude	CompetitionNumber	CooperationNumber
8	M2J	43.778517	-79.346556	0	3

Conclusion

- Goal:
  - Less competitors
  - More cooperation businesses
  - →The fittest cluster is cluster 2 with 0 competitors and 5 business to cooperate with at M2N.