TORONTO – WHERE TO OPEN A RESTAURANT

Ryszard Milewicz

Issued: 31.08.2020

INTRODUCTION

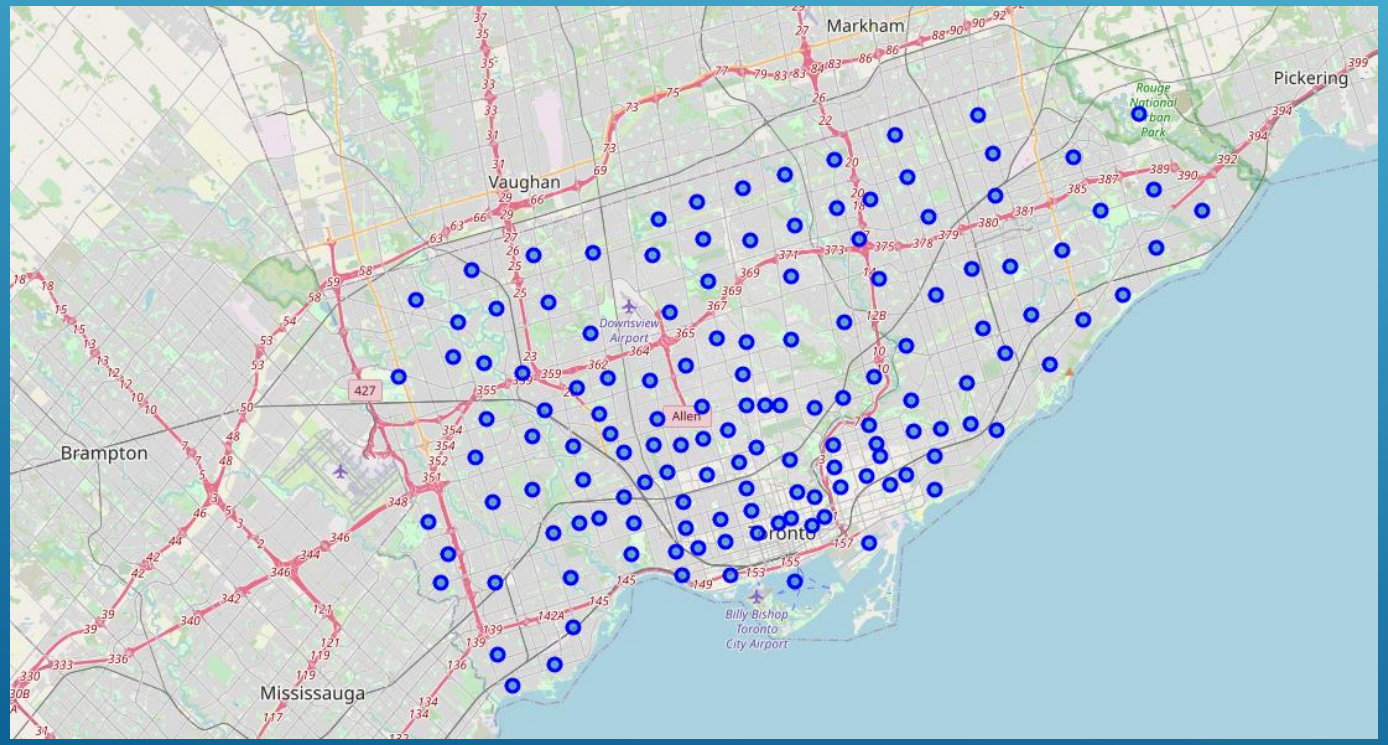
- Problem: where to open a restaurant in Toronto?
- Fact: 30 per cent of consumers' food budget spent on eating out, so there is a market for restaurants
- Fact: there are 140 neighborhoods, which one to choose from?
- Assumption: data science methods should give substantial support to take right decision about restaurant location
- Interested parties:
- - entrepreneurs from other parts of country who are looking to open a restaurant in Toronto,
- - business people from Toronto who are looking to change business area,
- - restaurants owners who are looking to expand their businesses.

