Report:
Clustering Toronto boroughs and neighborhoods by restaurants' cuisines

Coursera capstone project by Eve Belyaeva

1. Introduction

Previously we explored neighborhoods in Toronto and segmented and clustered them.

I thought that it would be interesting if we could add tags to the neighborhoods for people who are looking for a place to visit if they are looking for new tastes in food.

For example, combine different places by cuisines they have. We could use statistics about restaurant count at all and cuisines distribution across the places. So, when a customer wants a travel agent to advise the best place to spend time and eat great food it should be easier to provide comfortable variants according to the customer's tastes.

I think that it would be useful not only for me but for many people who travel to taste different food and experience the variety of flavors.

2. Data

To do the cuisine sorter we need to gather information about places from Foursquare such as restaurants and cuisines and make the table containing different cuisines.

Then we should find how many venues are contained in one place and calculate the most common venues for each. For example, if in place there are a lot of restaurants with Chinese cuisines then we could add it to other Chinese-oriented places. For each place we can add tags based on venue information and location: for example, combine them into great cluster with mark. And after that we should be able to get ranked list of neighborhoods on the choropleth map showing which places are suitable.

3. Methodology

For the research I composed a DataFrame which contained the following:

- Postalcodes, borough, neighbourhood and geospatial data;
- Venues: restaraunts with cuisines.

The first step was the obtaining data about Toronto, Canada. It resulted the way that is shown further.

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	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

Figure 1 Toronto Data

Secondly, I added this places to a map.

Figure 2 Toronto Boroughs map

And then I used Foursquares API to gather info about restaurants, with categories and coordinates for each neighborhood in Toronto.

	FSA	Borough	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	Federick Restaurant	43.774697	-79.241142	Hakka
3	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	Terry's Restaurant & Bar	43.774780	-79.241043	Restaurant
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	terry's restaurant	43.774969	-79.240872	Italian
5	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	Thai One On	43.774468	-79.241268	Thai
6	M1P	Scarborough	Dorset Park,Scarborough Town Centre,Wexford He...	43.757410	-79.273304	Karaikudi Chettinad South Indian Restaurant	43.756042	-79.276276	Indian

Figure 3 Restaurants in Toronto

For customer to understand the distribution of food related places it is good to show on a map how many entities are in each borough to visit.

	FSA	Count
0	M1H	3
1	M1P	2
2	M1R	1
3	M1S	2
4	M1T	3

	Postcode	Borough	Latitude	Longitude
0	M1B	Scarborough	43.806686	-79.194353
1	M1C	Scarborough	43.784535	-79.160497
2	M1E	Scarborough	43.763573	-79.188711
3	M1G	Scarborough	43.770992	-79.216917
4	M1H	Scarborough	43.773136	-79.239476

Figure 4 Restaurant count

Figure 5 Borough Data

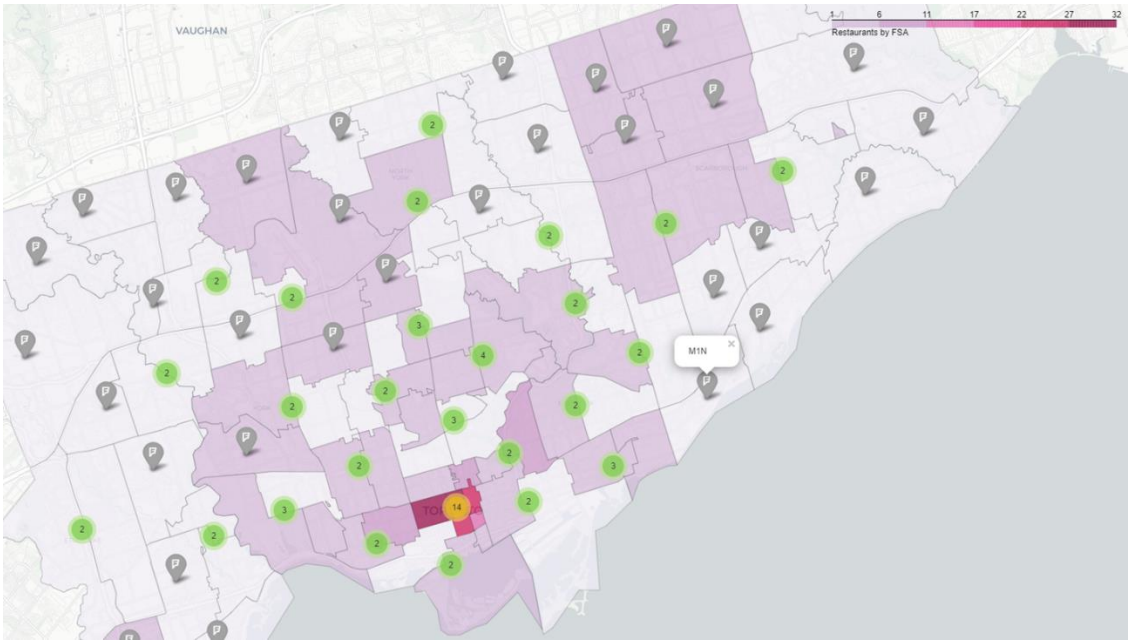

Figure 6 Choropleth map: Restaurant count

To understand how much distinct each neighborhood is I decided to plot some statistical data to show tops. Here are the cuisines and boroughs.

