Deciding where open a Baseball shop in Toronto

A recurrent business problem

- The neighborhoods in a city are very diversified, but we can cluster them in a way that are possible to identify patterns in what venues are most common in each group of similar neighborhoods. This is helpful to determine where should be open a shop.

- This give the entrepreneur the best chances to have success and a better return for him investment.

In our case we analyze where should be open a Baseball shop in the city of Toronto.

Data sources

- Wikipedia page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- Geospatial data provided in the csv http://cocl.us/Geospatial_data

- Venues data from Foursquare

Segmenting and Clustering Neighborhoods in Toronto

○ We cleaned our data droping postal codes without boroughs, joining neighborhoods contained in same postal code and assigning the borough name to neighborhoods with missing name.

○ And we append the latitude and longitude to the postal codes.

○ Below a example of the data we use to search the venues:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029

Exploring the neighborhoods in Toronto

Map of Toronto with neighborhoods superimposed on top:

# We use Foursquare to request a exploratory search of venues in the neighborhoods

○ We worked with requested data from Foursquare and analyze each neighborhood based on most common venue categories like example below:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Steakhouse	Thai Restaurant	Bar	Café
1	Agincourt	Lounge	Sandwich Place	Breakfast Spot	Skating Rink	Drugstore
2	Agincourt North, L'Amoreaux East, Milliken, St...	Playground	Park	Yoga Studio	Drugstore	Dim Sum Restaurant
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	Pizza Place	Fried Chicken Joint	Coffee Shop	Sandwich Place
4	Alderwood, Long Branch	Pizza Place	Skating Rink	Dance Studio	Pharmacy	Coffee Shop

Clustering neighborhoods using the k-means methodology

- We use a machine learning algorithm to cluster the neighborhoods based in the categories of venues contained in each of them to better understand how similar are the neighborhoods.

- We used k-means to form 5 clusters, so we can take a look and decide latter where a Baseball shop should be open to have the most likely public based on categories of venues more frequently in a cluster.

Map with resulting clusters in Toronto
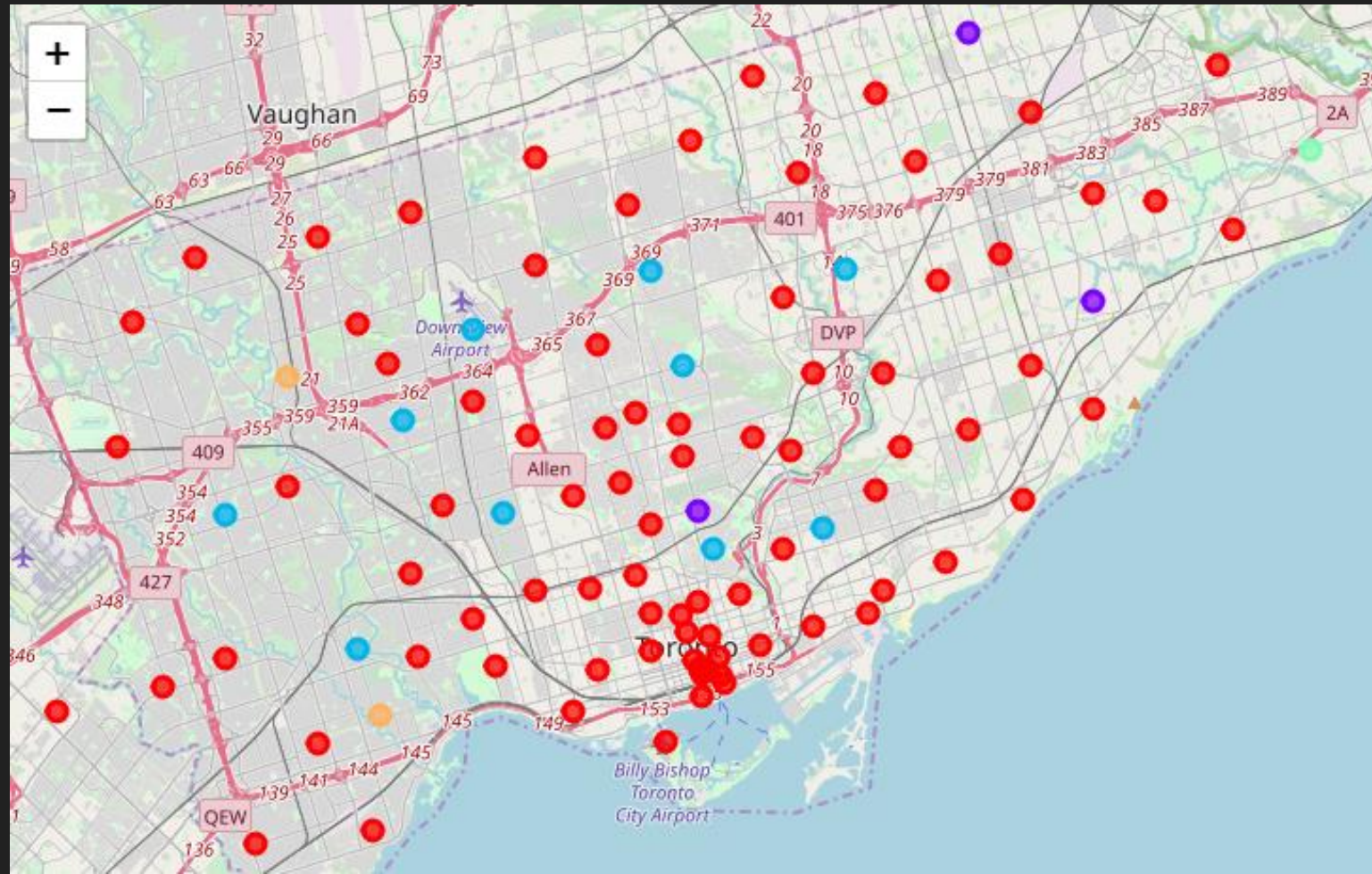
Results

