



Applied Data Science Capstone

Segmenting and Clustering Neighborhoods in Toronto, Canada
-Rishabh Tamhane

Introduction



When a Restaurant Chain plans on opening a chain of restaurants in a city, there are many factors that need to be considered:

- 1) The most commonly visited category of restaurants in the region.
- 2) The most optimal locations for setting up the restaurants so that they don't end up competing against each other, thereby reducing the profits of the restaurant chain.
- 3) The locations of the restaurants the Restaurant Chain plans on competing against so as to avoid cut throat competition against closely clustered restaurants.

Methodology



- 1) Web Scraping Data about the Neighborhoods in Toronto from Wikipedia
- 2) Web Scraping Data about the Restaurants in the Neighborhood
- 3) Cleaning and Organising the Data
- 4) Exploratory Data Analysis
- 5) Conclusion

Web Scraping Data about the Neighborhoods in Toronto from Wikipedia

	Postal Code	Borough	Neighborhood	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

Web Scrapping Data about the Restaurants

in Toronto Using FourSquare API

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs
0	Ted's Restaurant	Breakfast Spot	404 Old Kingston Rd.	CA	Scarborough	Canada	NaN	3194	[404 Old Kingston Rd., Scarborough ON, Canada]	[{"label": "display", "lat": 43.784467967441, "lon": -79.251111111111}]
1	Africa Restaurant	Food & Drink Shop	NaN	CA	Toronto Division	Canada	NaN	1636	[Toronto Division ON, Canada]	[{"label": "display", "lat": 43.81958724358, "lon": -79.251111111111}]
2	Rex's Den Restaurant	Burger Joint	1265 Military Tr	CA	Toronto	Canada	NaN	2485	[1265 Military Tr, Toronto ON, Canada]	[{"label": "display", "lat": 43.78500731279, "lon": -79.251111111111}]
3	Grapevine Restaurant	Restaurant	NaN	CA	NaN	Canada	NaN	3042	[Canada]	[{"label": "display", "lat": 43.78047180175, "lon": -79.251111111111}]
4	Sabby's Restaurant	Caribbean Restaurant	5780 Sheppard Ave E	CA	Toronto	Canada	Markham Rd	3276	[5780 Sheppard Ave E (Markham Rd), Toronto ON, ...]	[{"label": "display", "lat": 43.79554878818, "lon": -79.251111111111}]