DATA SOURCES

- Reliable data sources: Toronto Open Data database: https://www.toronto.ca/city-government/data-research-maps/open-data/
- Demographic data (neighborhood, population, income) from Toronto Open Data database
- Neighborhoods location also from Toronto Open Data database
- Information about existing restaurants in neighborhoods taken using Foursquare API

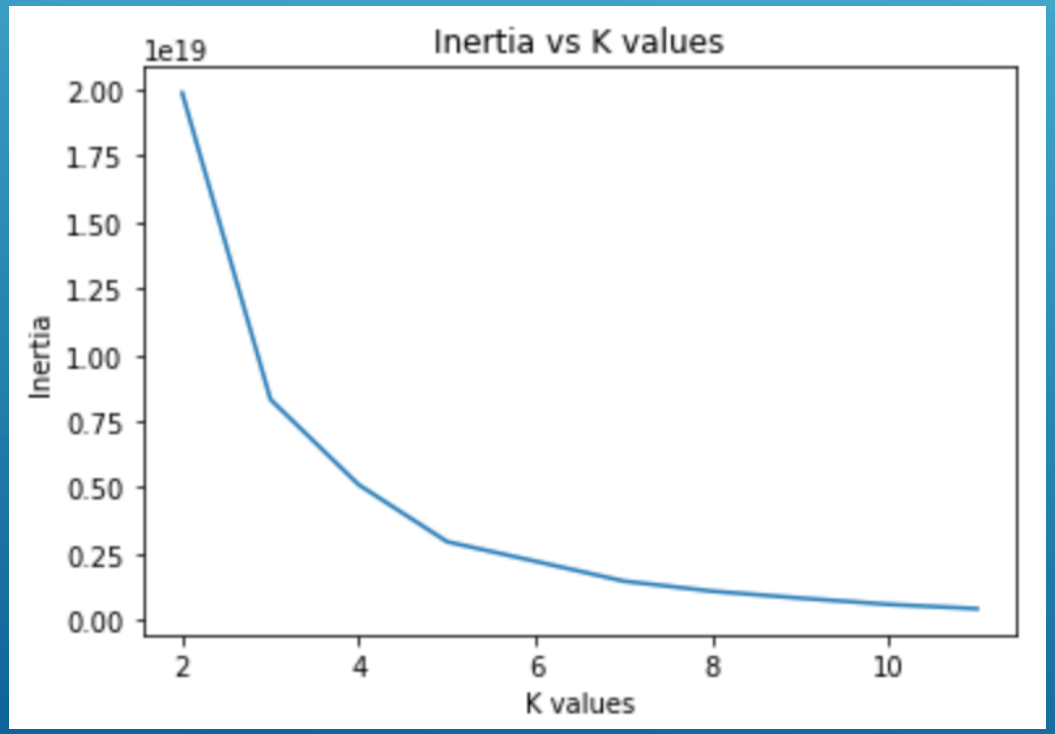
METHODOLOGY

- Prepare dataset:
- Population data: filter out interesting 5 columns out of 2383
- Location data: filter out interesting 3 columns out of 16
- Merge and visualise neighborhoods locations on the Toronto map:



