

Capstone Project: The Battle of the Neighborhoods

Introduction: Business Problem

- Target Audience: Aspiring Chefs or Business Owners
- Issue: Selecting the most optimal location to open a new Chinese Restaurant
- Consideration factors: location, demographics, preference of locals, existing venues etc.

Introduction: Business Problem

- 3 crucial factors:
 - Number of existing restaurants
 - Number of existing Chinese restaurants (potential competitors)
 - Number of Chinese residing in the location (customers)

Data Sources

- Neighborhood data of Toronto
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Data of existing venues like venues, venue categories, coordinates to be obtained from Foursquare API
- Coordinates of neighborhoods
http://cocl.us/Geospatial_data
- Population demographics of each neighborhood
<https://open.toronto.ca/dataset/wellbeing-toronto-demographics/>

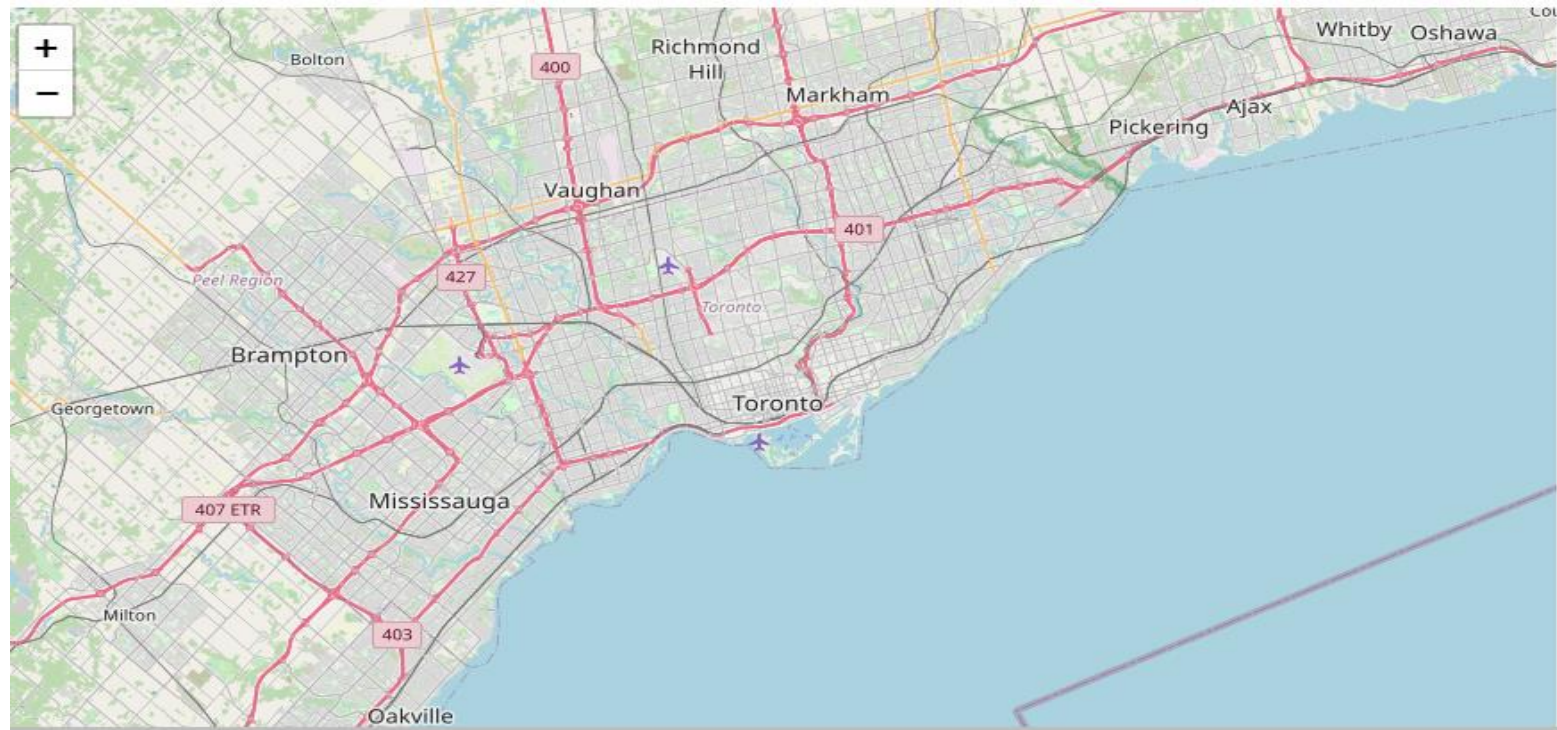
Consolidate data into a single dataframe