Web Scrapping Data about the Restaurants

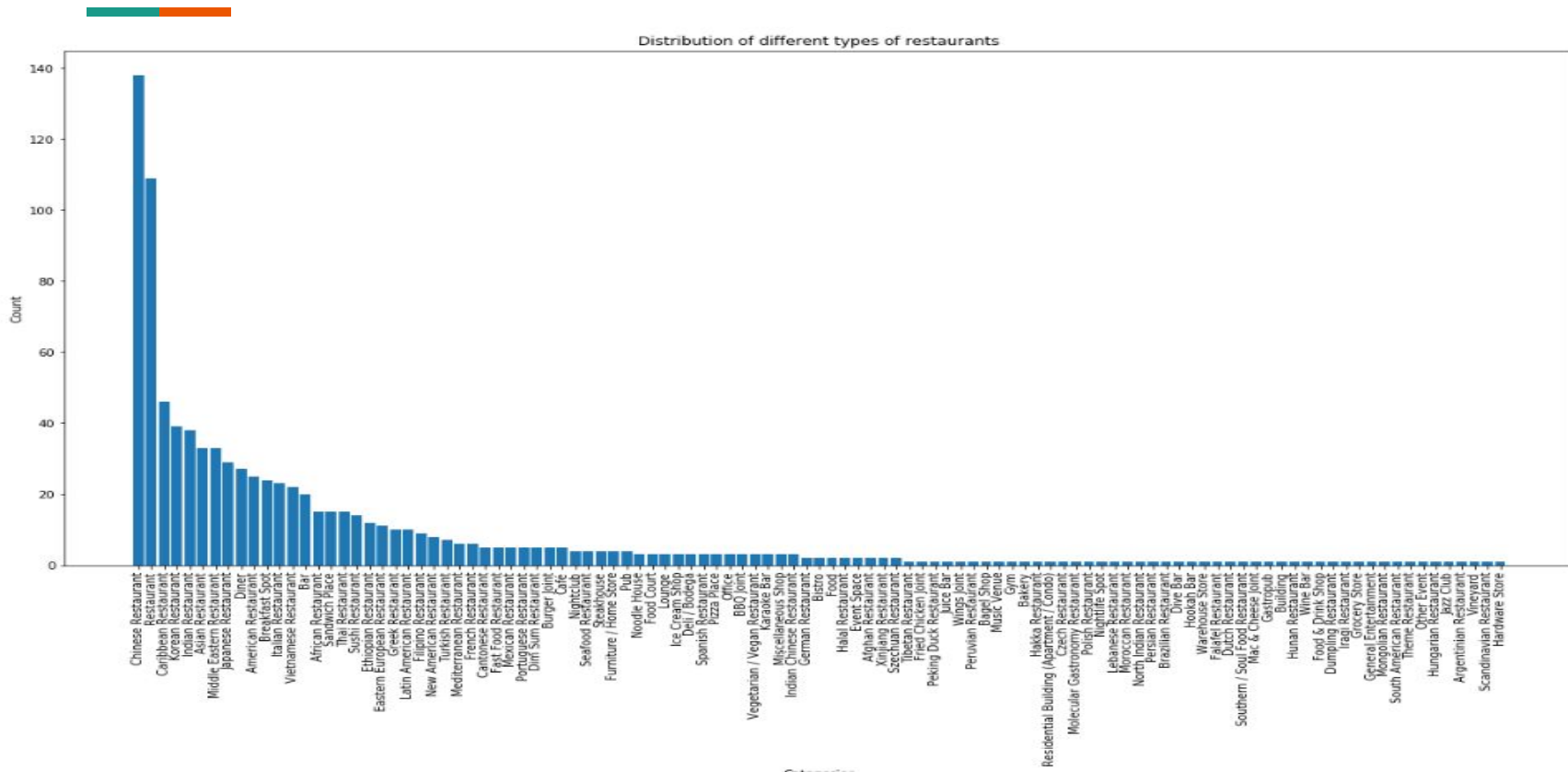
in Toronto Using FourSquare API

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs
0	Ted's Restaurant	Breakfast Spot	404 Old Kingston Rd.	CA	Scarborough	Canada	NaN	3194	[404 Old Kingston Rd., Scarborough ON, Canada]	[{"label": "display", "lat": 43.784467967441, "lon": -79.251111111111}]
1	Africa Restaurant	Food & Drink Shop	NaN	CA	Toronto Division	Canada	NaN	1636	[Toronto Division ON, Canada]	[{"label": "display", "lat": 43.81958724358, "lon": -79.251111111111}]
2	Rex's Den Restaurant	Burger Joint	1265 Military Tr	CA	Toronto	Canada	NaN	2485	[1265 Military Tr, Toronto ON, Canada]	[{"label": "display", "lat": 43.78500731279, "lon": -79.251111111111}]
3	Grapevine Restaurant	Restaurant	NaN	CA	NaN	Canada	NaN	3042	[Canada]	[{"label": "display", "lat": 43.78047180175, "lon": -79.251111111111}]
4	Sabby's Restaurant	Caribbean Restaurant	5780 Sheppard Ave E	CA	Toronto	Canada	Markham Rd	3276	[5780 Sheppard Ave E (Markham Rd), Toronto ON, ...]	[{"label": "display", "lat": 43.79554878818, "lon": -79.251111111111}]

