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 have indicated there is a demand for high quality authentic Japanese restaurant that serves delicious mouth-watering 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 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 10
5. Number of Japanese restaurant within the radius of 1 km less or equal to 4 depend on population and number business 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 authentic 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, 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	Birchcliffe Cliffsid
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	Designatio
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	22,29
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	21,85
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 and sliced. This is because for this project we are interested in information related to average family income after tax, land area, and population for each neighbourhood.

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.

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, transformed, merged, filtered, analyzed and visualized.

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 neighbourhoods 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. However, 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.

It would have been nice to have family median income value for each neighbourhood available on csv file for Toronto neighbourhood profiles. This is because median value is good representation of the majority family income. The average value on the other hand 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.

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. The unique postal codes after aggregation are 98.

Postcode	Borough	Neighbourhood	Population	Area square km	Density	Average Income
M1B	Scarborough	Malvern,Rouge	90290	45.74	1973	146431.388091
M1C	Scarborough	Highland Creek	12494	5.20	2402	45615.245662
M1C	Scarborough	Centennial Scarborough	13362	5.39	2479	49910.669695
M1E	Scarborough	Guildwood,Morningside,West Hill	54764	19.04	2876	59286.750198
M1G	Scarborough	Woburn	53485	12.31	4344	145933.058584
M1J	Scarborough	Scarborough Village	16724	3.10	5394	43143.781864
M1J	Scarborough	Eglinton East	22776	3.23	7051	79582.155095
M1K	Scarborough	Ionview,Kennedy Park	30764	5.53	5563	46217.973809
M1L	Scarborough	Oakridge	13845	1.86	7443	31157.890032
M1L	Scarborough	Clairlea-Birchmount	26984	7.43	3631	81379.203680

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	Borough	Neighbourhood	Population	Area square km	Density	Average Income	Latitude	Longitude
M4W	Downtown Toronto	Rosedale-Moore Park	20923	4.65	4499	327842.352420	43.679563	-79.377529
M4N	Central Toronto	Lawrence Park South	15179	3.24	4684	282244.042538	43.728020	-79.388790
M8X	Etobicoke	Kingsway South	9271	2.58	3593	208942.045617	43.653654	-79.506944
M4G	East York	Leaside-Bennington	16828	4.68	3595	195253.063975	43.709060	-79.363452
M4V	Central Toronto	Forest Hill South	10732	2.45	4380	182035.268195	43.686412	-79.400049

