Toronto Restaurant Market Opportunity Analysis

1.Introduction

There is always an opening of new business. Before opening any new business, the market opportunity analysis for that kind of business is necessary. This is the first step to ensure business success. A restaurant business will be interested in this if they plan to open a new restaurant in Toronto. Which kind of restaurant and where to open the restaurant are very important factors for the business's success. I use the neighborhood listings and the FourSquare venue listings from prior projects to obtain a picture of which and how many restaurants of different types can be found in the different neighborhoods of Toronto. We then compare the frequencies of different types of restaurants across neighborhoods to find out where there may be fewer restaurants than the market can support, which we can recommend as opportunities to open new restaurants of different types.

2. Data Acquisition and Cleaning

2.1 Data Sources

Toronto neighborhood data and geospatial data (latitude and longitude) can be found in the website https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and http://cocl.us/Geospatial_data . Venue listings are queried from FourSquare.

2.2 Data Preprocessing

First, the data from the above sources are merged on postal code to obtain the latitude and longitude for each neighborhood and then the Toronto neighborhood data was extracted.

Here is the snapshot of Toronto neighborhood data:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
15	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
19	M4E	East Toronto	The Beaches	43.676357	-79.293031
20	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306
24	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
25	M6G	Downtown Toronto	Christie	43.669542	-79.422564
30	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568
31	M6H	West Toronto	Dufferin, Dovercourt Village	43.669005	-79.442259
36	M5J	Downtown Toronto	Harbourfront East, Union Station, Toronto Islands	43.640816	-79.381752

Second, the neighborhood venue listings were queried from FourSquare API and the Toronto neighborhood venue listing were extracted. Here is some example data queried from foursquare API and merged with Toronto neighborhood data.

	Postal Code	Borough	Neighborhood	Latitude	Longitude	headerLocation	venue.location.lat	venue.location.lng	venue.id	venue.name	category	ne.lat	
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	Corktown	43.653447	-79.362017	54ea41ad498e9a11e9e13308	Roselle Desserts	Bakery	43.658760	-79.
1	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	Corktown	43.653559	-79.361809	53b8466a498e83df908c3f21	Tandem Coffee	Coffee Shop	43.658760	-79.
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	Corktown	43.653947	-79.361149	4ae5b91ff964a520a6a121e3	Morning Glory Cafe	Breakfast	43.658760	-79.
3	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	Corktown	43.653249	-79.358008	574c229e498ebb5c6b257902	Cooper Koo Family YMCA	Distributor	43.658760	-79.
4	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	Corktown	43.654735	-79.359874	50760559e4b0e8c7babe2497	Body Blitz Spa East	Spa	43.658760	-79.
5	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	Corktown	43.656369	-79.356980	5612b1cc498e3dd742af0dc8	Impact Kitchen	Restaurant	43.658760	-79.

After checking FourSquare's neighborhood and postal code neighborhood, I found that FourSquare's neighborhoods are not the same as the postal code neighborhood boundaries, so some venues will show up more than once. I assign each venue to the postal code neighborhood it is closest to.

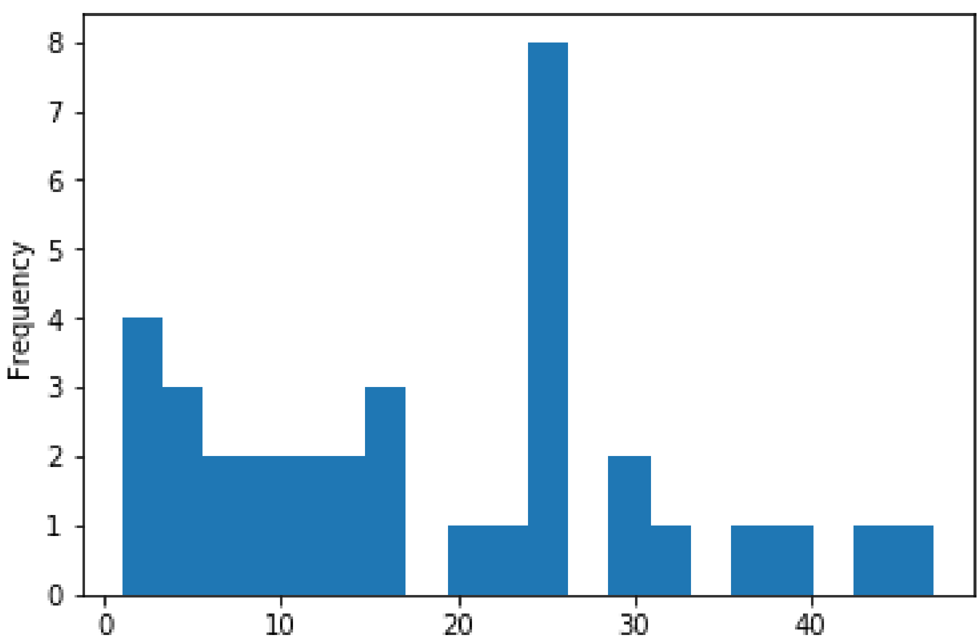
Third, I checked how many different kinds of venues we have in each neighborhood. Some of these are not restaurants, so I drop them. In addition, since this list is long, I drop any types that have 5 or fewer venues; it may not be feasible to open another one of these restaurants because they are not very popular. Now the data is ready to do analysis.

