Capstone - Ice Cream Analysis

1. Introduction

1.1. Description & Discussion of the Background

Toronto is one of the largest cities in North America. It is commonlu ranked as a "global city" [1], and is considered Canada's financial and economic hub. With a population of over 6 million people, many of which come from all over the world, Toronto has an evolving food scene. By the City of Toronto estimates, there are about 7500 restaurants [2] in the city. Needless to say, the restaurant industry can be competitive.

When setting up a restaurant, it's important to have great food, but also choose a great location. Finding a "good" location can be tricky, you don't want to be too densely populated with competitors, but you also want to be in an area where there is a certain amount of density that will draw customers in (ie the food neighbours). My friend operates an ice cream company in Toronto, eventually they may want to set up a physical location. I will try to identify the best neighbourhood to establish a restaurant in Toronto. Understanding which neighbourhoods and general area may be useful in choosing their future location based on competitor density and overall restaurant density. Most of the time when deciding a new location, a restaurant owner will use gut feeling and the direction of a commercial real estate agent, but we will use data to determine what's best. Keep in mind, many layers can be added to this analysis, but for simplicity's sake we will exclude property pricing from this analysis. As they say "you have to spend money to make money", we will remain under the assumption that if a restaurant owner chooses the right location, the cost will justify the traffic.

1.2. Data Description

To tackle this analysis we will utilize the datasets listed below:

- I used the postal code/neighbourhood data that can be found on Wikipedia, this data was used for grouping and clustering food venues [3]
- I also used a verified CSV with Toronto postal codes and their corresponding longitude and latitudes to plot the postal code clusters [4]
- I used the Foursquare API to create a list of the food venues in Toronto, the coordinates and restaurant type were used to plot and compare density[5].
- I found geospatial data for Toronto that allowed me to overlay colour on my map in order to better distinguish the clusters [6]

2. Methodology

I started the analysis by creating a rough course of action. This was my original plan:

1. Scape neighbourhood/postal code data for Toronto and surrounding areas
2. Run a cluster analysis to group them into higher level buckets
3. Pull restaurant and ice cream shop data from Foursquare
4. Join neighbourhood/cluster data with restaurant data
5. Analyze which neighbourhoods have many restaurants but few ice cream shops

Disclaimer: The intent of this analysis is not to select a specific street address but rather to gain a better understanding of which neighbourhood and clusters in which my friend should or shouldn't explore opening an ice cream shop.

2.1 Scraping

I began by scraping the Postal Code, Borough and Neighbourhood data from Wikipedia

	Postal Code	Borough	Neighbourhood
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
...	...	...	...
160	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
165	M4Y	Downtown Toronto	Church and Wellesley
168	M7Y	East Toronto	Business reply mail Processing Centre, South C...
169	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
178	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...

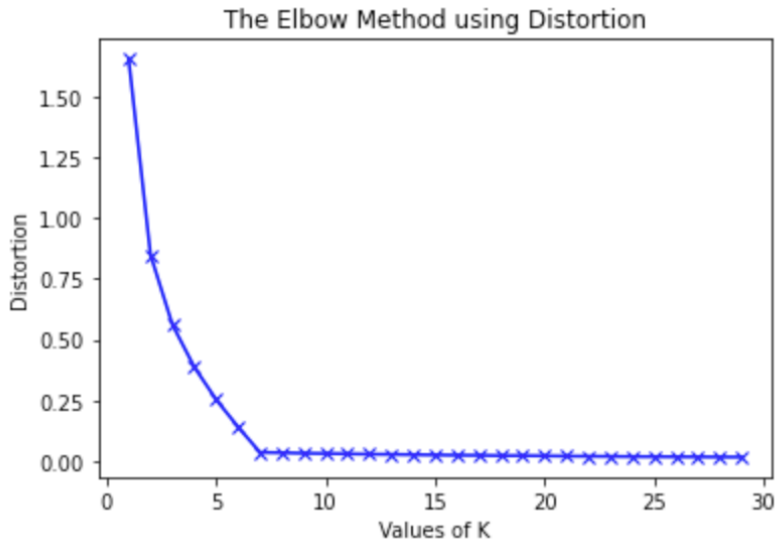
103 rows × 3 columns

I then used the geospatial data to add the longitude and latitude data to the dataframe.

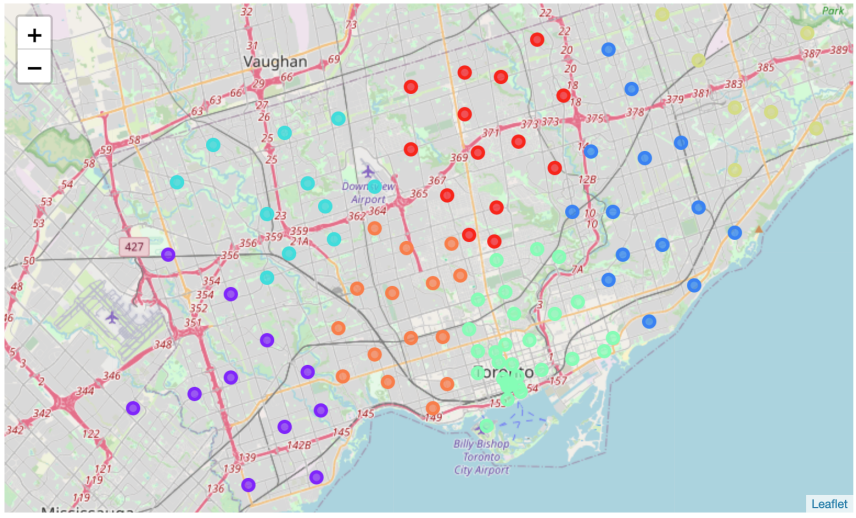
	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
...	...	...	...	...	...
98	M9N	York	Weston	43.706876	-79.518188
99	M9P	Etobicoke	Westmount	43.696319	-79.532242
100	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...	43.688905	-79.554724
101	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...	43.739416	-79.588437
102	M9W	Etobicoke	Northwest, West Humber - Clairville	43.706748	-79.594054

103 rows × 5 columns

The next step was to perform a cluster analysis on the neighbourhoods using the k-means method. I used the elbow method to determine how many clusters to use.



