A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green. They are positioned diagonally, with the blue one partially covering the green one.

Capstone Project The Battle of the Neighborhoods

Bar location analysis



Introduction

1. Bars are usually co-localized with other business like restaurants, dancing clubs and other bars as well.
2. Many cities have bars that are spatially clustered together in some streets, neighborhoods or districts.
3. Here, I clusterize the venues in Toronto, find their location, size and their venue composition in order to present options to clients who want to invest in a bar business.



Data acquisition

Data were scrapped from:

- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- https://cocl.us/Geospatial_data
- Foursquare API.

Venues in a radius of 500 m from each neighbourhood in Toronto were collected into a dataframe. This dataframe was then filtered using keywords from the 'Venue Category' column and further categorized into three broader categories (Drinks, Foods and Music).



Analytical methods and data representation

- K-means and DBSCAN unsupervised algorithms were used for clustering.
- Folium map was used to visualize venues and clusters on the Toronto map.
- For each cluster the mean latitude and longitude, area and venue composition was calculated.
- Bar plots were used to represent the data.



Results

Toronto's neighbourhoods postal codes were collected

	Postal Code	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government



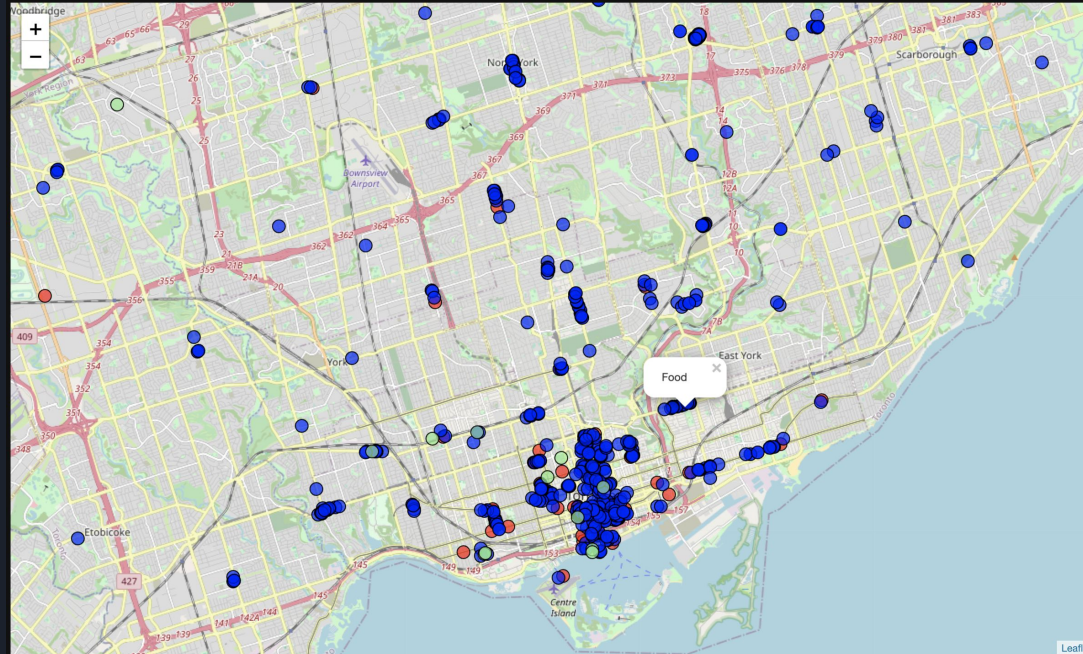
Results

Postal code data were used to match for spatial coordinates

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Drinks_Food_Music	Drinks_Food_Music_Code
0	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	Bar Volo	43.665462	-79.385692	Beer Bar	Drinks	0
1	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	SUDS	43.659880	-79.394712	Bar	Drinks	0
2	Garden District, Ryerson	43.657162	-79.378937	Duke's Refresher + Bar	43.658980	-79.382949	Bar	Drinks	0
3	Garden District, Ryerson	43.657162	-79.378937	Reds Midtown Tavern	43.659128	-79.382266	Wine Bar	Drinks	0
4	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar	Drinks	0