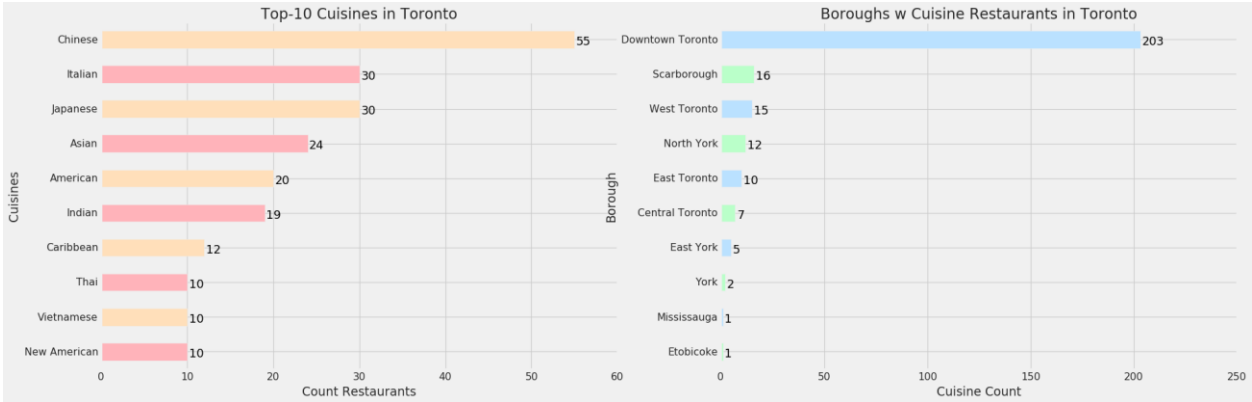

Figure 7 Top Cuisines and Boroughs

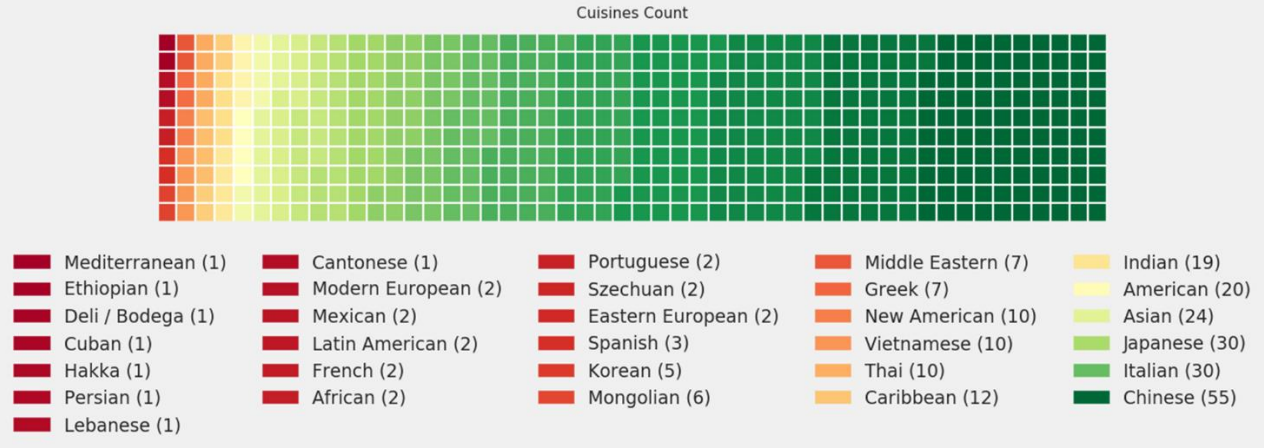

Figure 8 All cuisines relationship

The downtown Toronto neighborhood was too large to understand it's diversity so here is plot for only one neighborhood.

