IBM Applied Data Science Capstone Report

EVALUATION OF LOCATIONS FOR A NEW HIGH QUALITY JAPANESE RESTAURANT IN TORONTO BASED ON DEMOGRAPHICS, AND LOCATIONS

Hendrika Kuffar

Table of Contents

Introduction

The goal of the project is to create location shortlist for a new high quality Japanese restaurant in the City of Toronto, Ontario, Canada.

Business Problem

Toronto is a home to some of the best restaurant around the world. However, several sources [1][2][3] have indicated there is a demand for high quality Japanese restaurant that serves delicious meals. The plan is to have a restaurant that will provide excellent customer service and serve lunch and supper meals.

The restaurant industry in Toronto is very competitive and therefore the strategy is to find locations with high foot and car traffic, affluent clientele, and low competition. The restaurant is also expected to accommodate fifty sittings. The suitable location is supposed to meet the following criteria:

1. Median family income above $70,000 (due to availability of data for the preliminary analysis average family income is used)
2. Population greater than 4700

3. Population density above 2000 per square km
4. Number of popular established business on Top 100 most popular venues within radius of 1 km greater than 16
5. Number of Japanese restaurant within the radius of 1 km less or equal to 4 depend on population, number of business and restaurants existed.

Targeted customers: affluent clientele (family income $70000+) in most densely populated neighborhoods as well as corporate

Audience & Stakeholders: restaurateurs/investors, restaurant critics, other businesses and people living in the neighborhood and their surrounding neighborhoods

Data

This project extracts insight from data to create location shortlist for a new high quality Japanese restaurant in Toronto. The project will utilize data for Toronto locations and neighborhood demographics. The average income per households after tax for each neighbourhood is used on preliminary analysis, because the median income values for individual neighbourhoods are not available on Toronto neighbourhood profiles csv files for 2016 census.

Toronto demographics

The demographic data from [Toronto neighbourhood profiles csv files for 2016 census](). The data is cleaned and transformed before it's combined with other data to find a suitable location for a new restaurant.

	_id	Category	Topic	Data Source	Characteristic	City of Toronto	Agincourt North	Agincourt South-Malvern West	Alderwood	Annex	Banbury-Don Mills	Bathurst Manor	Bay Street Corridor	Bayview Village	Bayview Woods-Steeles	Bedford Park-Nortown	Beechborough-Greenbrook	Bendale	Birchcliff-Cliffsic
0	1	Neighbourhood Information	Neighbourhood Information	City of Toronto	Neighbourhood Number	NaN	129	128	20	95	42	34	76	52	49	39	112	127	12
1	2	Neighbourhood Information	Neighbourhood Information	City of Toronto	TSNS2020 Designation	NaN	No Designation	No Designation	No Designation	No Designation	No Designation	No Designation	No Designation	No Designation	No Designation	No Designation	NIA	No Designation	N Designatic
2	3	Population	Population and dwellings	Census Profile 98-316-X2016001	Population, 2016	2,731,571	29,113	23,757	12,054	30,526	27,695	15,873	25,797	21,396	13,154	23,236	6,577	29,960	22,2
3	4	Population	Population and dwellings	Census Profile 98-316-X2016001	Population, 2011	2,615,060	30,279	21,988	11,904	29,177	26,918	15,434	19,348	17,671	13,530	23,185	6,488	27,876	21,8
4	5	Population	Population and dwellings	Census Profile 98-316-X2016001	Population Change 2011-2016	4.50%	-3.90%	8.00%	1.30%	4.60%	2.90%	2.80%	33.30%	21.10%	-2.80%	0.20%	1.40%	7.50%	2.00

Other data set which are used for this project have neighbourhoods as rows. Thus, the data is transposed. The information related to average family income after tax, land area, and population for each neighbourhood are sliced from the data.

Neighbourhood	Population	Density	Area square km	Average Income
Agincourt North	29,113	3,929	7.41	427,037
Agincourt South-Malvern West	23,757	3,034	7.83	278,390
Alderwood	12,054	2,435	4.95	168,602
Annex	30,526	10,863	2.81	792,507
Banbury-Don Mills	27,695	2,775	9.98	493,486

Toronto postal code, borough and Neighbourhood

The sample of Toronto postal code data scrapped from Wikipedia table using beautiful soup and transformed to data frame is shown below. The data consist of combination of official and unofficial neighbourhood name.

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M5A	Downtown Toronto	Regent Park
6	M6A	North York	Lawrence Heights
7	M6A	North York	Lawrence Manor
8	M7A	Queen's Park	Not assigned
9	M8A	Not assigned	Not assigned
10	M9A	Etobicoke	Islington Avenue
11	M1B	Scarborough	Rouge
12	M1B	Scarborough	Malvern
13	M2B	Not assigned	Not assigned
14	M3B	North York	Don Mills North

The initial cleaning of Toronto postal code data was done to remove 'Not Assigned' Borough and renaming 'Not Assigned' Neighbourhood.