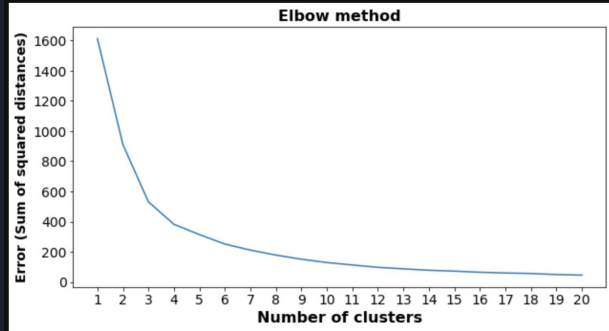
Results

Drinks, Food and Music venues were color coded and visualized on the Toronto map

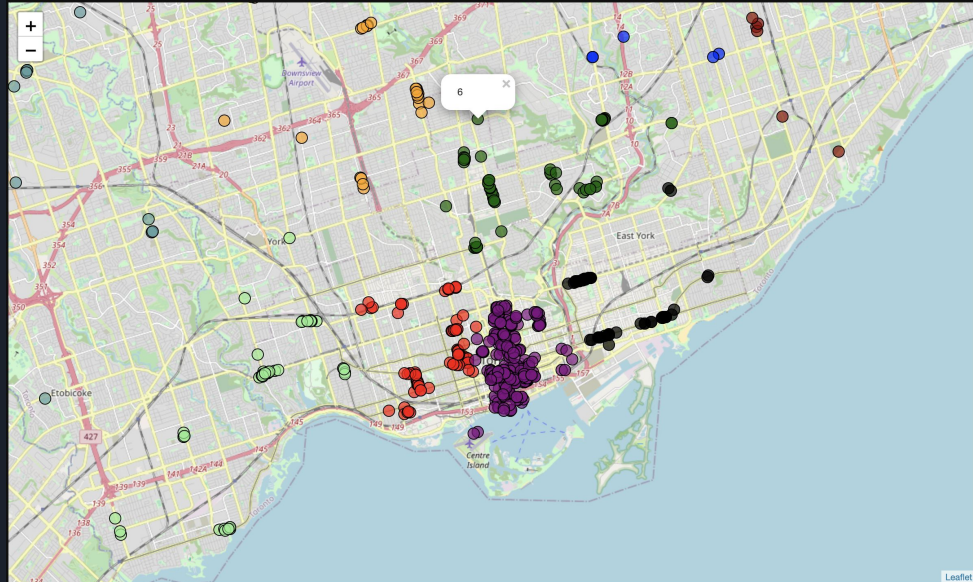


Results

To clusterize all venues, K-means algorithm was applied



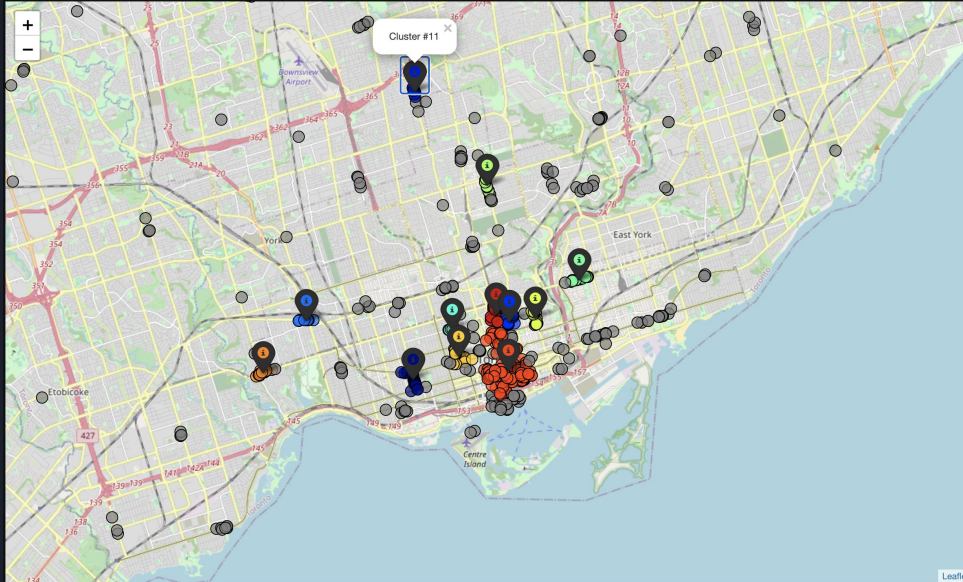
The elbow method was used to find the optimal number of clusters



The clusters were too extended. This strategy was not satisfactory.