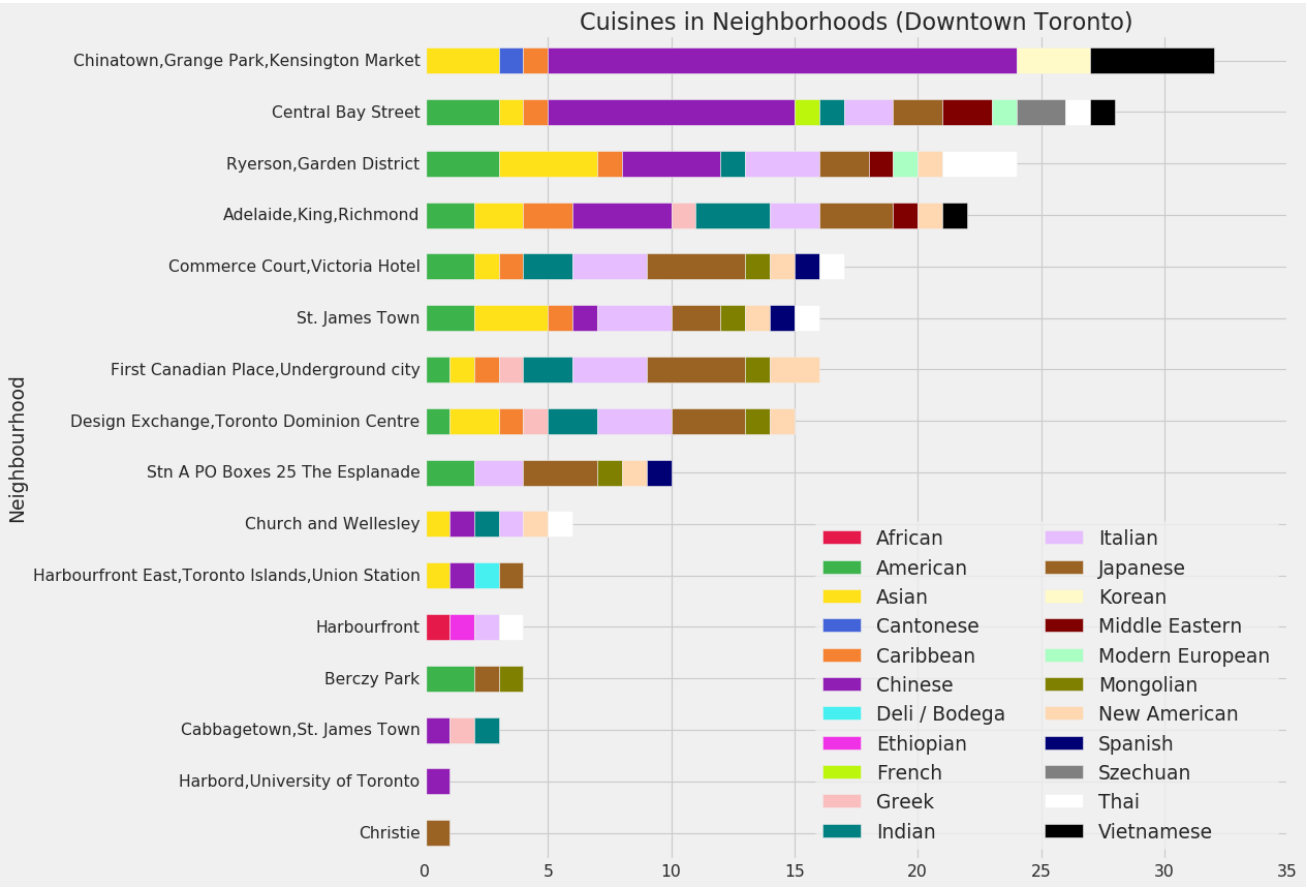

Figure 9 Downtown Toronto cuisines

So how to understand the relationship between each neighborhood? I chose to obtain clusters from the dataframe I combined. But first the data needed to be wrangled the right way. According to that I transformed the set I has to a table with common venue sorting after performing pandas one hot encoding.

	Borough	African	American	Asian	Cantonese	Caribbean	Chinese	Cuban	Deli / Bodega	Eastern European	Ethiopian	French	Greek	Hakka	Indian	Italian	Japanese	Korean	Latin American	Lebanese	Mediterranean	Mexican	Middle Eastern	Modern European	Mongolian	Am
0	Central Toronto	0.000000	0.000000	0.142857	0.000000	0.000000	0.142857	0.000000	0.000000	0.142857	0.000000	0.000000	0.000000	0.0000	0.142857	0.285714	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0
1	Downtown Toronto	0.004926	0.088670	0.093596	0.004926	0.044335	0.206897	0.000000	0.004926	0.000000	0.004926	0.004926	0.019704	0.0000	0.064039	0.113300	0.128079	0.014778	0.000000	0.0	0.0	0.000000	0.019704	0.009852	0.029557	0.0
2	East Toronto	0.000000	0.000000	0.100000	0.000000	0.100000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.300000	0.0000	0.100000	0.100000	0.000000	0.000000	0.000000	0.1	0.1	0.000000	0.000000	0.000000	0.000000	0.0
3	East York	0.200000	0.200000	0.000000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.200000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.200000	0.000000	0.000000	0.0
4	Etobicoke	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	1.000000	0.000000	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0
5	Mississauga	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0
6	North York	0.000000	0.000000	0.166667	0.000000	0.000000	0.083333	0.000000	0.000000	0.083333	0.000000	0.000000	0.000000	0.0000	0.083333	0.166667	0.250000	0.000000	0.000000	0.0	0.0	0.000000	0.083333	0.000000	0.000000	0.0
7	Scarborough	0.000000	0.000000	0.000000	0.000000	0.000000	0.562500	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0625	0.125000	0.062500	0.000000	0.062500	0.000000	0.0	0.0	0.000000	0.062500	0.000000	0.000000	0.0
8	West Toronto	0.000000	0.066667	0.066667	0.000000	0.000000	0.000000	0.066667	0.000000	0.000000	0.000000	0.066667	0.000000	0.0000	0.000000	0.066667	0.066667	0.000000	0.133333	0.0	0.0	0.133333	0.000000	0.000000	0.000000	0.0
9	York	0.000000	0.000000	0.000000	0.000000	0.500000	0.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0

