



Analysis of location for Russian Restaurant in Toronto



Coursera Data Science Capstone Project by Harri Hokkanen

Intoduction

Analysing suitable location for new Russian restaurant in Toronto

- Restaurant owners prefer good knowledge of competitive situation of new location
 - Places where there are lots of venues are source of customers
-
- Areas where similar businesses reside have more competition and differentiation is more difficult
-
- Cuisines from other main streams attract people and are not direct competition

01



Target Audience

This analytics is beneficial 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 Sources

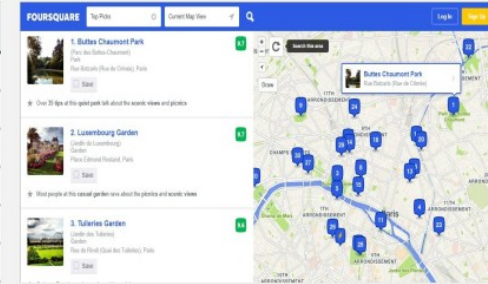
Data is read from multiple sources and converted to Pandas dataframes for manipulation and analysis.

Postcode	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Harbourfront
M5A	Downtown Toronto	Regent Park
M6A	North York	Lawrence Heights
M6A	North York	Lawrence Manor
M7A	Queen's Park	Not assigned
M8A	Not assigned	Not assigned

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 BeautifulSoup4 resulting dataframe which we can manipulate.

	A	B	C
1	Postal Code	Latitude	Longitude
2	M1B	43.8066863	-79.1943534
3	M1C	43.7845351	-79.1604971
4	M1E	43.7635726	-79.1887115
5	M1G	43.7709921	-79.2169174
6	M1H	43.773136	-79.2394761

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.



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

03.



Methodology

Data Cleaning and Manipulation

- 1) Data from Toronto postalcodes including borough and neighbourhood are combined with geospatial data to have postcodes with latitudes and longitudes.
- 2) Foursquare data has location data.

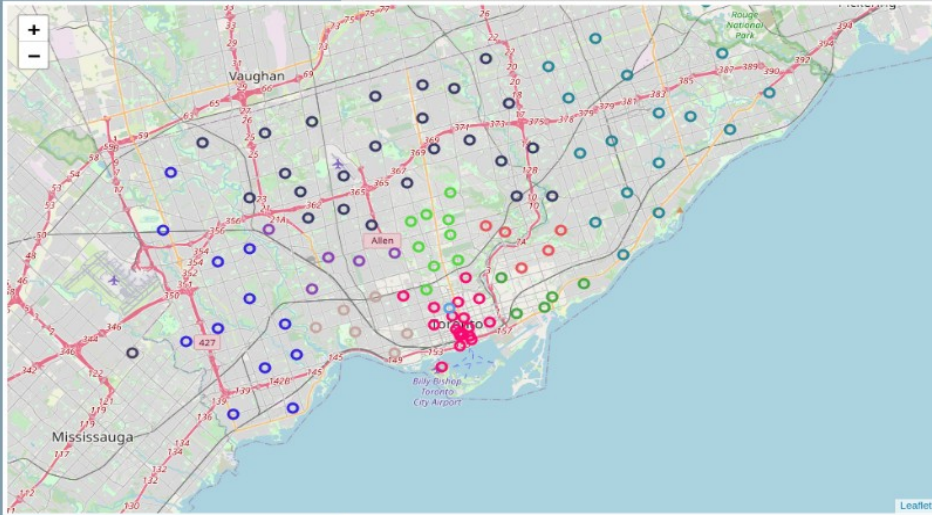
04.

HTML pages are processed with BeautifulSoup4 to select the table with data and inserted to Pandas dataframe.

	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

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Lawrence Park	43.72802	-79.38879	Sherwood Park	43.716551	-79.387776	Park
1	Lawrence Park	43.72802	-79.38879	Sheridan Nurseries	43.719005	-79.400500	Flower Shop
2	Lawrence Park	43.72802	-79.38879	Alexander Muir Memorial Gardens	43.721315	-79.400822	Garden
3	Lawrence Park	43.72802	-79.38879	De Mello Palheta Coffee Roasters	43.711791	-79.399403	Coffee Shop
4	Lawrence Park	43.72802	-79.38879	Himalayan Java	43.713486	-79.399811	Café

Map of Boroughs in Toronto

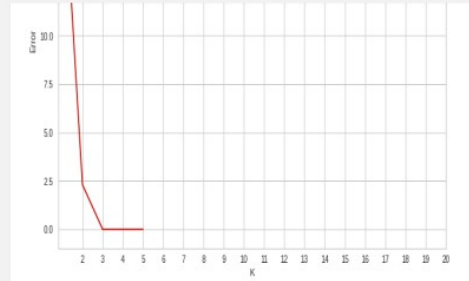


- Using folium map we get interactive map of boroughs and neighbourhoods.
- From Foursquare data we find venues that are “Eastern European Restaurant” with locations

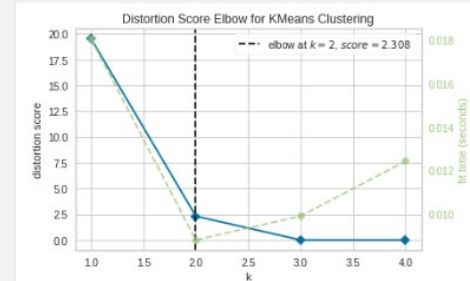
Clustering the data

With changing parameters we got also comparative data from Italian and Indian restaurants to improve the finan conclusion.

