Report - Opening a new restaurant in Toronto

1. Introduction / Business Problem

In this scenario, there is an investor that wants to start a new restaurant in Toronto. It should be decided in which place to start the restaurant. The investor is looking for a place with less competitiveness so it could be offered a new restaurant for an existing demand.

Given that, it will be analyzed all the regions of Toronto in terms of number of restaurants and type of them, so the current offer of food can be defined for each region. It will also be considered the average income of each neighborhood in this analysis.

2. Data

It will be used the list of neighborhoods of Toronto scrapped from the link "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M" and combine the information with location data from Foursquare.

In the wikipedia link we have useful information such as name, population and land area of each neighborhood. There is also information about average income that can also be useful in our analysis. These complementary information can be found on: https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods

Name	FM	Census Tracts	Population	Land area (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit Commuting %	% Renters	Second most common language (after English) by name	Second most common language (after English) by percentage	Map
Toronto CMA Average		All	**5,113,149**	**5903.63**	**866**	**9.0**	**40,704**	**10.6**	**11.4**			
Agincourt	S	0377.01, 0377.02, 0377.03, 0377.04, 0378.02, 0378.08, 0378.14, 0378.23, 0378.24	44,577	12.45	3580	4.6	25,750	11.1	5.9	Cantonese (19.3%)	19.3% Cantonese	
Alderwood	E	0211.00, 0212.00	11,656	4.94	2360	-4.0	35,239	8.8	8.5	Polish (6.2%)	06.2% Polish	
Alexandra Park	OCoT	0039.00	4,355	0.32	13,609	0.0	19,687	13.8	28.0	Cantonese (17.9%)	17.9% Cantonese	
Allenby	OCoT	0140.00	2,513	0.58	4333	-1.0	245,592	5.2	3.4	Russian (1.4%)	01.4% Russian	
Amesbury	NY	0280.00, 0281.01, 0281.02	17,318	3.51	4,934	1.1	27,546	16.4	19.7	Spanish (6.1%)	06.1% Spanish	

The idea is to get the number of restaurants for each neighborhood. This way it will be possible to find the neighborhood with less offer of restaurants or a certain cuisine type and use this information to define where to open a new one.

The list of restaurants will be gathered from Foursquare database.

3. Methodology, Results and Discussion

Once I have to get the number of restaurants for each postal code of Toronto, I first need to get the list of postal codes. This information can be found in the link https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, so it was scrapped from there.

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M6A	North York	Lawrence Heights,Lawrence Manor
4	M7A	Queen's Park	Queen's Park

I end up with a list of 103 postal codes from Toronto, and the analysis was based on this list.

So the next step should get the list of restaurants from Foursquare API. But if I had used the query to get the list of restaurants from Toronto, I would have only 50 results which is the limit of response from the query. Considering that, the solution was to do a query for each postal code, returning a list of up to 50 restaurants of each search.

To achieve that, I had to get the latitude and longitude of each postal code and added it to the dataframe. The geographical data was got from the link http://cocl.us/Geospatial_data used on the course and I had as a result the following dataframe:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Heights,Lawrence Manor	43.718518	-79.464763
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494
5	M9A	Queen's Park	Queen's Park	43.667856	-79.532242
6	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
7	M3B	North York	Don Mills North	43.745906	-79.352188
8	M4B	East York	Woodbine Gardens,Parkview Hill	43.706397	-79.309937
9	M5B	Downtown Toronto	Ryerson,Garden District	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073
11	M9B	Etobicoke	Cloverdale,Islington,Martin Grove,Princess Gar...	43.650943	-79.554724
12	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497

Given that, I was able to do a query for each postal code and get the information using the Foursquare API. The data gathered was stored in a new dataframe df_restaurants.