	Postcode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M5A	Downtown Toronto	Regent Park
4	M6A	North York	Lawrence Heights
5	M6A	North York	Lawrence Manor
6	M7A	Queen's Park	Queen's Park
7	M9A	Etobicoke	Islington Avenue
8	M1B	Scarborough	Rouge
9	M1B	Scarborough	Malvern
10	M3B	North York	Don Mills North
11	M4B	East York	Woodbine Gardens
12	M4B	East York	Parkview Hill
13	M5B	Downtown Toronto	Ryerson
14	M5B	Downtown Toronto	Garden District

There are about 211 neighbourhoods listed after initial clean up. However the official list of Toronto contains 140 neighbourhoods. The name for majority of neighbourhood does not match the actual list of official neighbourhoods on neighbourhood profiles and other sources. For example, only 2 neighbourhoods out of 5 are exactly a match on sorted data by designated neighbourhood from A-Z.

https://en.wikipedia.org/wiki/List_of_city-designated_neighbourhoods_in_Toronto

CDN number	City-designated area	Former city/borough	Neighbourhoods covered
129	Agincourt North	Scarborough	Agincourt and Brimwood
128	Agincourt South-Malvern West	Scarborough	Agincourt and Malvern
020	Alderwood	Etobicoke	Alderwood
095	Annex	Old City of Toronto	The Annex and Seaton Village
042	Banbury-Don Mills	North York	Don Mills

	Postcode	Borough	Neighbourhood
0	M5H	Downtown Toronto	Adelaide
1	M1S	Scarborough	Agincourt
2	M1V	Scarborough	Agincourt North
3	M9V	Etobicoke	Albion Gardens
4	M8W	Etobicoke	Alderwood

The scrapped pre-clean data was saved to object storage on IBM cloud as csv file for further cleanup. The saved file was updated using information from [Wikipedia list of designated Toronto neighbourhoods](). The update included adding missing neighbourhood and updates the one in the list to use designated neighbourhood names e.g. replace "The Annex" with Annex. After mapping the designated neighbourhoods to postal code, the top 5 neighbourhoods sorted in ascending order match the list from Wikipedia.

	Postcode	Borough	Neighbourhood
0	M1V	Scarborough	Agincourt North
1	M1S	Scarborough	Agincourt South-Malvern West
2	M8W	Etobicoke	Alderwood
3	M5R	Central Toronto	Annex
4	M3B	North York	Banbury-Don Mills

The change made easy to merge the demographic data with the postal code ones based on 'Neighbourhoods' without losing a lot of important information prior to analysis. Below is the sample of data after merging postal code and demographic data for the city of Toronto.

Postcode	Borough	Neighbourhood	Population	Density	Area square km	Average Income
M1V	Scarborough	Agincourt North	29113	3929	7.41	99071.293163
M1S	Scarborough	Agincourt South-Malvern West	23757	3034	7.83	64585.638490
M8W	Etobicoke	Alderwood	12054	2435	4.95	39115.154354
M5R	Central Toronto	Annex	30526	10863	2.81	183859.228430
M3B	North York	Banbury-Don Mills	27695	2775	9.98	114487.260303

Geolocation postal code longitude and latitude

The list of postal code geographical locations in Toronto is used. The geolocation data are merged with data for demographic and neighbourhoods. The location latitude and longitude values are needed in order to obtain information about the venue from foursquare.

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

Foursquare location venues information

The Foursquare API was used to retrieve the list of most popular top 100 venues for each of the selected popular neighborhoods location in Toronto within a radius of 1 km. The information is important in understanding different categories of venues that exist within the selected radius of the neighbourhood. Below is the sample for foursquare data for venues for Rosedale-Moore Park neighbourhood in Toronto.


```
{'meta': {'code': 200, 'requestId': '5d5dc7ecbf7dde002ce26574'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'Rosedale',
  'headerFullLocation': 'Rosedale, Toronto',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 27,
  'suggestedBounds': {'ne': {'lat': 43.68856260900001,
    'lng': -79.36510816548741},
   'sw': {'lat': 43.670562590999985, 'lng': -79.38995063451262}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '4adcb343f964a520e32e21e3',
       'name': 'Summerhill Market',
       'location': {'address': '446 Summerhill Ave',
        'crossStreet': 'btwn. MacLennan Ave. and Glen Rd.',
```


Methodology

The data were acquired from multiple sources, cleaned, scaled, aggregated, merged, filtered, analyzed and visualized. New features were also created based on information from the venues. The features were used to create locations clusters with k-means algorithm based on demographic and venue features.

Neighbourhood Demographic and Location

The data set was created for neighbourhoods that fit the demographic criteria for targeted customer and location based on 2016 census data for the city of Toronto. The first thing that was done is to check whether data require scaling or not. The comparison of average income per households after tax for city of Toronto listed on datasheet and the calculated average from neighbourhoods values indicated that scaling is required for neighbourhoods average family income values. The average family income from data sheet for the city of Toronto is $81,495 and estimated average from neighbourhood values is ~ $ 351,276. Moreover, the minimum average income for the neighbourhoods is about $ 102,259 as shown on statistical value below.

```
The Average Income for City of Toronto =  81495.0

Estimated Values of Income, Density, and Population for Toronto
The estimated Density for Neighbourhoods =  4334.175869510028

The estimated Population for Neighbourhoods =  2731571

The estimated Average Income for Neighbourhoods =  351276.1285714286
Minimum Average Income =  102259.0
```

Characteristic	Population	Density	Area square km	Average Income
count	140.000000	140.000000	140.000000	1.400000e+02
mean	19511.221429	6261.135714	4.501714	3.512761e+05
std	10033.589222	4840.359075	4.544665	2.309379e+05
min	6577.000000	1040.000000	0.420000	1.022590e+05
25%	12019.500000	3595.250000	1.852500	1.953375e+05
50%	16749.500000	5071.500000	3.275000	2.915495e+05
75%	23854.500000	7621.250000	5.382500	4.305408e+05
max	65913.000000	44321.000000	36.890000	1.413132e+06

The issue of average family income from 2016 census in Toronto is complex. The problem with average value is it could be inflated or deflated by few household with higher or lower income. The other issue is family income values could have been collected from the smaller sample (25% of the population) that is not representative of the whole population.

It would have been nice to have family median income value for majority family for each neighbourhood available on csv file for Toronto neighbourhood profiles. This is because median value is good representation of the majority family income. For simplicity the average income for each neighbourhood is divided by 4.31 (351,276/81,495). This factor will not take care of inflated or deflated average family income in some of the neighbourhoods.