Results

DBSCAN algorithm was used afterwards



The clusters are much more compact and therefore better suited to be places where clients can walk from one venue to another.



Results

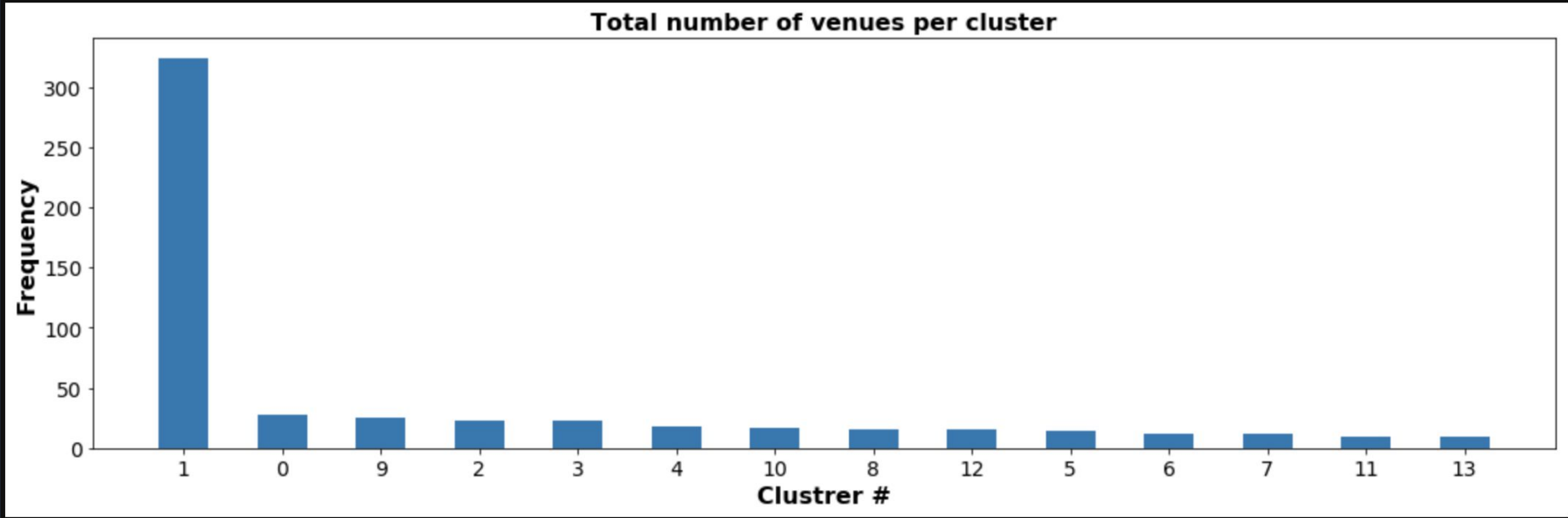
Relevant data from each cluster calculated and collected

DBSCAN_label	venue_count	mean_latitude	mean_longitude	Coordinates	Polygon	Area
6	6	12	43.665356	-79.465531	[[43.66529624102518, -79.46511783107383], [43....	POLYGON ((43.66533428942121 -79.46625298427486... 1.915263e-07
7	7	12	43.662697	-79.403960	[[43.66244824890102, -79.40470324675829], [43....	POLYGON ((43.66271377074943 -79.40433001263669... 3.105787e-07
4	4	18	43.677867	-79.350489	[[43.677438, -79.352683], [43.67733505610778, ...	(POLYGON ((43.6774845518262 -79.35307562861175... 6.614821e-07
11	11	10	43.735046	-79.419911	[[43.73517255469138, -79.41970208136466], [43....	POLYGON ((43.73619518432498 -79.42027081330127... 7.692195e-07
13	13	10	43.706572	-79.389300	[[43.70498511291793, -79.38847622354874], [43....	(POLYGON ((43.70808171836455 -79.3899946724600... 8.255057e-07

