

Opening an Italian Restaurant in Toronto

Note: I'm used to writing reports / journal articles for Chemistry journals, so I'll be talking in plural first-person most of the time

Business Problem

My wife and I will be moving to Toronto shortly, and we would like to open an Italian restaurant. However, Toronto is a big city with many neighbourhoods. How should we pick the right neighbourhood to start our business? This type of study is important for us but will also interest anyone who is trying to rank neighbourhoods for various purposes. With this problem, a key aspect is to determine the criteria over which neighbourhoods will be ranked. Why would one neighbourhood be better than another? Specifically, what metrics will be chosen?

We have chosen four key metrics in this aim:

1. Low percentage of Italian restaurants.
 - a. In Canada, we believe that Italian food is very much a staple. Therefore, the lack of Italian restaurants in an area may not mean that it is undesirable. In fact, we expect that opening a restaurant in an area with few Italian restaurants may result in less competition and greater sales.
2. High salary of residents
 - a. People with high salaries have the means of going out to eat. A restaurant in a more well-off area may be able to sustain higher prices for the food.
3. High population density
 - a. If there are more people in a given neighbourhood, then a given conversion percentage will result in more people eating at the restaurant.
4. Low average ratings of current Italian restaurants
 - a. If there are multiple Italian restaurants in a neighbourhood, people may look up ratings to decide where to get their food. Our restaurant may struggle more in a neighbourhood that already has very highly rated Italian restaurants.

Data

We will begin with data from Wikipedia on the different neighbourhoods of Toronto (https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods). This will include the name of the neighbourhood, population, land area, and average income.

	Name	Population	LandArea	PercentChangePopulation	AverageIncome
0	Agincourt	44577	12.45	4.6	25750
1	Alderwood	11656	4.94	-4.0	35239
2	Alexandra Park	4355	0.32	0.0	19687
3	Allenby	2513	0.58	-1.0	245592
4	Amesbury	17318	3.51	1.1	27546

We will then extract latitude and longitude coordinates using geolocator. From the coordinates, we will retrieve information from FourSquare, including the number of restaurants, the percentage of Italian restaurants, and ratings of each Italian restaurant.

Methodology

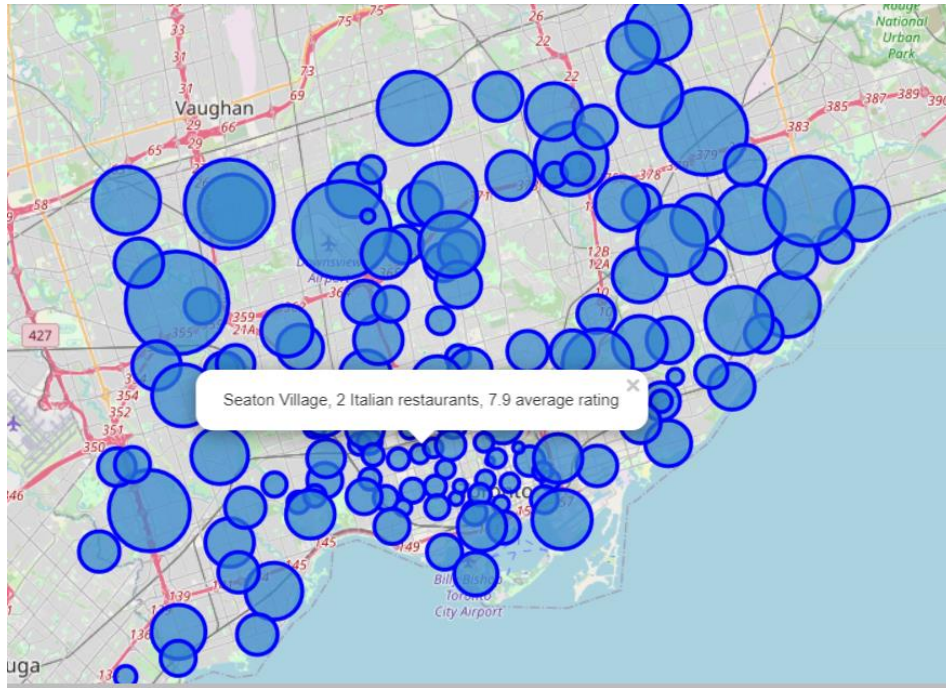
The data for the neighbourhoods had to be cleaned up slightly. A few of the neighbourhood names were not acknowledged by geolocator. We altered the name of a couple well-known neighbourhoods to circumvent this. We then removed from the study the few neighbourhoods whose coordinates could not be found easily. We also removed neighbourhoods whose distance to the city centre of Toronto was very far.