	Postal Code	Borough	Neighborhood	Latitude	Longitude	headerLocation	venue.location.lat	venue.location.lng	venue.id	venue.name	category	ne.lat	n
0	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	Bay Street Corridor	43.660391	-79.387664	4a8355bff964a520d3fa1fe3	Mercatto	Italian	43.666802	-79.38
1	M5W	Downtown Toronto	Stn A PO Boxes	43.646435	-79.374846	St. Lawrence	43.646964	-79.374403	4ada4b1ef964a520022121e3	The Old Spaghetti Factory	Italian	43.650935	-79.36
2	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280	Financial District	43.650243	-79.380820	4adc5c6af964a520da2b21e3	Mercatto	Italian	43.652929	-79.37
3	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	Greektown	43.677062	-79.353934	4af4e0d0f964a5202ff721e3	7 Numbers	Italian	43.684057	-79.34

3. Exploratory Data Analysis

Now that I have filtered the restaurants, let's get some descriptive statistics on the number of restaurants throughout Toronto and within each neighborhood. After that, I can start to compare the restaurant distributions in each neighborhood with the whole. Of course, this is clearly not a count of all of the restaurants in Toronto, but only of the most popular ones in each neighborhood and of the most popular cuisines. Nevertheless, since I used the same limits and parameters in the FourSquare API queries, I can expect the counts across neighborhoods to be comparable.

First let's look at the total number of restaurants (regardless of cuisine) in Toronto as a whole and compare it to the count in each neighborhood.

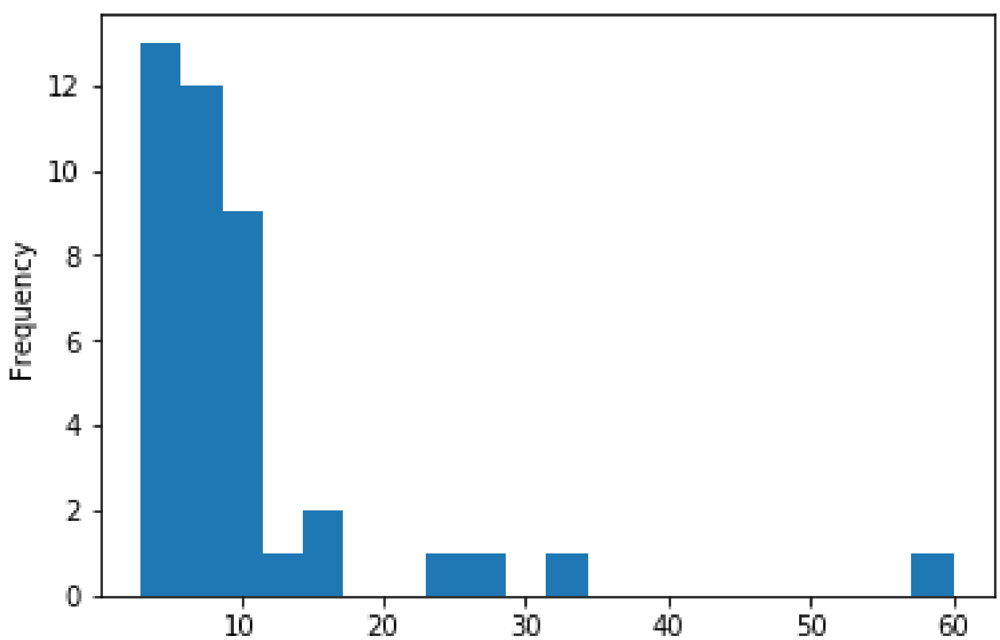


```
count    35.000000
mean     18.571429
std      12.455298
min       1.000000
25%       8.500000
50%      17.000000
75%      26.000000
max      47.000000
```

Some neighborhoods have many more restaurants than others. Let’s focus on the more populous neighborhoods. This will also have better data for modeling. According to the description above, just more than the top 25% of neighborhoods have 25 or more restaurants.

Having eliminated the less populous neighborhoods, let's look at the categories of venues. I will throw away from analysis the less popular categories, because we don't have enough data to determine whether or not it will be popular in a particular neighborhood. I now have 14 neighborhoods.

```
count    41.00000
mean      9.95122
std      10.30522
min       3.00000
25%       5.00000
50%       7.00000
75%      10.00000
max      60.00000
```



Let’s again consider only the top quarter of the categories. This mean, according to the quartiles above, those categories with 10 or more restaurants in the top neighborhoods

category Neighborhood	Bakery	Bar	Café	Coffee Shop	Greek	Italian	Japanese	Pizza	Restaurant	Sushi	All
Central Bay Street	0	1	4	10	0	2	2	0	1	1	21
Church and Wellesley	0	0	1	4	0	1	5	1	3	5	20
Commerce Court, Victoria Hotel	1	0	4	5	0	0	3	0	2	0	15
First Canadian Place, Underground city	0	3	2	3	1	2	0	1	2	1	15
Garden District, Ryerson	1	0	4	7	1	2	3	1	3	0	22
Harbourfront East, Union Station, Toronto Islands	1	2	4	10	0	3	1	2	4	1	28
Kensington Market, Chinatown, Grange Park	3	2	5	4	0	1	1	1	0	0	17
Little Portugal, Trinity	0	4	2	1	1	1	1	1	3	0	14
Richmond, Adelaide, King	1	0	2	3	0	0	1	1	1	2	11
Runnymede, Swansea	0	1	3	3	0	2	0	3	1	2	15
St. James Town, Cabbagetown	2	0	2	3	0	2	1	2	3	0	15
Stn A PO Boxes	2	0	1	2	0	5	0	0	2	0	12
The Danforth West, Riverdale	1	0	1	3	8	3	1	1	1	0	19
All	12	13	35	58	11	24	19	14	26	12	224