Figure 10 Mean values for the cuisines in Toronto

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Central Toronto	Italian	Asian	Chinese	Eastern European	Indian
1	Downtown Toronto	Chinese	Japanese	Italian	Asian	American
2	East Toronto	Greek	Indian	Mediterranean	Italian	Asian
3	East York	African	American	Caribbean	Middle Eastern	Indian
4	Etobicoke	Korean	Vietnamese	Italian	American	Asian

Figure 11 Common cuisines in Toronto boroughs

As a methodology I used k-means clustering. The first step was to count K-value that would become a number of the future clusters.

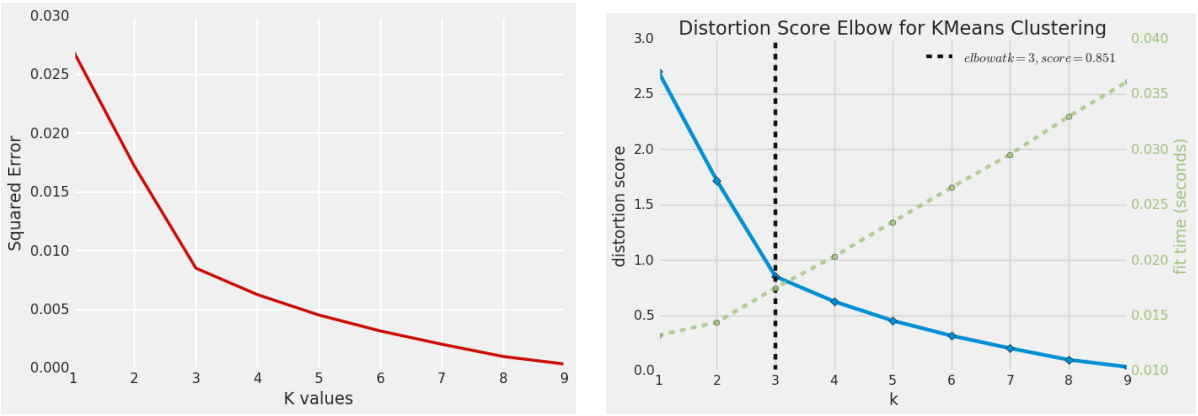

Figure 12 K-value

As the K vas calculated it was time to add cluster labels after calculating in which cluster each neighborhood would go.