	Name	Population	LandArea	PercentChangePopulation	AverageIncome	Latitude	Longitude	DistanceCityCentre
0	Agincourt	44577	12.45	4.6	25750	43.785353	-79.278549	16.933531
1	Alderwood	11656	4.94	-4.0	35239	43.601717	-79.545232	14.201221
2	Alexandra Park	4355	0.32	0.0	19687	43.650758	-79.404308	1.666843
3	Amesbury	17318	3.51	1.1	27546	43.706162	-79.483492	9.920269
4	Armour Heights	4384	2.29	2.0	116651	43.743944	-79.430851	10.742798

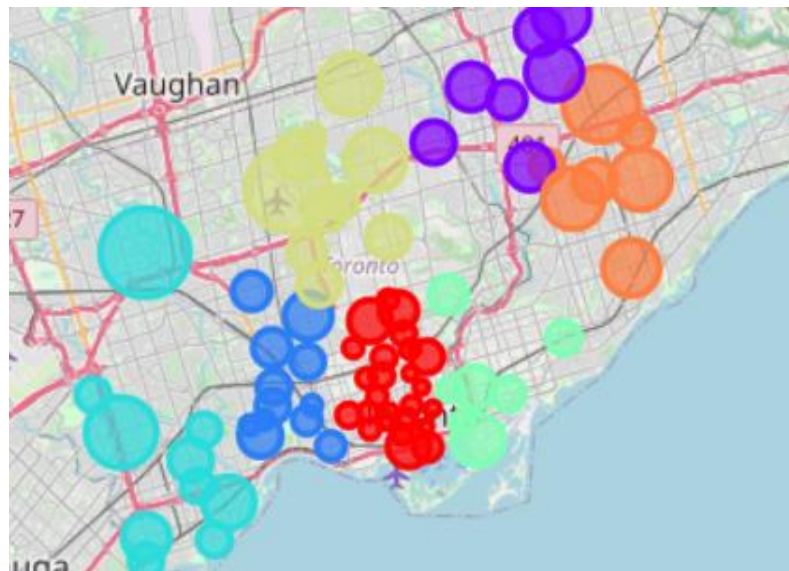
We then used FourSquare to collect restaurant information for each of the neighbourhoods in a given radius, which was defined using the land area. This introduced a small amount of error, as the land area for each restaurant is of course not a circle, but we estimated that it would not influence the results too much. For each restaurant, we determined the percentage of Italian restaurants by summing the venues whose category was either “Italian Restaurant” or “Pizza Place”. Then, for each Italian Restaurant, we accessed the rating. For a given neighbourhoods, we averaged the ratings of all Italian restaurants with 5 or more reviews to yield an average rating. We also kept the max rating for each neighbourhood.

	Name	Population	LandArea	PercentChangePopulation	AverageIncome	Latitude	Longitude	DistanceCityCentre	NumberRestaurants	NumberItalianRestaurants	AverageRating	MaxRating
0	Agincourt	44577	12.45	4.6	25750	43.785353	-79.278549	16.933531	67	6	6.00	6.2
1	Alderwood	11656	4.94	-4.0	35239	43.601717	-79.545232	14.201221	12	4	7.90	8.2
2	Alexandra Park	4355	0.32	0.0	19687	43.650758	-79.404308	1.666843	13	2	7.25	7.5
3	Amesbury	17318	3.51	1.1	27546	43.706162	-79.483492	9.920269	12	2	6.60	6.6
4	Armour Heights	4384	2.29	2.0	116651	43.743944	-79.430851	10.742798	5	2	6.80	6.8

We then wanted to visualise the neighbourhoods in Toronto. We used Folium to generate the map, but unlike the maps in the labs, we made the circles directly related to the size of the land area for each neighbourhood. This way, the circle of the folium map would correspond exactly to the radius from which we extracted the restaurants. Popup labels revealed the name of the neighbourhood, the number of Italian restaurants, and the average rating of those restaurants. Overall, there were 402 restaurants tallied; this was convenient, as premium FourSquare calls for ratings maxes out at 500 per day.