	Name	Category	City	Postal Code	Distance	PostalCode Searched
0	Allwyn's Bakery	Caribbean Restaurant	Toronto	M3A 1Z5	833	M3A
1	Tim Hortons	Café	Toronto	M3A 1Z5	866	M3A
2	A&W Canada	Fast Food Restaurant	Toronto	M3A 1Z5	852	M3A
3	High Street Fish & Chips	Fish & Chips Shop	Toronto	M3A 2J7	967	M3A
4	Pizza Pizza	Pizza Place	North York	M3A 1Z5	839	M3A
5	Spicy Chicken House	Chinese Restaurant	North York	M3A 1Z5	881	M3A
0	Portugril	Portuguese Restaurant	Toronto	NaN	224	M4A
1	The Frig	French Restaurant	NaN	NaN	197	M4A
2	Pizza Nova	Pizza Place	Toronto	M4A 1J8	218	M4A
3	JJ Bean	Café	NaN	NaN	649	M4A
0	Impact Kitchen	Restaurant	Toronto	M5A 4L3	376	M5A

The next step was to do some treatment to the data:

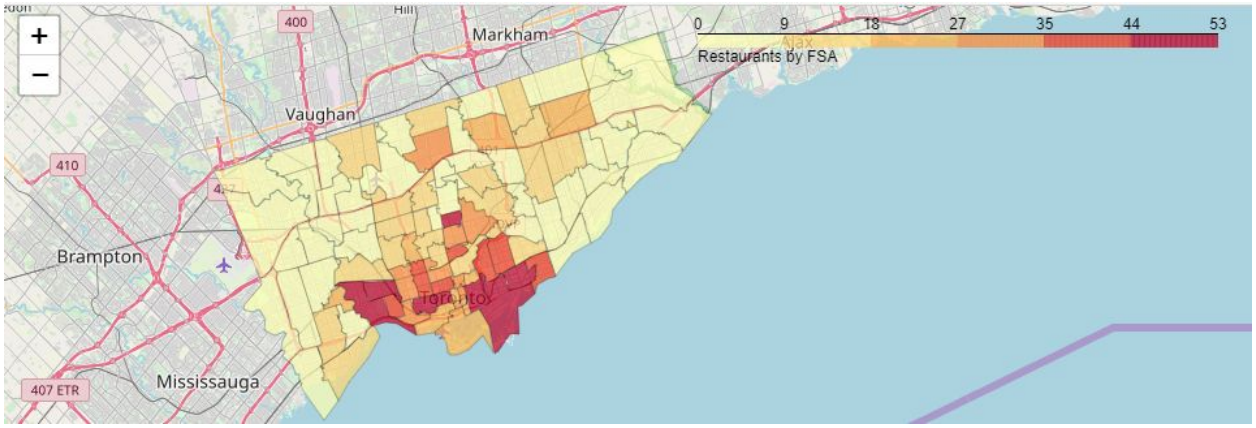
- Duplicated data were removed - once I did a search for each postal code and a fixed radius, I end up with duplicate data;
- The information of address got from Foursquare was incomplete for some venues (NaN value for postal code). In these cases I assumed that the postal code was the postal code searched;
- I was interested only in the first 3 digits of postal code, so the data was treated.

After the treatments to the data, I end up with a list of 1609 restaurants or other venues that deal with food (cafes for example).

The next step was then group the number of restaurants for each postal code and add this information to the first dataframe. I end up with the following dataframe.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Restaurants
0	M3A	North York	Parkwoods	43.753259	-79.329656	4.0
1	M4A	North York	Victoria Village	43.725882	-79.315572	2.0
2	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636	53.0
3	M6A	North York	Lawrence Heights,Lawrence Manor	43.718518	-79.464763	16.0
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494	8.0
5	M9A	Queen's Park	Queen's Park	43.667856	-79.532242	3.0
6	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353	8.0
7	M3B	North York	Don Mills North	43.745906	-79.352188	16.0
8	M4B	East York	Woodbine Gardens,Parkview Hill	43.706397	-79.309937	7.0

It was also plotted a map with chloropleth representing the number of restaurants for each region.