- The ultimate dataframe which we have created consists of the respective neighborhoods along with the number of each nationality, combined with the coordinates of each neighborhood

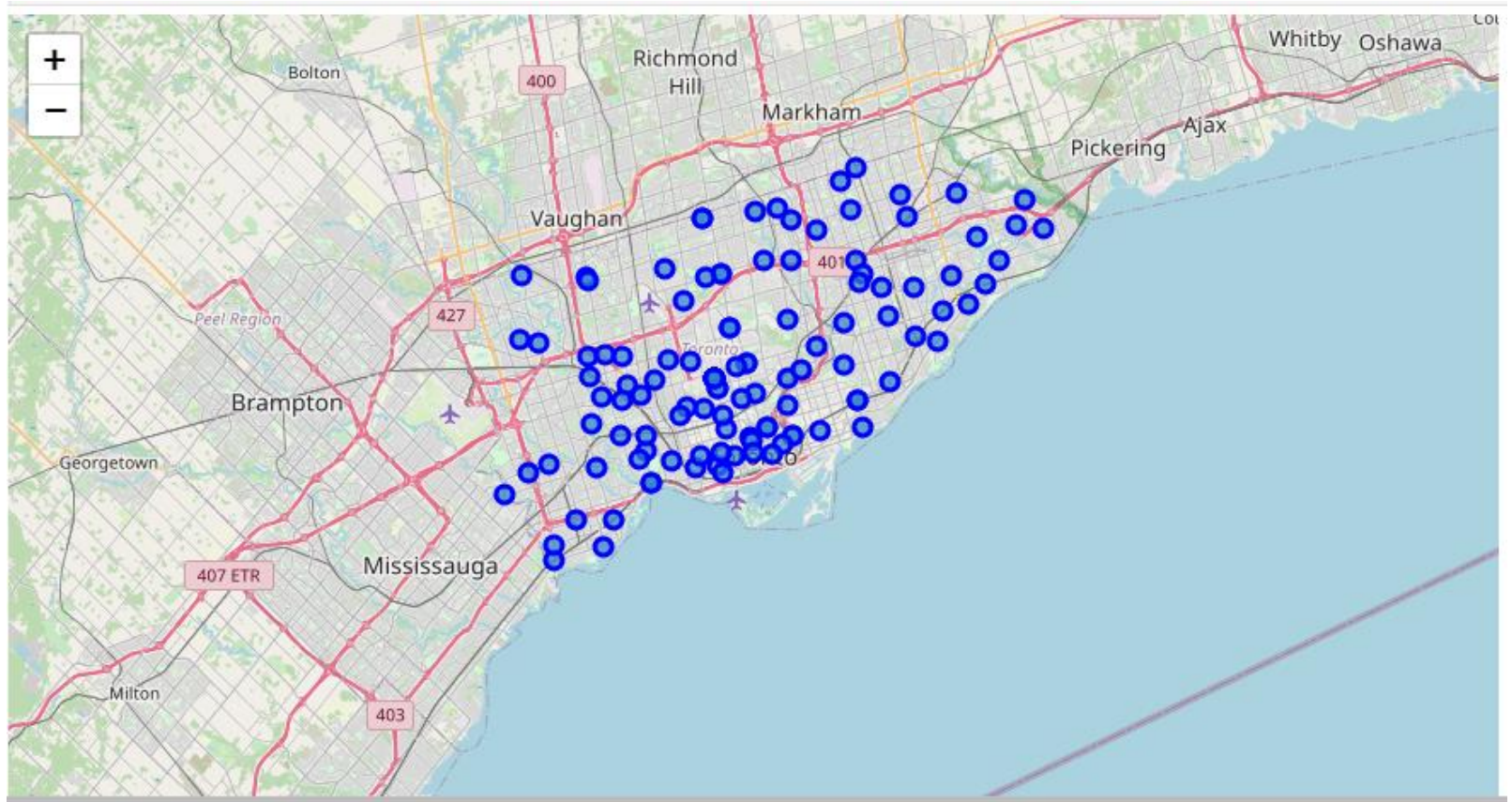
	Neighborhood	Chinese	South Asian	Black	Filipino	Latin American	Southeast Asian	Arab	West Asian	Korean	Japanese	Latitude	Longitude
1	Milliken	16790	3780	1365	1145	125	290	170	135	20	60	43.823174	-79.301763
2	Steeles	16705	1895	660	755	50	95	360	115	140	55	43.816178	-79.314538
3	Agincourt North	16565	5160	1530	1355	230	230	370	75	155	135	43.808038	-79.266439
4	L'Amoreaux	16455	8285	3875	1905	385	635	985	930	205	245	43.799003	-79.305967
5	Willowdale East	14860	1700	845	520	440	190	485	3395	4265	285	43.761510	-79.410923