Neighbourhood Maps

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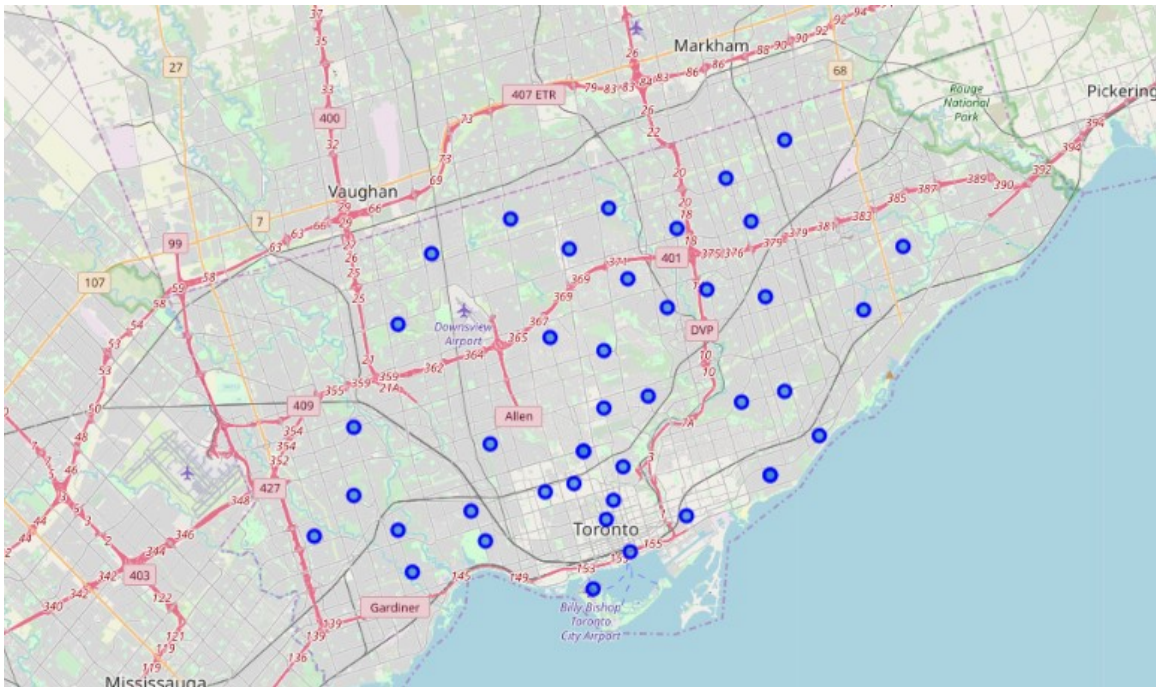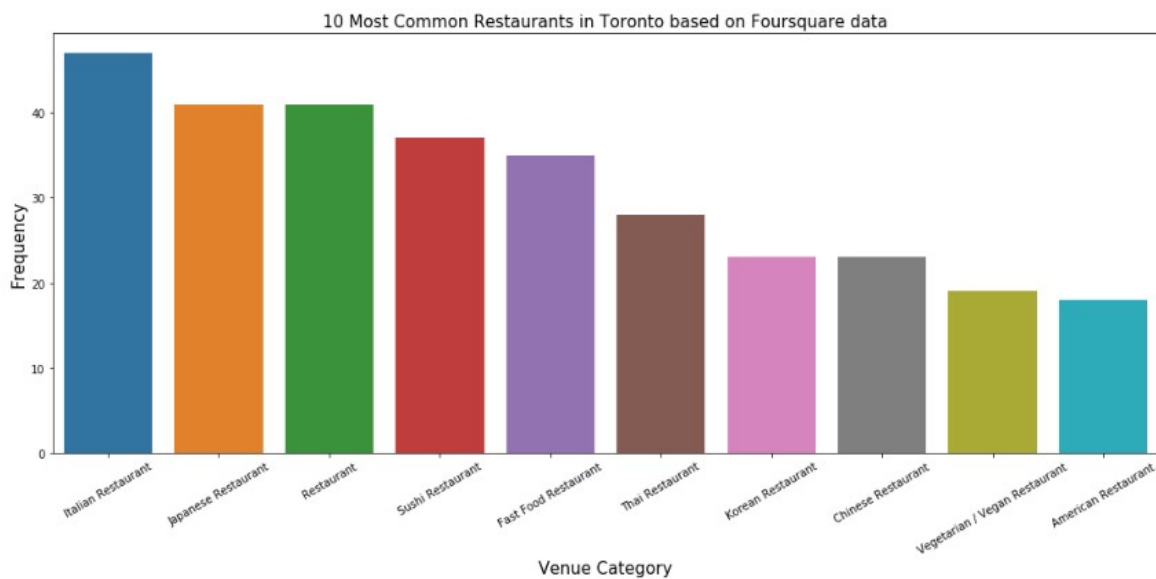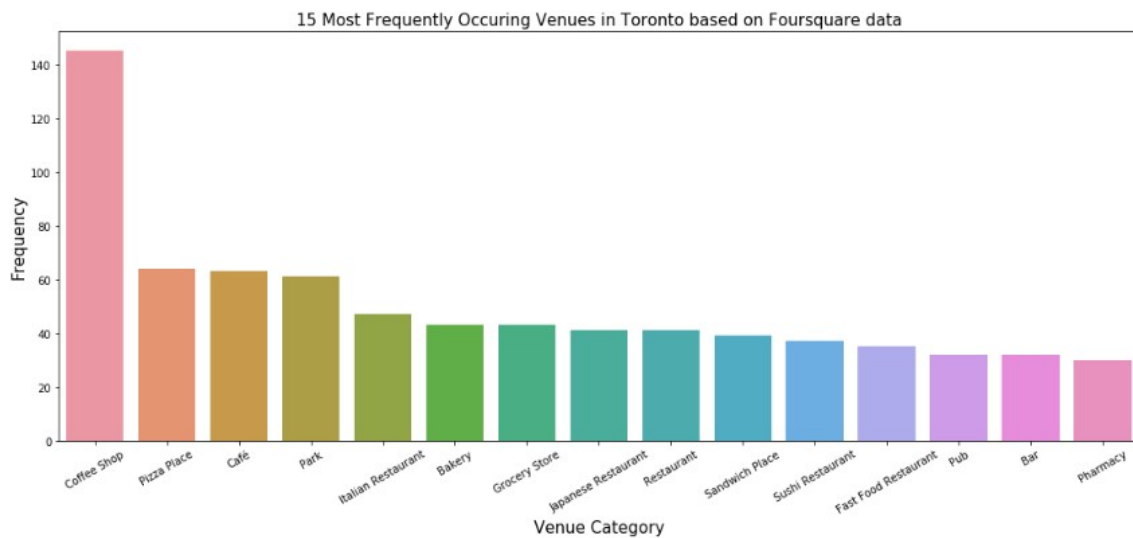