It appears the inflection point occurs around 7. After performing a Kmeans cluster analysis, I plotted the neighbourhoods to a map using the Python Folium library. I used latitude and longitude values to plot the map below:



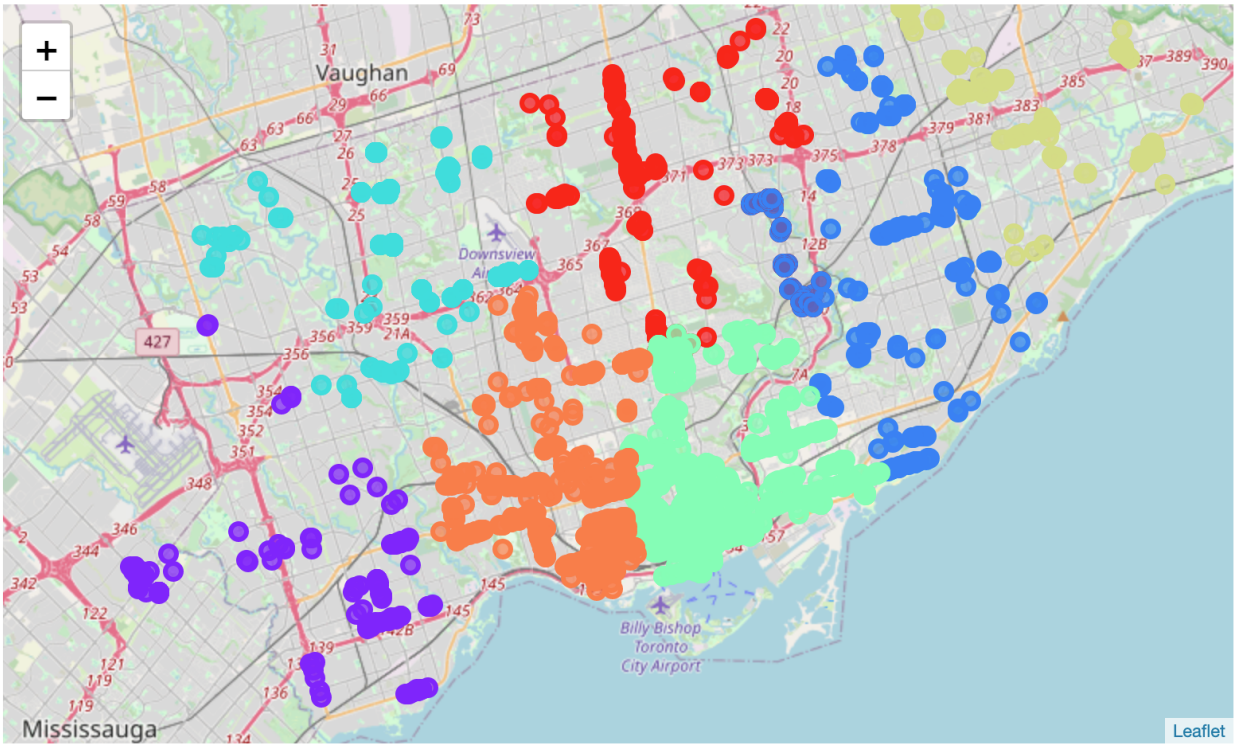
2.2. Pulling Restaurant and Ice Cream Shop Data

I used the Foursquare API to pull venues in each neighbourhood. I set the radius to 1KM and limited the search to only food venues. Using a 'for' statement, I ran through all the neighbourhoods, extracting the venue name, coordinates and category from each result. After completing the extraction, there were 3733 results returned. Here is a sample of the data that I extracted:

	Neighbourhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Malvern, Rouge	43.806686	-79.194353	Harvey's	43.800020	-79.198307	Restaurant
1	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.802008	-79.198080	Fast Food Restaurant
2	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
3	Malvern, Rouge	43.806686	-79.194353	Caribbean Wave	43.798558	-79.195777	Caribbean Restaurant
4	Malvern, Rouge	43.806686	-79.194353	Bella's Lechon	43.801291	-79.198378	Filipino Restaurant

Note: Interestingly enough this 3733 total food venues differs from the estimated 7500 restaurants from the City of Toronto. This variance isn't surprising because the number "includes licenced, non-licenced, nightclubs & bars in the City of Toronto." Additionally, due to the limitations of Foursquare's API, the results for each neighbourhood may not include EVERY single restaurant. Foursquare's API doesn't have pagination, which allows you to pull results page by page past the limit.

I merged the data with the cluster data I previously created, joining it on the neighbourhood key. I decided to cluster based on neighbourhood instead of creating restaurant clusters because due to the nature of the city, it's likely there would be too many clusters to be useful. The neighbourhood cluster data would be used to colour code the restaurants on the map. Here is a the quickly plotted map:



As a quick check, I performed an analysis of the unique food categories that were pulled from the Foursquare API.

```
There are 114 unique categories.
array(['Restaurant', 'Fast Food Restaurant', 'Caribbean Restaurant',
       'Filipino Restaurant', 'African Restaurant', 'Chinese Restaurant',
       'Greek Restaurant', 'Bakery', 'Sandwich Place',
       'Italian Restaurant', 'Burger Joint', 'Breakfast Spot',
       'Fried Chicken Joint', 'Pizza Place', 'Mexican Restaurant',
       'Donut Shop', 'Fish & Chips Shop', 'Indian Restaurant',
       'Hakka Restaurant', 'Thai Restaurant', 'Wings Joint',
       'Asian Restaurant', 'German Restaurant', 'Japanese Restaurant',
       'Café', 'Diner', 'Vietnamese Restaurant', 'American Restaurant',
       'Korean Restaurant', 'Middle Eastern Restaurant',
       'Seafood Restaurant', 'Indian Chinese Restaurant',
       'Sri Lankan Restaurant', 'Noodle House', 'Cantonese Restaurant',
       'Malay Restaurant', 'Latin American Restaurant',
       'Sushi Restaurant', 'Mediterranean Restaurant',
       'Shanghai Restaurant', 'Hong Kong Restaurant',
       'Dim Sum Restaurant', 'BBQ Joint', 'Taiwanese Restaurant',
       'Deli / Bodega', 'Vegetarian / Vegan Restaurant', 'Food Court',
       'Udon Restaurant', 'Hotpot Restaurant', 'Burrito Place',
       'Snack Place', 'Cafeteria', 'Hot Dog Joint', 'Halal Restaurant',
       'Ramen Restaurant', 'Dumpling Restaurant', 'Steakhouse',
       'Bagel Shop', 'French Restaurant', 'Eastern European Restaurant',
       'Salad Place', 'New American Restaurant', 'Falafel Restaurant',
       'Doner Restaurant', 'Turkish Restaurant', 'Food Truck',
       'Portuguese Restaurant', 'Gastropub', 'Peking Duck Restaurant',
       'Afghan Restaurant', 'Ethiopian Restaurant', 'Souvlaki Shop',
       'Tapas Restaurant', 'Cuban Restaurant', 'Tibetan Restaurant',
       'Churrascaria', 'Pakistani Restaurant', 'Bistro', 'Food',
       'Taco Place', 'Comfort Food Restaurant', 'Pide Place',
       'Poutine Place', 'Indonesian Restaurant', 'Syrian Restaurant',
       'Buffet', 'Irish Pub', 'Modern European Restaurant',
       'Theme Restaurant', 'Creperie', 'Persian Restaurant',
       'Polish Restaurant', 'Spanish Restaurant', 'Poke Place',
       'Molecular Gastronomy Restaurant', 'Gluten-free Restaurant',
       'Belgian Restaurant', 'Moroccan Restaurant',
       'Colombian Restaurant', 'Brazilian Restaurant', 'Soup Place',
       'Jewish Restaurant', 'Empanada Restaurant', 'Arepa Restaurant',
       'Israeli Restaurant', 'South American Restaurant',
       'Mac & Cheese Joint', 'Argentinian Restaurant',
       'Southern / Soul Food Restaurant', 'Tex-Mex Restaurant',
       'Hawaiian Restaurant', 'North Indian Restaurant',
       'Cajun / Creole Restaurant', 'Cambodian Restaurant'], dtype=object)
```