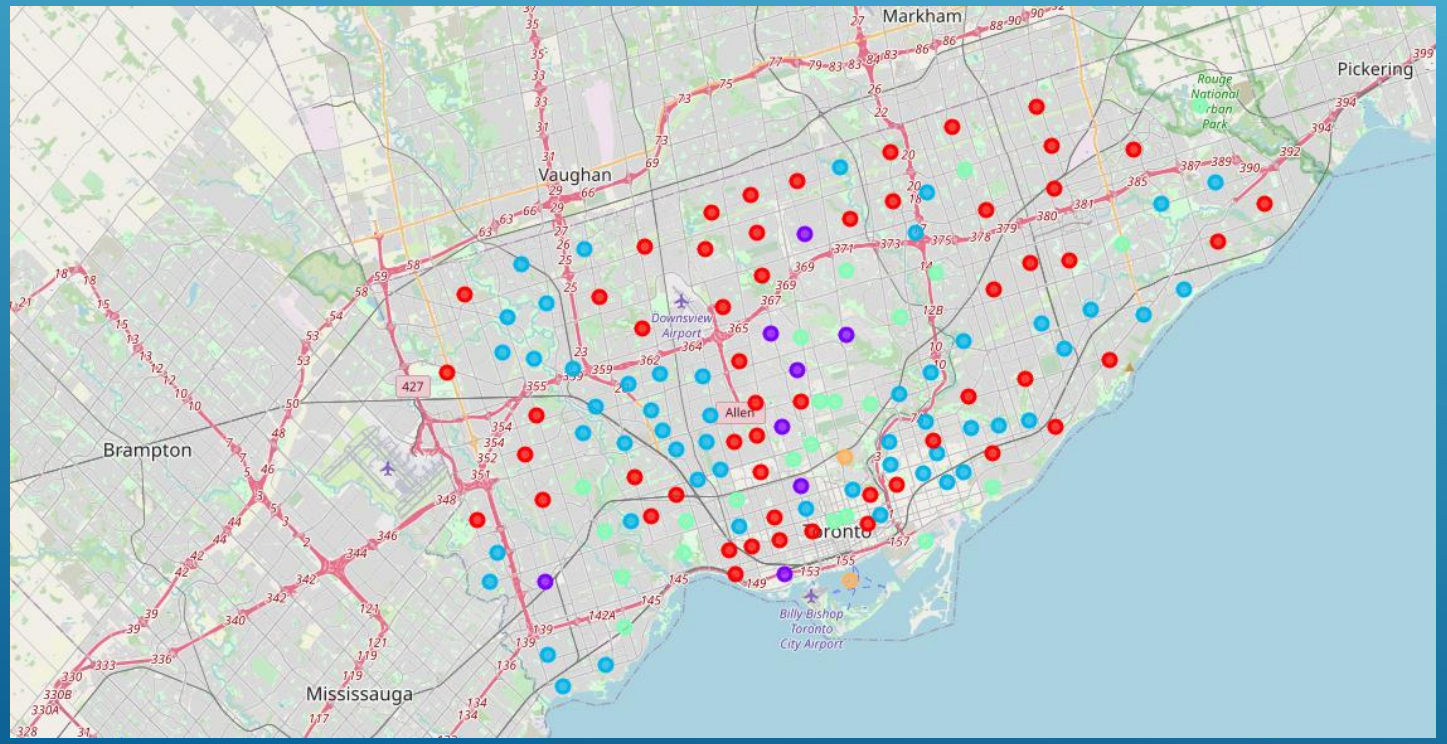
METHODOLOGY

- Add data about restaurants using Foursquare API: from none to 37 restaurants per neighborhood
- K-means clustering was performed with numer of clusters from 2 to 12
- Optimal numer of clusters: 5