On the other hand, the density value for city of Toronto is about 4334 and the calculated mean 6261 per square kilometer. The mean value for population density of neighbourhoods is not the same as the population density for the city. This is because mathematically:

$$\frac{\sum x}{\sum y} \neq \frac{1}{n}\sum\frac{x}{y}$$

Where:

x=Population of each neighbourhood
y=Land area per square km of each neighbourhood

The individual sum of population for all neighbourhoods divide by total land areas for these neighbourhoods will give us the 4334 per square kilometer. Hence, the new aggregated density value is calculated using formula below. In other word the recalculation of the population density for the neighbourhoods which share the postal code is done by divide the sum of population for all the neighbourhoods with the sum of land area.

$$\frac{\sum x}{\sum y}$$

The data are aggregated so that we can have one unique postal code per location. The postal codes are used to obtain geolocation.

Postcode	Neighbourhood	Population	Area square km	Density	Average Income
M1B	Malvern,Rouge	90290	45.74	1973	146431.388091
M1C	Centennial Scarborough,Highland Creek	25856	10.59	2441	47762.957679
M1E	Guildwood,Morningside,West Hill	54764	19.04	2876	59286.750198
M1G	Woburn	53485	12.31	4344	145933.058584
M1J	Eglinton East,Scarborough Village	39500	6.33	6240	61362.968479

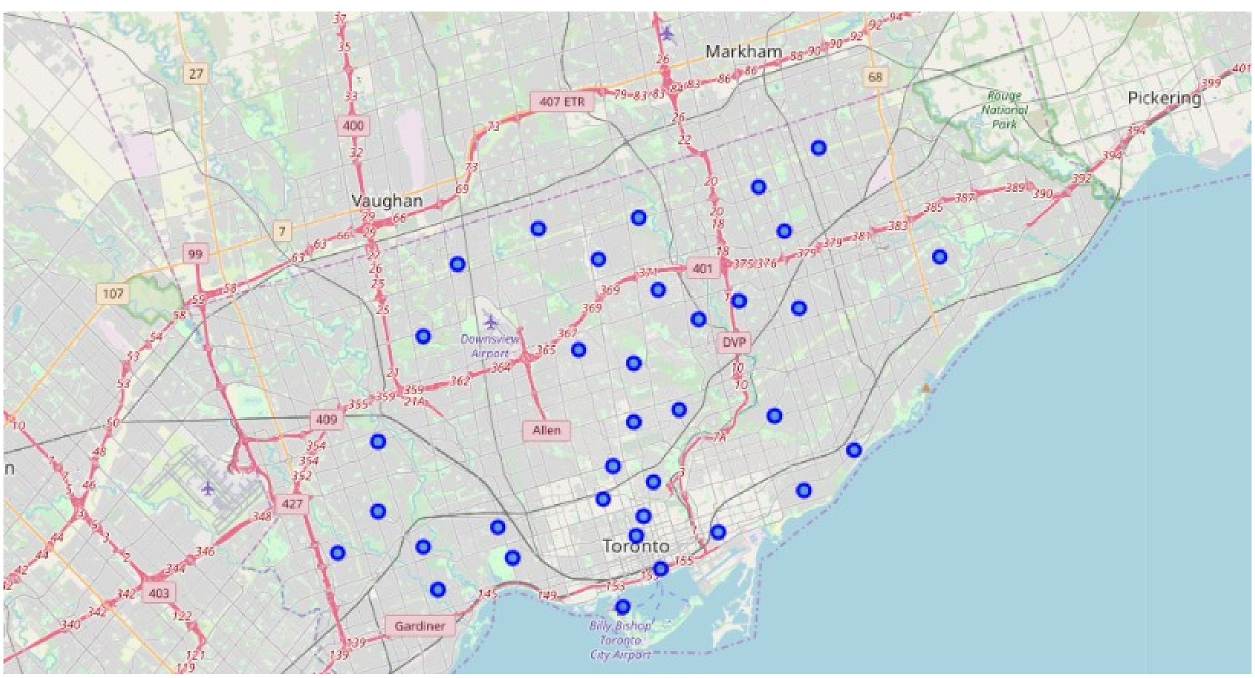
The scaled and aggregated data are then merged with geo location data set. The sample merged data that contain demographic and location data is shown below. The list corresponds to neighbourhoods which have met the demographic criteria for our suitable location and customers. The demographic criteria for suitable locations are:

1. Average Family Income > 70000
2. Population > 4700
3. Population Density > 2000

Postcode	Neighbourhood	Population	Area square km	Density	Average Income
M1G	Woburn	53485	12.31	4344	145933.058584
M1N	Birchcliffe-Cliffside	22291	5.92	3765	85644.236109
M1R	Wexford/Maryvale	27917	10.25	2723	78067.678827
M1T	Tam O'Shanter-Sullivan	27446	5.41	5073	110348.666226
M1V	Agincourt North,Milliken	55685	16.80	3314	110493.664337

Neighbourhood Maps

The visualization of the selected neighbourhoods on the map for the city of Toronto is shown below.