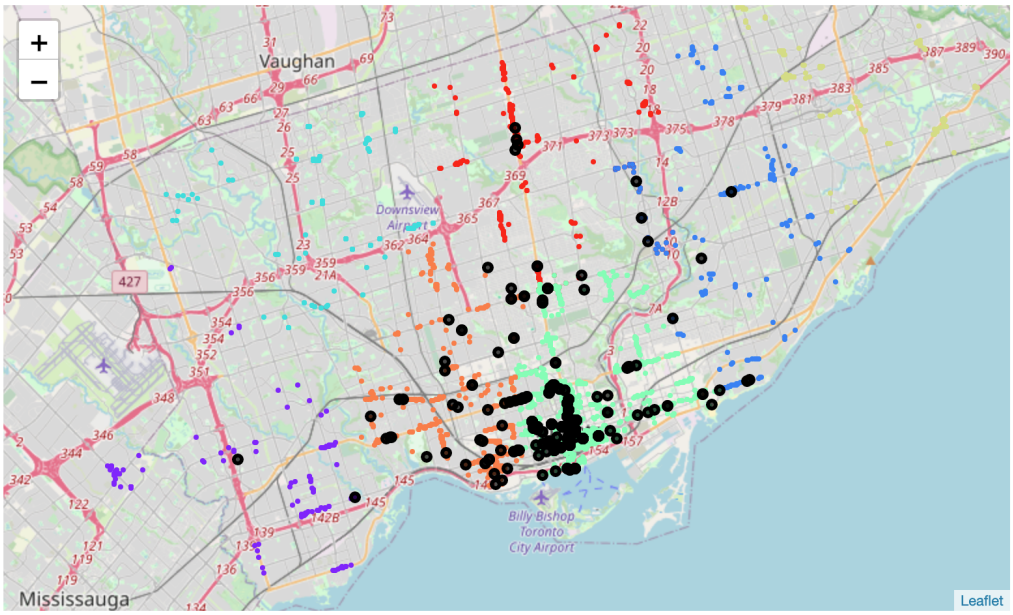
Upon further investigation, it appears that "Ice Cream Shops" are not a category listed. This was the first major obstacle I encountered in the analysis. At this point, I figured I'd run another pull from Foursquare that was only for ice cream. Using the category ID found here [7], I extracted all the Ice Cream Shops in Toronto which yielded 463 results. Here are a list of the unique names:

```
There are 130 unique Ice Cream Shops.
array(["McDonald's", 'Dairy Queen', 'Extreme Cream', 'Baskin-Robbins',
       'Tsujiri', 'carpriccio', 'Sweet Jesus', 'Nanashake',
       'Wooffles & Cream', 'Kitaya', "Ice 'n Cake", 'Dairy Queen (Treat)',
       'Tropical Treets', 'Baskin Robins', 'Lickadee Split', 'La Diperie',
       "Ed's Real Scoop", 'Hollywood Gelato', 'Candy Land',
       'Dolce Gelato', 'Zaza Espresso Bar', "Menchie's",
       "Brett's Ice Cream", 'The Lemonade Stand', 'Mister Soft Twist',
       'Piccolina Gelato E Altra Roba', 'Milky Way Ice Cream Parlour',
       'Milkcow', 'Les Epicuriens', 'Punto Gelato',
       'Gelato Simply Italian', 'Village Chill',
       'My Favourite Ice Cream Shop', 'Park Snacks', 'Grinning Face',
       'The Lansdowne Cone', 'Nani’s Gelato', 'Lab Sense',
       'The Gelato Shop', 'Zaza Espresso Bar & Gelato',
       'Il Gelato Artigianale', "Summer's Homemade Ice Cream",
       'Mr. Soft Treats', 'W Burger Bar', 'Gelato', "Menchies's",
       'Yogen-Fruz', 'Punto Gelato, Simply Italian', "Ben & Jerry's",
       'Cumberland Cafe', 'Mister Softee (G. Demos)', 'Roselle Desserts',
       'Foods88', 'SOMA chocolatemaker', 'Froshberg Gelato',
       'Taiyaki NYC - Toronto',
       'Master Soft & Delight Dairy Products Inc.',
       'Dairy Queen / Orange Julius', 'OMG Its Yogurt', 'Laura Secord',
       'Fugo Desserts', 'The Fix', 'Arctic Bites', 'Berry Mix',
       'Death In Venice', '00 Gelato', 'Mr. Soft And Delight',
       'Dairy Belle - Metro Square', 'Bow Tie Ice Cream Truck',
       'Lick It Gelato', 'Chavo Gelato House', 'Ice Creamonology',
       'Hotel Gelato', "Menchie's St. Clair West", 'Not Just Yogurt',
       'Cinavro', 'Village Chill 2', "Eva's Original Chimneys", 'Cefiore',
       'Sweet Fantasies', 'Frosty Roll', 'Nice Pan', 'Koishi', 'Eative',
       'Mr. Cream', 'Wrestlers', 'Dolce Gelato Kensington',
       'Mizzica Gelateria', 'Cool N2', 'Kekou Gelato House',
       'Ice Cream Cafe', 'Scoop Moovement', "Yogurty's", 'Palettamerica',
       'Nez', 'Jr Sweets', 'Erozen Yorguts', 'Ice Cream Truck',
       'The Milkshake Factory', 'Put A Cone On It', 'Poop Café',
       'Dairy Village', 'Gelateria', 'Knockout Ice Cream',
       'Dairy Belle Ice Cream Truck', 'La Paloma Gelateria & Cafe',
       'The Landsdown Scoop', 'Bang Bang Ice Cream & Bakery',
       'Death in Venice Gelato', 'iHalo Krunch', 'Sukoi',
       'Sicilian Sidewalk Cafe', 'White Squirrel', 'Sukoi Desserts',
       'Hollywood Cone', 'Twisted Shakes', 'The Port', 'BeaverTails',
       'The Perfect Scoop', 'Kobo Ice Cream Tuesdays', 'Star Ice Cream',
       'Ice Cream Junction', 'Madonna Mia', 'Mrs. Softy', 'Simply Yummy',
       'Baby Point Ice Cream & Espresso Bar', 'Ci Gusta',
       "Tom's Dairy Freeze", 'Yogen Früz', 'Red Mango'], dtype=object)
```

As we can see, there are several chain ice cream shops included. For the sake of this analysis, I've removed McDonald's, Dairy Queen, Baskin-Robbins, Yogen Fruz and Ben & Jerry's as these chain shops likely have a different demographic from boutique ice cream shops like my friend might set-up.

	Neighbourhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory	Cluster Labels
0	Malvern, Rouge	43.806686	-79.194353	Harvey's	43.800020	-79.198307	Restaurant	5
1	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.802008	-79.198080	Fast Food Restaurant	5
2	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant	5
3	Malvern, Rouge	43.806686	-79.194353	Caribbean Wave	43.798558	-79.195777	Caribbean Restaurant	5
4	Malvern, Rouge	43.806686	-79.194353	Bella's Lechon	43.801291	-79.198378	Filipino Restaurant	5

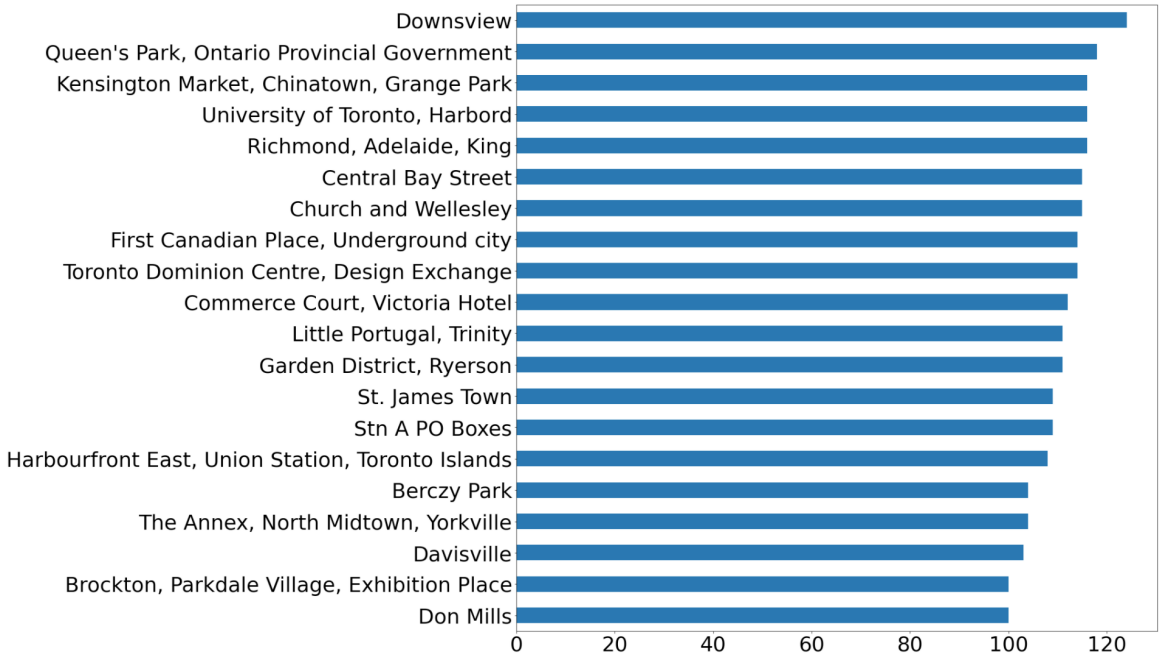
After removing the chain ice cream stores, I plotted the ice cream shops over the restaurants on the map. Now we are starting to get a visualization of areas of high restaurant density and areas of high ice cream shop density. You can see it below:



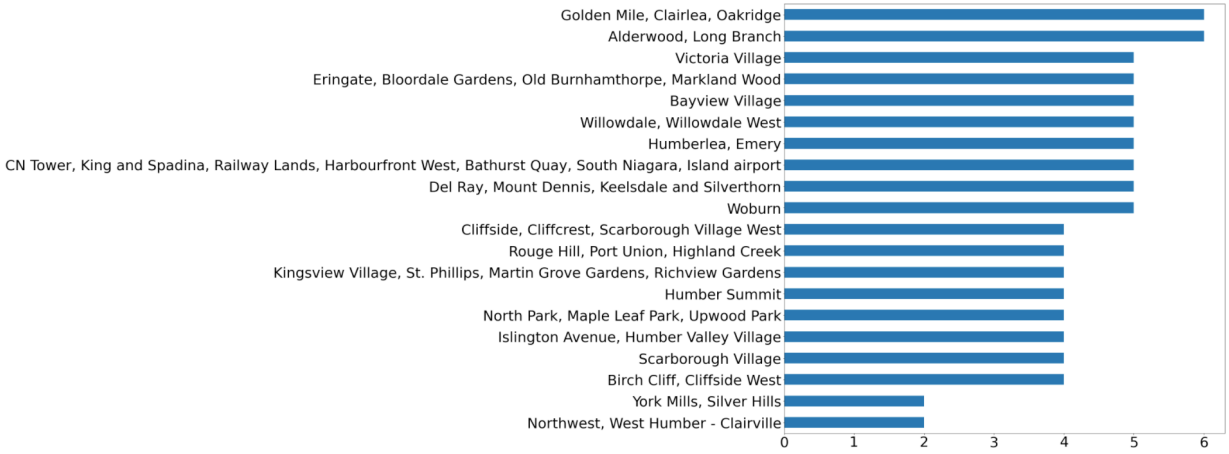
2.3. Digging into the numbers

After plotting the map we can see certain clusters and neighbourhoods have a higher density of restaurants as well as ice cream shops. I decided to run a quick analysis on the following:

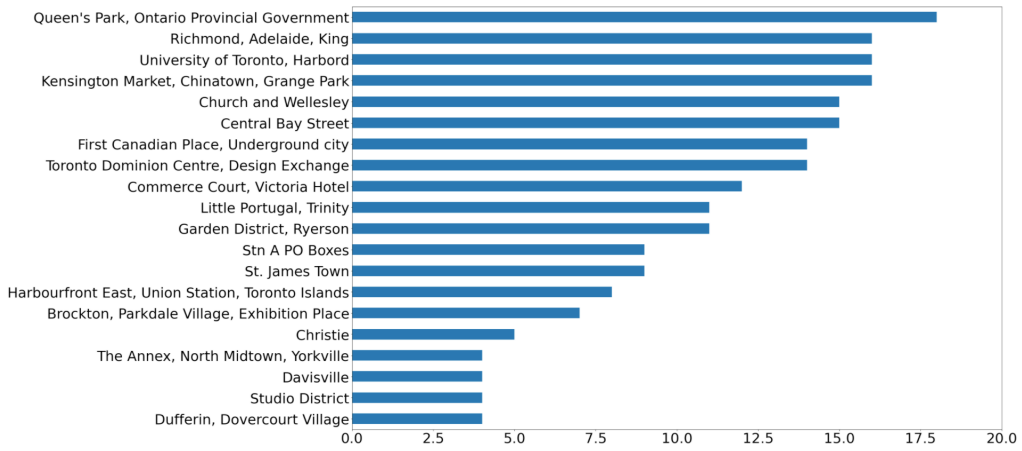
1. Neighbourhoods with the most restaurants