We then looked at some of the demographic information for each neighbourhood and tried to see if it could be used to predict either 1) the percentage of Italian restaurants or 2) the average rating of Italian restaurants. We also used a bit of machine learning, in case the small sample size of restaurants in some neighbourhoods were distorting the data. We performed K Clustering to lump the locations with rated Italian restaurants into seven clusters. From this, we then again tried to see how correlated the demographic data was.



After this, we then took the four main metrics and standardized them. We generated a score for each neighbourhood based on these metrics. For neighbourhoods lacking an average rating, we replaced these values with the mean value of those restaurants that did have a rating. We then made a restriction that we would only look at neighbourhoods whose population densities and average incomes were not too low (below 0.1 of the standardized score). We then sorted the neighbourhoods by their aggregated score to yield the most ideal locations to open an Italian restaurant.

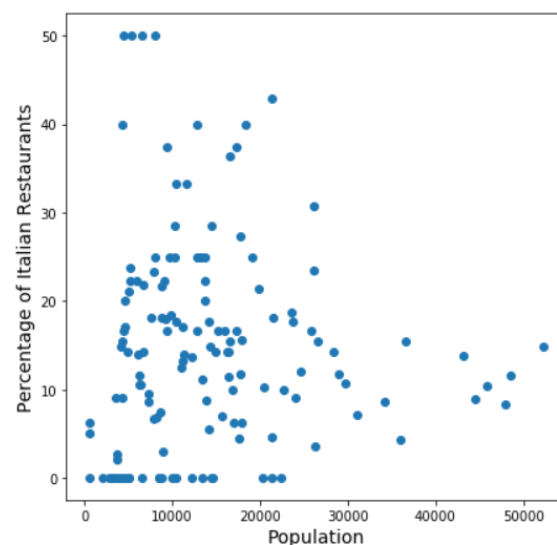
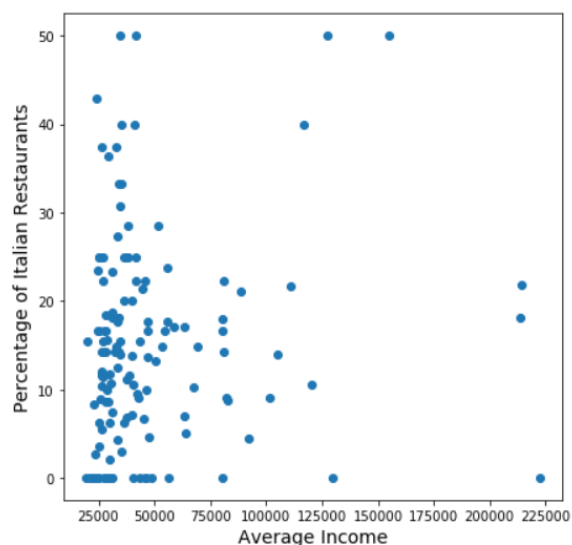
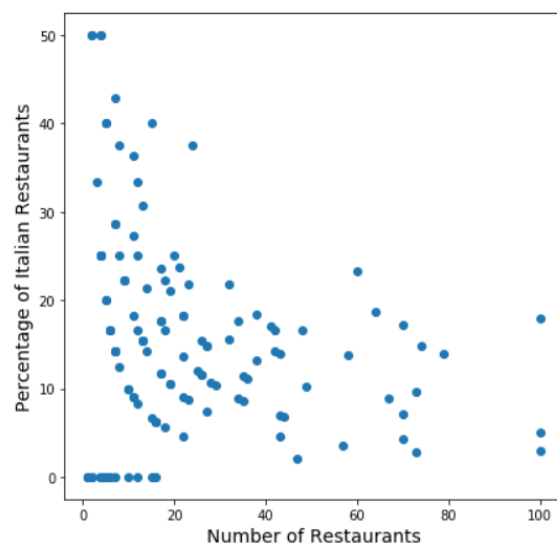
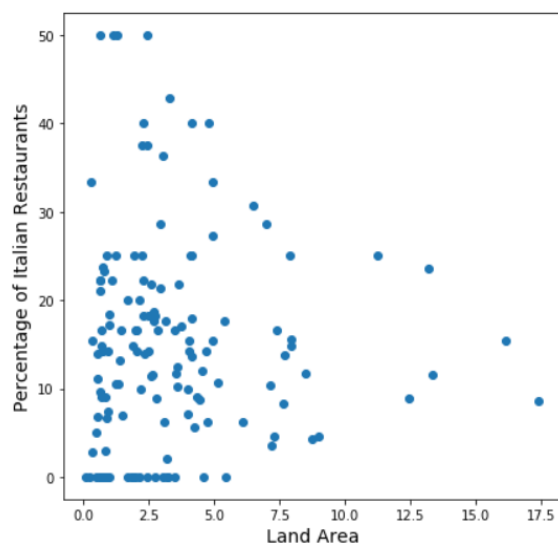
Results and Discussion

- Again, trained in chemistry/physics, where these two sections are usually combined. I think the flow tends to be better.