Neighbourhoods Venues

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.

name	categories	lat	lng
Summerhill Market	Grocery Store	43.686265	-79.375458
Toronto Lawn Tennis Club	Athletics & Sports	43.680667	-79.388559
Black Camel	BBQ Joint	43.677016	-79.389367
Craigleigh Gardens	Park	43.678099	-79.371586
Pie Squared	Pie Shop	43.672143	-79.377856
Tinuno	Filipino Restaurant	43.671281	-79.374920
Starbucks	Coffee Shop	43.671478	-79.380664
Manulife Financial	Office	43.672070	-79.382449
Booster Juice	Smoothie Shop	43.671566	-79.378581
Aroma Espresso Bar	Coffee Shop	43.672154	-79.377885
No Frills	Grocery Store	43.671616	-79.378187
Nijo Japanese Restaurant	Japanese Restaurant	43.671849	-79.378824
TD Canada Trust	Bank	43.686094	-79.376549
TD Canada Trust	Bank	43.672484	-79.377162
Eggsmart	Breakfast Spot	43.671158	-79.376240
Subway	Sandwich Place	43.672168	-79.377760
Mooredale House	Building	43.678631	-79.380091
Rosedale Park	Playground	43.682328	-79.378934
ALLDAYS 5 Selby St.	Hostel	43.671094	-79.377922
Whitney Park	Park	43.682036	-79.373788

The table 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. The data consist of about 257 unique categories and 1950 venues corresponding to 42 selected locations.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agincourt North,Milliken	29	29	29	29	29	29
Annex,Casa Loma	100	100	100	100	100	100
Banbury-Don Mills	31	31	31	31	31	31
Bay Street Corridor	100	100	100	100	100	100
Bayview Village,Bayview Woods-Steeles	15	15	15	15	15	15
Bedford Park-Nortown,Lawrence Park North	37	37	37	37	37	37
Birchcliffe-Cliffside	16	16	16	16	16	16
Church-Yonge Corridor	100	100	100	100	100	100
Clairlea-Birchmount	31	31	31	31	31	31
Corso Italia-Davenport	24	24	24	24	24	24
Don Valley Village	44	44	44	44	44	44
East End-Danforth	80	80	80	80	80	80
Edenbridge-Humber Valley	13	13	13	13	13	13
Eglinton East	11	11	11	11	11	11
Eringate-Centennial-West Deane,Islington-City Centre West,Princess-Rosethorn	16	16	16	16	16	16
Forest Hill South	77	77	77	77	77	77
Glenfield-Jane Heights	8	8	8	8	8	8
High Park North,High Park-Swansea,Junction Area	98	98	98	98	98	98
Kingsway South	46	46	46	46	46	46
L'Amoreaux,Steeles	24	24	24	24	24	24
Lansing-Westgate,Willowdale East	100	100	100	100	100	100