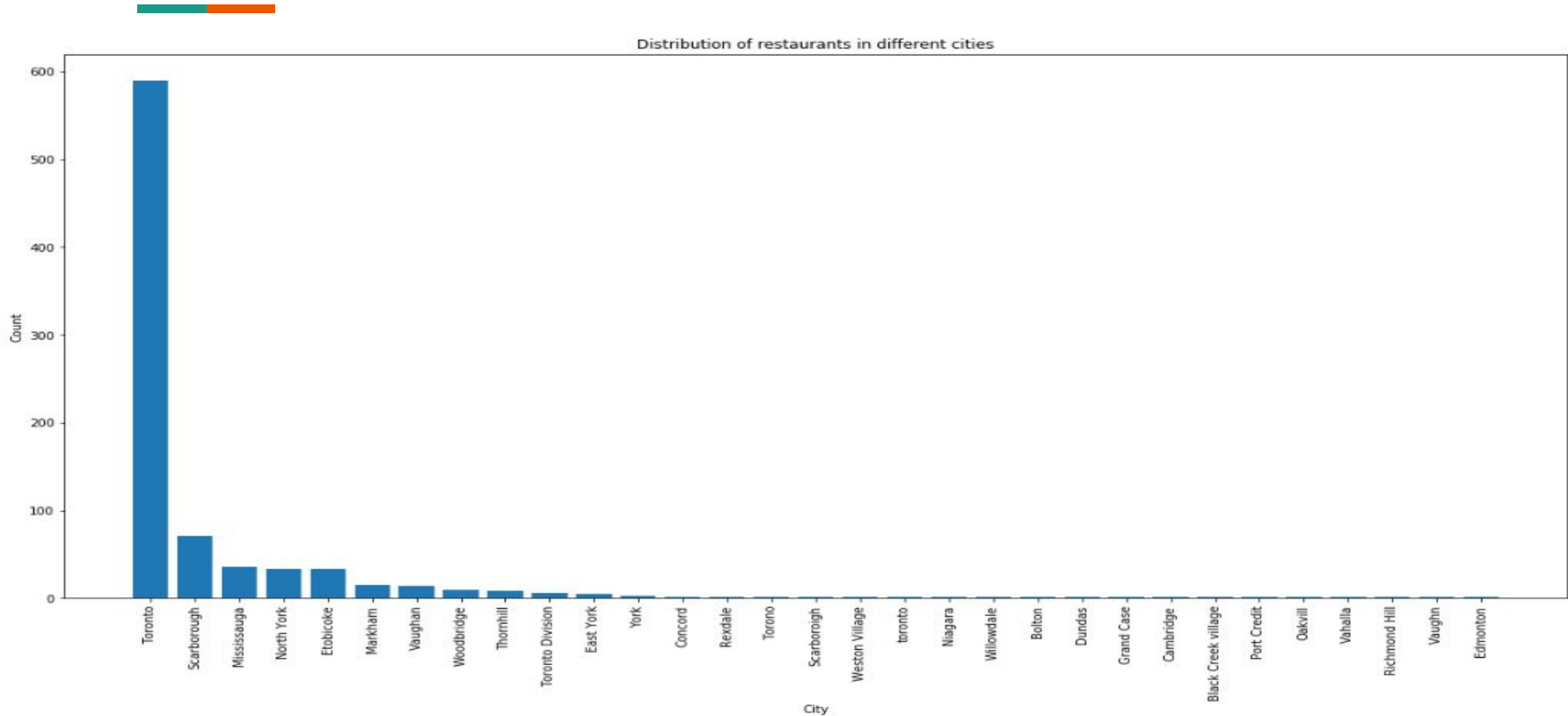


Exploratory Data Analysis