This is much more manageable! Now since some neighborhoods have more restaurants than others (even after I threw out the least popular neighborhoods and the least popular categories), let us normalize the counts by the number of restaurants per neighborhood. This will give us, for each neighborhood, the relative frequency of each type of restaurant. From there, we can start finding restaurant opportunities.

category Neighborhood	Bakery	Bar	Café	Coffee Shop	Greek	Italian	Japanese	Pizza	Restaurant	Sushi	All
Central Bay Street	0.000000	0.047619	0.190476	0.476190	0.000000	0.095238	0.095238	0.000000	0.047619	0.047619	1.0
Church and Wellesley	0.000000	0.000000	0.050000	0.200000	0.000000	0.050000	0.250000	0.050000	0.150000	0.250000	1.0
Commerce Court, Victoria Hotel	0.066667	0.000000	0.266667	0.333333	0.000000	0.000000	0.200000	0.000000	0.133333	0.000000	1.0
First Canadian Place, Underground city	0.000000	0.200000	0.133333	0.200000	0.066667	0.133333	0.000000	0.066667	0.133333	0.066667	1.0
Garden District, Ryerson	0.045455	0.000000	0.181818	0.318182	0.045455	0.090909	0.136364	0.045455	0.136364	0.000000	1.0
Harbourfront East, Union Station, Toronto Islands	0.035714	0.071429	0.142857	0.357143	0.000000	0.107143	0.035714	0.071429	0.142857	0.035714	1.0
Kensington Market, Chinatown, Grange Park	0.176471	0.117647	0.294118	0.235294	0.000000	0.058824	0.058824	0.058824	0.000000	0.000000	1.0
Little Portugal, Trinity	0.000000	0.285714	0.142857	0.071429	0.071429	0.071429	0.071429	0.071429	0.214286	0.000000	1.0
Richmond, Adelaide, King	0.090909	0.000000	0.181818	0.272727	0.000000	0.000000	0.090909	0.090909	0.090909	0.181818	1.0
Runnymede, Swansea	0.000000	0.066667	0.200000	0.200000	0.000000	0.133333	0.000000	0.200000	0.066667	0.133333	1.0
St. James Town, Cabbagetown	0.133333	0.000000	0.133333	0.200000	0.000000	0.133333	0.066667	0.133333	0.200000	0.000000	1.0
Stn A PO Boxes	0.166667	0.000000	0.083333	0.166667	0.000000	0.416667	0.000000	0.000000	0.166667	0.000000	1.0
The Danforth West, Riverdale	0.052632	0.000000	0.052632	0.157895	0.421053	0.157895	0.052632	0.052632	0.052632	0.000000	1.0
All	0.053571	0.058036	0.156250	0.258929	0.049107	0.107143	0.084821	0.062500	0.116071	0.053571	1.0

I now convert the relative proportions to percentiles. The categories with low percentiles are opportunities to open new restaurants. There are two reasons for this logic:

1. I have already selected the most popular kinds of cuisines and the most popular neighborhoods within the entire city of Toronto. Therefore, for every combination of (Neighborhood, Category) in this final table, there is probably at least some demand for that kind of restaurant in that neighborhood.
2. I have normalized the counts by the total number of restaurants in each neighborhood. Therefore, we identify the cuisine categories that are *relatively* uncommon in that particular neighborhood. So even a popular neighborhood could have opportunity for a different kind of restaurant.

I now ranked each cuisine in different neighborhood.

category	Bakery	Bar	Café	Coffee Shop	Greek	Italian	Japanese	Pizza	Restaurant	Sushi	All
Neighborhood											
Central Bay Street	0.214286	0.571429	0.785714	1.000000	0.357143	0.500000	0.785714	0.142857	0.142857	0.642857	0.535714
Church and Wellesley	0.214286	0.285714	0.071429	0.392857	0.357143	0.214286	1.000000	0.357143	0.785714	1.000000	0.535714
Commerce Court, Victoria Hotel	0.714286	0.285714	0.928571	0.857143	0.357143	0.107143	0.928571	0.142857	0.535714	0.285714	0.535714
First Canadian Place, Underground city	0.214286	0.928571	0.321429	0.392857	0.857143	0.785714	0.142857	0.642857	0.535714	0.785714	0.535714
Garden District, Ryerson	0.500000	0.285714	0.678571	0.785714	0.714286	0.428571	0.857143	0.285714	0.642857	0.285714	0.535714
Harbourfront East, Union Station, Toronto Islands	0.428571	0.785714	0.464286	0.928571	0.357143	0.607143	0.285714	0.750000	0.714286	0.571429	0.535714
Kensington Market, Chinatown, Grange Park	1.000000	0.857143	1.000000	0.571429	0.357143	0.285714	0.428571	0.500000	0.071429	0.285714	0.535714
Little Portugal, Trinity	0.214286	1.000000	0.464286	0.071429	0.928571	0.357143	0.571429	0.750000	1.000000	0.285714	0.535714
Richmond, Adelaide, King	0.785714	0.285714	0.678571	0.714286	0.357143	0.107143	0.714286	0.857143	0.357143	0.928571	0.535714
Runnymede, Swansea	0.214286	0.714286	0.857143	0.392857	0.357143	0.785714	0.142857	1.000000	0.285714	0.857143	0.535714