Neighbourhoods Venues

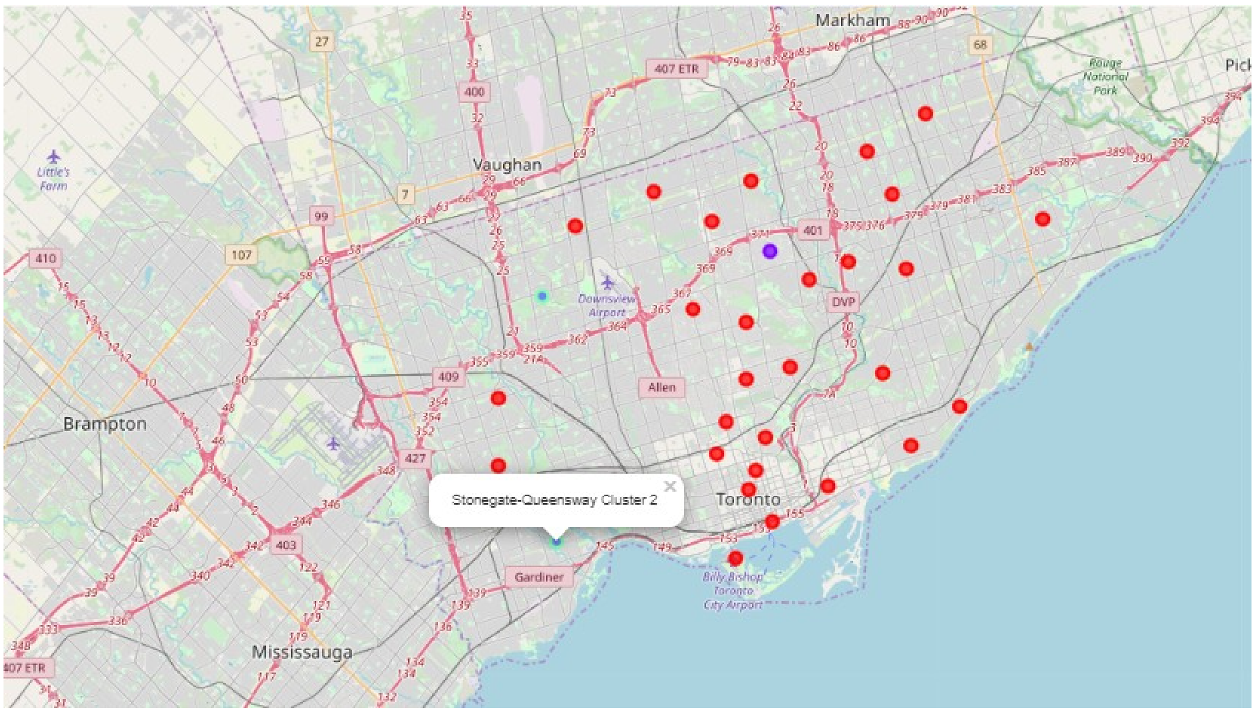
The venues data are obtained from foursquare. The venues information is crucial to determining the locations that are deemed suitable to open a new restaurant. The table below is the sample of venues for some of the selected locations in Toronto. The search was for top 100 popular venues within radius of one kilometer. However, some of the location had less than 100 venues. For example Agincourt North, Milliken had 29 venues within radius of 1 km.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agincourt North,Milliken	31	31	31	31	31	31
Annex,Casa Loma	100	100	100	100	100	100
Banbury-Don Mills	32	32	32	32	32	32
Bay Street Corridor	100	100	100	100	100	100
Bayview Village,Bayview Woods-Steeles	14	14	14	14	14	14
Bedford Park-Nortown,Lawrence Park North	41	41	41	41	41	41
Birchcliffe-Cliffside	15	15	15	15	15	15
Church-Yonge Corridor	100	100	100	100	100	100
East End-Danforth,The Beaches	80	80	80	80	80	80
Edenbridge-Humber Valley	13	13	13	13	13	13
Eringate-Centennial-West Deane,Islington-City Centre West,Princess-Rosethorn	15	15	15	15	15	15
Forest Hill South,Yonge-St.Clair	78	78	78	78	78	78
Glenfield-Jane Heights	8	8	8	8	8	8
High Park North,High Park-Swansea,Junction Area	100	100	100	100	100	100
Kingsway South	44	44	44	44	44	44
L'Amoreaux,Steeles	24	24	24	24	24	24
Lansing-Westgate,Willowdale East	100	100	100	100	100	100
Lawrence Park South	8	8	8	8	8	8
Leaside-Bennington	62	62	62	62	62	62
Mount Pleasant East,Mount Pleasant West	100	100	100	100	100	100
Niagara	15	15	15	15	15	15
O'Connor-Parkview	21	21	21	21	21	21
Parkwoods-Donalda	29	29	29	29	29	29
Roncesvalles	100	100	100	100	100	100
Rosedale-Moore Park	26	26	26	26	26	26
South Riverdale	100	100	100	100	100	100
St.Andrew-Windfields	4	4	4	4	4	4
Stonegate-Queensway	7	7	7	7	7	7
Tam O'Shanter-Sullivan	35	35	35	35	35	35
Waterfront Communities-The Island	100	100	100	100	100	100
Westminster-Branson,Willowdale West	12	12	12	12	12	12
Wexford/Maryvale	28	28	28	28	28	28
Willowridge-Martingrove-Richview	14	14	14	14	14	14
Woburn	10	10	10	10	10	10
York University Heights	22	22	22	22	22	22

Neighbourhoods Clusters

The k-means unsupervised machine learning algorithm is used to cluster the locations in two different forms. First to create **3 clusters based on similarity of venues** out for the selected locations. Then to **create 9 clusters of locations which have similar demographic characteristics and aggregated venues features**. The second set of clusters involved the creation and analysis of the new features.

Toronto map which has 3 clusters based similarity of venues correspond to each location are display below. Where cluster 0 is display in red, cluster 1 in purple and cluster 2 in cyan.

