Optimal Restaurant Location Finder Toronto

Data

The sources of data used for this project was:

- Wikipedia Toronto "Borough","Neighbourhood", and "Postal Code" data.
- geopy geocoder, convert to obtain Coordinates
- Foursquare Venue data

	Postal Code	Borough	Neighbourhood
3	M3A	North York	Parkwoods
4	M4A	North York	Victoria Village
5	M5A	Downtown Toronto	Regent Park, Harbourfront
6	M6A	North York	Lawrence Manor, Lawrence Heights
7	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

1 - Wikipedia Scraped DataFrame

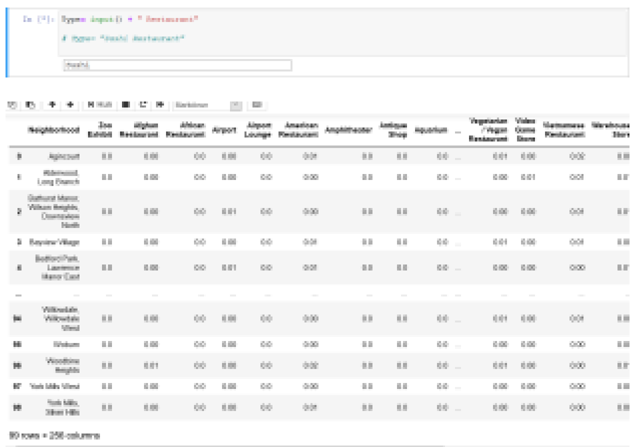
Method

- Once the Toronto Location data is cleaned combine with API Data
  - API data provides venue information like venue category, venue locations, venue frequency.
- The goal is to prepare the data to prepare and visualize the relationship and activity between location and restaurants.

2 - Neighbourhood Maps

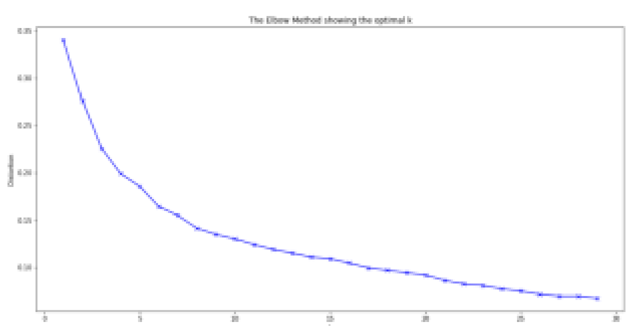
Data is cleaned and prepared for Machine Learning, and Features extracted for Machine Learning.

- Filtered all columns for only Restaurants
- User input to index specific type of restaurant to be analyzed



Machine Learning

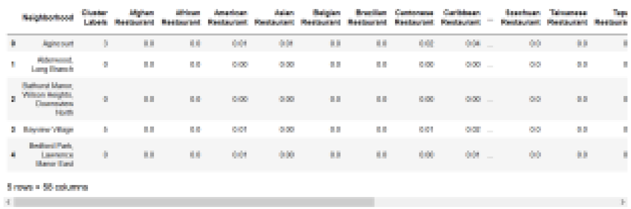
- Found Optimal cluster number was 6 using Elbow Method
- Used KMeans Unsupervised learning to create 6 clusters based off restaurant frequency and location data



Results

Cluster 0 had the neighbourhoods with the most potential to open up a new Sushi Restaurant. Mores specifically the sum of the mean frequency for cluster 0 was the largest out of all the other clusters.

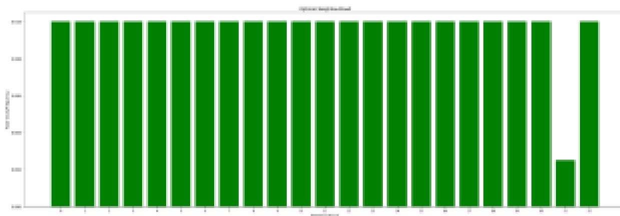
Cluster Labels	0	1	2	3	4	5
Sushi Restaurant	0.31	0.20	0.670001	0.10	0.11	0.17



	Neighborhood	Cluster Labels	Afghan Restaurant	African Restaurant	American Restaurant	Asian Restaurant	Belgian Restaurant	Brazilian Restaurant	Cantonese Restaurant	Caribbean Restaurant	...	Szechuan Restaurant	Taiwanese Restaurant	Tap Restaura
0	Agincourt	[illegible]	0.0	0.0	0.01	0.01	0.0	0.0	0.02	0.04	...	0.0	0.0	[illegible]
1	Alderwood, Long Branch	0	0.0	0.0	0.00	0.00	0.0	0.0	0.00	0.00	...	0.0	0.0	[illegible]
2	Bathurst Manor, Wilson Heights, Downsview North	0	0.0	0.0	0.00	0.00	0.0	0.0	0.00	0.00	...	0.0	0.0	[illegible]
3	Bayview Village	[illegible]	0.0	0.0	0.01	0.00	0.0	0.0	0.01	0.02	...	0.0	0.0	[illegible]
4	Bedford Park, Lawrence Manor East	0	0.0	0.0	0.01	0.00	0.0	0.0	0.00	0.01	...	0.0	0.0	[illegible]

5 rows × 56 columns

The Best neighbourhoods to open a Sushi Restaurant in cluster 0 would are shown in the bar graph. The y axis is based off Sushi Restaurant Frequency/ visits per neighbourhood, to find the nieghbourhood that had the most frequency and the least amount of restaurants.



Next Step

A good extension for this project in the future, is to find data that would able to create a time series plot of the frequency in visits, and use a regression algorithm to predict what the demand for such a restaurant is in the future.

Another good project would be to find the most optimal places to rent/share commercial kitchens. That mixed with data on food delivery may be promising for business's to survive situations like Covid where they can pay a fraction of the cost to share existing kitchens and have the bare minimum of staff. With chef's spread around neighbourhood's with the highest demand, not keeping all of the chefs "in the same basket" they would generate more revenue, and be more disaster resistant.