

Location for a new restaurant

Using regional data from Foursquare.com to identify the optimal location

Coursera IBM Data Science Capstone Project

Martin Kos

Business problem and question

- ▶ Success of a restaurant depends very much on its location. Today, availability of large amount of information about different places can provide basis for identification of the optimal place to open a new restaurant.
- ▶ **Question: How to identify the best location for a new restaurant?**
- ▶ Use data science approaches to analyse location data Foursquare.com and other sources to predict optimal regions. In this example I use Toronto as a model city.

Criteria used for modelling of the problem

- ▶ Population size - positive factor
- ▶ Population dynamics - growth (positive) vs decline (negative)
- ▶ Income - positive (people with larger income are more likely to eat outside.)
- ▶ Number of restaurants in vicinity - negative factor (direct competition)
- ▶ Non-native population - affects selection of the type (cuisine) of a restaurant.

Data used for analysis

- ▶ Toronto demographic data from https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods
- ▶ Geocoding data obtained from Here Maps API.
- ▶ List of restaurants in 2 km radius obtained from Foursquare.com

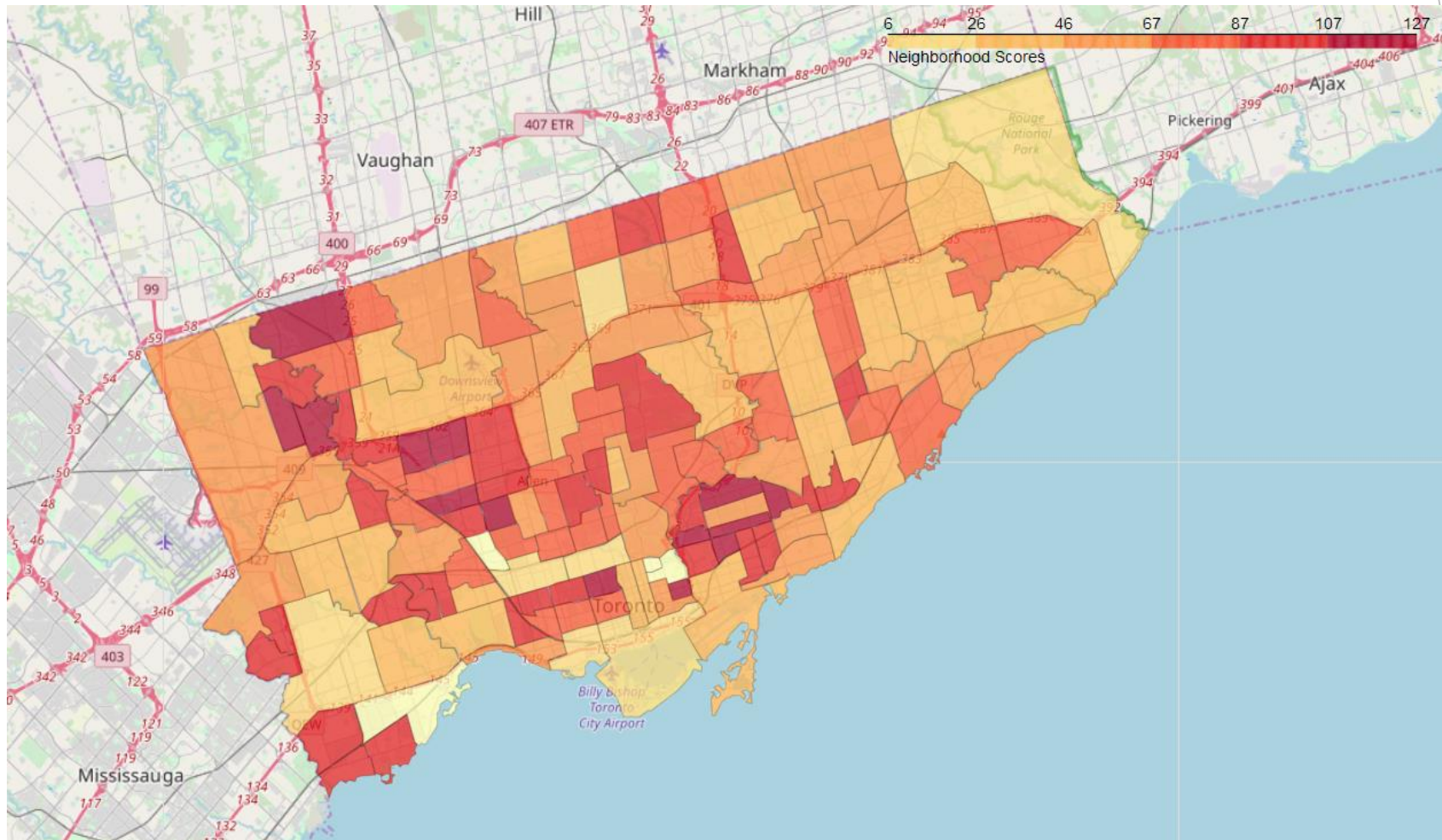
Snapshot of the data table

	Neighborhood	Population	Population_growth%	Average Income	Second language	Latitude	Longitude
0	Agincourt North	29113	-0.039	26955	Cantonese	43.8093	-79.2671