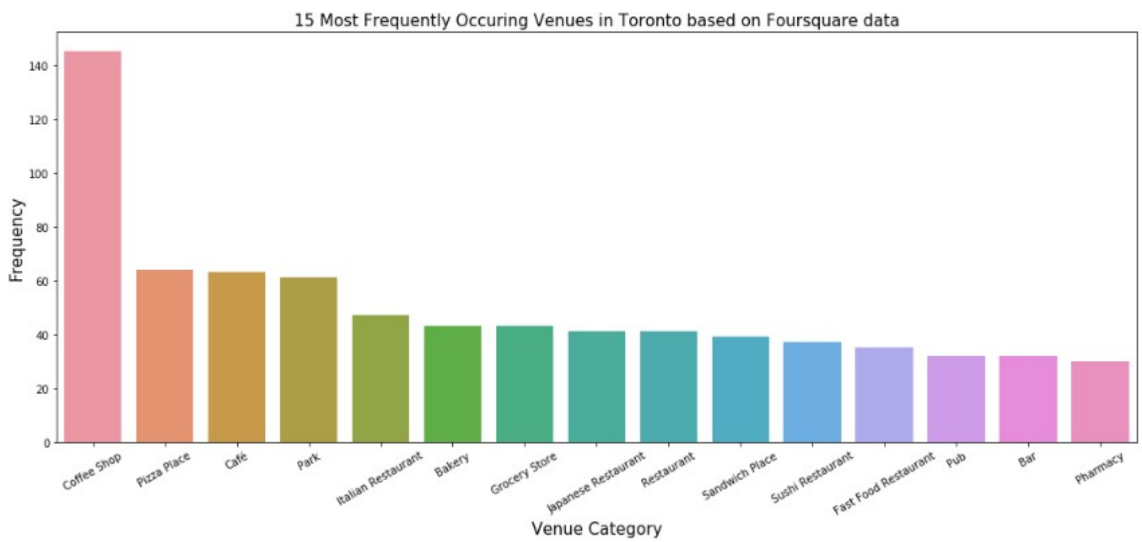


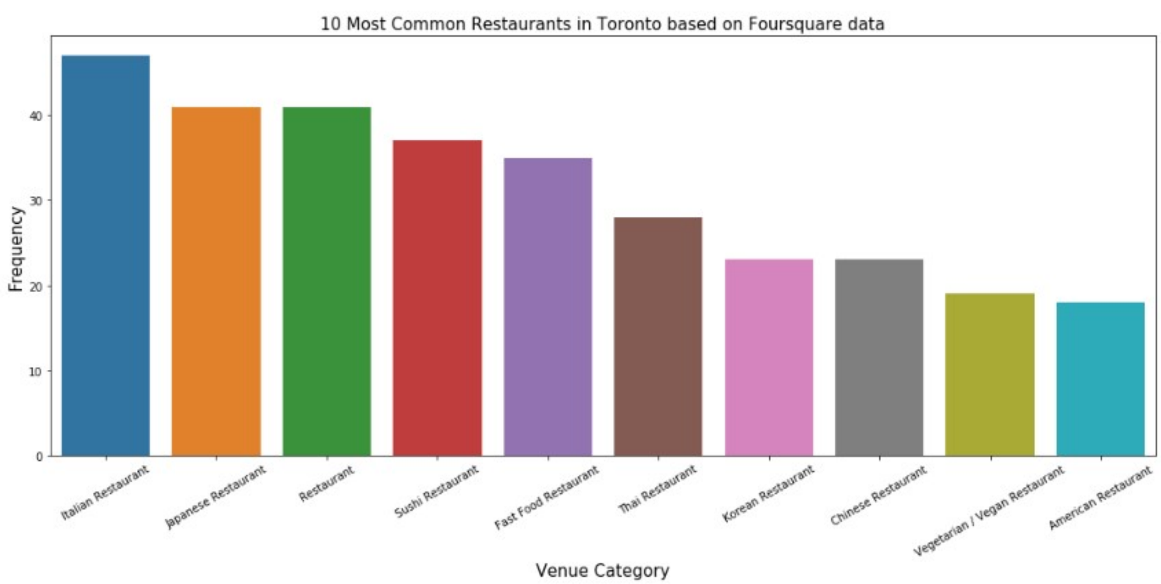
Cluster Labels 0: The **coffee shop/Cafe** is among the most popular venues for this cluster. The other venues categories on most common venues are stores, restaurants, banks, pubs, bars, gyms, bakeries and pizza places.

Cluster Labels 1: It only have one location St.Andrews-Windfield which has 4 venue around the radius of 1 km. The most frequent visited places are the **park** followed by the **pool**.

Cluster Labels 2: The **Park** is among the most popular venues for all locations corresponding to this cluster. There are also banks, bakeries, coffee shops, café, malls, pharmacies, stores and other restaurants in most of these locations.

Using the data we found the top 15 most popular venues and Top 10 most popular restaurants.





The clustering above is based on popularity of the venues and not demographic data. The number of total venue categories is over 240. From the venue categories new features are created by aggregating the venues data to represent[1]:

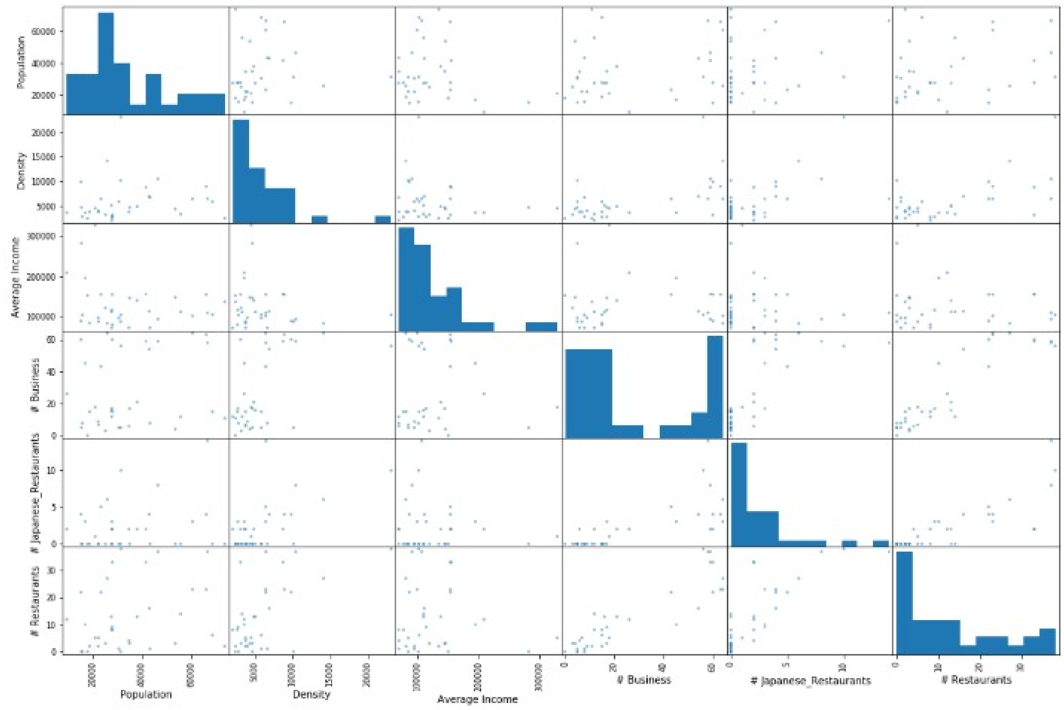
1. Number of Businesses
2. Number of Restaurants
3. Number of Japanese Restaurants

The new features were merged to demographic data for locations. Before creating clusters the features are analyzed and information is used to make adjustment. In order to get good insight from the clustering based on demographic and aggregated features, the correlated feautures are eliminated.

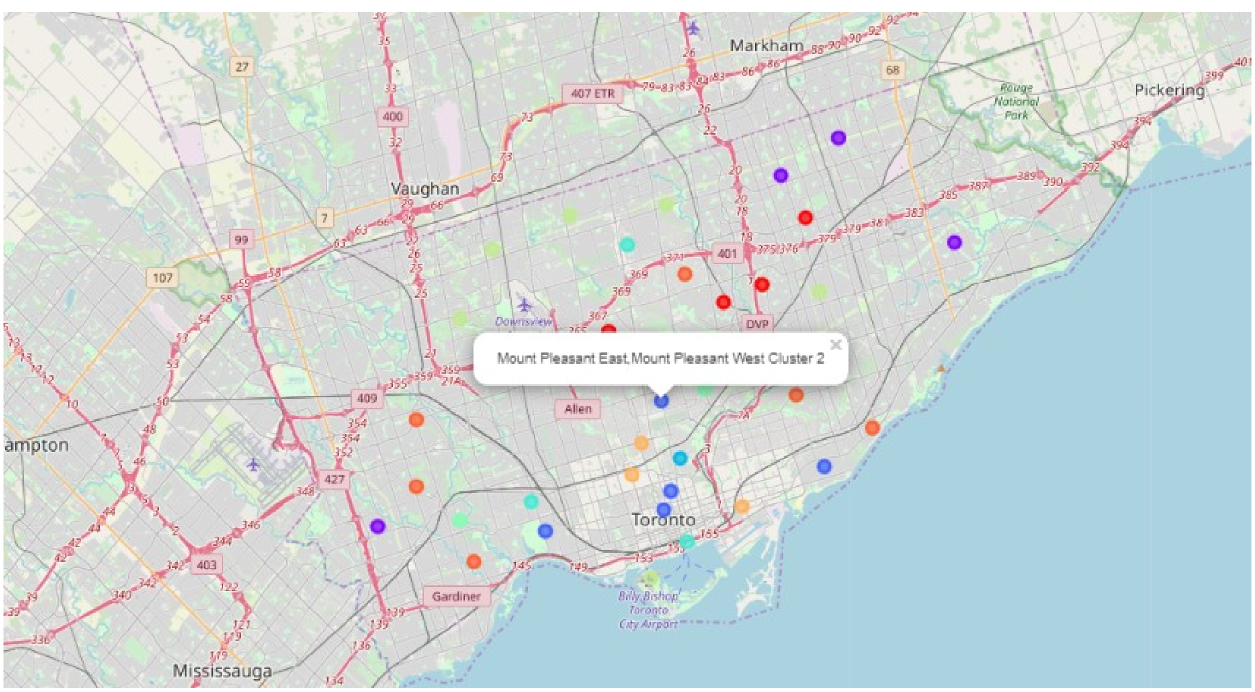
The pearson correlation is calculated, and the values of correlation are used to determine the features that are not correllated. The correlation matrix below show that there is **very strong correlation** between number of businesses and restaurants (0.9), and number of restaurants and Japanese restaurants (0.81). There is also **strong correlation** between the number of businesses and Japanese restaurants (0.73), population density and number of restuarants (0.6). Moreover, there is **moderate correlation** between population density and number of businesses.