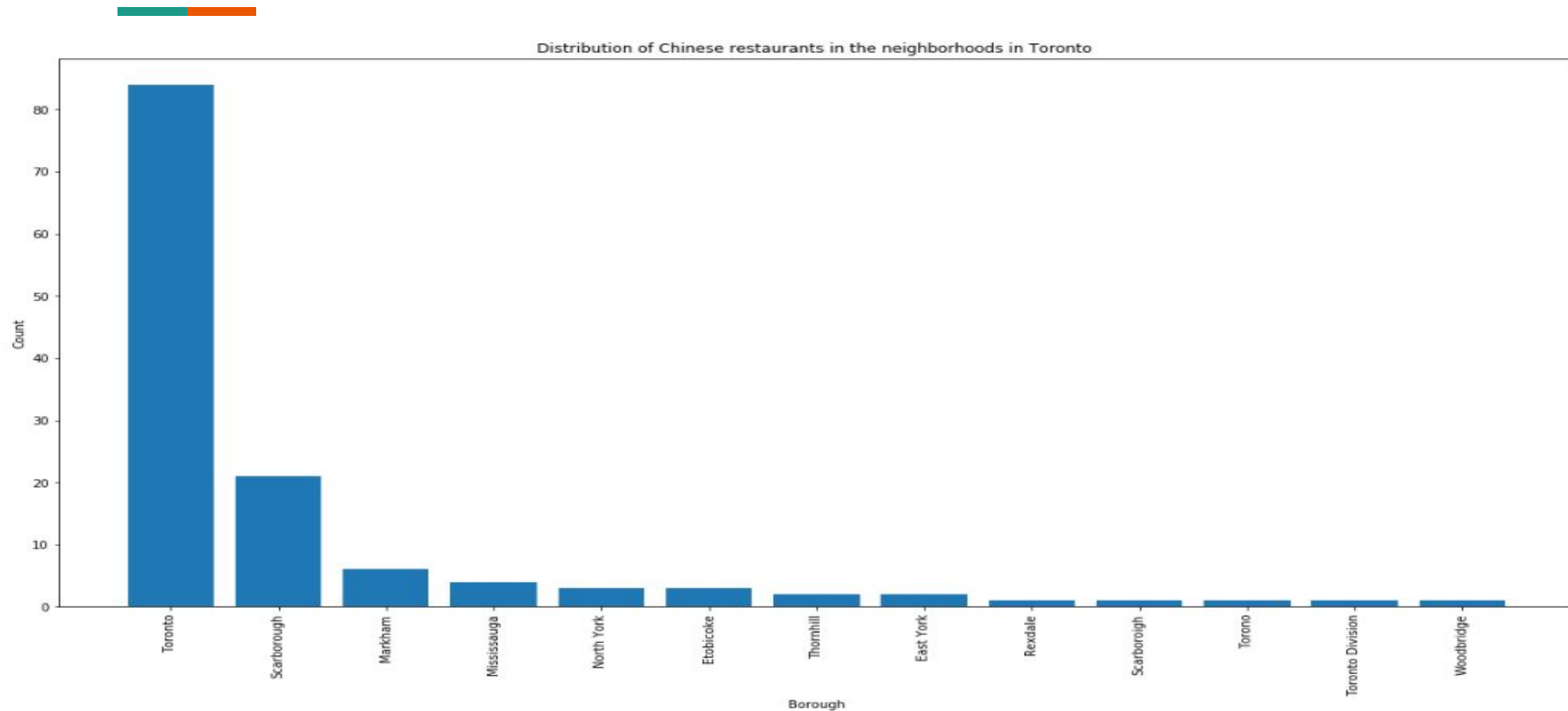
Distribution of different categories of restaurants



