The opening of Restaurant.

1. Introduction.

1.1 Background and Problem.

When starting a business, such as opening a restaurant, it is important to evaluate the characteristics of the place where the object will be located. For example, population density, contingent, average income, the presence of competitors with similar business - all this will affect the success of your business.

Using data on places and city districts that provide web resources, we will be able to assess the presence of possible competitors for future business and select the best area for placing an object in the city of Toronto.

1.2 Interest.

This study may be of interest to potential owners of chains of catering establishments or people just starting your own business. This study affects only part of the assessment of the choice of location for the facility. Other economic calculations for business are not carried out here.

2. Data collection and assessment.

Data for this study is provided by Foursquare. The data are location coordinates and categories of various public places, classified into various categories. The essence of the data assessment is to calculate the places of catering (search for popular categories related to food, or kitchen) for each district of the city of Toronto and determining the minimum number of possible competitors in the vicinity of the area where the business will be located - a restaurant.

3. Methodology.

The data on the names of the districts in the city of Toronto are taken from the source https://en.wikipedia.org /wiki/ List_of_postal_codes_of_Canada:_M as a table. Data is cleared from zero values, sorted alphabetically. The latitude and longitude for each district are provided in the form of a CSV table at http://cocl.us/Geospatial_data and compared with data from Wikipedia.

After processing and compiling a single table with names and geolocation data, it became possible to create a map with marked areas:

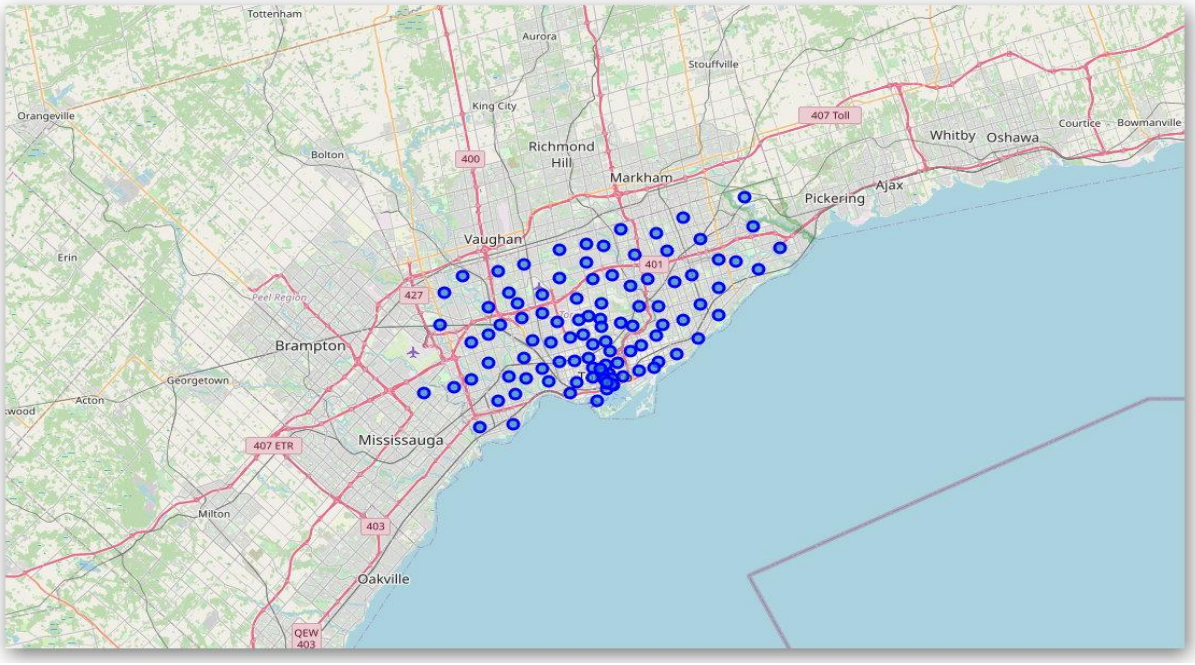

Figure 1. Toronto and marked areas.

The study further focused on one of the central parts of the city - North York - one of the largest areas. From the general data table, those data were identified that belong to the North York area . We focused on the latitude and longitude of the small areas that make up North York and studied them:

Figure 2. Neighborhoods of North York.