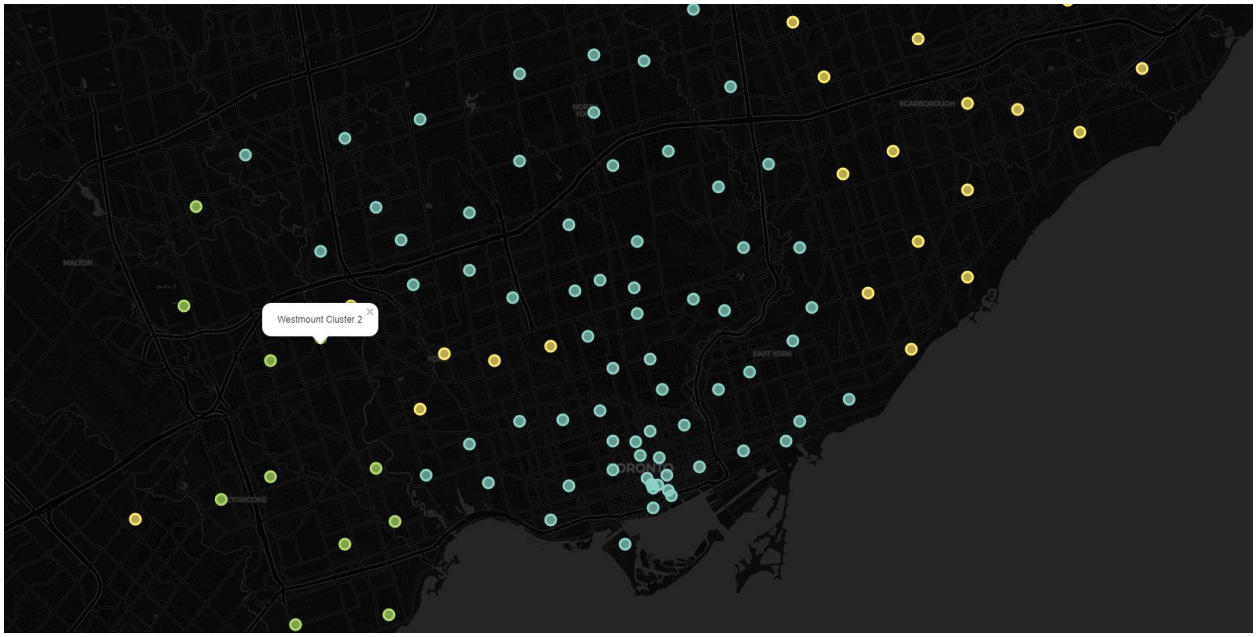

Figure 13 Cuisine clusters in Toronto

4. Results

As a result, I got three clusters with boroughs and neighborhoods. Let's look at them closer.

The first cluster is mostly Chinese cuisine at a first glance, but it is not so clear. For more understanding I plotted all the results additionally.

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Rouge,Malvern	0	Chinese	Indian	Italian	Thai	Korean
1	Highland Creek,Rouge Hill,Port Union	0	Chinese	Indian	Italian	Thai	Korean
2	Guildwood,Morningside,West Hill	0	Chinese	Indian	Italian	Thai	Korean
3	Woburn	0	Chinese	Indian	Italian	Thai	Korean
4	Cedarbrae	0	Chinese	Indian	Italian	Thai	Korean
5	Scarborough Village	0	Chinese	Indian	Italian	Thai	Korean
6	East Birchmount Park,Ionview,Kennedy Park	0	Chinese	Indian	Italian	Thai	Korean
7	Clairlea,Golden Mile,Oakridge	0	Chinese	Indian	Italian	Thai	Korean
8	Cliffcrest,Cliffside,Scarborough Village West	0	Chinese	Indian	Italian	Thai	Korean
9	Birch Cliff,Cliffside West	0	Chinese	Indian	Italian	Thai	Korean
10	Dorset Park,Scarborough Town Centre,Wexford He...	0	Chinese	Indian	Italian	Thai	Korean
11	Maryvale,Wexford	0	Chinese	Indian	Italian	Thai	Korean
12	Agincourt	0	Chinese	Indian	Italian	Thai	Korean
13	Clarks Corners,Sullivan,Tam O'Shanter	0	Chinese	Indian	Italian	Thai	Korean
14	Agincourt North,L'Amoreaux East,Milliken,Steel...	0	Chinese	Indian	Italian	Thai	Korean
15	L'Amoreaux West	0	Chinese	Indian	Italian	Thai	Korean
16	Upper Rouge	0	Chinese	Indian	Italian	Thai	Korean
73	Humewood-Cedarvale	0	Caribbean	Chinese	Vietnamese	Italian	American
74	Caledonia-Fairbanks	0	Caribbean	Chinese	Vietnamese	Italian	American
80	Del Ray,Keelesdale,Mount Dennis,Silverthorn	0	Caribbean	Chinese	Vietnamese	Italian	American
81	The Junction North,Runnymede	0	Caribbean	Chinese	Vietnamese	Italian	American
86	Canada Post Gateway Processing Centre	0	Chinese	Vietnamese	Italian	American	Asian
98	Weston	0	Caribbean	Chinese	Vietnamese	Italian	American

Figure 14 CLuster 0

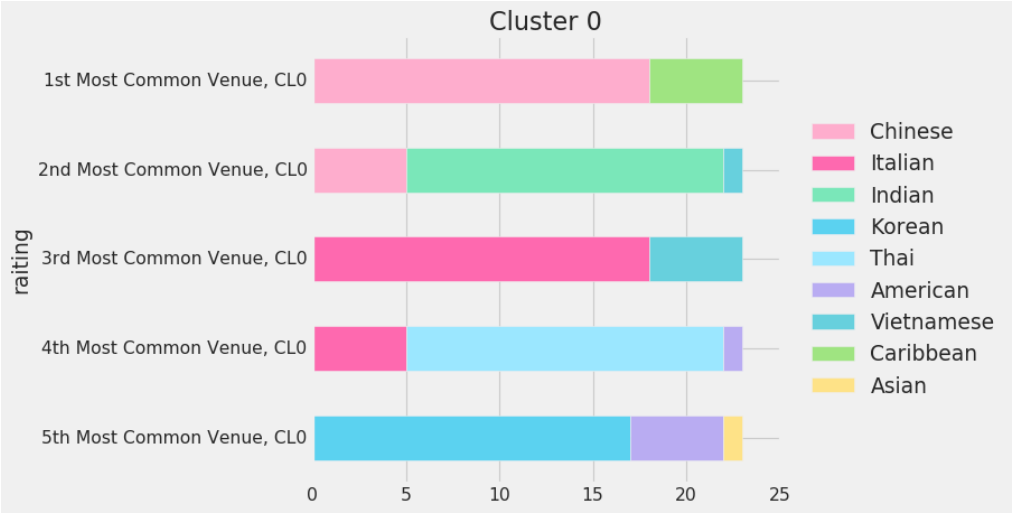

Figure 15 Cluster 0 plot

Yes, the Chinese cuisine is the most popular here, but Italian weight not less.