Explore and Cluster Neighborhoods in Toronto

- We extracted the map of Toronto and overlaid it with the various neighborhoods



Explore and Cluster Neighborhoods in Toronto



Exploring Venues in a Neighborhood

- Using Foursquare API, we extracted the json results of nearby venues of each neighborhood

```
{'meta': {'code': 200, 'requestId': '5ed323729da7ee001b6f27ce'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
 'filters': [{'name': 'Open now', 'key': 'openNow'}]}},
 'headerLocation': 'Scarborough',
 'headerFullLocation': 'Scarborough',
 'headerLocationGranularity': 'city',
 'totalResults': 40,
 'suggestedBounds': {'ne': {'lat': 43.8276743045, 'lng': -79.29553706206312},
 'sw': {'lat': 43.818674295499996, 'lng': -79.30798813793689}},
 'groups': [{'type': 'Recommended Places',
 'name': 'recommended',
 'items': [{'reasons': {'count': 0,
 'items': [{'summary': 'This spot is popular',
 'type': 'general',
 'reasonName': 'globalInteractionReason'}]}]},
 'venue': {'id': '56945d1c498e11466e96405f',
 'name': 'Planet Fitness North Scarborough',
 'location': {'lat': 43.824095167666584,
 'lng': -79.30141064389495,
 'labeledLatLngs': [{'label': 'display',
 'lat': 43.824095167666584,
 'lng': -79.30141064389495}],
 'distance': 106,
```


Exploring Venues in a Neighborhood

- We converted it into a pandas dataframe
- The venues and categories for each neighborhood can be clearly seen