	Population	Area square km	Density	Average Income	# Business	# Japanese_Restaurants	# Restaurants
Population	1.000000	0.637909	0.095340	-0.213518	0.215131	0.225134	0.253807
Area square km	0.637909	1.000000	-0.448334	-0.123913	-0.269283	-0.209849	-0.204087
Density	0.095340	-0.448334	1.000000	-0.168971	0.563686	0.600832	0.603548
Average Income	-0.213518	-0.123913	-0.168971	1.000000	-0.010136	-0.103332	-0.080902
# Business	0.215131	-0.269283	0.563686	-0.010136	1.000000	0.731745	0.900716
# Japanese_Restaurants	0.225134	-0.209849	0.600832	-0.103332	0.731745	1.000000	0.814420
# Restaurants	0.253807	-0.204087	0.603548	-0.080902	0.900716	0.814420	1.000000

[1]**Keywords for Business Categories:** Bakery, Bank, Bar, Brewery, Building, Business, Breakfast, Taco, Burrito, Butcher, store, Boutique, Coffee, Café, Creperie, Club, Chiropractor, Concert, Deli, Entertainment, Athletics, Service, Pharmacy, Hotel, Hostel, Health, Storage, Soup, Snack, Sandwich, Salon, Tea, Grocery, Store, pub, Historic, Gym, Studio, Rock, Shop, Museum, Plaza, Gallery, Pub, Office, Theater, Mall, Market, Wings, Gas, Rental, Pizza
Keywords for Restaurant Categories: Restaurant, BBQ, Bistro, Steakhouse, Noodle, Diner, Poke
Keywords for Japanese restaurant Categories: Ramen, Sushi, and Japanese



The population, average family income, and number of buisnesses are the features which were used to create clusters after eliminating the features which have correlation coefficient greater than 0.4. The values of these features is standardized first before the locations data are fitted into nine clusters using k-means clustering algorithm. Below are the colour coded markers correspond to different clusters superimposed on the map for Toronto based on their location.



Cluster Labels 0: locations with few restaurants, and population 27448 -37843

Cluster Labels	Neighbourhood	Population	Area square km	Density	Average Income	# Business	# Japanese_Restaurants	# Restaurants
0	Tam O'Shanter-Sullivan	27446.0	5.41	5073.0	110348.666226	18.0	0.0	12.0
0	Parkwoods-Donalda	34805.0	7.42	4690.0	144639.211442	17.0	0.0	3.0
0	Banbury-Don Mills	27695.0	9.98	2775.0	114487.260303	18.0	3.0	10.0
0	Bedford Park-Nortown,Lawrence Park North	37843.0	7.80	4851.0	138167.075770	26.0	1.0	12.0

Cluster Labels 1: locations with few numbers of businesses and restaurants, and high population

Cluster Labels	Neighbourhood	Population	Area square km	Density	Average Income	# Business	# Japanese_Restaurants	# Restaurants
1	Woburn	53485.0	12.31	4344.0	145933.058584	5.0	0.0	3.0
1	Agincourt North,Milliken	55685.0	16.80	3314.0	110493.664337	13.0	0.0	14.0
1	L'Amoreaux,Steeles	68616.0	11.69	5869.0	103520.299196	13.0	0.0	8.0
1	Eringate-Centennial-West Deane,Islington-City ...	73604.0	29.94	2458.0	134993.473120	10.0	0.0	2.0

Cluster Labels 2: locations with a lot of businesses and high number of Japanese restaurants

Cluster Labels	Neighbourhood	Population	Area square km	Density	Average Income	# Business	# Japanese_Restaurants	# Restaurants
2	East End-Danforth,The Beaches	42948.0	6.22	6904.0	111729.976229	54.0	4.0	15.0
2	Mount Pleasant East,Mount Pleasant West	46433.0	4.45	10434.0	91801.667882	59.0	8.0	37.0
2	Church-Yonge Corridor	31340.0	1.36	23044.0	102944.946692	56.0	10.0	38.0
2	Bay Street Corridor	25797.0	1.83	14096.0	81713.511324	64.0	6.0	27.0
2	Roncesvalles	14974.0	1.52	9851.0	87583.034848	61.0	4.0	22.0

Cluster Labels 3: locations with high income neighbourhood which are densely populated with few business and restaurants.

Cluster Labels	Neighbourhood	Population	Area square km	Density	Average Income	# Business	# Japanese_Restaurants	# Restaurants
3	Lawrence Park South	15179.0	3.24	4684.0	282244.042538	5.0	0.0	0.0
3	Rosedale-Moore Park	20923.0	4.65	4499.0	327842.352420	17.0	1.0	5.0

Cluster Labels 4: locations with a lot of businesses, high population, and high numbers of restaurants

Cluster Labels	Neighbourhood	Population	Area square km	Density	Average Income	# Business	# Japanese_Restaurants	# Restaurants
4	Lansing-Westgate,Willowdale East	66598.0	10.32	6453.0	107678.729014	58.0	14.0	37.0
4	Waterfront Communities-The Island	65913.0	7.37	8943.0	153659.253917	63.0	4.0	23.0
4	High Park North,High Park-Swansea,Junction Area	60453.0	9.42	6417.0	99839.976481	67.0	3.0	23.0

Cluster Labels 5: locations with population 9271 - 16828, density ~3595, and high average family income

Cluster Labels	Neighbourhood	Population	Area square km	Density	Average Income	# Business	# Japanese_Restaurants	# Restaurants
5	Leaside-Bennington	16828.0	4.68	3595.0	195253.063975	45.0	3.0	10.0
5	Kingsway South	9271.0	2.58	3593.0	208942.045617	24.0	2.0	12.0

Cluster Labels 6: locations with low number of businesses and restaurants

Cluster Labels	Neighbourhood	Population	Area square km	Density	Average Income	# Business	# Japanese_Restaurants	# Restaurants
6	Wexford/Maryvale	27917.0	10.25	2723.0	78067.678827	15.0	0.0	8.0
6	Bayview Village,Bayview Woods-Steeles	34550.0	9.16	3771.0	70518.961182	6.0	2.0	4.0
6	Westminster-Branson,Willowdale West	43210.0	6.49	6657.0	70267.476459	9.0	0.0	1.0
6	York University Heights	27593.0	13.23	2085.0	70146.142040	14.0	1.0	8.0
6	Glenfield-Jane Heights	30491.0	5.20	5863.0	95121.544626	4.0	0.0	1.0
6	Niagara	31180.0	3.07	10156.0	85765.338532	5.0	0.0	0.0