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
17	Hillcrest Village	1	Japanese	Asian	Italian	Middle Eastern	Chinese
18	Fairview,Henry Farm,Oriole	1	Japanese	Asian	Italian	Middle Eastern	Chinese
19	Bayview Village	1	Japanese	Asian	Italian	Middle Eastern	Chinese
20	Silver Hills,York Mills	1	Japanese	Asian	Italian	Middle Eastern	Chinese
21	Newtonbrook,Willowdale	1	Japanese	Asian	Italian	Middle Eastern	Chinese
22	Willowdale South	1	Japanese	Asian	Italian	Middle Eastern	Chinese
23	York Mills West	1	Japanese	Asian	Italian	Middle Eastern	Chinese
24	Willowdale West	1	Japanese	Asian	Italian	Middle Eastern	Chinese
25	Parkwoods	1	Japanese	Asian	Italian	Middle Eastern	Chinese
26	Don Mills North	1	Japanese	Asian	Italian	Middle Eastern	Chinese
27	Flemingdon Park,Don Mills South	1	Japanese	Asian	Italian	Middle Eastern	Chinese
28	Bathurst Manor,Downsview North,Wilson Heights	1	Japanese	Asian	Italian	Middle Eastern	Chinese
29	Northwood Park,York University	1	Japanese	Asian	Italian	Middle Eastern	Chinese
30	CFB Toronto,Downsview East	1	Japanese	Asian	Italian	Middle Eastern	Chinese
31	Downsview West	1	Japanese	Asian	Italian	Middle Eastern	Chinese
32	Downsview Central	1	Japanese	Asian	Italian	Middle Eastern	Chinese
33	Downsview Northwest	1	Japanese	Asian	Italian	Middle Eastern	Chinese
34	Victoria Village	1	Japanese	Asian	Italian	Middle Eastern	Chinese
35	Woodbine Gardens,Parkview Hill	1	African	American	Caribbean	Middle Eastern	Indian
36	Woodbine Heights	1	African	American	Caribbean	Middle Eastern	Indian
37	The Beaches	1	Greek	Indian	Mediterranean	Italian	Asian
38	Leaside	1	African	American	Caribbean	Middle Eastern	Indian
39	Thorncliffe Park	1	African	American	Caribbean	Middle Eastern	Indian
40	East Toronto	1	African	American	Caribbean	Middle Eastern	Indian
41	The Danforth West,Riverdale	1	Greek	Indian	Mediterranean	Italian	Asian
42	The Beaches West,India Bazaar	1	Greek	Indian	Mediterranean	Italian	Asian
43	Studio District	1	Greek	Indian	Mediterranean	Italian	Asian
44	Lawrence Park	1	Italian	Asian	Chinese	Eastern European	Indian
45	Davisville North	1	Italian	Asian	Chinese	Eastern European	Indian
46	North Toronto West	1	Italian	Asian	Chinese	Eastern European	Indian

Figure 16 Cluster 1

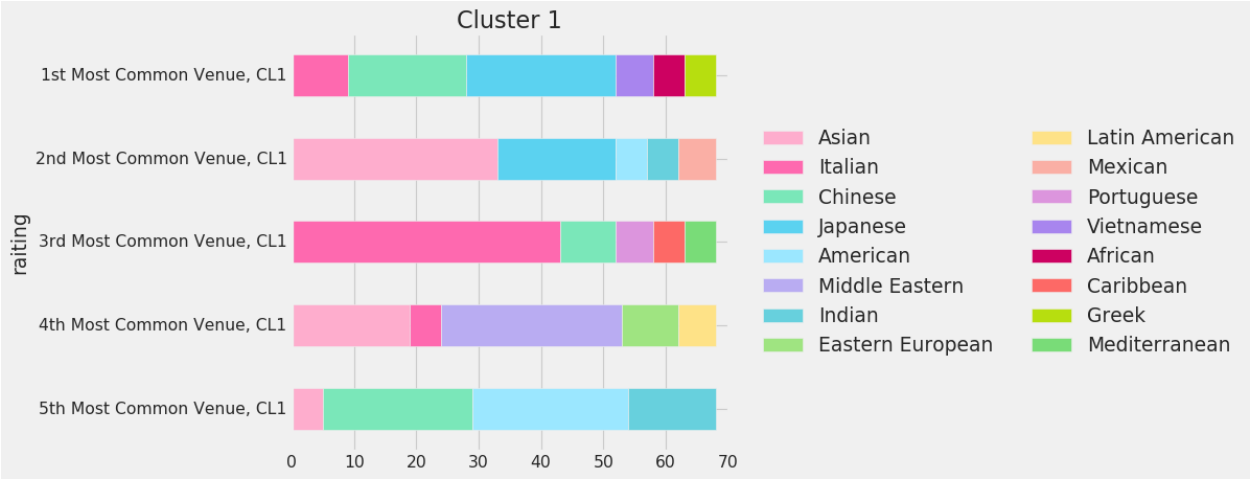

Figure 17 Cluster 1 plot

And in second cluster is much more heterogeneity. There are a lot cuisines but most common Italian and Asian.