	name	categories	lat	lng
0	Planet Fitness North Scarborough	Gym	43.824095	-79.301411
1	Deer Garden Signatures 鹿園魚湯米線	Noodle House	43.821898	-79.298857
2	Nichiban Sushi	Sushi Restaurant	43.823172	-79.306064
3	Aka-Oni Izakaya	Japanese Restaurant	43.822372	-79.298905
4	Sun's Kitchen 拉麵王	Chinese Restaurant	43.825282	-79.306231
5	Allan's Pastry Shop	Bakery	43.820953	-79.304564
6	Uncle Tetsu's Japanese Cheesecake	Bakery	43.825150	-79.305954
7	New Northern Dumplings 新北方餃子館	Dumpling Restaurant	43.821886	-79.298751
8	Kim Po Vietnamese Cuisine - 金寶越南美食	Vietnamese Restaurant	43.823292	-79.305257
9	Fish Ball Place 真之味小食屋	Snack Place	43.825290	-79.306202
10	Tim Hortons	Coffee Shop	43.825423	-79.297787
11	Pacific Heritage Town 太古民族村	Miscellaneous Shop	43.825796	-79.306242
12	RBC Royal Bank	Bank	43.825195	-79.299575
13	Brothers Bakery 兄弟餅店	Bakery	43.825443	-79.306984
14	Tak Fu Seafood Restaurant 德福點心皇	Chinese Restaurant	43.822633	-79.298958
15	Fat Ninja Bite	Japanese Restaurant	43.823043	-79.306446

Exploring Venues in a Neighborhood

- This was eventually explored for all neighborhoods in Toronto

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Milliken	43.823174	-79.301763	Planet Fitness North Scarborough	43.824095	-79.301411	Gym
1	Milliken	43.823174	-79.301763	Deer Garden Signatures 鹿園魚湯米線	43.821898	-79.298857	Noodle House
2	Milliken	43.823174	-79.301763	Nichiban Sushi	43.823172	-79.306064	Sushi Restaurant
3	Milliken	43.823174	-79.301763	Aka-Oni Izakaya	43.822372	-79.298905	Japanese Restaurant
4	Milliken	43.823174	-79.301763	Sun's Kitchen 拉麵王	43.825282	-79.306231	Chinese Restaurant

Analyzing each Neighborhood

- Using onehot function, we provide a binary view of the overview of the venue categories for each neighborhood

	Neighborhood	Longitude	ATM	Accessories Store	Afghan Restaurant	American Restaurant	Animal Shelter	Antique Shop	Arcade	Art Gallery	Art Museum	Pub & Crafts Store	Asian Restaurant
0	Agincourt North	-2140.253999	0	0	0	0	0	0	0	0	0	0	0
1	Agincourt South-Malvern West	-317.040209	1	0	0	0	0	0	0	0	0	0	0
2	Alderwood	-795.453404	0	0	0	0	0	0	0	0	0	0	0
3	Annex	-3493.906322	0	0	0	0	0	0	0	0	0	0	0
4	Banbury-Don Mills	-396.783017	0	0	0	0	0	0	0	0	0	0	0

- This was subsequently converted into a table listing the top 5 venues for each neighborhood

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Agincourt North	Bakery	Bank	Chinese Restaurant	Fast Food Restaurant	Frozen Yogurt Shop
1	Agincourt South-Malvern West	Breakfast Spot	Lounge	Latin American Restaurant	ATM	Fish & Chips Shop
2	Alderwood	Pizza Place	Coffee Shop	Dance Studio	Pub	Skating Rink
3	Annex	Pizza Place	Coffee Shop	Thai Restaurant	Indian Restaurant	Ice Cream Shop
4	Banbury-Don Mills	Botanical Garden	Park	Trail	Coffee Shop	Intersection