Further, using the Foursquare service and API requests, we managed to get all the significant public places within a radius of 1000 meters from the latitude and longitude coordinates of each area in North York . Data is provided as a JSON file and processed by software libraries. At the end we get this table with the names of places and categories.

	name	categories	lat	lng
0	Tastee	Bakery	43.807722	-79.356798
1	Galati	Grocery Store	43.797831	-79.369410
2	Cummer Park	Park	43.799564	-79.371175
3	고려삼계탕 Korean Ginseng Chicken Soup & Bibimbap	Korean Restaurant	43.798391	-79.369187
4	Tim Hortons	Coffee Shop	43.798945	-79.369644

Figure 3. Data on public places

Since our task is to choose a place to open a restaurant, then of all the objects received, we will be interested in those categories that are connected, one way or another, with food. Such objects may be potential competitors for our restaurant. And the less of them will be concentrated in the considered areas, the better. We have identified the most common words that may be in the name of the institutions of potential competitors. We identified such key words: “Restaurant”, “Cafe”, “Coffee”, “Food”, “Pizza”, “Bar”, “Ice Cream”, “ Burger ”, “Sandwich”. All objects in the names of which these key words occur were allocated in the corresponding categories.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Main Venue Category
0	Hillcrest Village	43.803762	-79.363452	Tastee	43.807722	-79.356798	Bakery	Others
1	Hillcrest Village	43.803762	-79.363452	고려삼계탕 Korean Ginseng Chicken Soup & Bibimbap	43.798391	-79.369187	Korean Restaurant	Restaurant
2	Hillcrest Village	43.803762	-79.363452	Cummer Park	43.799564	-79.371175	Park	Others
3	Hillcrest Village	43.803762	-79.363452	Tim Hortons	43.798945	-79.369644	Coffee Shop	Coffee
4	Hillcrest Village	43.803762	-79.363452	TD Canada Trust	43.798466	-79.368832	Bank	Others
5	Hillcrest Village	43.803762	-79.363452	Subway	43.799059	-79.368946	Sandwich Place	Sandwich
6	Hillcrest Village	43.803762	-79.363452	Pizza Pizza	43.799079	-79.369449	Pizza Place	Pizza
7	Hillcrest Village	43.803762	-79.363452	New York Fries	43.803664	-79.363905	Fast Food Restaurant	Restaurant
8	Hillcrest Village	43.803762	-79.363452	Shoppers Drug Mart	43.798341	-79.369804	Pharmacy	Others
9	Hillcrest Village	43.803762	-79.363452	Cliffwood I.D.A. Pharmacy	43.807909	-79.357777	Pharmacy	Others

Figure 4. The main categories. Potential competitors.

After further processing and sorting of the data, it was possible to compile a table that displays the total number of objects of the main categories for each district:

	Neighborhood	Total Amount	Restaurants	Coffees	Foods	Pizza	Bars	Ice Cream	Burgers	Sandwiches	Others
0	Bathurst Manor, Wilson Heights, Downsview North	26	4.0	2.0	0.0	2.0	0.0	1.0	0.0	1.0	16
1	Bayview Village	9	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6
2	Bedford Park, Lawrence Manor East	37	11.0	3.0	0.0	1.0	1.0	0.0	0.0	2.0	19
3	Don Mills	34	13.0	2.0	0.0	0.0	0.0	0.0	0.0	1.0	18
4	Downsview	52	13.0	5.0	2.0	4.0	0.0	0.0	0.0	2.0	26

Figure 5. Summary table. The sum of all potential competitors. Grouped by district.

Now we have the opportunity to apply one of the methods of machine learning - Clustering. Because this method allows you to create groups of objects with similar characteristics . For us, it will look like clustering areas with a large and small number of potential competitors for the restaurant.

Borough	Neighborhood	Latitude	Longitude	Cluster Labels	Total Amount	Restaurants	Coffees	Foods	Pizza	Bars	Ice Cream	Burgers	Sandwiches	Others
North York	Hillcrest Village	43.803762	-79.363452	0	17	4.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	9
North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	3	39	7.0	5.0	1.0	0.0	2.0	0.0	1.0	1.0	22
North York	Bayview Village	43.786947	-79.385975	0	9	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6
North York	York Mills, Silver Hills	43.757490	-79.374714	0	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1
North York	Willowdale, Newtonbrook	43.789053	-79.408493	0	18	6.0	2.0	0.0	1.0	1.0	0.0	0.0	1.0	7
North York	Willowdale	43.770120	-79.408493	1	90	34.0	6.0	0.0	6.0	3.0	1.0	0.0	3.0	37
North York	York Mills West	43.752758	-79.400049	0	8	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7
North York	Willowdale	43.782736	-79.442259	1	90	34.0	6.0	0.0	6.0	3.0	1.0	0.0	3.0	37
North York	Parkwoods	43.753259	-79.329656	0	5	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	4
North York	Don Mills	43.745906	-79.352188	3	34	13.0	2.0	0.0	0.0	0.0	0.0	0.0	1.0	18
North York	Don Mills	43.725900	-79.340923	3	34	13.0	2.0	0.0	0.0	0.0	0.0	0.0	1.0	18
North York	Bathurst Manor, Wilson Heights, Downsview North	43.754328	-79.442259	3	26	4.0	2.0	0.0	2.0	0.0	1.0	0.0	1.0	16
North York	Northwood Park, York University	43.767980	-79.487262	0	16	6.0	1.0	0.0	1.0	2.0	0.0	0.0	2.0	4
North York	Downsview	43.737473	-79.464763	2	52	13.0	5.0	2.0	4.0	0.0	0.0	0.0	2.0	26
North York	Downsview	43.739015	-79.506944	2	52	13.0	5.0	2.0	4.0	0.0	0.0	0.0	2.0	26
North York	Downsview	43.728496	-79.495697	2	52	13.0	5.0	2.0	4.0	0.0	0.0	0.0	2.0	26
North York	Downsview	43.761631	-79.520999	2	52	13.0	5.0	2.0	4.0	0.0	0.0	0.0	2.0	26
North York	Victoria Village	43.725882	-79.315572	0	8	2.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	5
North York	Bedford Park, Lawrence Manor East	43.733283	-79.419750	3	37	11.0	3.0	0.0	1.0	1.0	0.0	0.0	2.0	19
North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	2	43	10.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	31
North York	Glencairn	43.709577	-79.445073	0	20	5.0	1.0	0.0	2.0	0.0	1.0	0.0	0.0	11
North York	North Park, Maple Leaf Park, Upwood Park	43.713756	-79.490074	0	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5
North York	Humber Summit	43.756303	-79.565963	0	4	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	3
North York	Humberlea, Emery	43.724766	-79.532242	0	6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6