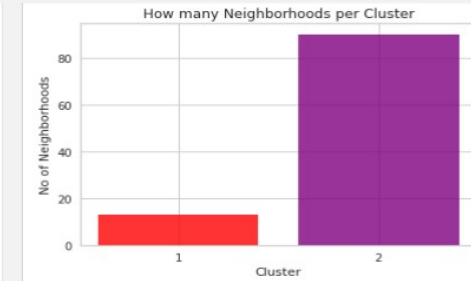
06.



First to search K values we get graph



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



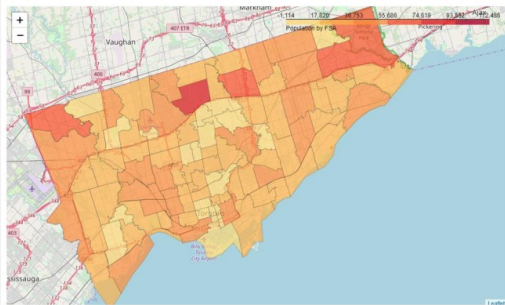
Checking how many neighbourhoods are per cluster

Results

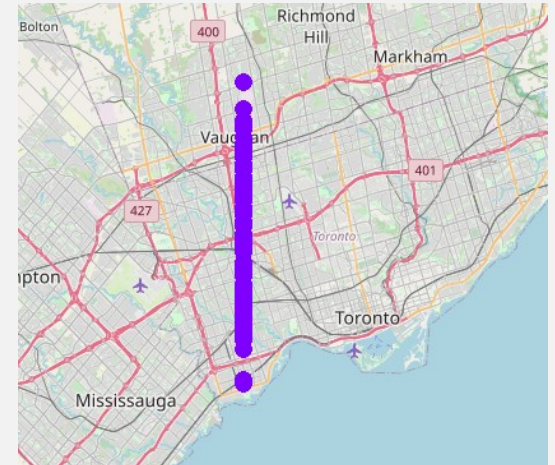
From data we have from "Demographics of Toronto" section "Ethnic diversity" These are the boroughs where Russians and Ukrainians live:

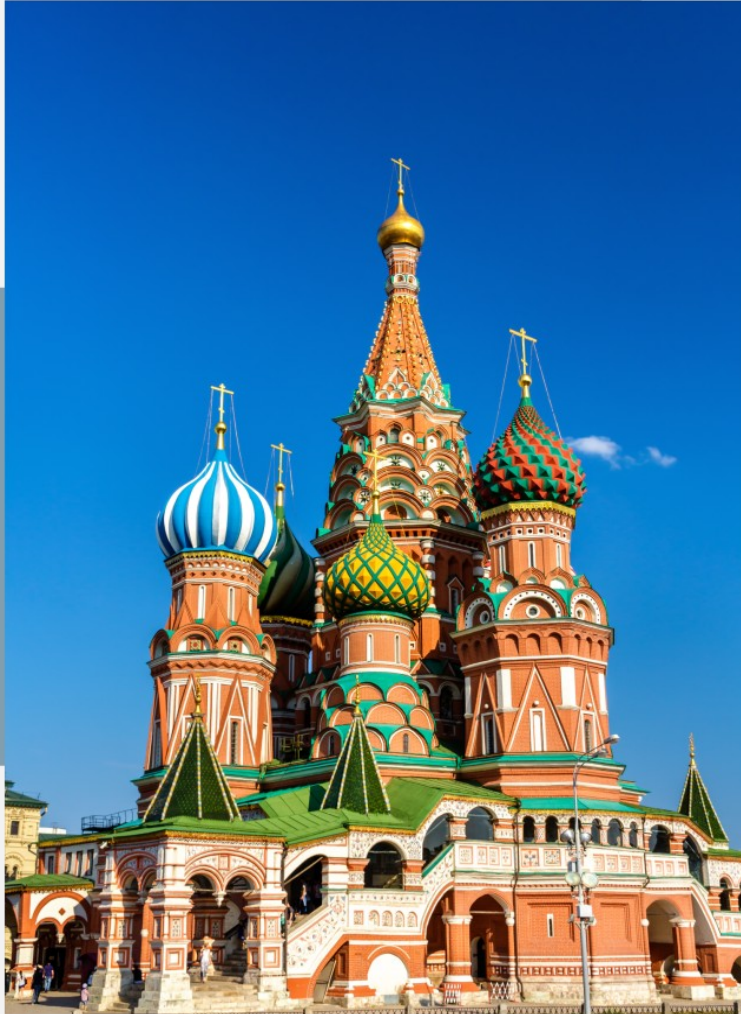
- 1) Russian 74,465 : 2.8 % of city : York Centre 9.5 % of locals
 - 2) Ukrainian 72,340 : 2.7 % of city : Etobicoke Centre 8.1 % of locals
-

- “Eastern European Restaurants” all 24 of them are located in these Boroughs
- As comparison in collorleph map Indian and Italian restaurants are scattered to wider are.
- There is no competition for Russian restaurant outside of York and Etobicoke.



Restaurants are in two boroughs and data is very limited.





Discussions

Additional Question Risen From
Presentation and Analysis ?

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. Comparative analytics by changing restaurant criterion to Italian or Indian gave additional information of restaurant landscape in toronto.

Conclusion

So where to put the restaurant or not?

With this data I would recommend 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.



Project documents and notebooks can be found:
https://github.com/harrioh/Coursera_Capstone