

Report: Identification of location to open a new restaurant in Toronto.

1. Introduction and description of the business problem.

One of the most important factors determining a success in opening and running a restaurant business is the location. The most suitable location can be characterised by the following criteria:

- sizable population
- positive population growth in the last years
- higher income
- low number of restaurants in the vicinity (restaurants per person).

In addition, other regional characteristics, such as the biggest minority population (other cultures) can also affect the suitability for a restaurant with a particular cuisine. Therefore, the question to be solved is to:

Identify neighborhoods with the lowest ratio of restaurants per population, ideally with high income and still growing population (e.g. newly developing areas).

2. Data sources, analysis and handling.

In order to answer identify the best place to open a restaurant, I will rank all the neighborhoods based on the following criteria:

- number of restaurants within 2 km per population (lower is better)
- median income in the neighborhood (higher is better)
- population growth in recent years (higher is better).

The lowest sum of the positions in the three rankings above will identify the best place.

I will obtain the data about restaurants within the 2 km radius of each neighborhood from the Foursquare location data. In addition, the demographics data for the Toronto neighborhoods will be obtained from the Wikipedia page [Demographics of Toronto neighbourhoods](#). This data provides information about the population, growth/decline of population, average income, second most common language for all the neighborhoods.

I will fuse the data obtained from the two sources into a master table containing all the relevant data. This will be used for ranking. A clustering analysis can be also performed in order to explore the similarity between different neighborhoods.

3. Methodology and data analysis

After import and installation of all necessary libraries, the data obtained from the Wikipedia page was imported. The data was wrangled to remove unnecessary columns, check and remove missing values and cleaned by removing hyperlinks and white spaces.

Analysis and visual presentation of data revealed that there are 140 unique neighborhoods. To further understand the similarity of between the neighborhoods, a machine learning using K-means clustering was used. The neighborhoods were clustered based on the population size, population growth, average income and a language spoken at home other than English (to account for distinct cultural aspect of the residents). The clusters were also plotted on the map to provide visual information about their geographical relation.

To identify neighborhoods most suitable for opening of a new restaurant, the data about existing restaurants in each neighborhood and 2 km radius were queried from the Foursquare.com location data. The number of restaurants and their respective types (cuisine served) were added to the neighborhoods data table. Finally, a neighborhood score was calculated, using weighted ranks of neighborhoods in the categories: number of restaurants per population, population size, population growth and average income. The neighborhoods with lowest scores are the best candidates to open a new restaurant. The neighborhoods and their scores were visually presented on a choropleth map.

4. Results

In total 140 neighborhoods of Toronto were included in the analysis.