4. Results

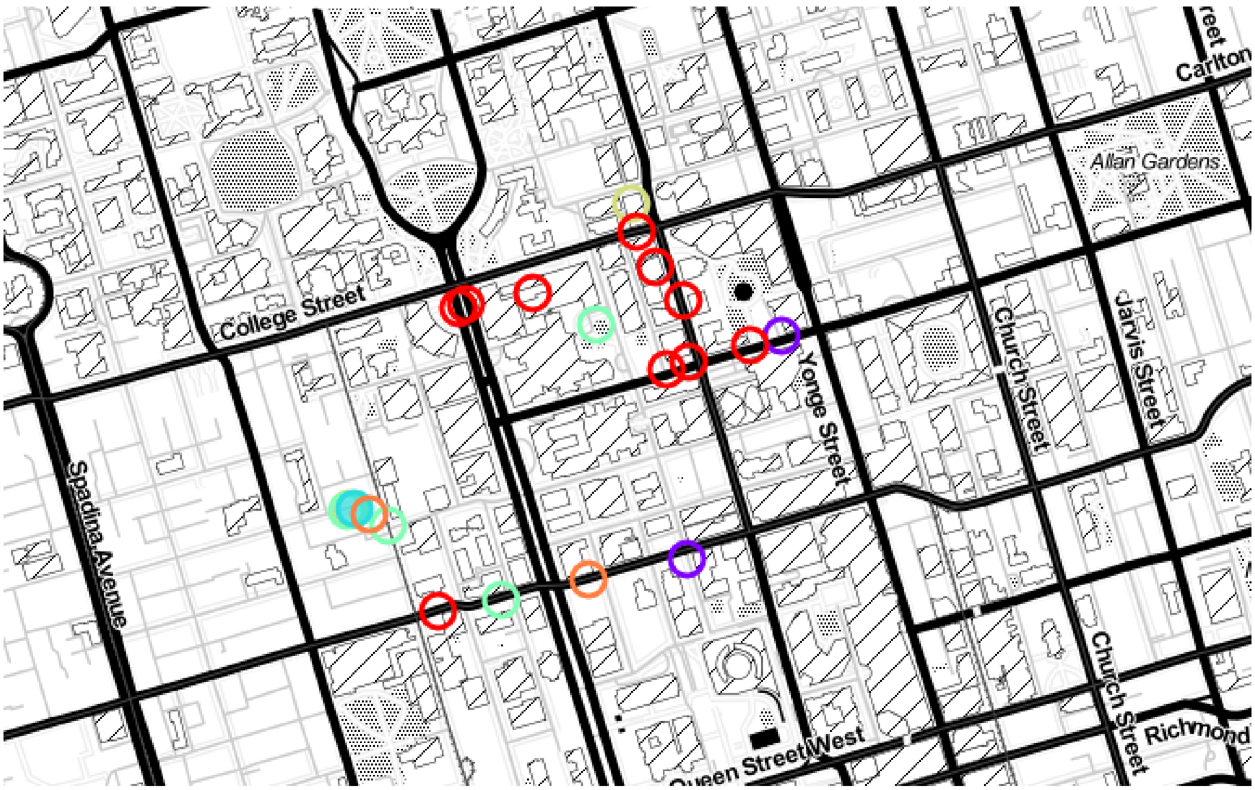
Let us consider, for each cuisine category, the neighborhoods in the bottom 20% for that cuisine. Those are neighborhoods where we could open a new restaurant of that type with little competition.

Therefore, we recommend opening the following types of restaurants in the following neighborhoods:

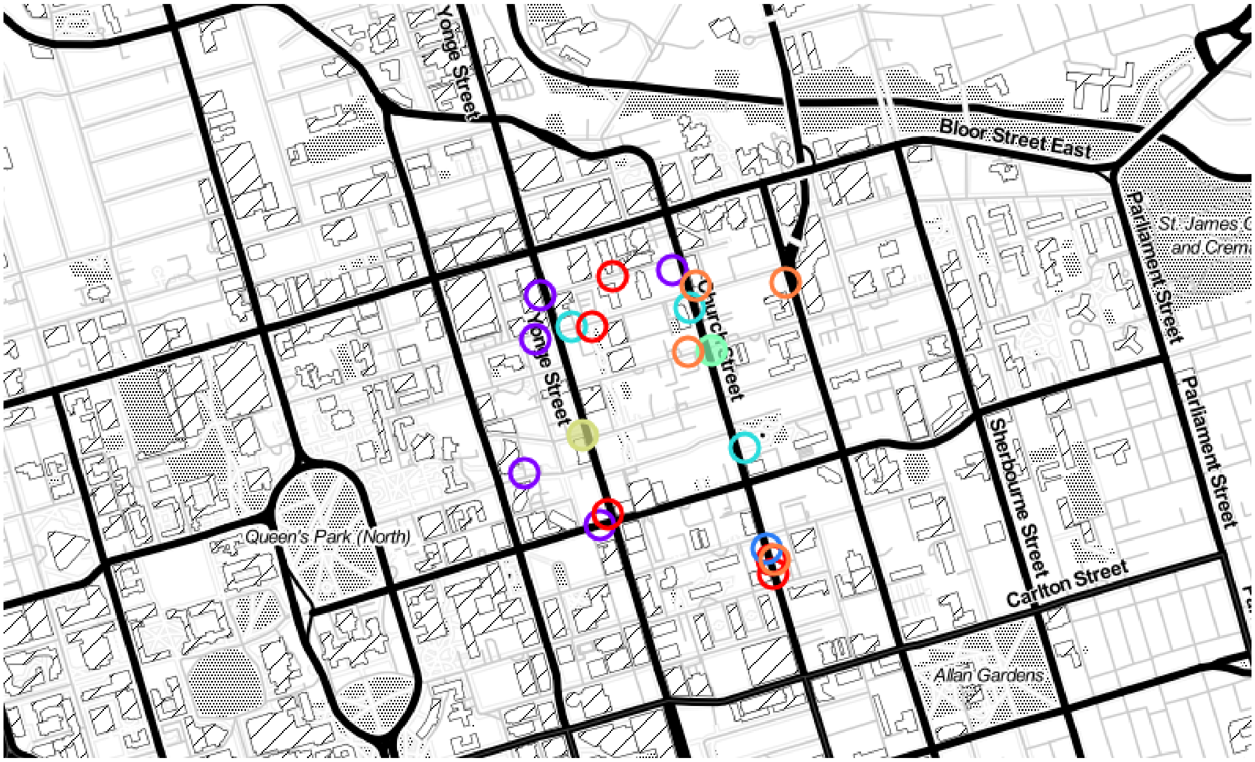
	Neighborhood	category	value
91	Central Bay Street	Pizza	0.153846
104	Central Bay Street	Restaurant	0.153846
27	Church and Wellesley	Café	0.076923
66	Church and Wellesley	Italian	0.153846
67	Commerce Court, Victoria Hotel	Italian	0.076923
93	Commerce Court, Victoria Hotel	Pizza	0.153846
110	Kensington Market, Chinatown, Grange Park	Restaurant	0.076923
46	Little Portugal, Trinity	Coffee Shop	0.076923
86	Runnymede, Swansea	Japanese	0.115385
88	Stn A PO Boxes	Japanese	0.115385
101	Stn A PO Boxes	Pizza	0.153846
37	The Danforth West, Riverdale	Café	0.153846
50	The Danforth West, Riverdale	Coffee Shop	0.153846

I draw a map for each neighborhood in which we propose to open a business. We have color coded by the category. If a dot is FILLED, then this is a competitor for the type of restaurant we want to open in this neighborhood. If the dot is EMPTY, then this restaurant is NOT a competitor--it serves a different kind of cuisine.

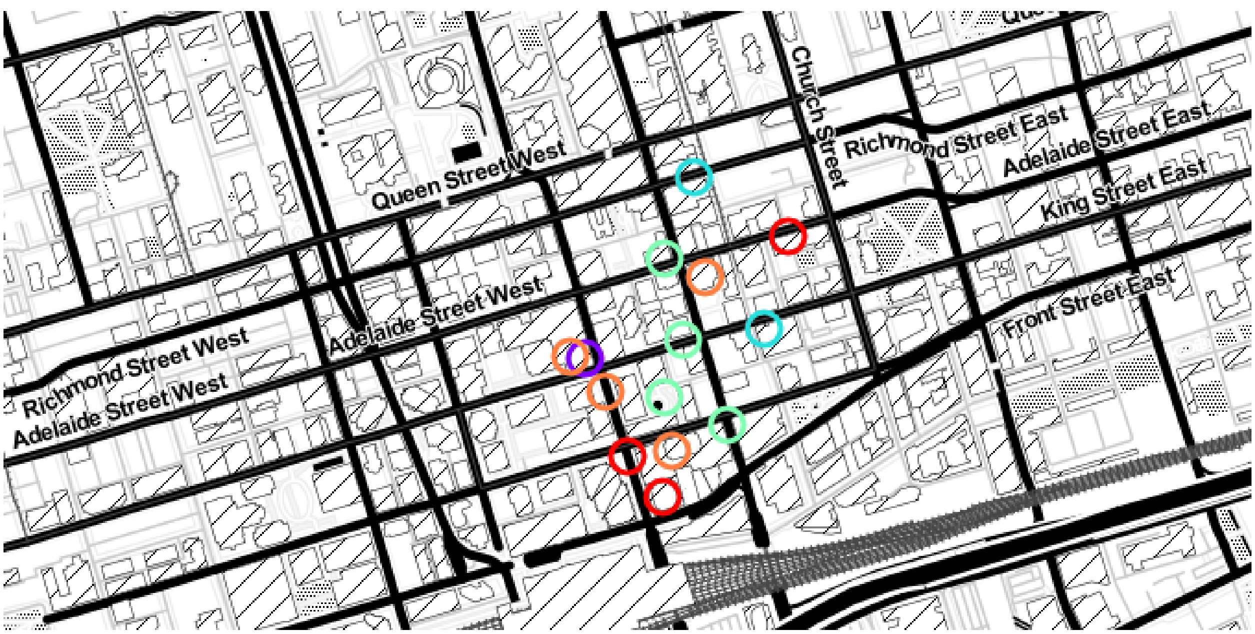
Central Bay Street Recommendations: **pizza, Restaurant**