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

15 Most Frequently Occuring Venues in Toronto based on Foursquare data

10 Most Common Restaurants in Toronto based on Foursquare data

Neighbourhoods Clusters

The k-means clustering method is used to create 3 clusters out of the selected neighbourhoods. The high foot and car traffic of the location that is based on number of most popular business exists on a particular location will be used to create location shortlist. Toronto map which has cluster of locations display in 3 different colors:
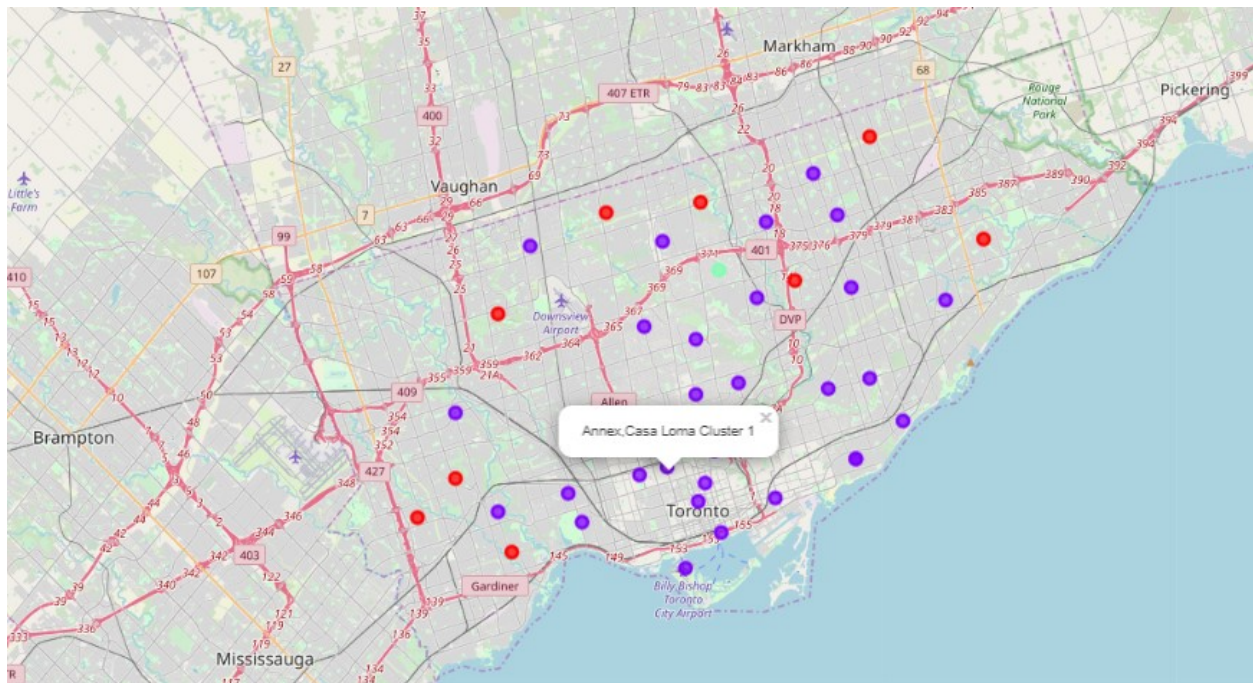
Cluster Labels 0

The **Park** is among the top ten 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. The existence of other established popular business venues is an indication of high foot and car traffic. Hence some of the locations in this cluster are suitable for a new location of restaurant.