Figure 6. Clustering of areas.

From this table you can see that areas with cluster 0 have the least number of potential competitors. Areas from cluster 3 have the largest number of potential competitors. Let's make a map with clusters:

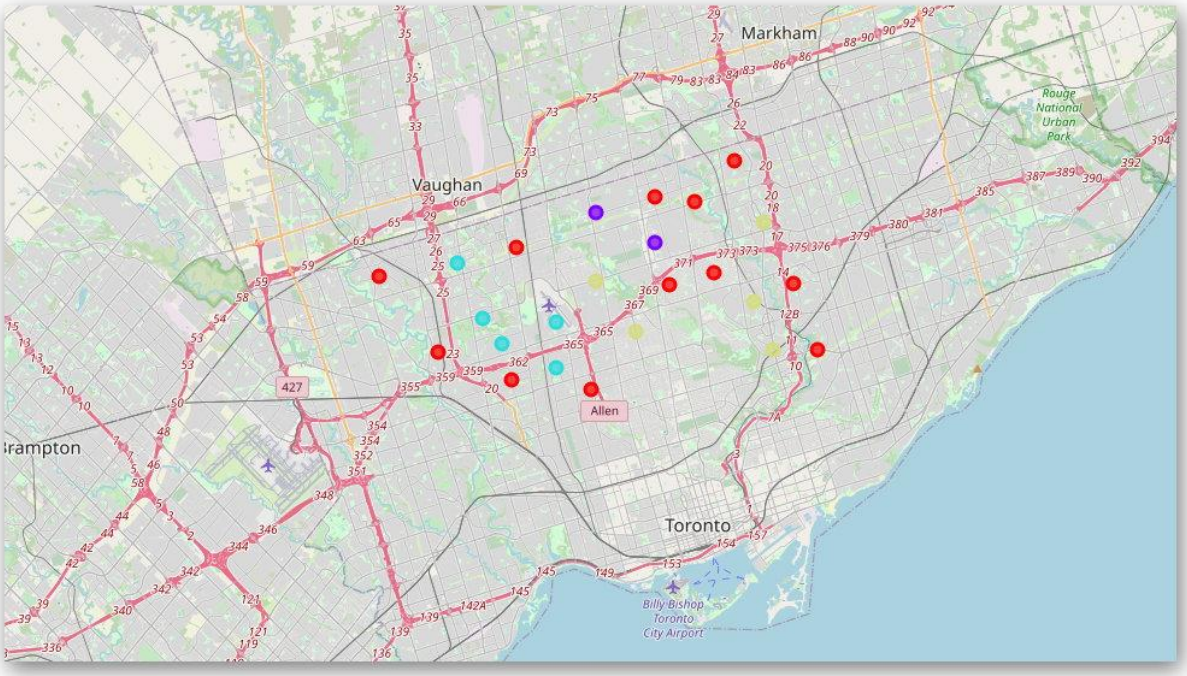

Figure 7. North York areas stained in groups.

4. Results.

The red marks are the cluster with the number 0. These are the areas that are primarily recommended for the new restaurant. Yellow labels are cluster number 3. They should be considered last, as they have the largest number of competitors. The data is current as of May 2020.

5. Conclusion section.

For further research and a more detailed selection of the restaurant location, it is worth considering data of a different nature. For example, to study the population of each region by income. Or you can explore the logistics component, which will be responsible for supplying the business. These data are beyond the scope of this study.