The Battle of Neighbourhoods

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.

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.

- Geographical coordinates of the neighborhoods from CSV file from https://cocl.us/Geospatial_data

- Wikipedia page "Demographics of Toronto" to identify the neighborhoods which are densely populated with Finnish (if any statistical significance) Russians.

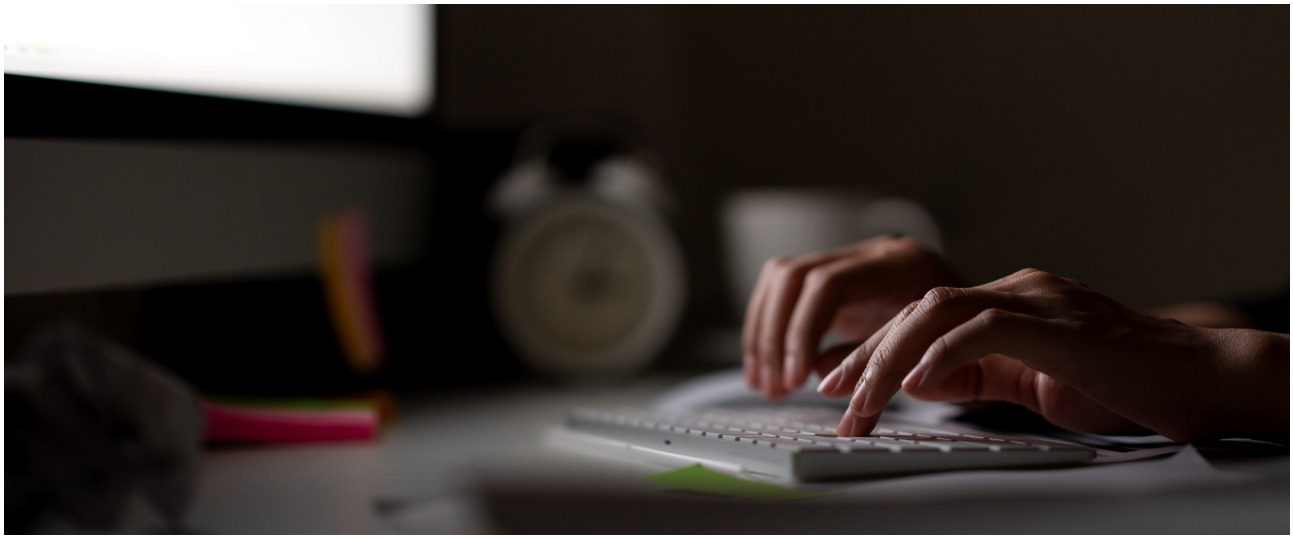
- Foursquare's explore API to get venues in Toronto and collected their names, categories locations (latitude and longitude).

Data Cleaning:

- Toronto Neighborhoods Table from Wikipedia
  - Clean the collected web page and extract PostalCode, Borough, and Neighbourhood to Pandas dataframe
  - Process the dataframe to have non empty
  - with geospatial data and dataframe to include latitude and longitude of the postalcode

- Scrap the distribution of population from Wikipedia
  - From “Demographics of Toronto” all the neighborhoods in Toronto below given neighborhoods only 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
- Foursquare data of neighbourhoods to dataframe. This is nice when you have very popular subjects but get more or less usefull to less general targets.
  Scraping or not assigned values
- Merging geographical coordinates to the neighborhoods

About Groud Research and Data Manipulation



I have to comment on this workflow… but that will come later...

Afterword to Linux and Open Source from experince since 1996...

… there is a part here missing (published later) about Ubuntu and RHEL/CentOS/Rocky Linux…

Fedora is excellent to use on newest and fastest hardware you use on convertible or detachable devices like HP Spectre 360 or HP Elite X2; you can ditch your Android tablet and also finally get mobile phone freedom with SailfishX and freedom without corporate spying.

Data analysis is just one part of the problem; you have to have other skills like business and analytical skill to come to potential suggestion.