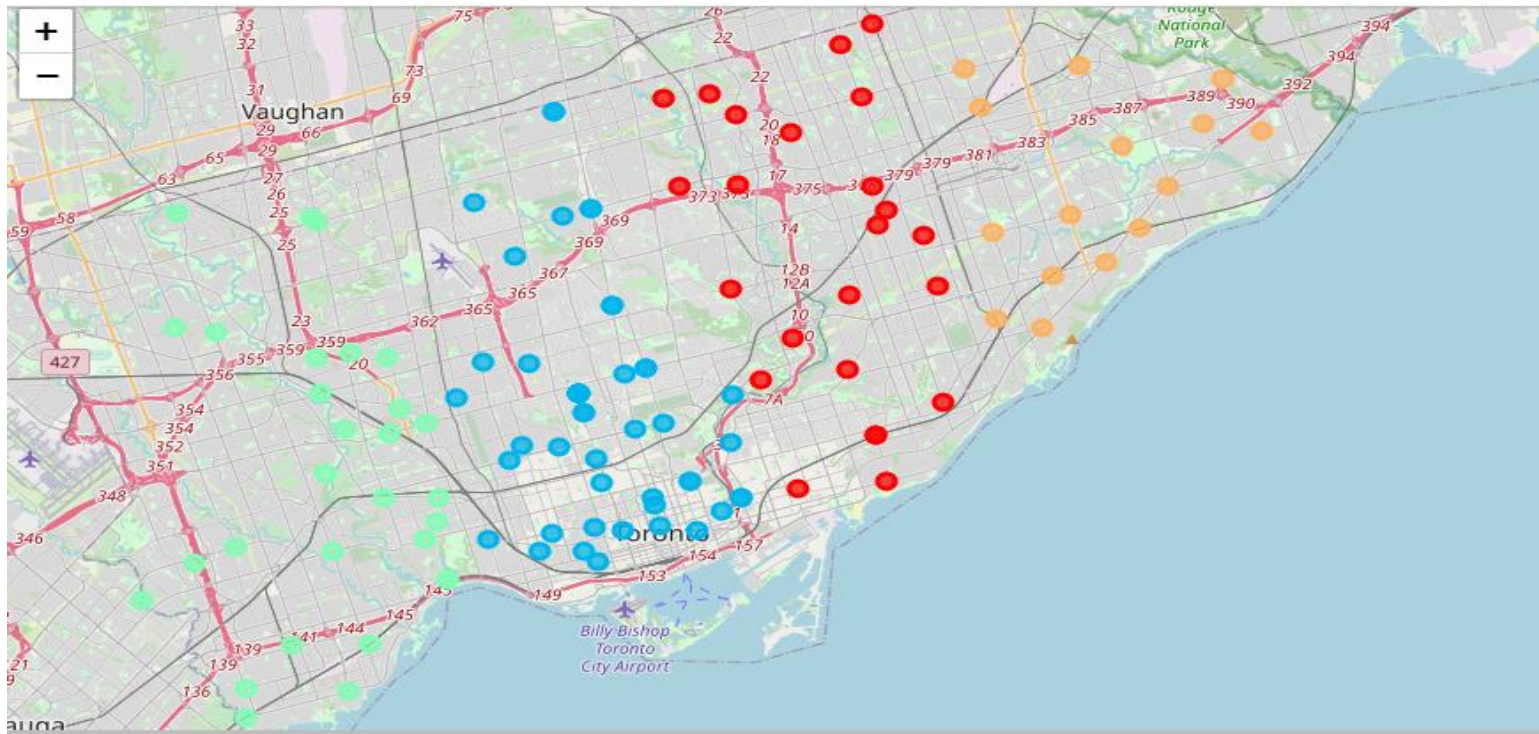
Model

- K-Means clustering was selected to cluster the neighborhoods into 5 clusters, plotted based on their latitude and longitude. Each neighborhood was clustered into 1 of the 5 clusters as shown below.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	No of Chinese	Latitude	Longitude	Cluster Labels
0	Agincourt North	Bakery	Bank	Chinese Restaurant	Fast Food Restaurant	Frozen Yogurt Shop	16565	43.808038	-79.266439	4
1	Agincourt South-Malvern West	Breakfast Spot	Lounge	Latin American Restaurant	ATM	Fish & Chips Shop	9810	43.795223	-79.260241	4
2	Alderwood	Pizza Place	Coffee Shop	Dance Studio	Pub	Skating Rink	70	43.601717	-79.545232	3
3	Annex	Pizza Place	Coffee Shop	Thai Restaurant	Indian Restaurant	Ice Cream Shop	1695	43.670338	-79.407117	2
4	Banbury-Don Mills	Botanical Garden	Park	Trail	Coffee Shop	Intersection	3535	43.734804	-79.357243	0