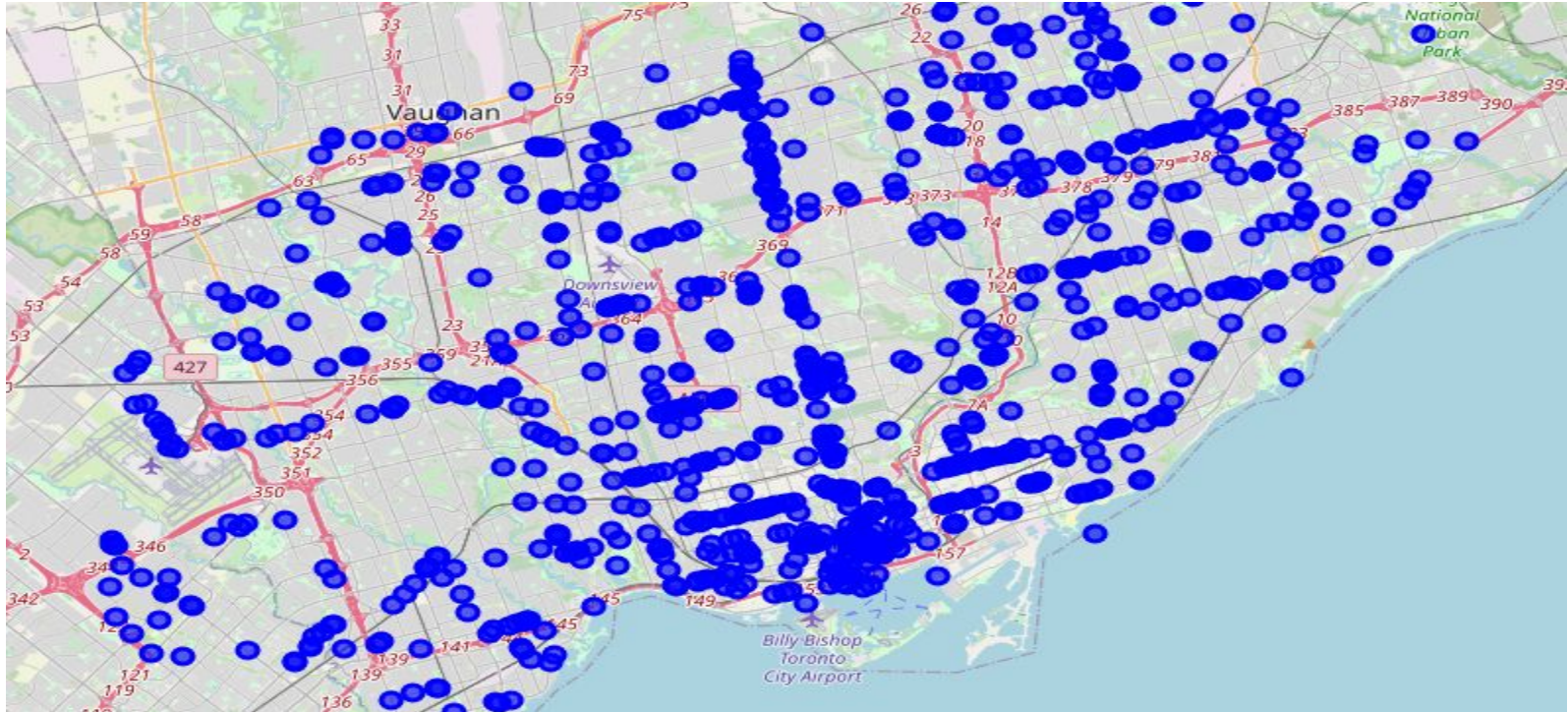
Distribution of restaurants



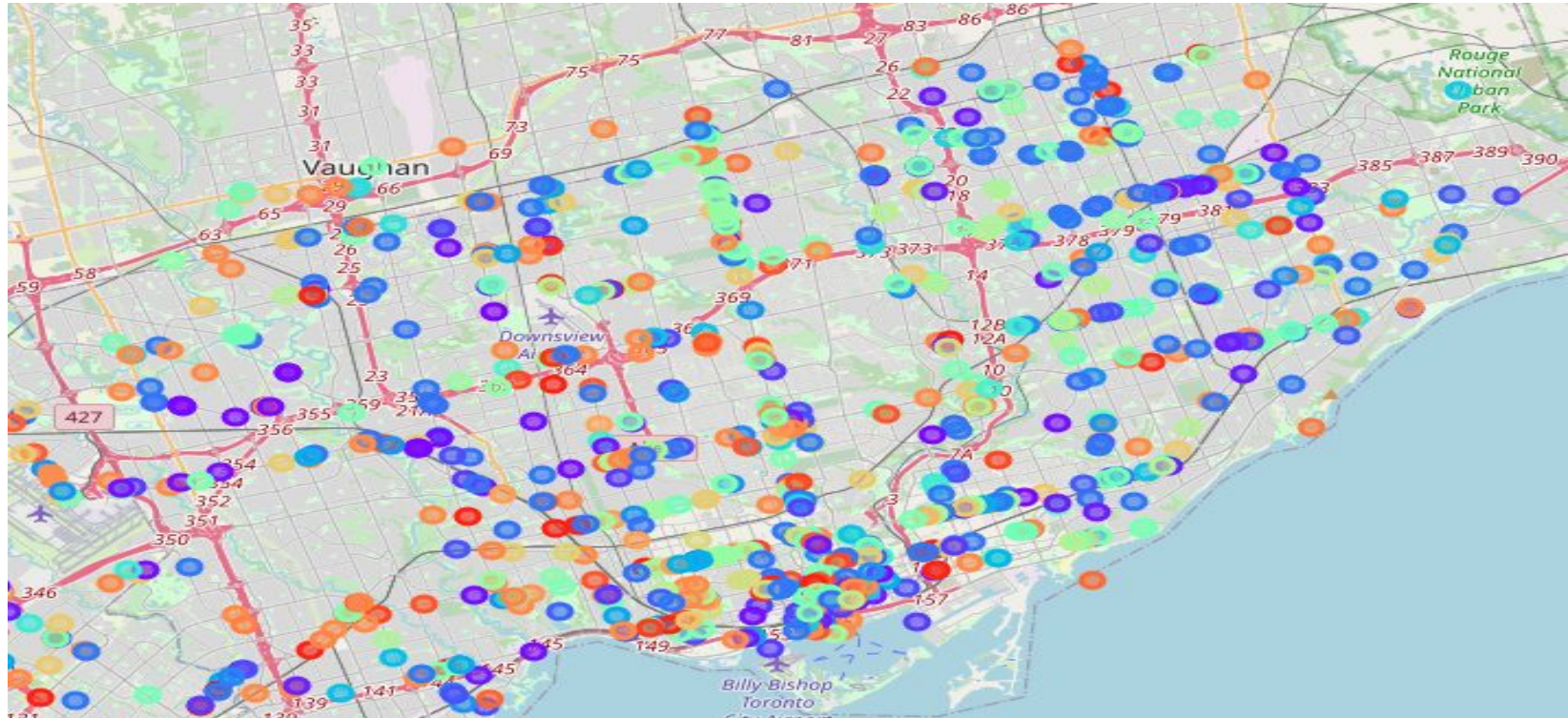
Distribution of Chinese Restaurants



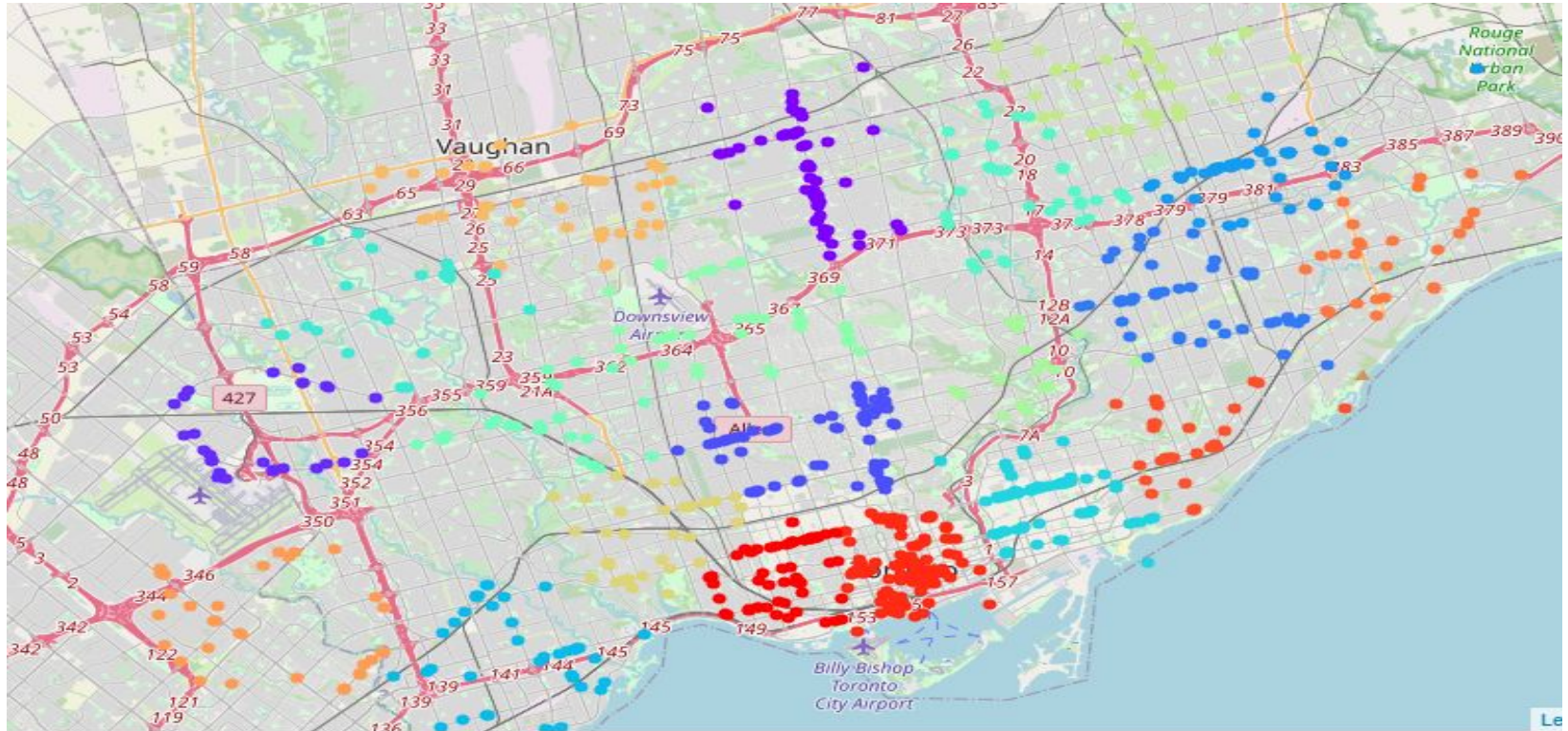
Map Plot of Restaurants



Clustering of Restaurants Based on Categories



Clustering of Restaurants Based on Location



Problem Statement and Solution



1) Let us say I have a client who is looking to set up a restaurant in Canada. I have to make a recommendation about which is the most common category of restaurants in the neighborhood.