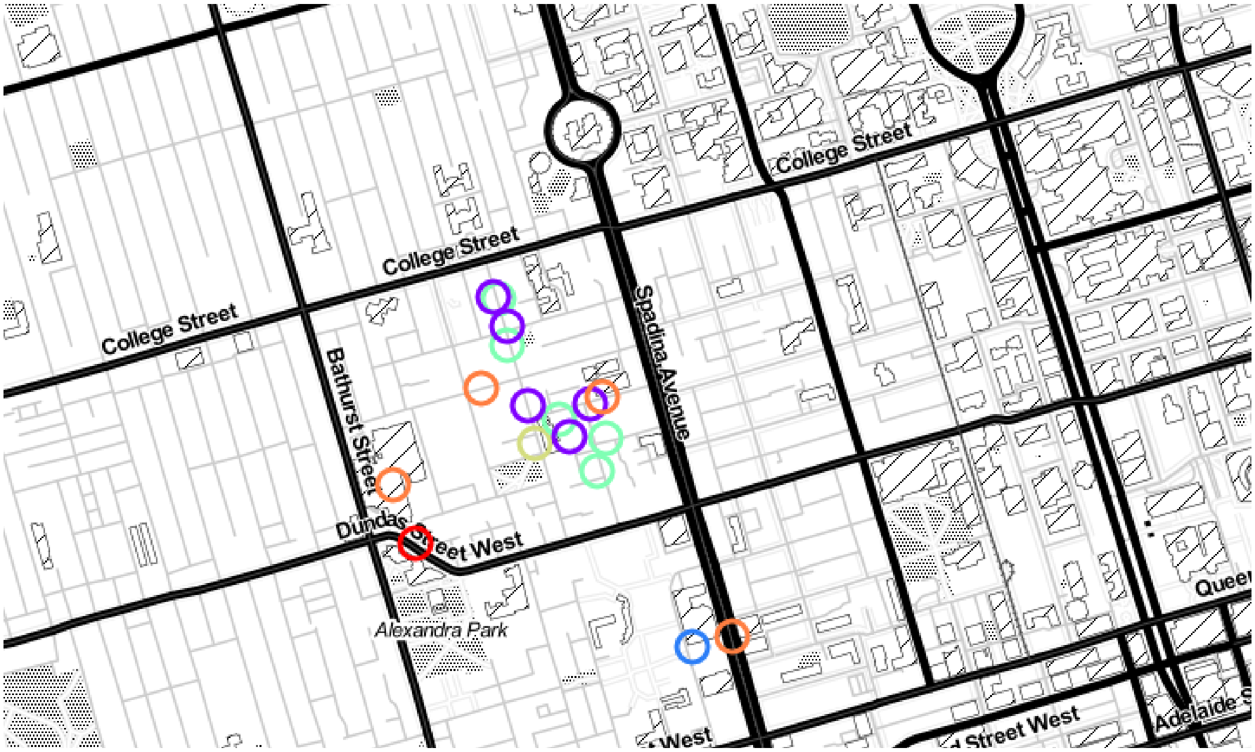
Church and Wellesley Recommendations: **Café, Italian**



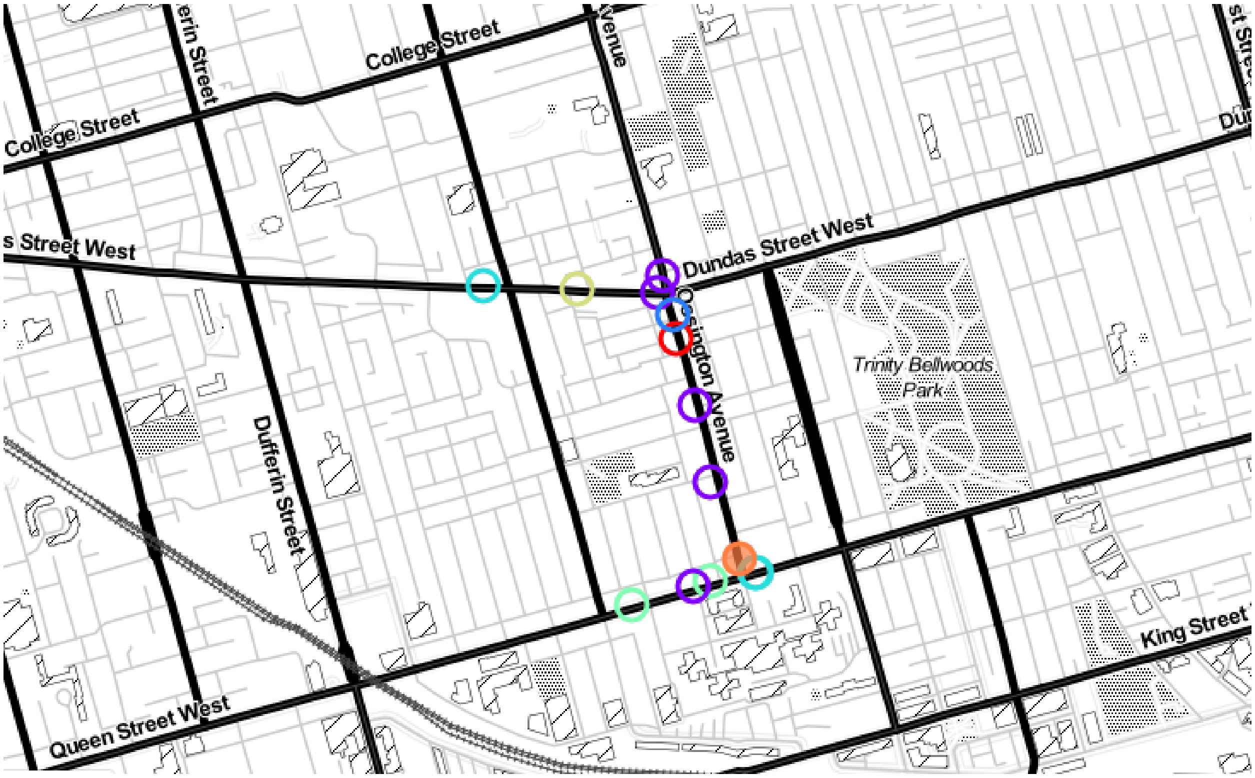
Commerce Court, Victoria Hotel Recommendations: **Italian, Pizza**