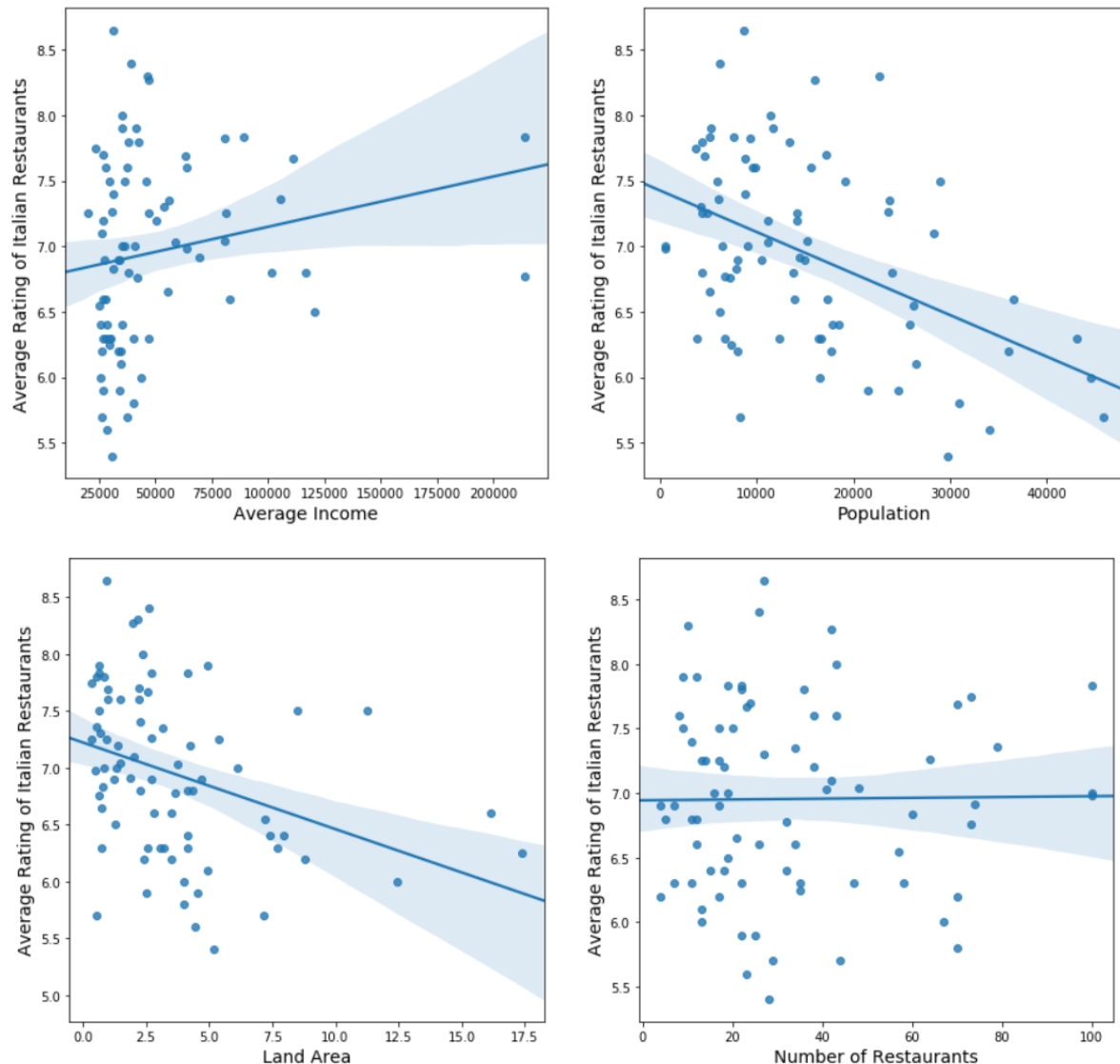
The first series of results obtained were both the total number of restaurants and the number of Italian restaurants in each neighbourhood. Expectedly, Little Italy was among those that had the largest number of Italian restaurants.