Cluster Labels 7: locations with a lot of restaurants and businesses

Cluster Labels	Neighbourhood	Population	Area square km	Density	Average Income	# Business	# Japanese_Restaurants	# Restaurants
7	South Riverdale	27876.0	8.89	3135.0	153733.028955	60.0	2.0	33.0
7	Forest Hill South,Yonge-St.Clair	23260.0	3.62	6425.0	153797.988109	42.0	5.0	22.0
7	Annex,Casa Loma	41494.0	4.74	8754.0	155187.997998	59.0	2.0	33.0

Cluster Labels 8: locations with few businesses, population 15535 - 25051, and density 2430- 4005

Cluster Labels	Neighbourhood	Population	Area square km	Density	Average Income	# Business	# Japanese_Restaurants	# Restaurants
8	Birchcliffe-Cliffside	22291.0	5.92	3765.0	85644.236109	7.0	0.0	5.0
8	St.Andrew-Windfields	17812.0	7.33	2430.0	150633.549340	0.0	0.0	0.0
8	O'Connor-Parkview	18675.0	4.94	3780.0	82248.728351	16.0	0.0	2.0
8	Stonegate-Queensway	25051.0	7.83	3199.0	119843.838510	2.0	0.0	3.0
8	Edenbridge-Humber Valley	15535.0	5.47	2840.0	102322.730799	8.0	0.0	0.0
8	Willowridge-Martingrove-Richview	22156.0	5.53	4006.0	95652.817704	9.0	0.0	2.0

Results and Discussion

The results from clustering indicated that some of the clusters (Cluster Labels 0, 2, 4 & 7) contained locations with high foot and car traffic. However, some of these locations have a lot of restaurants already and therefore high competition. The table below contain locations which have high population density, number of businesses and restaurants. These locations have high competition and therefore they may not be ideal for the location of new restaurants.

Cluster Labels	Neighbourhood	Population	Area square km	Density	# Business	# Japanese_Restaurants	# Restaurants
2	Church-Yonge Corridor	31340.0	1.36	23044.0	56.0	10.0	38.0
2	Bay Street Corridor	25797.0	1.83	14096.0	64.0	6.0	27.0
2	Mount Pleasant East,Mount Pleasant West	46433.0	4.45	10434.0	59.0	8.0	37.0
4	Lansing-Westgate,Willowdale East	66598.0	10.32	6453.0	58.0	14.0	37.0
7	South Riverdale	27876.0	8.89	3135.0	60.0	2.0	33.0
7	Annex,Casa Loma	41494.0	4.74	8754.0	59.0	2.0	33.0

The suitable locations selection was based on number of existing Japanese restaurant, total number of other businesses, total number of restaurants, family income, population density, and population. Below is the location shortlist for a new Japanese restaurant that needs to undergo further analysis in order to pick the best location.

1. **Bedford Park-Nortown and Lawrence Park North** are among the neighbourhoods with high family medium income.
2. **High Park North, High Park-Swansea, and Junction Area** have a combined population of over sixty thousands, a lot of established businesses and fewer restaurants relative to other locations.
3. **Parkwoods-Donalda** has no Japanese restaurant within 1 km radius, it has very few restaurants and over fifteen other businesses based on venue data from foursquare.
4. **Rosedale-Moore Park** is high income neighbourhood with few restaurants, and more than fifteen businesses.
5. **Waterfront Communities-The Island** has a lot of businesses, and population of over sixty five thousands

Neighborhood	Family Median Income	Density	Population	Number of Businesses	Number of Restaurants	Number of Japanese Restaurants
Bedford Park-Nortown, Lawrence Park North	$ 184,168	4,851	37,843	21	13	2
High Park North,High Park-Swansea,Junction Area	$ 107,012	6,417	60,453	64	23	3
Parkwoods-Donalda	$ 76,898	4,690	34,805	17	3	0
Rosedale-Moore Park	$ 179,068	4,499	20,923	18	5	1
Waterfront Communities-The Island	$ 108,199	8,943	65,913	63	23	4

Note: The family median income values were extracted from neighborhood documents which are available through Toronto wellbeing map. The locations which had more than one neighbourhood used mean value of individual neighbourhood median family income.

Conclusion and Recommendations

The purpose of this project was to create a location shortlist for a new restaurant based on location and demographics data. The process began with data acquisition from multiple sources, cleansing, transforming, analyzing and visualizing the data. The result is we were able to create location shortlist. The selected locations have high foot and car traffic, and are densely populated with family median income above $70,000.

Further analysis is required in order to pick one location for a new Japanese restaurant. The analysis should include cost analysis for location that factor in renting or buying cost, labour cost and other related fees. It should take into consideration among other things information about availability of parking space, accessibility and crime rate.

References

[1] Jason Chow, 2018. Toronto's food halls point to a grown-up city with a demand for quality, https://www.theglobeandmail.com/life/food-and-wine/restaurant-reviews/article-torontos-food-halls-point-to-a-grown-up-city-with-a-demand-for/ Last Accessed August 10, 2019

[2] Amy Carlberg, 2018. Toronto ramen joint can't keep overwhelming demand. https://www.blogto.com/eat_drink/2018/09/toronto-ramen-joint-cant-keep-overwhelming-demand/, Last Accessed August 15, 2019

[3] Walter Muschenheim, 2018. A Toronto Tradition Returns, http://bentoboxmag.ca/2018/11/18/a-toronto-tradition-returns/, Last Accessed August 10, 2019

[4] Tom Larkin, 201 8. Factors for Choosing a New Restaurant Location, https://www.foodnewsfeed.com/fsr/expert-insights/8-factors-choosing-new-restaurant-location, Last Accessed August 17, 2019

[5] http://map.toronto.ca/wellbeing, by City of Toronto

[6] https://thepostalcodelookup.com/