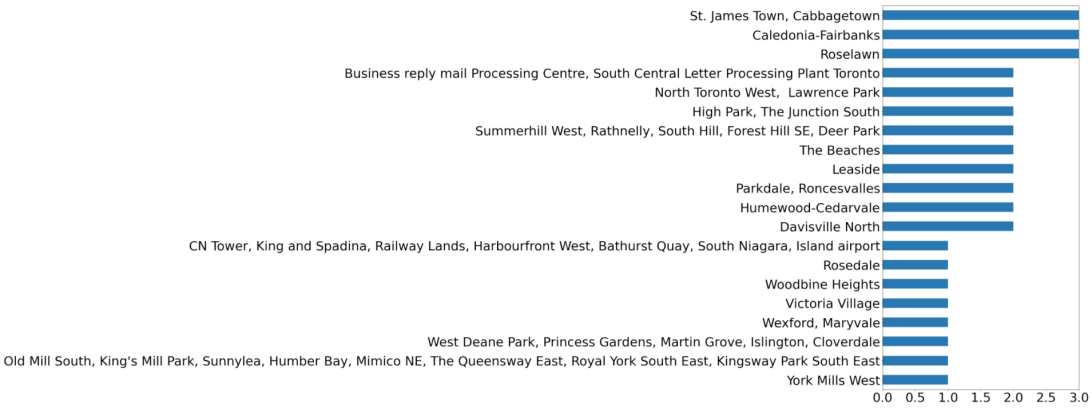
2. Neighbourhoods with least restaurants



3. Neighbourhoods with the most ice cream shops



4. Neighbourhoods with least ice cream shops

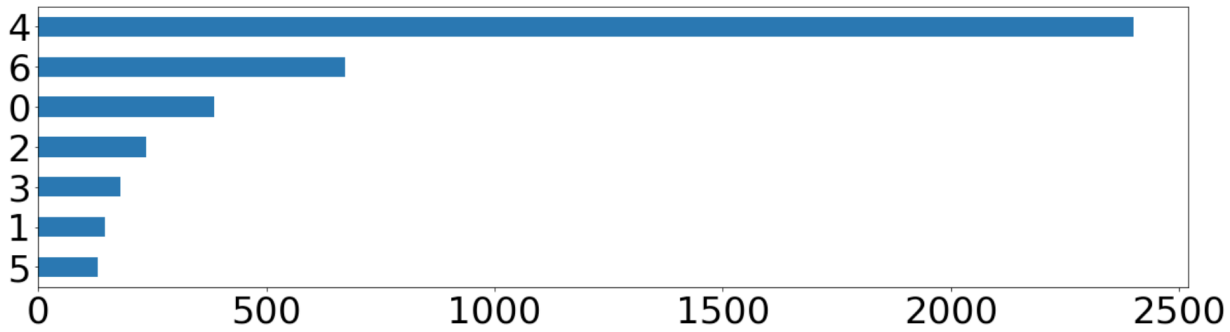


5. Neighbourhoods with no ice cream shops

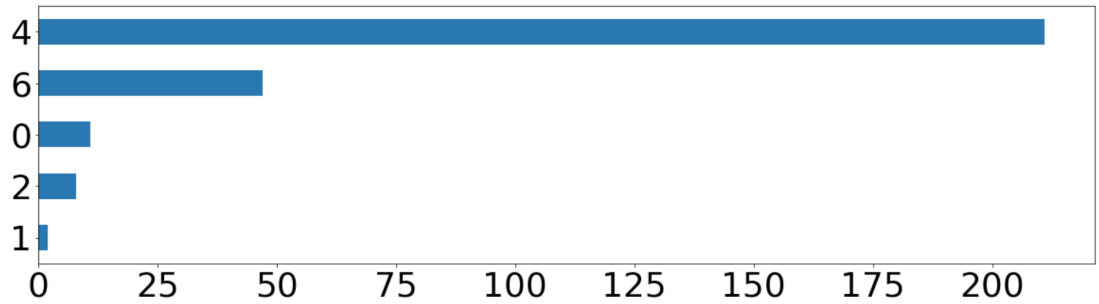
```
['Malvern, Rouge', 'Rouge Hill, Port Union, Highland Creek',
 'Guildwood, Morningside, West Hill', 'Woburn', 'Cedarbrae',
 'Scarborough Village',
 'Kennedy Park, Ionview, East Birchmount Park',
 'Golden Mile, Clairlea, Oakridge',
 'Cliffside, Cliffcrest, Scarborough Village West',
 'Birch Cliff, Cliffside West',
 'Dorset Park, Wexford Heights, Scarborough Town Centre',
 'Agincourt', "Clarks Corners, Tam O'Shanter, Sullivan",
 "Milliken, Agincourt North, Steeles East, L'Amoreaux East",
 "Steeles West, L'Amoreaux West", 'Upper Rouge',
 'Hillcrest Village', 'Fairview, Henry Farm, Oriole',
 'Bayview Village', 'York Mills, Silver Hills',
 'Willowdale, Newtonbrook', 'Willowdale, Willowdale West',
 'Parkwoods', 'Bathurst Manor, Wilson Heights, Downsview North',
 'Northwood Park, York University', 'Downsview',
 'Parkview Hill, Woodbine Gardens', 'Thorncliffe Park',
 'East Toronto, Broadview North (Old East York)', 'Lawrence Park',
 'Moore Park, Summerhill East', 'Bedford Park, Lawrence Manor East',
 'Lawrence Manor, Lawrence Heights', 'Glencairn',
 'North Park, Maple Leaf Park, Upwood Park',
 'Del Ray, Mount Dennis, Keelsdale and Silverthorn',
 'Runnymede, The Junction, Weston-Pellam Park, Carlton Village',
 'Canada Post Gateway Processing Centre',
 'New Toronto, Mimico South, Humber Bay Shores',
 'Alderwood, Long Branch',
 'The Kingsway, Montgomery Road, Old Mill North',
 'Mimico NW, The Queensway West, South of Bloor, Kingsway Park South West, Royal York South West',
 'Islington Avenue, Humber Valley Village',
 'Eringate, Bloordale Gardens, Old Burnhamthorpe, Markland Wood',
 'Humber Summit', 'Humberlea, Emery', 'Weston', 'Westmount',
 'Kingsview Village, St. Phillips, Martin Grove Gardens, Richview Gardens',
 'South Steeles, Silverstone, Humbergate, Jamestown, Mount Olive, Beaumond Heights, Thistletown, Albion Garden

 'Northwest, West Humber - Clairville'], dtype=object)
```

6. Clusters with the most restaurants



7. Clusters with the most ice cream shops



It appears that clusters 3 and 5 don't have any ice cream shops

Choosing a neighbourhood based on this data might not be sufficient given that it's possible that a neighbourhood has a high restaurant density but also a high ice cream shop density. Better metrics might be to consider:

1. Ice cream shop to restaurant ratio (ice cream ratio): calculated by dividing the total ice cream shops by the number of restaurants in any given neighbourhood or cluster. A lower ratio will reveal neighbourhoods or clusters with higher chance of having many restaurants but few ice cream shops.
2. % of total restaurants (total share): calculated by dividing the number of restaurants in any given neighbourhood or cluster by the total number of restaurants pulled in. This metric will give us insight into neighbourhoods with potential for more restaurant clusters. A higher ratio reveals a greater potential for restaurant clusters.