	Name	Population	LandArea	PercentChangePopulation	AverageIncome	Latitude	Longitude	DistanceCityCentre	NumberRestaurants	NumberItalianRestaurants
102	Princess Gardens	9288	4.13	1.0	80607	43.640466	-79.391224	1.561557	100	18
79	Little Italy	7917	0.81	-6.1	31231	43.655208	-79.414877	2.496738	60	14
78	Leslieville	23567	2.69	-7.3	30886	43.662700	-79.332815	4.238242	64	12
45	Fashion District	4642	0.98	123.0	63282	43.645456	-79.394994	1.260231	70	12
57	Harbourfront / CityPlace	14368	1.87	94.3	69232	43.640080	-79.380150	1.520997	74	11

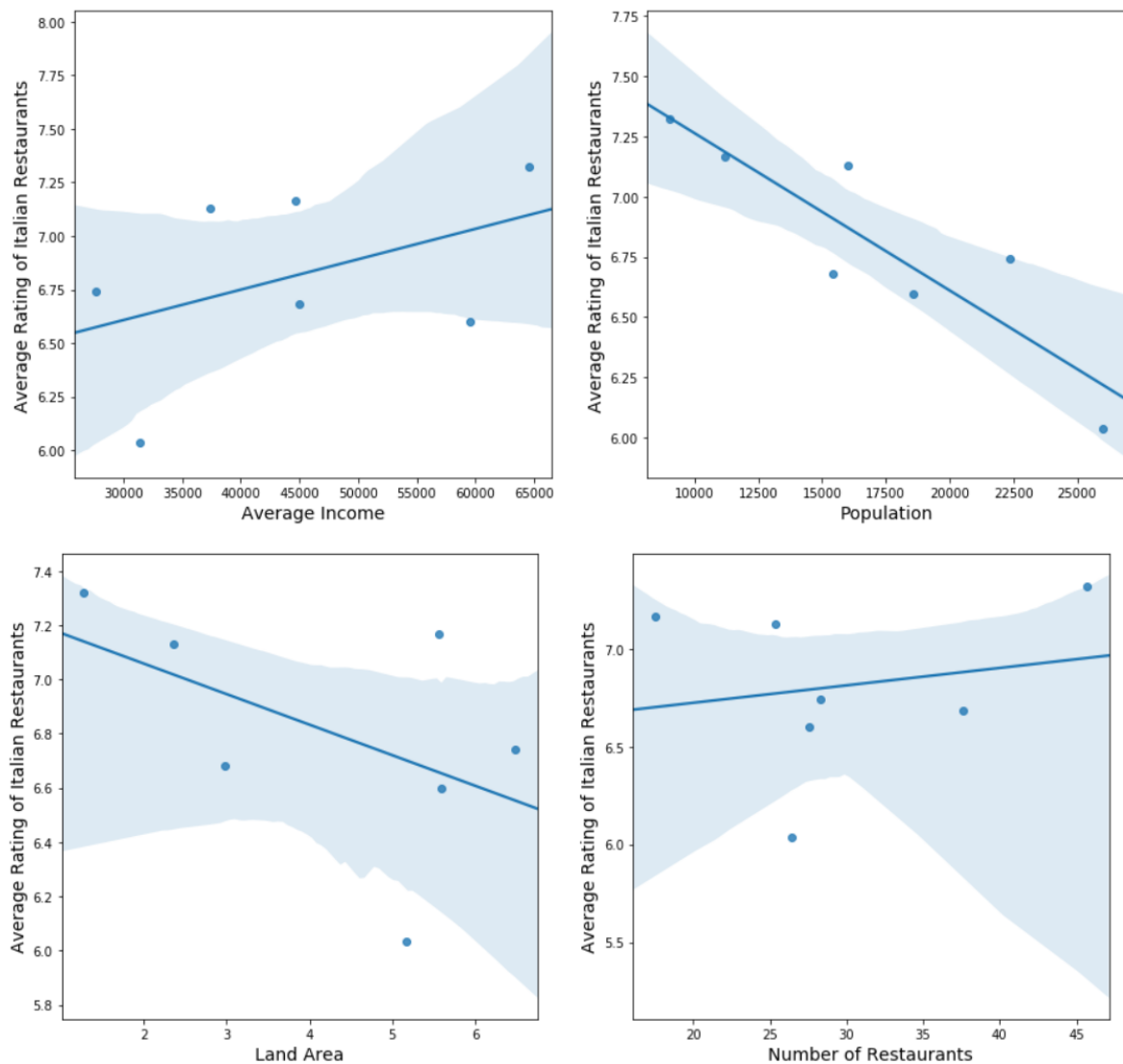
None of the demographic data had much correlation with the percentage of Italian restaurants. We tried land area, number of total restaurants, average income, and population. The graph with total number of restaurants almost appears to show a trend, but upon closer inspection, it really is only diverging as $x \rightarrow 0$; diverging to lower values get more steplike and is less noticeable. Overall, it seems that the percentage of Italian restaurants is completely uncoupled to all of these parameters.