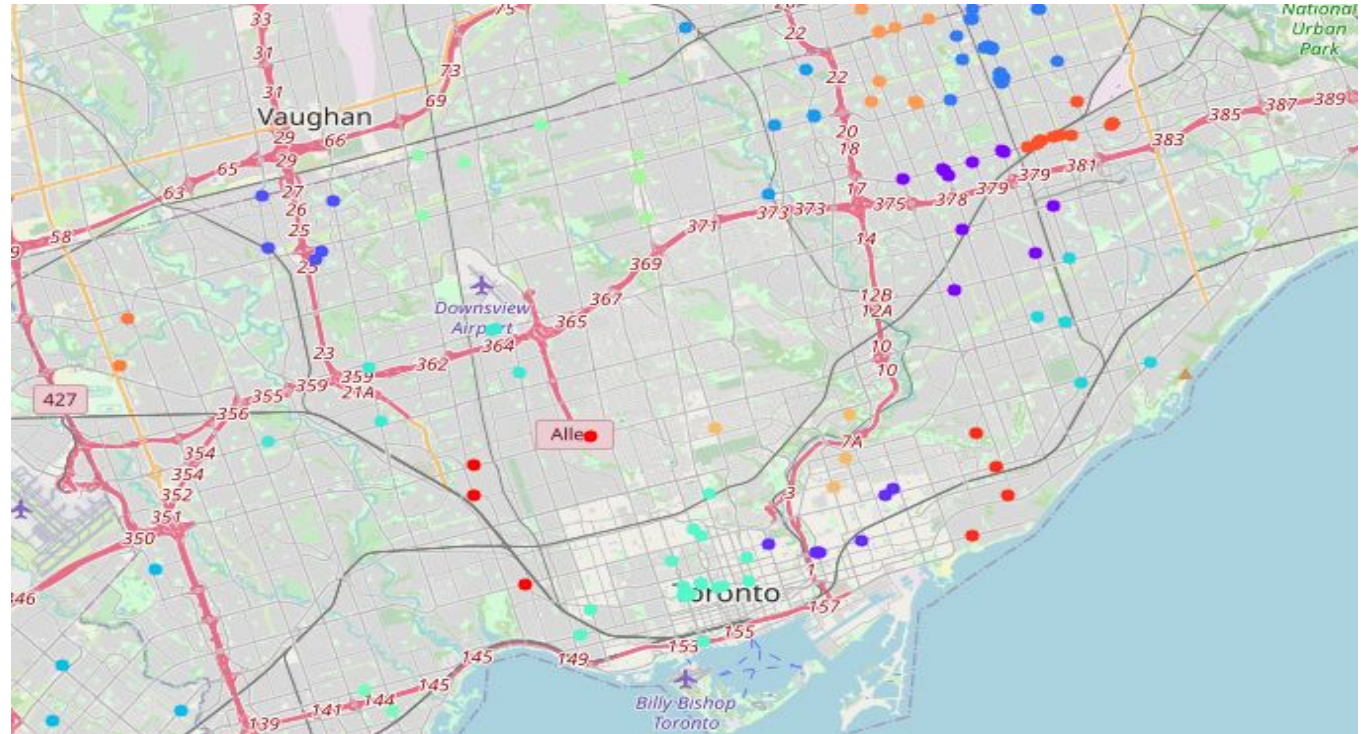
Since, there are 139 Chinese Restaurants, they are the most common restaurants in the neighborhoods.

2) Now the client is planning to set up a chain of 20 Chinese restaurants in and around Toronto. So, I have to make a recommendation about the location of these restaurants so that they don't cannibalise each other

Using K-Means Clustering which is an Unsupervised Machine Learning Algorithm, I clustered the Chinese Restaurants in 20 clusters (as shown in next slide). Now, I can recommend the Restaurant Chain to set up restaurants in each of these clusters, so that while catering to the customers, it will reduce the chances of cannibalisation amongst the restaurants belonging to the same restaurant chain. This approach also ensures that we cover a larger customer base while being aware of the competition prevalent in each area. It also makes sure that more restaurants are set up in the areas where Chinese Restaurants are more commonly visited than other areas.

Clustering of Chinese Restaurants Based on

Location





THE END