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
88	Humber Bay Shores,Mimico South,New Toronto	2	Korean	Vietnamese	Italian	American	Asian
89	Alderwood,Long Branch	2	Korean	Vietnamese	Italian	American	Asian
90	The Kingsway,Montgomery Road,Old Mill North	2	Korean	Vietnamese	Italian	American	Asian
91	Humber Bay,King's Mill Park,Kingsway Park Sout...	2	Korean	Vietnamese	Italian	American	Asian
92	Kingsway Park South West,Mimico NW,The Queensw...	2	Korean	Vietnamese	Italian	American	Asian
94	Cloverdale,Islington,Martin Grove,Princess Gar...	2	Korean	Vietnamese	Italian	American	Asian
95	Bloordale Gardens,Eringate,Markland Wood,Old B...	2	Korean	Vietnamese	Italian	American	Asian
99	Westmount	2	Korean	Vietnamese	Italian	American	Asian
100	Kingsview Village,Martin Grove Gardens,Richvie...	2	Korean	Vietnamese	Italian	American	Asian
101	Albion Gardens,Beaumond Heights,Humbergate,Jam...	2	Korean	Vietnamese	Italian	American	Asian
102	Northwest	2	Korean	Vietnamese	Italian	American	Asian

Figure 18 Cluster 2

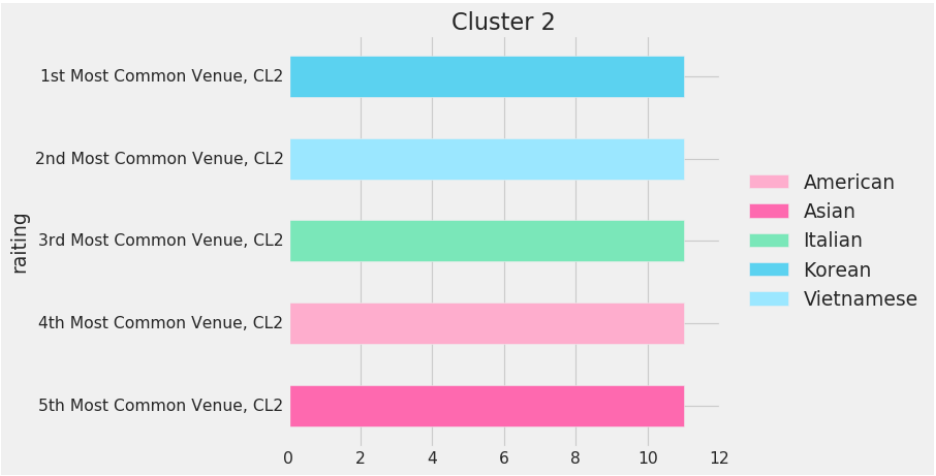

Figure 19 Cluster 2 plot

The most homogeneity cluster at all. All cuisines are equal here.

All clusters could be shown in one plot to explore.

