The Battle of Neighbourhoods
Location analysis for new Russian Restaurant

by Harri Hokkanen

Intoduction:

Toronto is the Canada's most populated metropolitan area by 2011 Census havin population of 5,583,064. The mild weather in Ontario is challenged only by Montréal in Quebec. Toronto has average 105 days with a maximum temperature between 20 and 30 °C (68 - 86 °F) [https://www.currentresults.com/Weather-Extremes/Canada/mildest-temperature-cities.php]. Land of ooportunity and good weather has gathered very diverse population having neighbourhoods as Chinatown, Corso Italia, Little India, Kensington Market, Little Italy, Koreatown and many more. Entrepreneurs can find meny opportunities to their businesses in downtown.

Ontario has other less known feature in its history. North from Toronto is mining sity sudbury and originally 20% of the population there were from Finland and most of the Ontario farms were first established by Finns. They were the people who came from the land of forests and were not afraid to settle outside the cities and start farming in middle of nowhere.

The objective is to make decision whether it is a good idea to have Russian restaurant in Toronto and secondly what would be best location. Now you ask why not Finnish restaurant but I have to say that excluding berries, mushrooms, fish and wild game the Finnish cuisine is not so interesting. Russian cuisine among Italian, Indian, French, Chinese and Tex-Mex are most popular in Finland. Anyhow East Finland and Karelia were part of Russian Empire from beginning of 1700 and West Finland from beginning of 1800. Karelia was joined with Finland 1917 after Soviet Revolution and Finnish independence.

Did you know that Ontario has more bears in woods than Finland has moose in same area? Well, didi you know that you can not have a gun in forest in Ontario after dark outside of hunting season but in Finland you can have? I have a picture from Ontario outback whith seven blackbears (we have brownboear in Finland) in one frame beacus two were trying to circle to our back. That was in Canada, Ontario and no right to protect yourself with a gun. "Yeah!"

Target Audience:

This project is benefitial to following groups:

- Entrepreneurs who want to open an Russian Restaurant in Toronto
- Russian, Ukrainian and Finnish travellers and migrants who want to find neighborhoods with lots of option for Eastern European restaurants
- Business Analyst or Data Scientists who can with few tweaks change the target parameters and create similar report to other cultures and business types.

Data acquisition and cleaning:

Data Sources

The View From Above

- Wikipedia page "List of Postal code of Canada: M" to postal code, borough & the name of all the neighborhoods present in Toronto. Data is extracted from website and the data matrix is cleaned from html with BeatifulSoup4 resulting dataframe which we can manipulate.

- Geographical coordinates of the neighborhoods from CSV file from https://cocl.us/Geospatial_data is read to dataframe to be connected with previous dataframe to conclude usable dataframe with postcodes, boroughs, neighbourhoods, longitudes and latitudes.

- Wikipedia page "Demographics of Toronto" to identify the neighborhoods which are densely populated with Russians and ethnic Russians like Kievan Russ and Bielarussians. History books have more info on that but Russian Patriarch did not talk to Roman Pope in 1000 years so there are some historical issues. Like Greek who are from down of times and Italians (West Rome) was dominating power 2000 years ago. Indians have longer history than we have human collective memory…

- Foursquare's explore API to get venues in Toronto and collected their names, categories, locations (latitude and longitude).

# Methodology == Data Cleaning and Manipulation

- First we get Toronto Neighborhoods Table from Wikipedia

  - Clean the collected web page and extract PostalCode, Borough, and Neighbourhood to Pandas dataframe. You have to have extra attention to writing of cell topics like "Postal Code", "PostalCode" and "PostCode". Topicks have to be unified during the process.

  - Process the dataframe to have non empty cells. Filter all Postal codes with NaN values and Neighbourhoods with NaN values. Empty cells and zero values are cleaned.

  - with geospatial data and dataframe to include latitude and longitude of the postalcode

	PostalCode	Borough	Neighbourhood	Latitude	Longitude
0	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790
1	M4P	Central Toronto	Davisville North	43.712751	-79.390197
2	M4R	Central Toronto	North Toronto West	43.715383	-79.405678
3	M4S	Central Toronto	Davisville	43.704324	-79.388790
4	M4T	Central Toronto	Moore Park,Summerhill East	43.689574	-79.383160