Visualizing the Cluster

- Clusters were plotted onto the map of Toronto



Analysis

- The number of neighborhoods per cluster is shown in the dataframe below

[illegible]

Analyzing Cluster 1

- The neighborhoods in Cluster 1 are 'Banbury-Don Mills', 'Bayview Village', 'Bayview Woods-Steeles', 'Danforth', 'Danforth-East York', 'Don Valley Village', 'Dorset Park', 'East End-Danforth', 'Flemingdon Park', 'Henry Farm', 'Hillcrest Village', 'Ionview', 'L'Amoreaux', 'Milliken', 'O'Connor-Parkview', 'Oakridge', 'Pleasant View', 'Steeles', 'Tam O'Shanter-Sullivan', 'Taylor-Massey', 'The Beaches', 'Thorncliffe Park', 'Victoria Village', 'Wexford/Maryvale'.

1st Most Common Venue in Cluster 1	2nd Most Common Venue in Cluster 1	3rd Most Common Venue in Cluster 1	4th Most Common Venue in Cluster 1	5th Most Common Venue in Cluster 1
Italian Restaurant	Chinese Restaurant	Mobile Phone Shop	Pharmacy	Cafe

- Max number of Chinese: 14860
- Min number of Chinese: 75
- Mean number of Chinese: 1447

Analyzing Cluster 2

- The neighborhood in Cluster 2 is 'St.Andrew-Windfields'.

As there is only 1 neighborhood under Cluster 2, we will not be performing the statistical analysis of number of chinese and the mode of each most common venue.

1st Most Common Venue in Cluster 2	2nd Most Common Venue in Cluster 2	3rd Most Common Venue in Cluster 2	4th Most Common Venue in Cluster 2	5th Most Common Venue in Cluster 2
Grocery Store	Shopping Mall	Japanese Restaurant	Furniture/Home Store	Print Shop

- Number of Chinese: 3065

Analyzing Cluster 3

- The neighborhoods in Cluster 3 are 'Annex', 'Bathurst Manor', 'Bay Street Corridor', 'Bedford Park-Nortown', 'Beechborough-Greenbrook', 'Birchcliffe-Cliffside', 'Blake-Jones', 'Briar Hill-Belgravia', 'Bridle Path-Sunnybrook-York Mills', 'Broadview North', 'Brookhaven-Amesbury', 'Cabbagetown-South St.James Town', 'Caledonia-Fairbank', 'Casa Loma', 'Church-Yonge Corridor', 'Clairlea-Birchmount', 'Clanton Park', 'Corso Italia-Davenport', 'Dovercourt-Wallace Emerson-Juncti', 'Downsview-Roding-CFB', 'Dufferin Grove', 'Englemount-Lawrence', 'Eringate-Centennial-West Deane', 'Forest Hill North', 'Forest Hill South', 'Greenwood-Coxwell', 'Humbermede', 'Humewood-Cedarvale', 'Kensington-Chinatown', 'Kingsview Village-The Westway', 'Lansing-Westgate', 'Lawrence Park North', 'Lawrence Park South', 'Leaside-Bennington', 'Little Portugal', 'Moss Park', 'Mount Olive-Silverstone-Jamestown', 'Mount Pleasant East', 'Mount Pleasant West', 'Newtonbrook East', 'Newtonbrook West', 'Niagara', 'North Riverdale', 'North St.James Town', 'Oakwood Village', 'Old East York', 'Palmerston-Little Italy', 'Parkwoods-Donalda', 'Playter Estates-Danforth', 'Princess-Rosethorn', 'Regent Park', 'Rockcliffe-Smythe', 'Roncesvalles', 'Rosedale-Moore Park', 'South Riverdale', 'Thistletown-Beaumond Heights', 'Trinity-Bellwoods', 'University', 'Waterfront Communities-The Island', 'West Humber-Clairville', 'Westminster-Branson', 'Weston-Pellam Park', 'Willowdale East', 'Willowdale West', 'Willowridge-Martingrove-Richview', 'Woodbine Corridor', 'Woodbine-Lumsden', 'Wychwood', 'Yonge-Eglinton', 'Yonge-St.Clair', 'Yorkdale-Glen Park'.