We can combine these metrics into one to determine ideal neighbourhoods or clusters to explore. We can do so by dividing the ice cream ratio but the total share, which we will call the "opportunity score" - the lower the number the better. Here we can see a histogram of the neighbourhoods and clusters by opportunity score.

Opportunity score by neighbourhood:

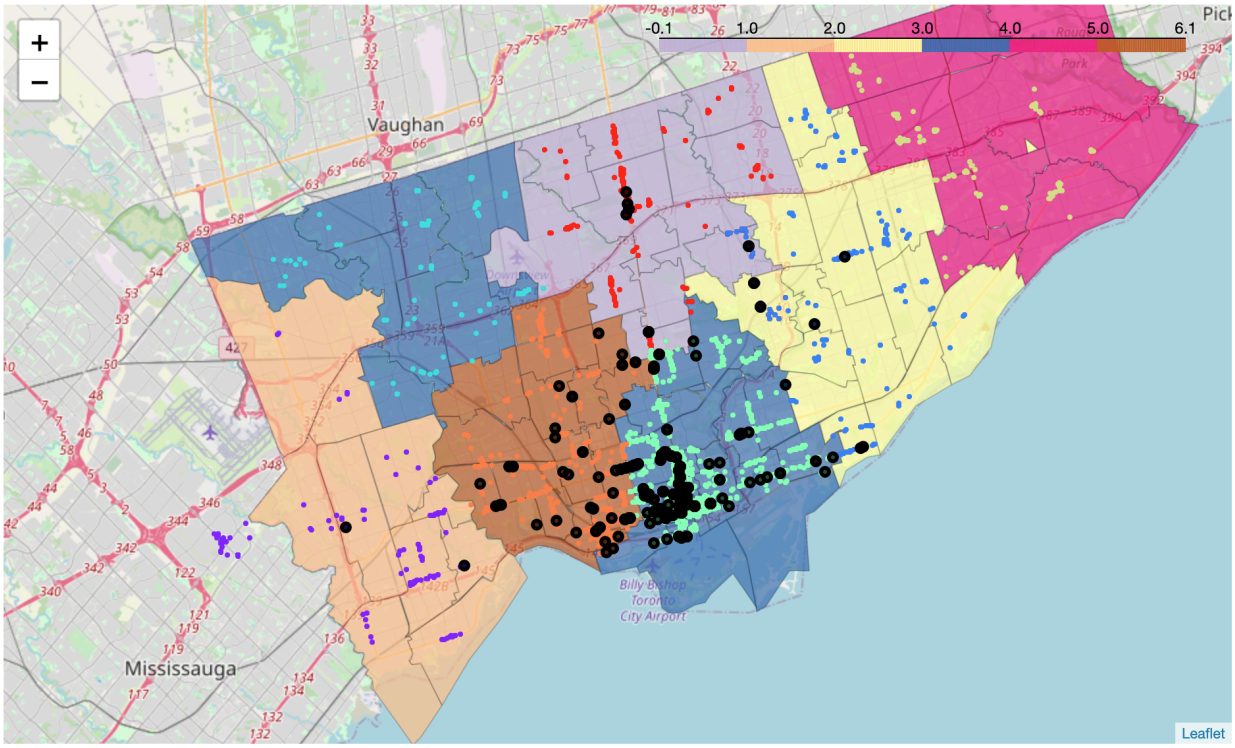
	Neighbourhood	Ice Cream Count	Restaurant Count	Ice Cream Ratio	% of Total	Opportunity Score
0	Berczy Park	4	104	0.038462	0.027860	1.380547
1	Brockton, Parkdale Village, Exhibition Place	7	100	0.070000	0.026788	2.613100
2	Business reply mail Processing Centre, South C...	2	31	0.064516	0.008304	7.768991
3	CN Tower, King and Spadina, Railway Lands, Har...	1	5	0.200000	0.001339	149.320000
4	Caledonia-Fairbanks	3	14	0.214286	0.003750	57.137755

Opportunity score by cluster:

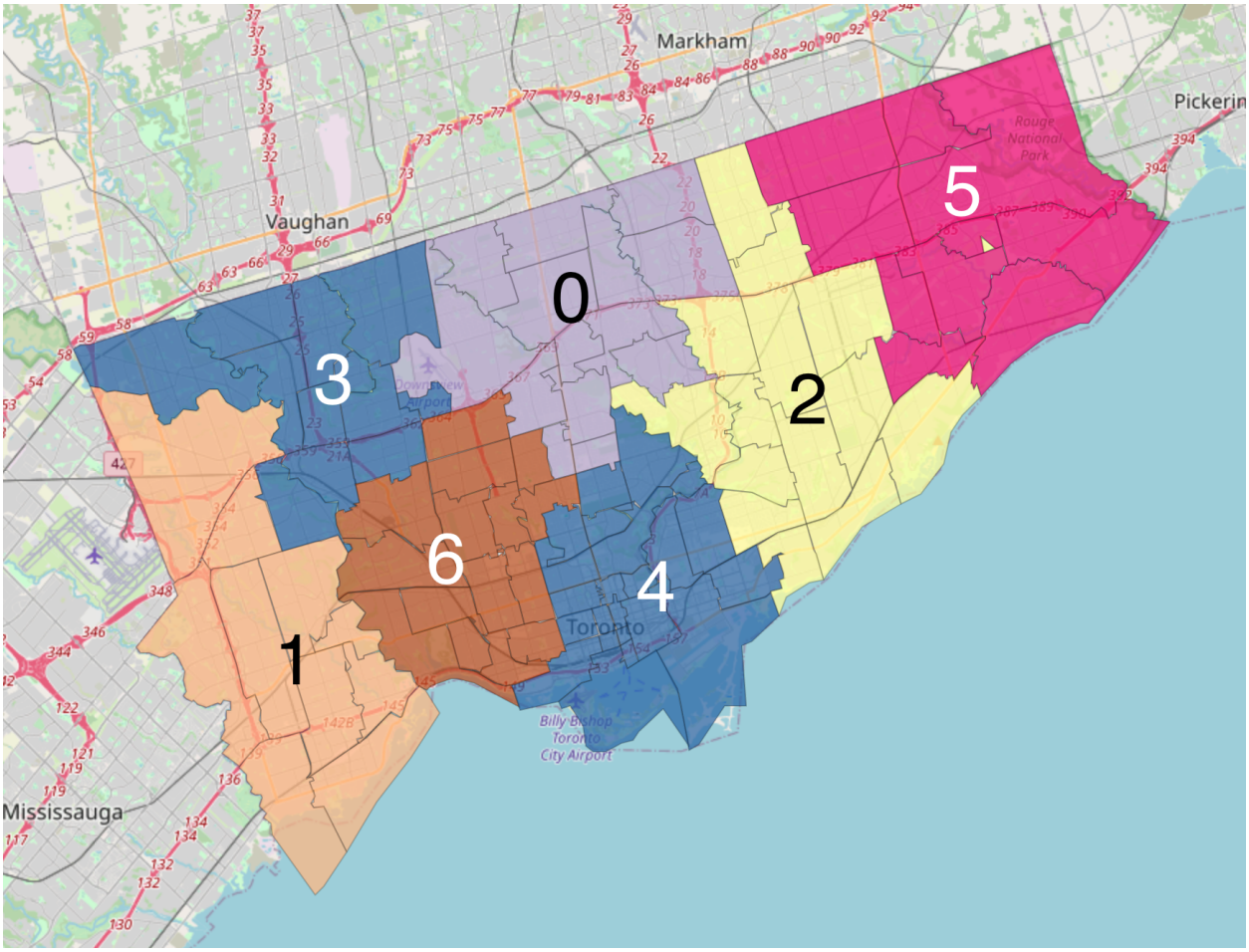
	Cluster Labels	Ice Cream Count	Restaurant Count	Ice Cream Ratio	% of Total	Opportunity Score
0	0	11	385	0.028571	0.103134	0.277032
1	1	2	146	0.013699	0.039111	0.350253
2	2	8	237	0.033755	0.063488	0.531681
3	4	211	2400	0.087917	0.642915	0.136747
4	6	47	673	0.069837	0.180284	0.387370