Resulting dataframe including location data to postal code

From there we end up with map of Boroughs in Toronto
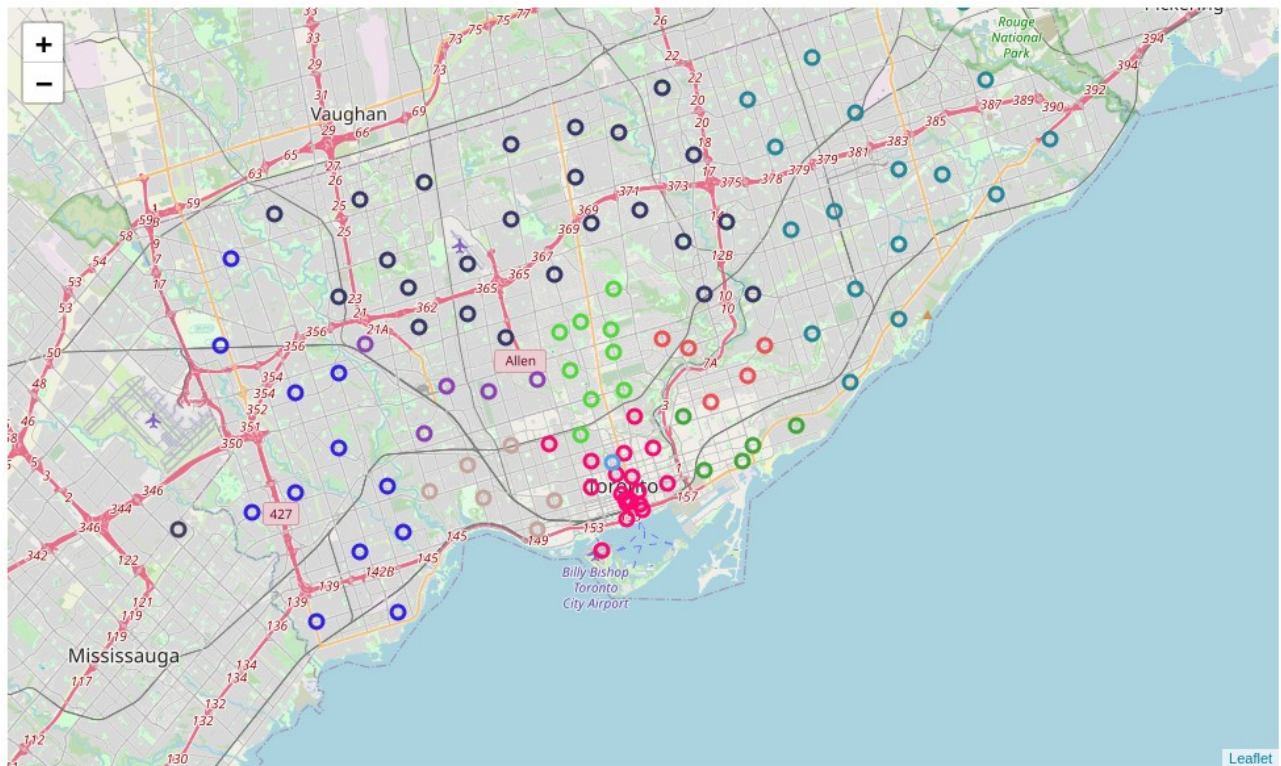
- Scrape the distribution of population from Wikipedia

   - From "Demographics of Toronto" all the neighborhoods in Toronto few had considerable amount of Russian and Ukrainian (part of Russia about 1000 years before collapse of USSR) people and identify the densely populated neighbourhoods. *https://en.m.wikipedia.org/wiki/Demographics_of_Toronto#Ethnic_diversity*

- For analysis what is happening in Toronto we use Foursquare data and sort it to neighbourhoods to dataframe for further analysis. This is nice when you have very popular subjects but get more or less usefull to less general targets.
   Scraping or not assigned values

- From Foursquare we collect all Toronto data using 10km radius and 200 event limit. This comes from preliminary results for "Russian Restaurant" == "NaN" → "Eastern European Restaurant" == "Exists" . Like in every investigation when you run in deadend but you have gut feeling and basic knowledge that the results are wrong you change the analysing method and this case as usually I changed to paper, whiteboard and pen to add-hoc analytics. This also resulted to investigate the data section of Toronto demographics.

   So we got lots of venues which we had to sort out by grouping them to Neighbourhoods and the getting uniques venues. To get the venue gategory which we needen due the

missing "Russian Restaurant" resulted a list where "Easten European Restaurant" is part of. Well, qoite large group of cuisines; just like putting all Western European cuisines to one category and USA to one single category. !?!?!?

• Merging geographical coordinates to the neighborhoods. Then we can start clustering the data. First to search K values we get graph:

And then we analyse distortion score for Kmeans clustering with result of 2.

So we get two clusters which starts to rise eyebrows.

Checking how meny neigbourhoods are per cluster we see clearly that the our targets are in limited area.