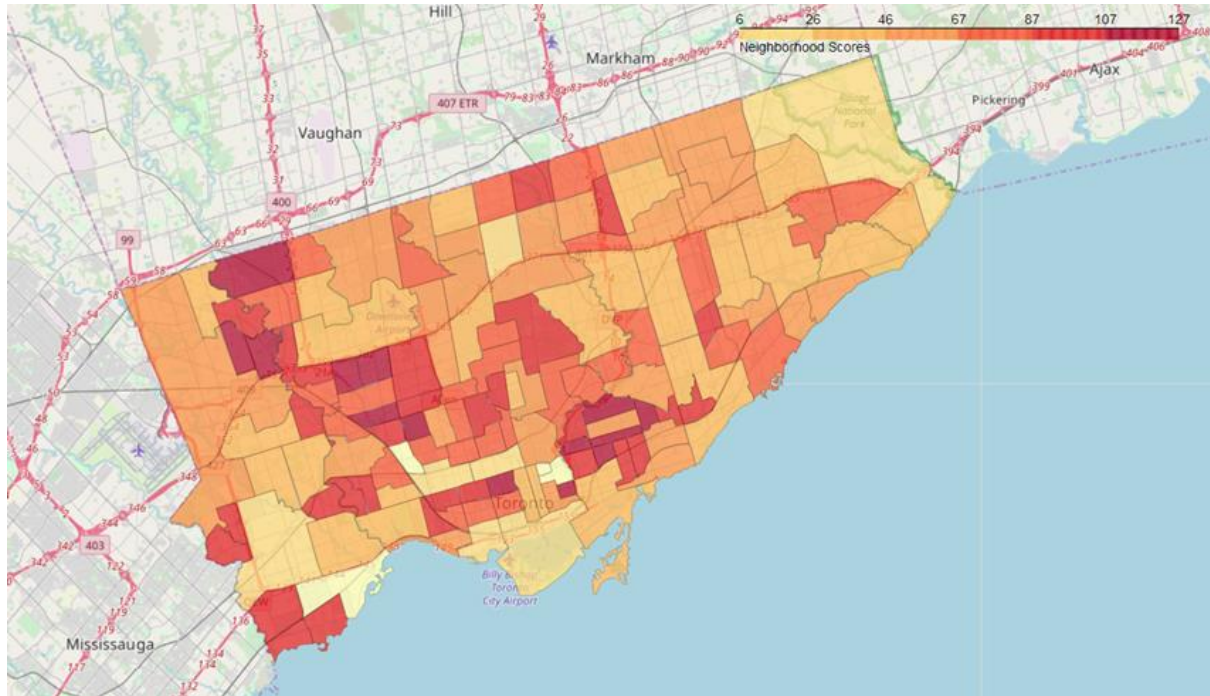
Scoring analysis

Scoring analysis revealed that the best neighborhoods tend to have a large population, above average income and positive growth.

	Neighborhood	Population	Population_growth%	Average Income	Second language	Latitude	Longitude	Restaurants count	Restaurants/1000residents	Restaurants rank	Population rank	Income rank	Population growth	Score
122	Waterfront Communities-The Island	65913	0.520	54292	Mandarin	43.643	-79.3716	18	0.273087	4	1	25	1	7.3
58	Islington-City Centre West	43965	0.154	42550	Korean	43.6361	-79.543	17	0.386671	7	6	48	10	15.3
87	Niagara	31180	0.466	54803	Spanish	43.6429	-79.4079	18	0.577293	24	15	24	2	20.0
129	Willowdale East	50434	0.120	36740	Mandarin	43.7725	-79.4004	20	0.396558	9	3	69	13	20.2
77	Mimico	33964	0.280	43971	Polish	43.6173	-79.4989	18	0.529973	21	11	37	6	20.7
3	Annex	30526	0.046	80138	Mandarin	43.6694	-79.4028	19	0.622420	27	16	10	32	21.9
32	Dovercourt-Wallace Emerson-Junction	36625	0.058	33777	Portuguese	43.666	-79.4369	13	0.354949	6	8	80	28	23.4
105	Rouge	46496	0.013	33736	Tamil	43.8077	-79.174	5	0.107536	1	4	82	76	25.3
23	Church-Yonge Corridor	31340	0.106	43039	Mandarin	43.6602	-79.3787	20	0.638162	29	14	45	15	27.8
36	Edenbridge-Humber Valley	15535	0.040	72156	Spanish	43.6699	-79.5213	7	0.450595	13	83	12	38	29.3