As stated above, I was interested in the area with the minimum number of restaurants. Considering that, I had 5 postal codes with 0 restaurant collected:

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Restaurants
76	M7R	Mississauga	Canada Post Gateway Processing Centre	43.636966	-79.615819	0.0
92	M5W	Downtown Toronto	Stn A PO Boxes 25 The Esplanade	43.646435	-79.374846	0.0
93	M8W	Etobicoke	Alderwood,Long Branch	43.602414	-79.543484	0.0
94	M9W	Etobicoke	Northwest	43.706748	-79.594054	0.0
95	M1X	Scarborough	Upper Rouge	43.836125	-79.205636	0.0

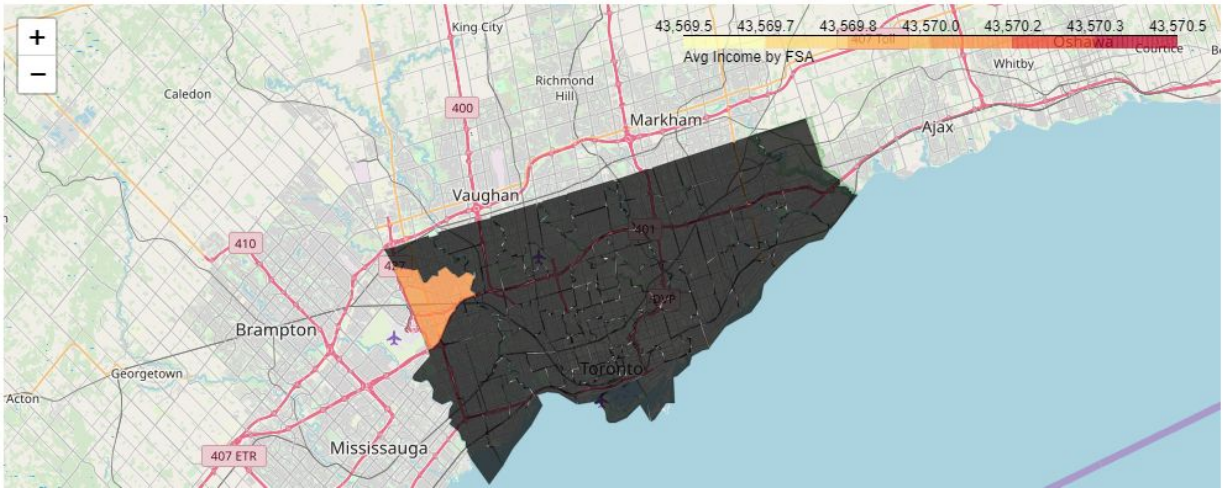
To help decide whether to choose one or another, I considered then the average income for each region. The information was got from the link https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods.

The first two rows represent a Post Centre and a Station, which are not areas of interest. So I assumed value 0 for average income.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Restaurants	Avg Income
76	M7R	Mississauga	Canada Post Gateway Processing Centre	43.636966	-79.615819	0.0	0
92	M5W	Downtown Toronto	Stn A PO Boxes 25 The Esplanade	43.646435	-79.374846	0.0	0
93	M8W	Etobicoke	Alderwood,Long Branch	43.602414	-79.543484	0.0	35239
94	M9W	Etobicoke	Northwest	43.706748	-79.594054	0.0	43570
95	M1X	Scarborough	Upper Rouge	43.836125	-79.205636	0.0	29230

Given that, I got as a result that the best region to open a new restaurant would be the M9W postal code because it is a region with 0 restaurant and the best average income.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Restaurants	Avg Income
94	M9W	Etobicoke	Northwest	43.706748	-79.594054	0.0	43570



4. Conclusion

I started this case looking for the best place to open a new restaurant in Toronto. The idea was to consider the region with less competitiveness, in this approach represented as the minimum number of restaurants. I end up with a list of 5 regions with 0 restaurant and used a complementary criteria to choose the best one: the average income. I assumed that people with higher average income could spend more with restaurants so it could be a better area to open a new one. I have to note that it is a bit strange that there are postal codes with zero restaurant, but it is a result from the used approach using the Foursquare API. So, given this approach, the best place to open a new restaurant would be the M9W postal code, which refers to a Etobicoke region (Northwest).