1	Agincourt South-Malvern West	23757	0.080	27928	Cantonese	43.7874	-79.2694
2	Alderwood	12054	0.013	39159	Polish	43.605	-79.5412
3	Annex	30526	0.046	80138	Mandarin	43.6694	-79.4028
4	Banbury-Don Mills	27695	0.029	51874	Mandarin	43.7404	-79.3485
5	Bathurst Manor	15873	0.028	37927	Russian	43.7638	-79.4548
6	Bay Street Corridor	25797	0.333	43427	Mandarin	43.6577	-79.3862
7	Bayview Village	21396	0.211	41440	Mandarin	43.7771	-79.3796
8	Bayview Woods-Steeles	13154	-0.028	38196	Mandarin	43.7949	-79.3822
9	Bedford Park-Nortown	23236	0.002	85678	Tagalog	43.7307	-79.4245

Methodology and processing of data

- ▶ import and installation of necessary libraries,
- ▶ Import of data
- ▶ Data wrangling (remove unnecessary columns, remove missing values and white spaces.
- ▶ Analysis and visual presentation of data:
 - ▶ 140 unique neighborhoods.
 - ▶ machine learning using K-means clustering based on the population size, population growth, average income and a language spoken at home other than English
 - ▶ the clusters we also plotted on the map to provide visual information about their geographical relation.
- ▶ **Scoring:** the neighbourhood score was calculated using weighted ranks of the following categories:
 - ▶ number of restaurants per population,
 - ▶ population size,
 - ▶ population growth
 - ▶ average income.
- ▶ The neighborhoods with lowest scores are the best candidates to open a new restaurant.
- ▶ Neighborhood score visualization using choropleth map.