	DBSCAN_label	Drinks	Food	Music
1	0	0.107143	0.892857	0.000000
2	1	0.185185	0.808642	0.006173
3	2	0.130435	0.869565	0.000000
4	3	0.260870	0.739130	0.000000
5	4	0.111111	0.888889	0.000000

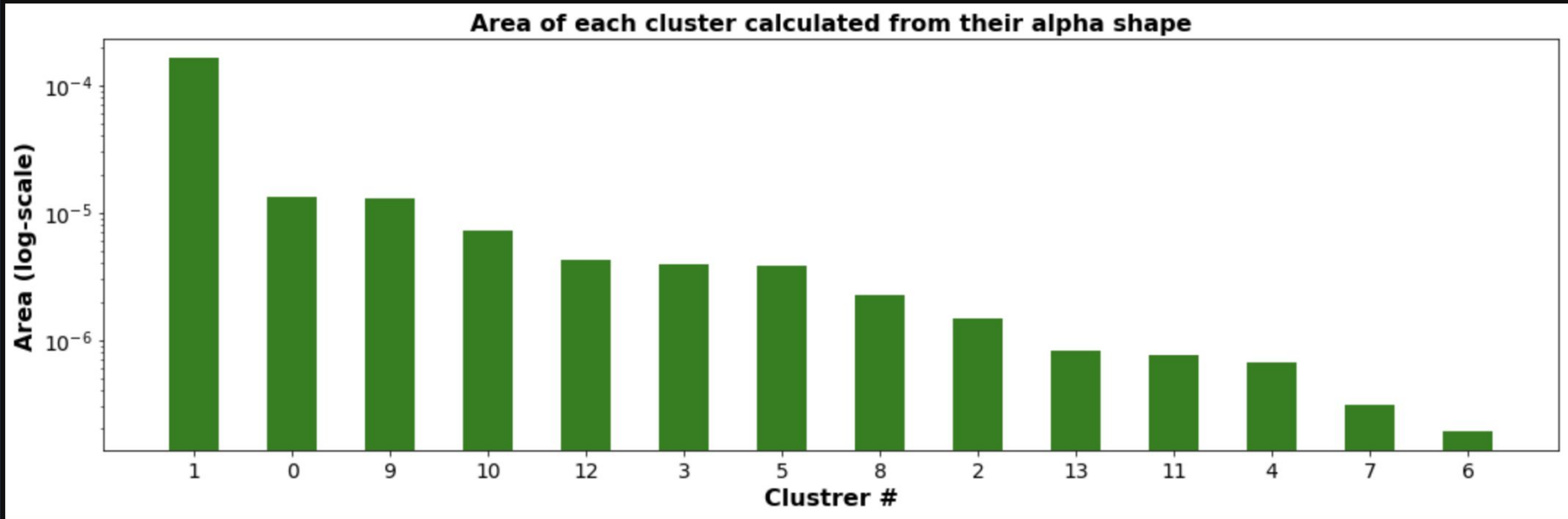
Results

Frequency analysis: Total number of venues per cluster



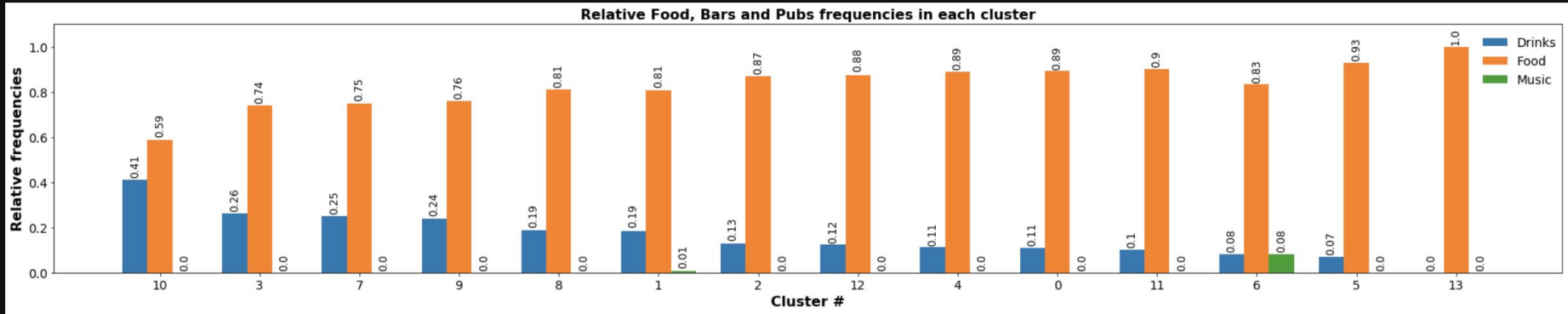
Results

Size of each cluster



Results

Relative frequency analysis of Drinks, Food and Music categories for each cluster





Results

Relative frequency analysis of other categories showing the ten more frequent venues per cluster

DBSCAN_label		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
1	0	Sushi Restaurant	Japanese Restaurant	Beer Bar	Mediterranean Restaurant	Chinese Restaurant	Persian Restaurant	Restaurant	Diner	Mexican Restaurant	Indian Restaurant
2	1	Restaurant	Seafood Restaurant	Italian Restaurant	Japanese Restaurant	American Restaurant	Beer Bar	Thai Restaurant	Bar	Fast Food Restaurant	Pub
3	2	Fast Food Restaurant	Japanese Restaurant	Restaurant	Juice Bar	Food Court	Burrito Place	Chinese Restaurant	American Restaurant	Asian Restaurant	Bar
4	3	Bar	Restaurant	Vietnamese Restaurant	Asian Restaurant	Italian Restaurant	Cuban Restaurant	Malay Restaurant	Cocktail Bar	New American Restaurant	Korean Restaurant
5	4	Greek Restaurant	Italian Restaurant	Restaurant	Juice Bar	Indian Restaurant	Sushi Restaurant	Caribbean Restaurant	Pizza Place	Pub	Falafel Restaurant



Results

Relative frequency analysis of other categories showing the ten more frequent venues per cluster

DBSCAN_label		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