3. Results

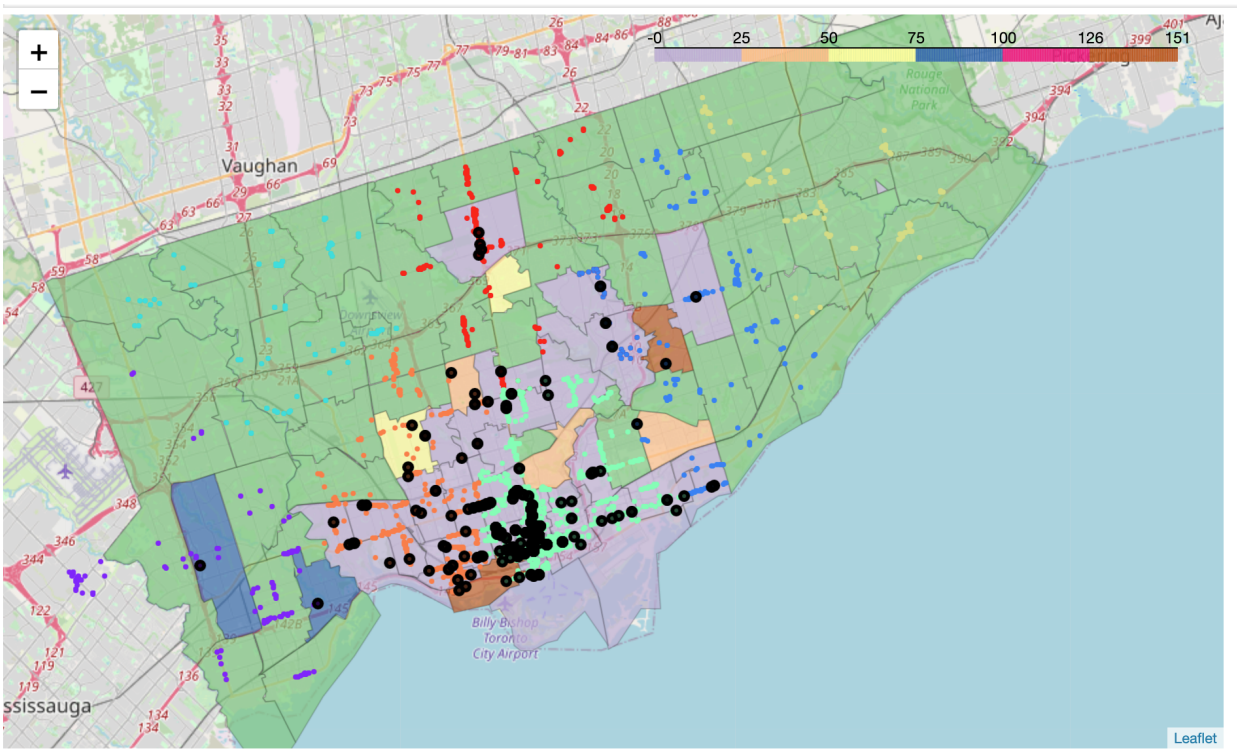
From the data I scraped and cleaned above, I was able to plot choropleth summary maps that overlay the restaurants, ice cream shops and clusters for exploration. I exported the maps to html for exploration here. The exploration map is screenshotted below:



Here is a reference map with cluster labels:



I also created a visualization with the opportunity score overlayed on the choropleth map.



As we can see from the map, there are plenty of opportunities in the neighbourhoods in green (no ice cream shops) or in purple.

4. Discussion

Based on these opportunity scores, my friend could start exploring the map looking at different locations that may be an opportunity to set up his ice cream shop. He will be able to take into consideration things like distance to where he lives, cost of property, construction, parking, etc. but the analysis performed will allow him to make a better informed decision.