	Neighbourhood	Population	Average Income	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Woburn	53485	145933.058584	43.770992	-79.216917	0	Park	Coffee Shop	Fast Food Restaurant	Pharmacy	Chinese Restaurant	Indian Restaurant	Electronics Store	Event Space	Ethiopian Restaurant	Design Studio
1	Parkwoods-Donalda	34805	144639.211442	43.753259	-79.329656	0	Park	Pharmacy	Shopping Mall	Convenience Store	Bus Stop	Skating Rink	Tennis Court	Food & Drink Shop	Café	Supermarket
2	Eringate-Centennial-West Deane,Islington-City ...	73604	134993.473120	43.650943	-79.554724	0	Park	Pizza Place	Hotel	Bank	Gym	American Restaurant	Café	Theater	Mexican Restaurant	Fish & Chips Shop
3	Stonegate-Queensway	25051	119843.838510	43.636258	-79.498509	0	Italian Restaurant	Park	Shopping Mall	Ice Cream Shop	Gym / Fitness Center	Eastern European Restaurant	Diner	Discount Store	Dive Bar	Dog Run
4	Agincourt North,Milliken	55685	110493.664337	43.815252	-79.284577	0	Chinese Restaurant	Korean Restaurant	Pizza Place	Pharmacy	Bakery	Noodle House	Park	Malay Restaurant	Dessert Shop	Caribbean Restaurant
5	Edenbridge-Humber Valley	15535	102322.730799	43.667856	-79.532242	0	Pharmacy	Playground	Shopping Mall	Bakery	Café	Baseball Field	Golf Course	Park	Convenience Store	Grocery Store
6	Glenfield-Jane Heights	30491	95121.544626	43.739015	-79.506944	0	Bank	Coffee Shop	Vietnamese Restaurant	Pizza Place	Grocery Store	Spa	Park	Shopping Mall	Dog Run	Dessert Shop
7	Corso Italia-Davenport	14133	70527.081077	43.689026	-79.453512	0	Pharmacy	Mexican Restaurant	Park	Bakery	Thai Restaurant	Market	Fast Food Restaurant	Bus Stop	Sporting Goods Shop	Beer Store
8	Bayview Village,Bayview Woods-Steeles	34550	70518.961182	43.786947	-79.385975	0	Japanese Restaurant	Grocery Store	Bank	Park	Café	Intersection	Chinese Restaurant	Trail	Shopping Mall	Skate Park
9	Westminster-Branson,Willowdale West	43210	70267.476459	43.782736	-79.442259	0	Pharmacy	Pizza Place	Bakery	Discount Store	Butcher	Eastern European Restaurant	Park	Bus Line	Convenience Store	Coffee Shop

Cluster Labels 1

The **coffee shop** is among the most popular venues for this cluster. The other venues categories on top ten most common venues are stores, restaurants, banks, pubs, bars, gyms, bakeries and pizza places.

	Neighbourhood	Population	Average Income	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Rosedale-Moore Park	20923	327842.352420	43.679563	-79.377529	1	Coffee Shop	Park	Grocery Store	Bank	Filipino Restaurant	Metro Station	Bistro	Breakfast Spot	Building	Smoothie Shop
1	Lawrence Park South	15179	282244.042538	43.728020	-79.388790	1	College Quad	Coffee Shop	Café	Park	Bus Line	Bookstore	Trail	Gym / Fitness Center	College Gym	Falafel Restaurant
2	Kingsway South	9271	208942.045617	43.653654	-79.506944	1	Coffee Shop	Park	Italian Restaurant	Burger Joint	French Restaurant	Breakfast Spot	Sushi Restaurant	Dessert Shop	Pub	Gourmet Shop
3	Leaside-Bennington	16828	195253.063975	43.709060	-79.363452	1	Coffee Shop	Sporting Goods Shop	Electronics Store	Furniture / Home Store	Brewery	Sushi Restaurant	Restaurant	Sports Bar	Burger Joint	Sandwich Place
4	Forest Hill South	10732	182035.268195	43.686412	-79.400049	1	Coffee Shop	Sushi Restaurant	Park	Italian Restaurant	Thai Restaurant	Grocery Store	Pizza Place	Liquor Store	Sandwich Place	Restaurant
5	Annex,Casa Loma	41494	155187.997998	43.672710	-79.405678	1	Café	Vegetarian / Vegan Restaurant	Coffee Shop	Italian Restaurant	Bakery	Restaurant	Museum	Gym	Pizza Place	Pub
6	South Riverdale	27876	153733.028955	43.659526	-79.340923	1	Coffee Shop	Bar	Café	Bakery	Vietnamese Restaurant	American Restaurant	Brewery	Italian Restaurant	Diner	French Restaurant

Cluster Labels 2

It only have one location. Unlike other locations this one has a farm and farmers market on its list of top ten most common venues.