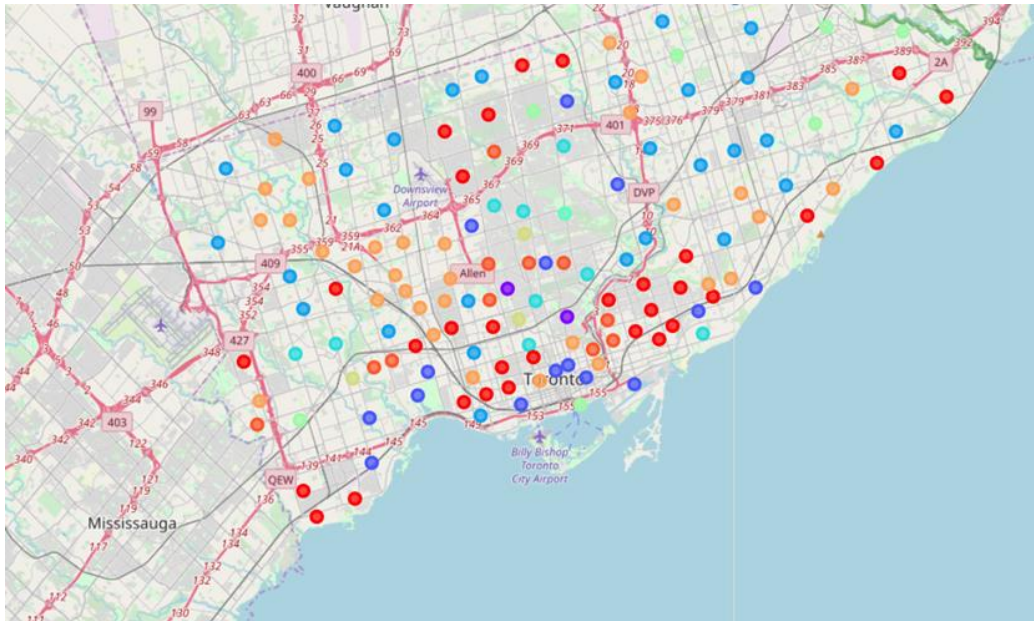
Choropleth map of scores

Plotting of the scores using a choropleth map of Toronto shows that the best neighborhood are not closely geographically related.



Clustering analysis

Clustering and plotting of the clusters show their geographical relation. Examining the clusters (examples below) can provide information about the type of customers that could be expected in different neighborhood. This can provide direction towards the selection of what type of cuisine might be best in a specific area.



Examination of example of clusters

Cluster 1 = neighborhoods with small population size and mixed foreign minority.

	Neighborhood	Population	Population_growth%	Average Income	Second language
2	Alderwood	12054	0.013	39159	Polish
5	Bathurst Manor	15873	0.028	37927	Russian
8	Bayview Woods-Steeles	13154	-0.028	38196	Mandarin
14	Blake-Jones	7727	-0.005	39515	Cantonese
17	Broadview North	11499	-0.006	37018	Greek
22	Centennial Scarborough	13362	0.021	43854	Tamil
25	Clanton Park	16472	0.127	41837	Tagalog
26	Cliffcrest	15935	0.015	36364	Bengali
27	Corso Italia-Davenport	14133	0.028	35165	Portuguese
28	Danforth	9666	0.024	43924	Greek
29	Danforth East York	17180	0.028	42455	Greek
40	Eringate-Centennial-West Deane	18588	-0.012	39203	Ukrainian

Cluster 2 = affluent neighborhoods with Chinese population.

	Neighborhood	Population	Population_growth%	Average Income	Second language
44	Forest Hill South	10732	-0.018	142627	Mandarin
104	Rosedale-Moore Park	20923	0.014	134865	Mandarin

Cluster 3 = middle class neighborhoods with mixed population.

	Neighborhood	Population	Population_growth%	Average Income	Second language
4	Banbury-Don Mills	27695	0.029	51874	Mandarin
6	Bay Street Corridor	25797	0.333	43427	Mandarin
7	Bayview Village	21396	0.211	41440	Mandarin
12	Birchcliffe-Cliffside	22291	0.020	43027	Bengali
23	Church-Yonge Corridor	31340	0.106	43039	Mandarin
35	East End-Danforth	21381	0.026	43567	Bengali
39	Englemount-Lawrence	22372	0.013	39911	Tagalog
49	High Park North	22162	0.041	45893	Russian
50	High Park-Swansea	23925	0.101	54872	Polish
77	Mimico	33964	0.280	43971	Polish
79	Moss Park	20506	0.258	45877	Mandarin
83	Mount Pleasant West	29658	0.037	45441	Persian

5. Discussion and recommendation

Based on the scoring analysis, the five best neighborhoods to open a new restaurant are:

Waterfront Communities-The Island, Islington-City Centre West, Niagara, Willowdale East and Mimico.

Several observations can be made about the optimal locations:

- ▶ the optimal neighborhoods tend to have above average income and larger population size.
- ▶ positive population growth, securing inflow of new customers in future.
- ▶ Based on the cluster analysis, apart from standard cuisine, Asian or Spanish or Polish restaurants might be the best choice.

6. Conclusions

In conclusion, the use of data analysis and machine learning and data from location information services such as Foursquare.com can provide relevant information and identify suitable location for opening of a new business. More detailed analysis would provide further direction about which type of customers can be expected and therefore which type of restaurant (business) to open in the selected area.