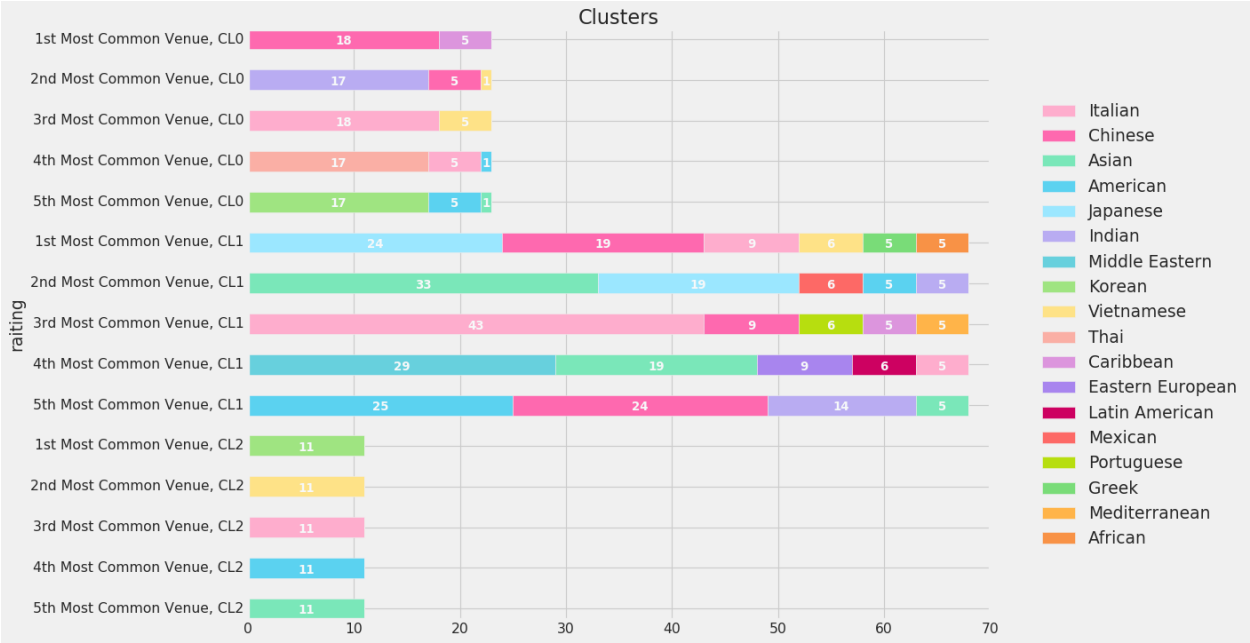

Figure 20 Clustered cuisines in Toronto

As a result to the traveler I decided to make a map with choropleth part with restaurant count and clusters with differentiation.

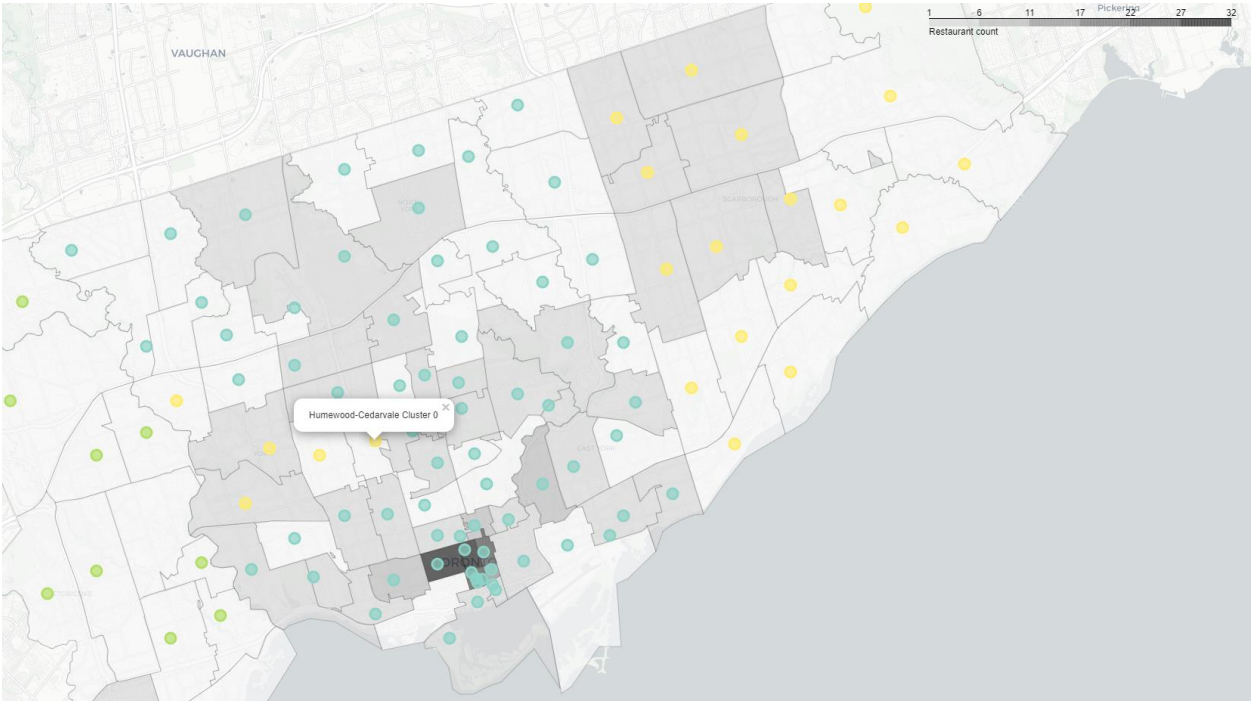

Figure 21 Toronto cuisines on a map

5. Discussion

Based on my observation results I understood that if you decided to explore all variety of food tastes in Toronto, you won't.

Much more likely you end up in common Chinese or Italian food and search for other interesting locations.

But as they are so much restaurants it will be good to know what the Chinese kitchen in Toronto is: meals, receipts and ingredients. Maybe it is not all the same everywhere.

6. Conclusion

In conclusion there is a useful map for choosing meal location if you live in Canada and like to travel in your homeland.

Many statistics was gathered based on cuisines, I find it interesting.

Ideas to add further:

- ratings for restaurant;
- tips for chosen restaurants;
- user profile to help with choice and predict tastes.

It would be useful to do the same research with France or Spanish neighborhoods and compare them to Toronto. Maybe there is a place in the world in which you could try and taste everything.