Results - choropleth map of the neighborhood scores



Lower score is better

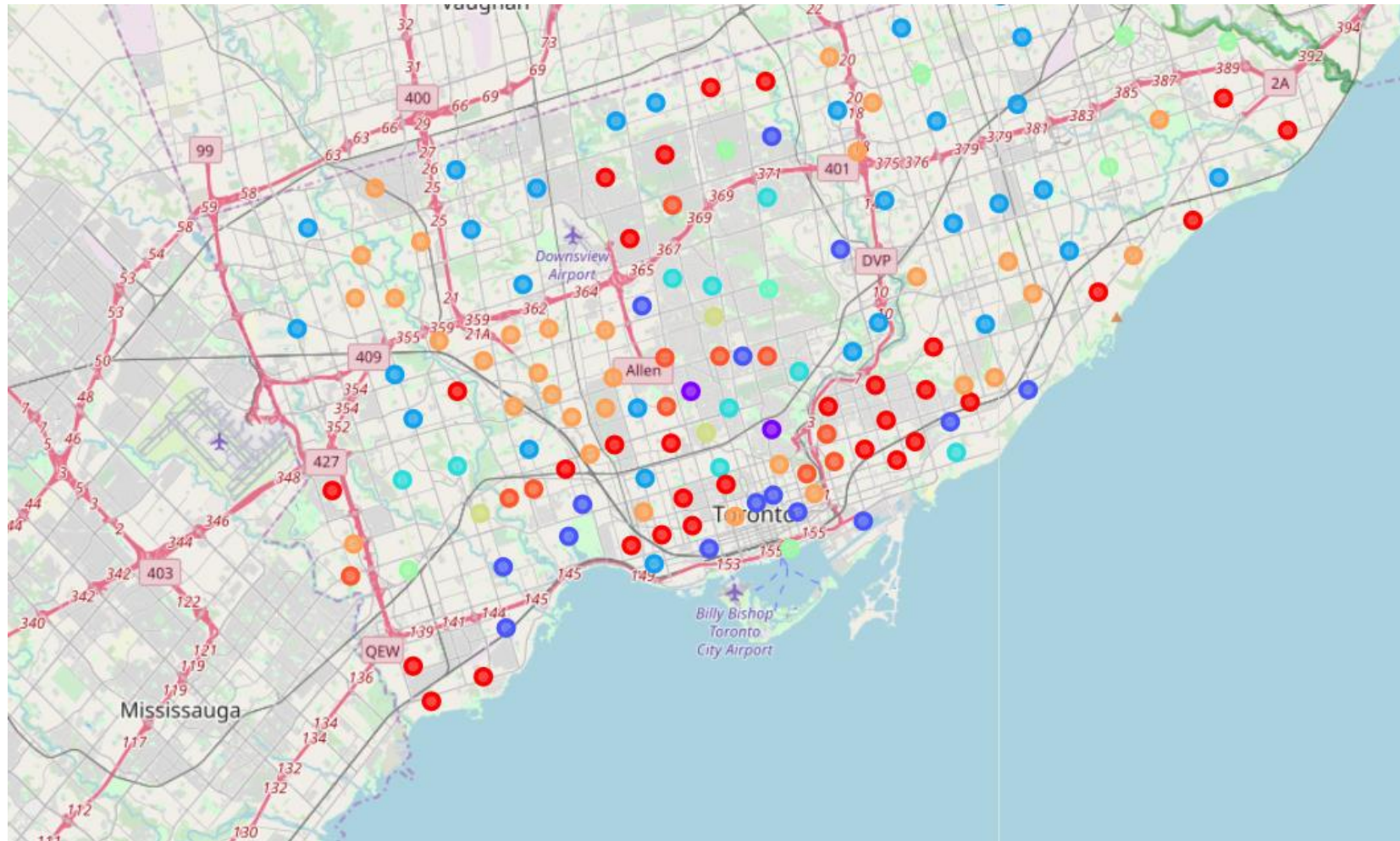
Results of neighbourhood scoring

- The 10 most suitable neighbourhoods using the score.

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

Result of neighborhood clustering

- Neighborhoods assigned to 10 clusters.



Results: Example cluster analysis

- ▶ **Cluster 1** = neighborhoods with small population size and mixed foreign minority.
- ▶ **Cluster 2** = affluent neighborhoods with Chinese population.
- ▶ **Cluster 3** = middle class neighborhoods with mixed population.
- ▶ **Cluster 4** = lower income and primarily Asian minority population.

Conclusions

- ▶ The best neighborhoods to open a restaurant are Waterfront Communities-The Island, Islington-City Centre West, Niagara, Willowdale East and Mimico.
- ▶ The optimal neighborhoods tend to have above average income and larger population size.
- ▶ The best neighborhoods also have 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.