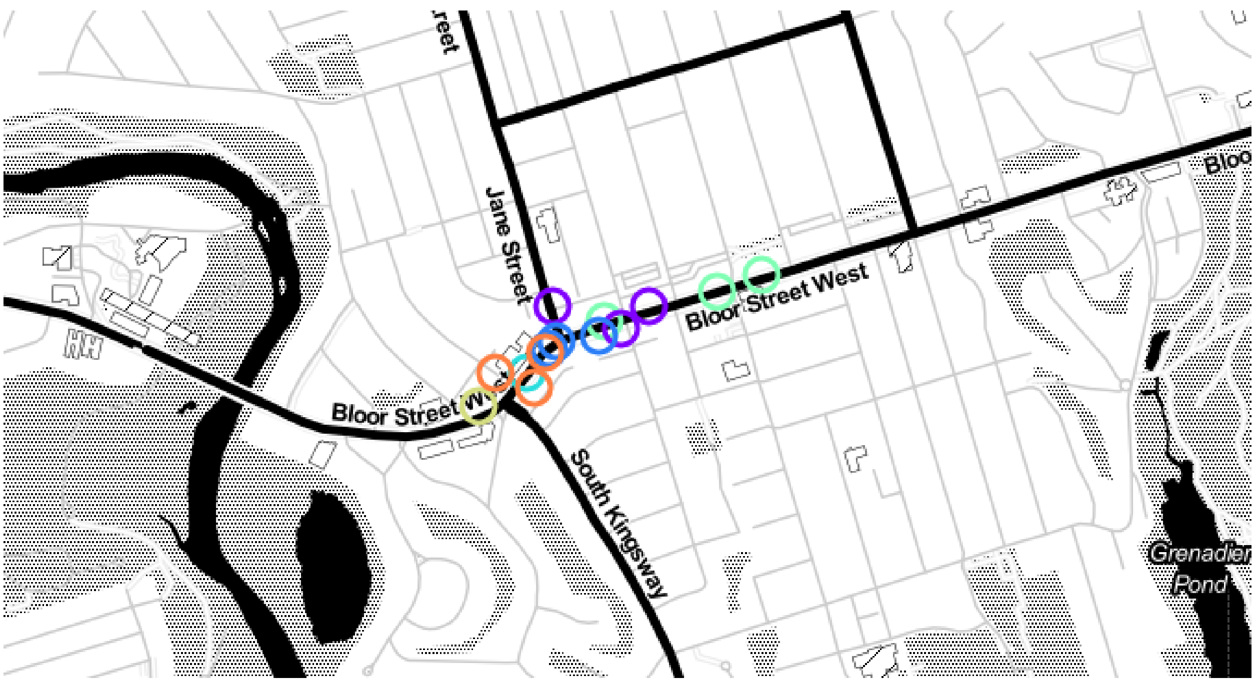
Kensington Market, Chinatown, Grange Park Recommendations**: Restaurant**