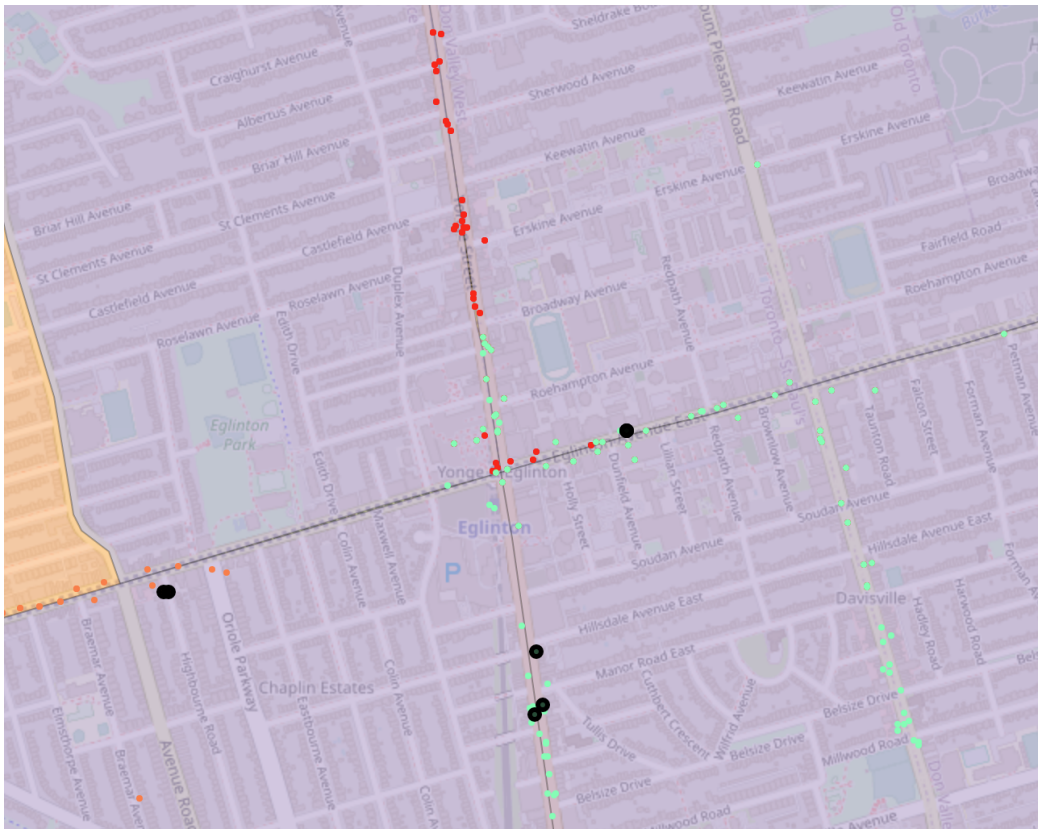
Toronto like most major cities has a diverse food scene. As a restaurant (or in this case ice cream shop owner), it can be daunting to choose a location. As I mentioned before, the decision is often left to the intuition and gut feeling of the owner or real estate agent. However, with some data and a little bit of elbow grease, that intuition can be strengthened and informed. The power of data allows us to turn a 'good guess' into an 'educated guess'.

Using some simple machine learning, we were able to cluster neighbourhoods together. Although the clustering doesn't directly affect the decision, it helps our mind to organize the information presented. Using k-means clustering and the elbow method we created 7 distinct clusters that helped divide up the city. Additionally, we used the neighbourhoods for a more granular level of grouping.

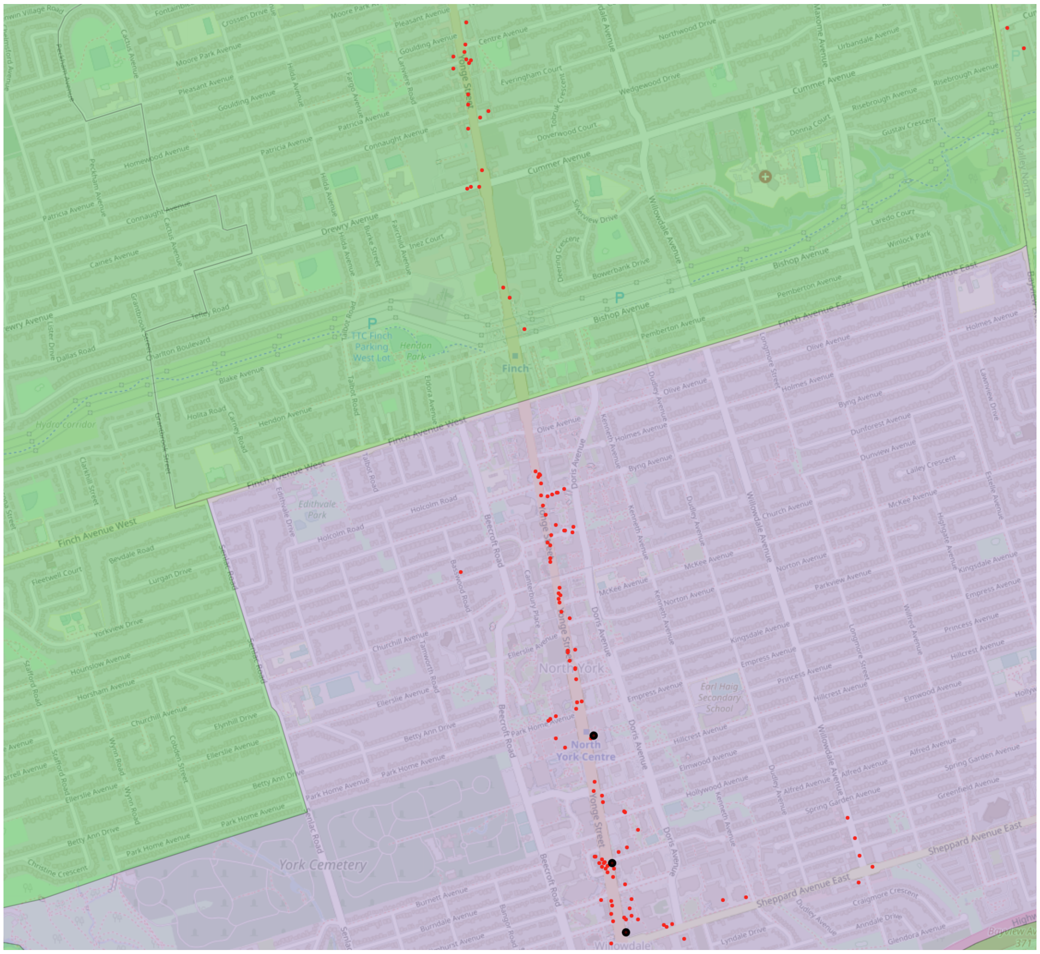
By layering in venues, location data, restaurant and ice cream shop density, we were able to determine an estimated "opportunity score" which would direct the exploration of neighbourhoods to set up a new ice cream shop. The data was used to create an interactive map that can be used to visually explore restaurant clusters.

We see from the analysis that there are many opportunities. At a quick glance without context of commute, and property prices the following restaurant clusters seem promising.

Eglington:



North York:


North York
North York Centre
York Cemetery
Finch
Sheppard Avenue East
Finch Avenue West
Finch Avenue East

Avenue Road:


Avenue Road
Yonge Street
Bathurst Street
Lawrence
Lawrence Park
Old Toronto

Roncesvalles:

Danforth:

5. Conclusion

While this exercise is not definitive in providing an answer, the output gives a clear idea of the potential of data to inform decisions. Through it all, I hope that my friend is more informed as he works through making a decision. Watch out Toronto, there may be a new ice cream shop popping up in a neighbourhood near you!

References:

1. https://en.wikipedia.org/wiki/Global_city
2. https://www.toronto.ca/311/knowledgebase/kb/docs/articles/economic-development-and-culture/program-support/number-of-restaurants-in-toronto.html
3. https://cocl.us/Geospatial_data
4. https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=1011037969
5. https://developer.foursquare.com/docs/places-api/
6. https://raw.githubusercontent.com/ag2816/Visualizations/master/data/Toronto2.geojson
7. https://developer.foursquare.com/docs/build-with-foursquare/categories/