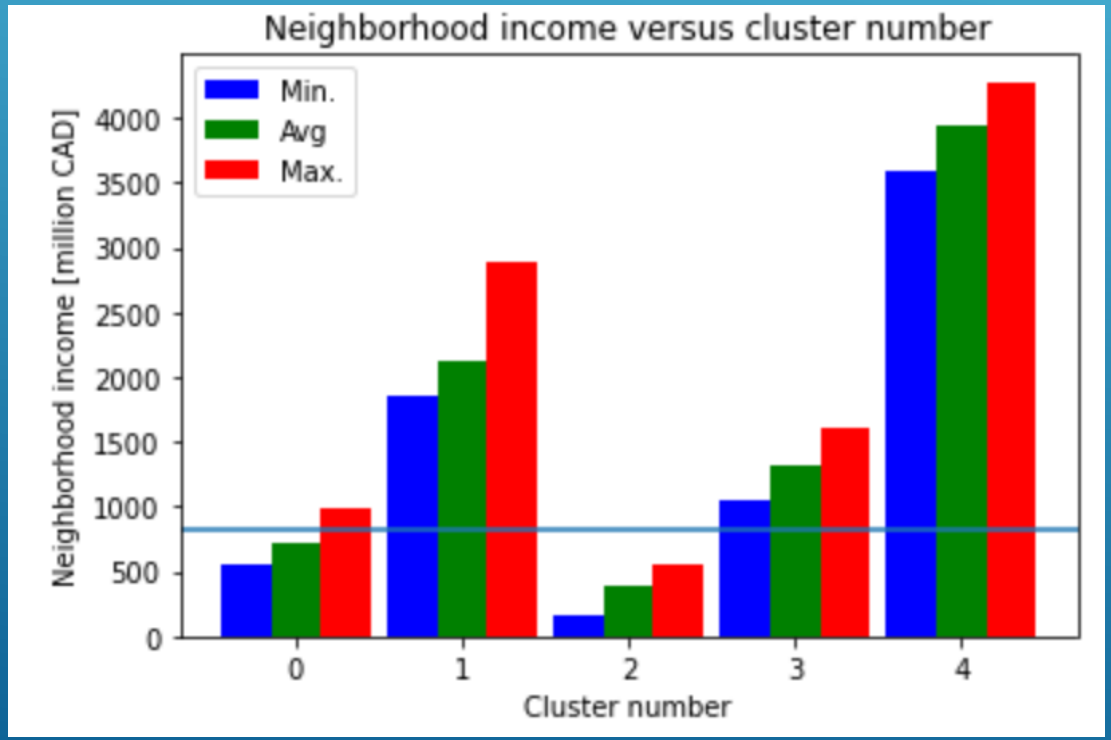
METHODOLOGY

- Visualisation of clustered neighborhoods on the Toronto map:



METHODOLOGY

- Analyze clusters: neighborhood income in cluster.
- Average neighborhood income is 816.5 million $
- Clusters 0 and 2 are below average, so we will concentrate on other clusters as potential customers have more free resources to spend in restaurants.