Distortion Score Elbow for KMeans Clustering

How many Neighborhoods per Cluster
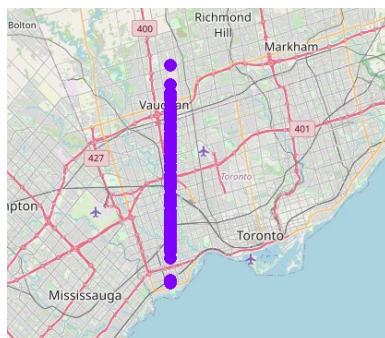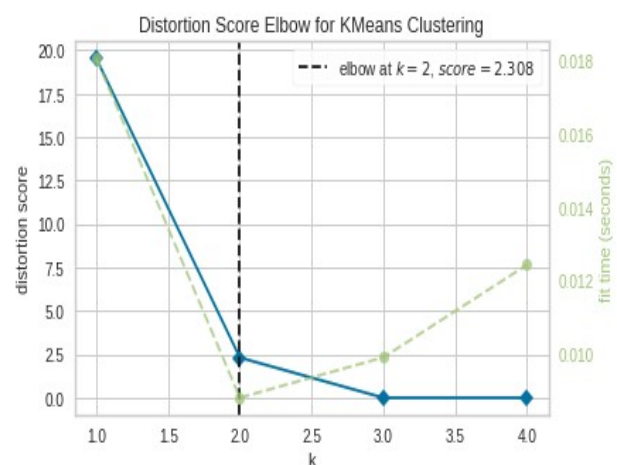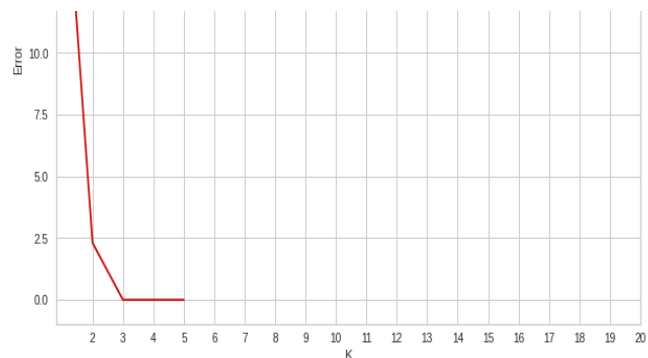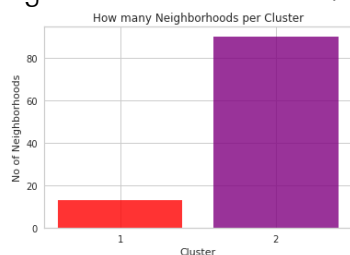
Next step is to look where the "Eastern European Restaurants" (==EER) are located. We get 24 restaurants from venues where total number of restaurants is 150 with in the sample and Foursquare database. So there are 16% of EER of total restaurants but we have only two clusters.

The maaping the cluster got very eery but actually other data support because all EER:s are located to two Borought where Russians and Ukrainian mainly live.

• Next I did a check and changed to target to my favourite "Indian Restaurant" (our family loves them on London) and got result of cluster distribution:

Number of Indian Resturants for each Neighborhood in Toronto City
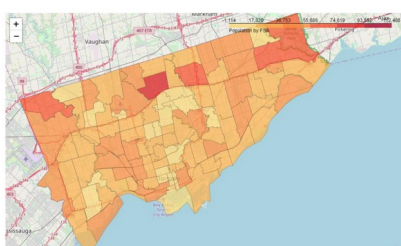
And that gives totally different distribution. Next I ran the same with European and American easy food favourite (not fast food) "Italian Restaurant" and the results
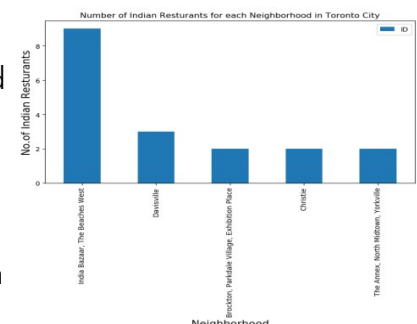
were similar to "Indian Restaurant" cross examination. Results are not conclusive and not definitive.

Results

From data we have from "Demographics of Toronto" section "Ethnic diversity" These are the boroughs where Russians and Ukrainians live:

- Russian 74,465 : 2.8 % of city : York Centre 9.5 % of locals
- Ukrainian 72,340 : 2.7 % of city : Etobicoke Centre 8.1 % of locals

All Eastern Eruopean Restaurants are located in these two boroughs. Comapred to as example Italian and Indian restaurants they are more scatterd around the city to multiple boroughs. This bring up reality that either existing restaurant owners like to be within their own people are there is not much requirement for eastern european cuisine among Toronto residents and tourists. This would require more marketing survey.

Discussions

In this scenario basic analytics run in trouble because of limited data and very biased results. Analytics formulas run in to errors due to too few clusters or sample size. These are easy to fix but would need more code to measure parameters to make Notebook or similar analytics generic. This "Russian" or "EER" analytics gave out

```
         if index is None:
---> 80          index = extract_index(arrays)
    81      else:
    82          index = ensure_index(index)

~/anaconda3/envs/tf-gpu/lib/python3.8/site-packages/pandas/co
    399          lengths = list(set(raw_lengths))
    400          if len(lengths) > 1:
---> 401              raise ValueError("arrays must all be
    402
    403          if have_dicts:

ValueError: arrays must all be same length
```

shortcomings in this kind of analysis and programming practises. If you are workin on ly in S&P500 landscape you will not run in to problems with these formulas and the segments. But with SME level data you end up with *"random segments of code"* with unsuspected results.

Conclusion

So where to put the restaurant or not? With this data I would rocommend to locate new Russian restaurant to city center where most of the venues and people are moving. Also those locations draw most of the tourist traffic. As e competition perspective diversification between other major cuisines is distinctive enough. If new restaurant would be put to areas where existiing EEA restaurants are then competition is agains similar entrepreneurs.

As we see from this analysis there should be new rounds of analytics after consulting with the restaurant ownew and how their first reaction is to this situation.