- In the resulting map we can see that are a very large cluster, one medium size and other three are small clusters.

- We need to examine clusters to understand what are the similarities between each group.

- Next we will present each cluster and the most common venue categories contained in the neighborhoods that distinguish each cluster.

Cluster 1

Example of five neighborhoods of Cluster 1 with respective top 5 common venues:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Rouge, Malvern	0	Fast Food Restaurant	Dumpling Restaurant	Diner	Discount Store	Dog Run
2	Guildwood, Morningside, West Hill	0	Pizza Place	Spa	Electronics Store	Breakfast Spot	Rental Car Location
3	Woburn	0	Coffee Shop	Korean Restaurant	Yoga Studio	Dumpling Restaurant	Discount Store
4	Cedarbrae	0	Hakka Restaurant	Caribbean Restaurant	Bank	Bakery	Fried Chicken Joint
6	East Birchmount Park, Ionview, Kennedy Park	0	Discount Store	Chinese Restaurant	Bus Station	Department Store	Coffee Shop

Cluster 2

The 3 neighborhoods of Cluster 2 with respective top 5 common venues:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Scarborough Village	1	Playground	Drugstore	Dim Sum Restaurant	Diner	Discount Store
14	Agincourt North, L'Amoreaux East, Milliken, St...	1	Playground	Park	Yoga Studio	Drugstore	Dim Sum Restaurant
48	Moore Park, Summerhill East	1	Playground	Gym	Drugstore	Dim Sum Restaurant	Diner

Cluster 3

Example of five neighborhoods of Cluster 3 with respective top 5 common venues:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
23	York Mills West	2	Park	Bank	Yoga Studio	Dumpling Restaurant	Discount Store
25	Parkwoods	2	Fast Food Restaurant	Food & Drink Shop	Park	Diner	Discount Store
30	CFB Toronto, Downsview East	2	Bus Stop	Airport	Park	Yoga Studio	Dumpling Restaurant
40	East Toronto	2	Park	Coffee Shop	Convenience Store	Yoga Studio	Dumpling Restaurant
44	Lawrence Park	2	Gym / Fitness Center	Park	Bus Line	Swim School	Yoga Studio

Cluster 4

The unique neighborhood of Cluster 4 with respective top 5 common venues:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Highland Creek, Rouge Hill, Port Union	3	Bar	Yoga Studio	Discount Store	Dog Run	Doner Restaurant

Cluster 5

The two neighborhoods of Cluster 5 with respective top 5 common venues:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
91	Humber Bay, King's Mill Park, Kingsway Park So...	4	Baseball Field	Pool	Yoga Studio	Diner	Discount Store
97	Emery, Humberlea	4	Baseball Field	Paper / Office Supplies Store	Yoga Studio	Discount Store	Dog Run

Insights

- CLUSTER 1: we observe that the cluster number 1 is the largest of them, so is hard to find particular venues that determines the link to this group. In the others clusters we have less neighborhoods and is possible see more clearly what venues is more common between the group

- CLUSTER 2: is distinguished by playgrounds and drugstore

- CLUSTER 3: is more common to a neighborhood have park, yoga studio, discount store and dog run

- CLUSTER 4: the most common venue is bars

- CLUSTER 5: we have the baseball field as most common venue of the group and second most common is yoga studio

So where is the better city area to open a Baseball shop?

We can properly say that the better area in the city of Toronto to open this shop is in any neighborhood of cluster number 5

Conclusion

○ Clustering neighborhoods of a city considering the most common venues categories is a very powerful tool to understanding how similar or dissimilar are the areas. Using this we can decide for example where should be open a new restaurant or barber shop in the city improving a business to have better chances to success.

○ We use the k-means to grouping the neighborhoods based in their similarities.

○ We conclude the answer is cluster number 5 because there are baseball fieds as most common venues and this is a attraction for public that are potential clients for the new Baseball shop, although the entrepreneur will have better financial expectatives, resulting in more salles and consequently more revenues.

Possible improvements to this analyzes

- A more accurate result is possible searching for data of other Baseball shops and making linear regressions with the revenues of this shops and the venues categories most commons in the respective neighborhoods of this shops.

- This way would tell what venues categories contribute more on less in revenues to a Baseball shop.