1	0	Sushi Restaurant	Japanese Restaurant	Beer Bar	Mediterranean Restaurant	Chinese Restaurant	Persian Restaurant	Restaurant	Diner	Mexican Restaurant	Indian Restaurant
2	1	Restaurant	Seafood Restaurant	Italian Restaurant	Japanese Restaurant	American Restaurant	Beer Bar	Thai Restaurant	Bar	Fast Food Restaurant	Pub
3	2	Fast Food Restaurant	Japanese Restaurant	Restaurant	Juice Bar	Food Court	Burrito Place	Chinese Restaurant	American Restaurant	Asian Restaurant	Bar
4	3	Bar	Restaurant	Vietnamese Restaurant	Asian Restaurant	Italian Restaurant	Cuban Restaurant	Malay Restaurant	Cocktail Bar	New American Restaurant	Korean Restaurant
5	4	Greek Restaurant	Italian Restaurant	Restaurant	Juice Bar	Indian Restaurant	Sushi Restaurant	Caribbean Restaurant	Pizza Place	Pub	Falafel Restaurant



Conclusions

- Clusters with bars, restaurants and music venues were identified in Toronto city.
- Each cluster present different size, shape and venue composition.
- The size of the clusters was highly variable as shown in the total amount of venues per cluster, ranging from 10 to more than 300 and in the cluster areas determination. There are some clusters more enriched in bars and pubs, like cluster number 10, while others seems to be predominantly populated by food venues. Music venues seems to be scarce and only the cluster 6 present music venues.



Perspectives

- The client will be able to select in which part of the city he/she wishes to invest in a bar according to this findings. Evenmore, depending on the composition of the venues provided, the type of bar can be finely tuned to fit in the cluster surroundings.
- This type of analysis can be performed for other cities as well following a similar pipeline.