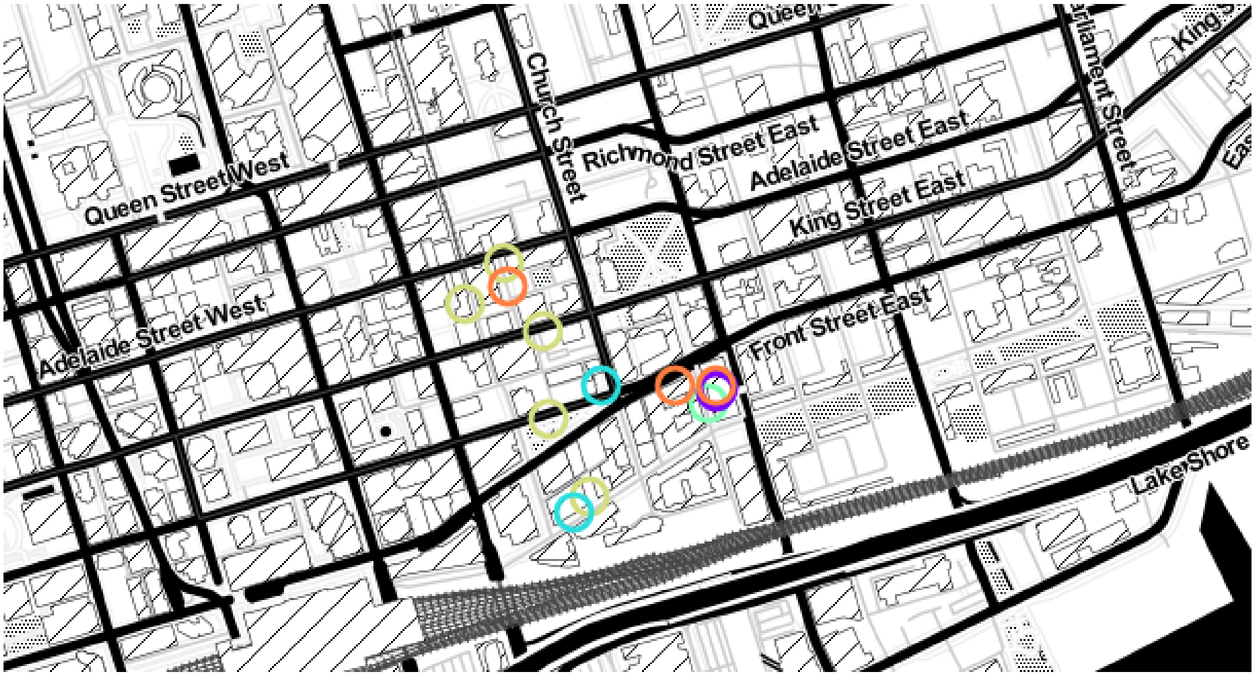
Little Portugal, Trinity Recommendations: **Coffee Shop**



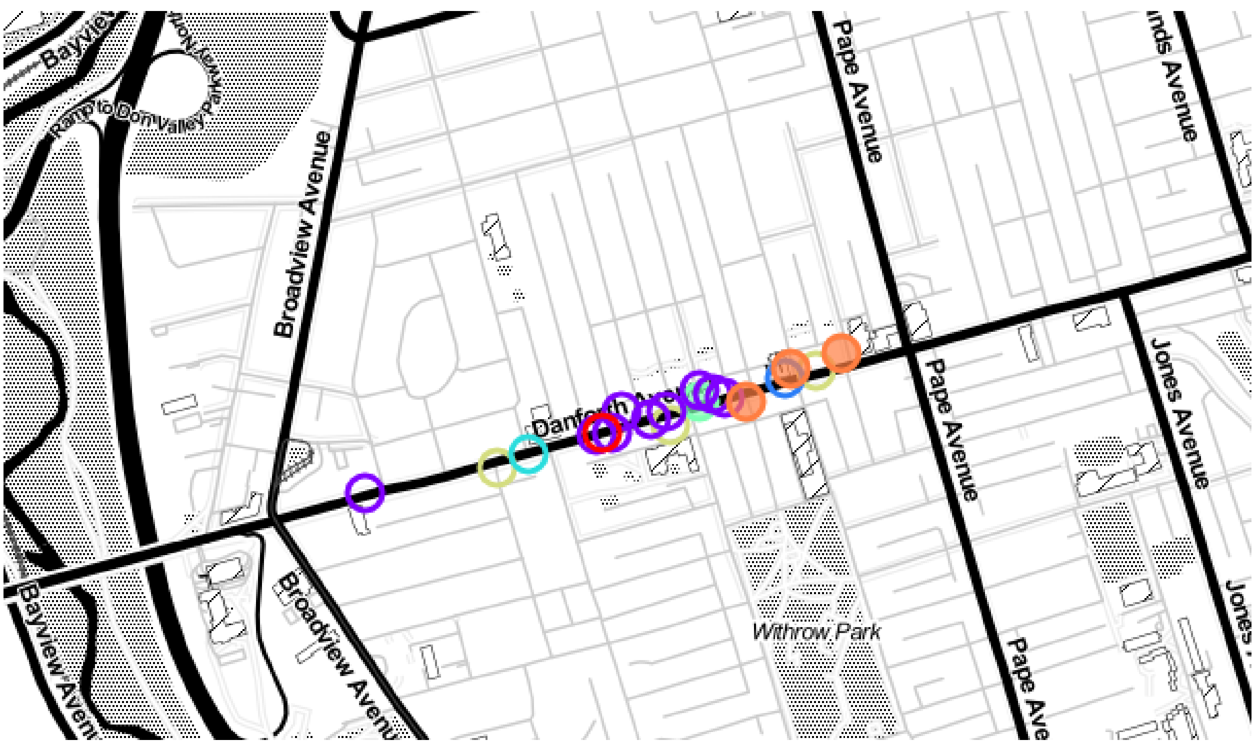
Runnymede, Swansea Recommendations: **Japanese**



Stn A PO Boxes Recommendations: **Japanese, Pizza**



The Danforth West, Riverdale Recommendations: **Café, Coffee Shop**



5. Discussion

I made this recommendation sole based on the existing restaurant on the market. I chose the neighborhood in the bottom 20% of that cuisine. This kind of cuisine could be less competitive. One can argue that this cuisine could also be less popular in that neighborhood. If I had data about the demographic of each neighborhood and survey data about people's taste, I could include that information into my recommendations. Then I will have the full picture of the restaurant market and my recommendation will be more accurate.