The comparison to average rating for Italian restaurants was slightly more illuminating. Both population and land area appeared to be inversely proportional to the average rating. However, the trends were very noisy and could be influenced by small sample size (some neighbourhoods had very few Italian restaurants).



The clustered data was meant to help with this. All restaurants were divided into seven different clusters. Here, the inverse correlation of rating and population is more apparent, while the other trends are barely above baseline. We speculate that this inverse correlation may be due to the following: in neighbourhoods with a lot of people, restaurants may not have to try as hard to maintain good quality and service. However, more in-depth analysis would be needed to test this thoroughly. Overall, the demographic data did not correlate that well with the average rating. Therefore, we did not pursue any further predictive modelling but instead chose to begin ranking the neighbourhoods based on our metrics.



The four metrics we chose were 1) low percentage of Italian restaurants, 2) high average income, 3) high population density, and 4) low average score of current Italian restaurants. For this analysis, we chose to keep the relative weights of the four equal. However, we did require that the standardized score of population density and average income must not be too low ($x > 0.1$). The highest rated neighbourhoods are listed below, with Bay Street Corridor, Riverdale and Yorkville rounding out the top three. To answer which neighbourhood is most ideal, one could simply pick the highest scored one. However, realistically, there would be other factors for a particular instance, and different weights may be applied to the different metrics.

	HighestPopulationDensity	HighestAverageIncome	LeastItalianRestaurants%	LowestAverageRating	Score	Name	Latitude	Longitude	LandArea
6	0.682008	0.103808	1.000000	0.521759	2.307575	Bay Street Corridor	43.664457	-79.387189	0.11
105	0.120581	0.101547	0.857143	0.876923	1.956194	Riverdale	43.665470	-79.352594	3.99
142	0.168066	0.422175	0.721519	0.395804	1.707564	Yorkville	43.671386	-79.390168	0.56
30	0.171339	0.110703	0.808219	0.581538	1.671799	Discovery District	43.657556	-79.389480	0.66
121	0.140223	0.125385	0.866667	0.521759	1.654033	The Danforth	43.686433	-79.300355	0.87
29	0.161664	0.301336	0.666667	0.495385	1.625051	Deer Park	43.688090	-79.394093	1.46
57	0.119203	0.244835	0.702703	0.535043	1.601783	Harbourfront / CityPlace	43.640080	-79.380150	1.87
119	0.165223	0.217274	0.860465	0.323077	1.566038	The Annex	43.670338	-79.407117	1.47
25	0.103640	0.173169	0.666667	0.521759	1.465234	Corktown	43.657370	-79.356513	0.67
16	0.123277	0.152074	0.736842	0.446154	1.458347	Cabbagetown	43.664473	-79.366986	1.40

However, this is a good starting point for the investigation. We've listed the top neighbourhoods on the map of Toronto below. Interestingly, a few of the high-scoring neighbourhoods almost border each other. Picking one of these neighbourhoods could be a good idea, as you could get potential customers from any of these bass.



Conclusion

In summary, we have provided a brief investigation into which Toronto neighbourhoods might serve as ideal locations for an Italian restaurant. Key metrics chosen were high population density, high average income, low percentage of Italian restaurants, and low average scores of current Italian restaurants. The demographic data of the neighbourhoods does not correlate well with average ratings and percentage of Italian restaurants, even when the neighbourhoods are broken into clusters. However, the neighbourhoods were able to be ranked according to the metrics, with Bay Street Corridor, Riverdale and Yorkville taking the top three spots.