	Neighbourhood	Population	Average Income	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	St.Andrew-Windfields	17812	150633.54934	43.75749	-79.374714	2	Park	Pool	Design Studio	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store

Results and Discussion

The results from clustering indicated that two out of the three clusters 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 meet one or both criteria below:

1. Less than 3 Japanese restaurants
2. More than 10 businesses excluding restaurant venue categories

That means the locations included on this list have to meet at least one of the two criteria. The locations which have a lot of Japanese restaurants and fewer businesses were eliminated due to high competition.

	Number of Japanese restaurants and businesses within radius of 1km of the location			
Neighborhood	Japanese Cuisine restaurants	Number of Businesses	Total of All Restaurants	Population
Annex,Casa Loma	2	47	31	41,494
Bay Street Corridor	2	53	24	25,797
Bedford Park-Nortown,Lawrence Park North	2	16	13	37,843
Corso Italia-Davenport	1	11	6	14,133
Church-Yonge Corridor	10	45	35	31,340
Don Valley Village	2	29	8	27,051
East End-Danforth	4	43	15	21,381
Eglinton East	1	3	5	22,776
Forest Hill South	5	36	22	10,732
High Park North,High Park-Swansea,Junction Ar	3	50	22	60,453
Kingsway South	2	23	12	9,271
Lansing-Westgate,Willowdale East	14	48	36	66,598
Leaside-Bennington	3	36	10	16,828
Mount Pleasant East,Mount Pleasant West	8	45	35	46,433
Palmerston-Little Italy	3	50	35	13,826
Parkwoods-Donalda	0	15	3	34,805
Roncesvalles	4	43	22	14,974
Rosedale-Moore Park	1	14	4	20,923
South Riverdale	2	42	30	27,876
The Beaches	4	43	15	21,567
Waterfront Communities-The Island	4	44	19	65,913
Yonge-St.Clair	5	36	22	12,528
York University Heights	2	12	8	27,593

Keywords for Business: Bank, Mall, Grocery, Store, pub, Historic, Building, Shop, Museum, Plaza, Gallery, Business, Pub, Bar, Office, Theater, Studio, Gym, Café, Coffee, Pizza, Burger, Storage, Tea, Breakfast
Keywords for Restaurant: BBQ, Restaurant
Keywords for Japanese restaurant: Ramen, Sushi, and Japanese

The locations which are highlighted in green are the one we want to keep. The selection of these locations 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. **Annex and Casa Loma** are among the locations in Toronto with high foot and car traffic. Together the two neighborhoods have a population of over forty thousand, and a lot of businesses.

2. **Bay Street Corridor** has a lot of businesses in comparison to others and one of its most popular venues is Ramen Restaurant which has a lot of customers in the city of Toronto.
3. **Don Valley Village** it has fewer restaurants overall, about twenty nine other businesses, and Japanese restaurant is among the top ten most popular venue in this location.
4. **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.
5. **Parkwoods-Donalda** has no Japanese restaurant; it has three restaurants and fifteen businesses based on venue data from foursquare.
6. **Rosedale-Moore Park** is high income neighbourhood with four restaurants, fifteen businesses and one Japanese restaurant
7. **Waterfront Communities-The Island** has a lot of businesses, and population of sixty five thousands

Neighborhood	Family Median Income	Density	Population
Annex,Casa Loma	$ 137,345	8,754	41,494
Bay Street Corridor	$ 78,348	14,096	25,797
Don Valley Village	$ 71,474	6,440	27,051
High Park North,High Park-Swansea,Junction Area	$ 107,012	6,417	60,453
Parkwoods-Donalda	$ 76,898	4,690	34,805
Rosedale-Moore Park	$ 179,068	4,499	20,923
Waterfront Communities-The Island	$ 108,199	8,943	65,913

Note: The family median income values were extracted from neighborhood documents which are available through [Toronto wellbeing map](#).

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 reduce the number of locations to seven. The selected locations have high foot and car traffic, family income above $70,000, and fewer Japanese restaurants.

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 and other operational

cost including 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

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

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

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

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

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

https://thepostalcodelookup.com/