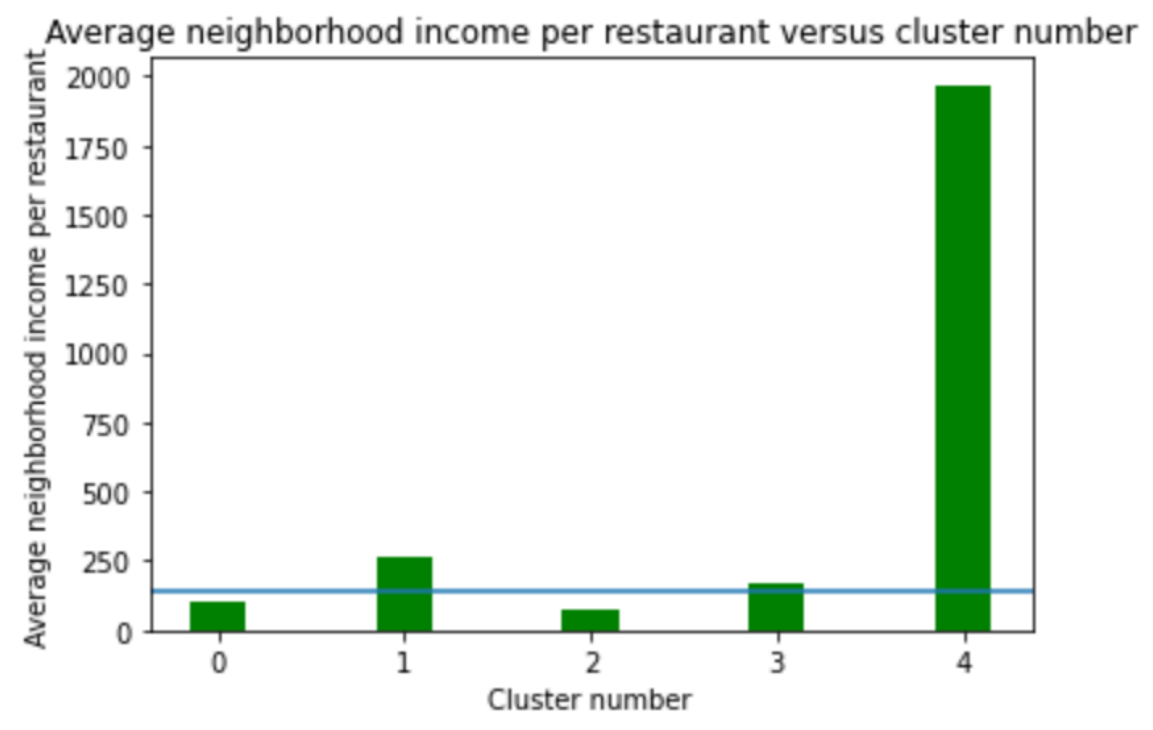


METHODOLOGY

- Analyze clusters: average number of restaurants in neighborhood in cluster.
- Average number of restaurants in neighborhood in Toronto is 6
- Average number of restaurants in neighborhood for clusters 0 to 3 is not far from the average, for cluster 4 is significantly lower – this means that there is less competition.

METHODOLOGY

- Analyze clusters: average neighborhood income per restaurant.
- Average neighborhood income per restaurant in Toronto is mln $
- Clearly cluster 4 is far above the average meaning that potential income of the new restaurant will be also above the average.



METHODOLOGY

- Analyze clusters: examine cluster 4.
- There are only two neighborhoods in cluster 4: Rosedale-Moore Park and Waterfront Communities-The Island.

	Cluster Labels	Neighborhood	Neighbourhood Number	Population	Average income	Neighborhood Income	Longitude	Latitude	Number_of_restaurants
104	4	Rosedale-Moore Park	98	17285	207903	3593603355	-79.379669	43.68282	0
122	4	Waterfront Communities-The Island	77	60620	70600	4279772000	-79.377202	43.63388	5

RESULTS

- Neighborhoods joined as cluster 4 are the potential best places to open an restaurant.
- Cluster 4 contains only two neighborhoods out of 140 in Toronto.
- Rosedale-Moore Park has no restaurants at all, so it is the first choice (no competition, so highest neighborhood income per restaurant if there will be one).
- Looking at average income, it can be quite expensive place, so as the second choice it is recommended to check possibilities in Waterfront Communities-The Island.

DISCUSSION

- Joined data from Toronto Open Data database about Toronto population with venues data from Foursquare API created a dataset which with help of K-mean clustering gave clear recommendation to choose location of an restaurant.
- Risk: last census data comes from year 2016, it is not up to date. Unfortunately, no newer data is available.
- Risk: venues data is changing fast, in four days max. number of restaurants per neighborhood changed from 36 to 37.
- For final selection out of two neighborhoods it would be recommended to check other data, e.g. perform on-site reviews about spending on eating-out.
- It would be also interesting to check why in some neighborhoods there are no restaurants.

CONCLUSION

- Publicly available data sources can give enough data to build solid dataset as the ground for data science methods.
- Local government databases gives rich set of information about the population of Toronto.
- There must be performed much preparation, filtering, cleaning but finally it is possible to get right dataset.
- Foursquare API is helpful to get information about venues, however it is dynamically changed and updated.
- Filtered and joined data is analyzed using data science methods to obtain answer to real life question like “where to open a restaurant”.