Analyzing Cluster 3

1st Most Common Venue in Cluster 3	2nd Most Common Venue in Cluster 3	3rd Most Common Venue in Cluster 3	4th Most Common Venue in Cluster 3	5th Most Common Venue in Cluster 3
Fast Food Restaurant	Bank	Discount Store	Beer Store	Department Store

- Max number of Chinese: 4620
- Min number of Chinese: 320
- Mean number of Chinese: 1465

Analyzing Cluster 4

- The neighborhoods in Cluster 4 are 'Alderwood', 'Black Creek', 'Edenbridge-Humber Valley', 'Elms-Old Rexdale', 'Etobicoke West Mall', 'Glenfield-Jane Heights', 'High Park North', 'High Park-Swansea', 'Humber Heights-Westmount', 'Humber Summit', 'Islington-City Centre West', 'Junction Area', 'Keele-Edenbridge', 'Keele-Islington', 'Keele-Islington West', 'Kingsway South', 'Lambton Baby Point', 'Long Branch', 'Maple Leaf', 'Markland Wood', 'Mimico', 'Mount Dennis', 'New Toronto', 'Pelmo Park-Humberlea', 'Rexdale-Kipling', 'Runnymede-Bloor West Village', 'Rustic', 'South Parkdale', 'Stonegate-Queensway', 'Weston', 'York University Heights'.

1st Most Common Venue in Cluster 4	2nd Most Common Venue in Cluster 4	3rd Most Common Venue in Cluster 4	4th Most Common Venue in Cluster 4	5th Most Common Venue in Cluster 4
Fast Food Restaurant	Coffee Shop	Bus Line	Park	Chinese

- Max number of Chinese: 16790
- Min number of Chinese: 595
- Mean number of Chinese: 5324

Analyzing Cluster 5

- The neighborhoods in Cluster 5 are 'Agincourt North', 'Agincourt South-Malvern West', 'Bendale', 'Centennial Scarborough', 'Cliffcrest', 'Eglinton East', 'Guildwood', 'Highland Creek', 'Kennedy Park', 'Malvern', 'Morningside', 'Rouge', 'Scarborough Village', 'West Hill', 'Woburn'.

1st Most Common Venue in Cluster 5	2nd Most Common Venue in Cluster 5	3rd Most Common Venue in Cluster 5	4th Most Common Venue in Cluster 5	5th Most Common Venue in Cluster 5
Coffee Shop	Coffee Shop	Discount Store	Cafe	Doner Restaurant

- Max number of Chinese: 9810
- Min number of Chinese: 50
- Mean number of Chinese: 850

Conclusion

- For cluster 1, one of the glaring features is that Chinese Restaurants is the top 2 most common venues, coupled with the acceptable number of Chinese (Mean of 1447), does not make it an ideal cluster to set up another Chinese Restaurant. Competition is expected to be tough and the restaurants have to compete with an average number of Chinese populace.
- Even though Cluster 5 does not have any Chinese Restaurants as its top 5 most common venues, the mean number of Chinese populace is only 850. This also makes it not an ideal location as the pool of potential customers will be low.

Conclusion

- Cluster 3 does not have any Chinese Restaurants as its top 5 most common venues but its mean number of Chinese populace of 1465 does not make it an ideal location as well.
- Eventually, we narrowed our choices to either Cluster 2 or 4. Cluster 2 has a respectable number of Chinese at 3065 with no Chinese Restaurants at its top venues. Cluster 4, albeit having Chinese Restaurants as its 5th most common venue, this is more than compensated by high average number of Chinese of 5324. Hence, the aspiring Chinese chef or business owner can select neighborhoods from either Cluster 2 or 4.