LOCATION FINDER

1.Introduction

1.1 Background

Toronto is a great city where cultures and people cross path with variety of language, foods, costumes and tradition. One can find restaurants offering cuisines ranging from American to Indian, Chinese to Japanese, Korean to Thai.

1.2 Problem statement

Hueng-min-son a football star want to open a Korean restaurant. He is so passionate about Korean food but it is not an easy task to do so because opening a restaurant involves a lot of capital and knowledge. for a restaurant to do well ,a lot of factor come to play and location is an important factor this project aims to predict a few locations that he can choose to open a Korean restaurant

2.Data acquisition and cleaning

2.1 Data acquisition

Neibourhoods data , borough data,postal codes of Toronto were scraped from :

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

the data was later merge with geospatial data over the postal code restaurant data was fetched over foursquare api

Postal Code	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue, Humber Valley Village
M1B	Scarborough	Malvern, Rouge
M2B	Not assigned	Not assigned
M3B	North York	Don Mills
M4B	East York	Parkview Hill, Woodbine Gardens
M5B	Downtown Toronto	Garden District, Ryerson
M6B	North York	Glencairn
M7B	Not assigned	Not assigned
M8B	Not assigned	Not assigned
M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale

2.2 data cleaning

the scraped data was first cleaned. Empty value cells and cell with the value “not assigned” were dropped. The dataframe was merged with the dataframe of geospatial data. With that the co-ordinates were assigned to every neighborhood. The cell having borough names and neighborhoods names were allotted borough names under the neighborhood column.

	Postal Code	Borough	Neighbourhood_x	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
...	...	...	...	...	...
98	M9N	York	Weston	43.706876	-79.518188
99	M9P	Etobicoke	Westmount	43.696319	-79.532242
100	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...	43.688905	-79.554724
101	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...	43.739416	-79.588437
102	M9W	Etobicoke	Northwest, West Humber - Clairville	43.706748	-79.594054

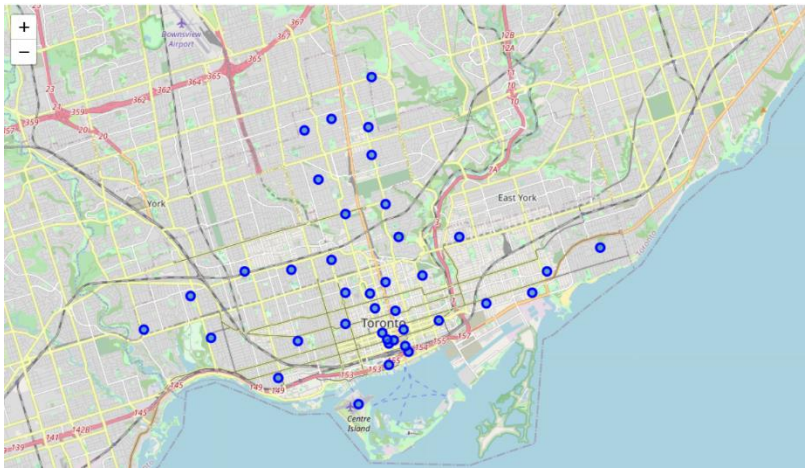
3.Methodology

3.1 Feature selection

Since our primary focus area is the restaurant. We filtered out the data returned by Foursquare API to include the venue category having the term “ restaurant”. Thus the data which were not required (historical monument, bakery, cafe etc.) were dropped from the dataset

3.2 Exploratory data analysis

3.2.2 Plotting of neighborhoods on a map



3.2.3 Onehot encoding

Onehot encoding was carried out on the data to create dummy variables for the venue category as it is a categorical variable

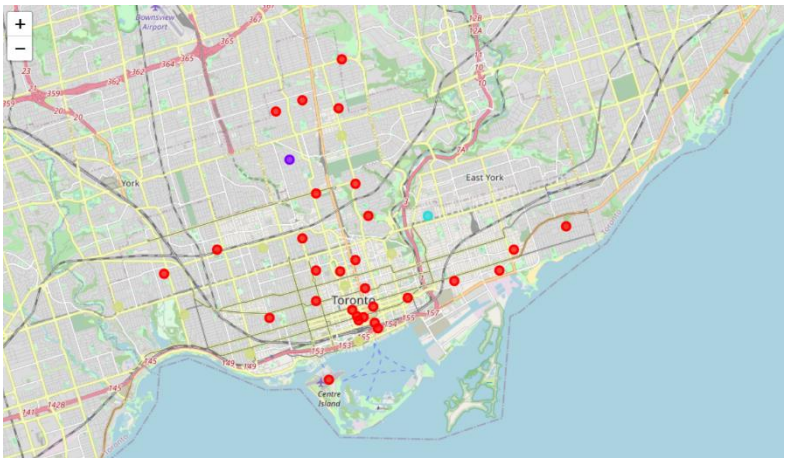
	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	...	Theme Restaurant	Tibetan Restaurant	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1600	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
1601	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
1602	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
1603	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0

3.2.3 create the most common venue base on restaurant category

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Berczy Park	Seafood Restaurant	Restaurant	Greek Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	Sushi Restaurant	Japanese Restaurant	French Restaurant	Comfort Food Restaurant	Eastern European Restaurant
1	Brockton, Parkdale Village, Exhibition Place	Italian Restaurant	Restaurant	Vietnamese Restaurant	Doner Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
2	Business reply mail Processing Centre, South C...	Fast Food Restaurant	Restaurant	Vietnamese Restaurant	Doner Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Falafel Restaurant	Ethiopian Restaurant	Eastern European Restaurant
3	CN Tower, King and Spadina, Railway Lands, Har...	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant	Eastern European Restaurant
4	Central Bay Street	Italian Restaurant	Indian Restaurant	Portuguese Restaurant	French Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Korean Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	Modern European Restaurant

3.4 classification of restaurant using KMeans clusterings

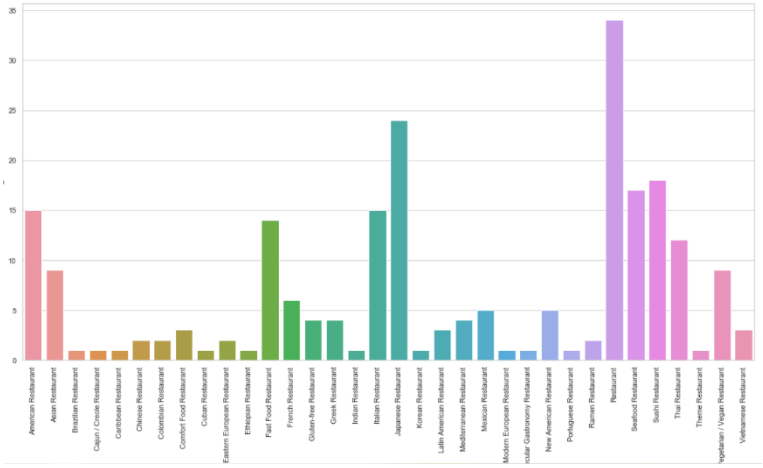
the restaurant were classified using KMeans clusterings. They were cluster based on their density. There will be 5 cluster



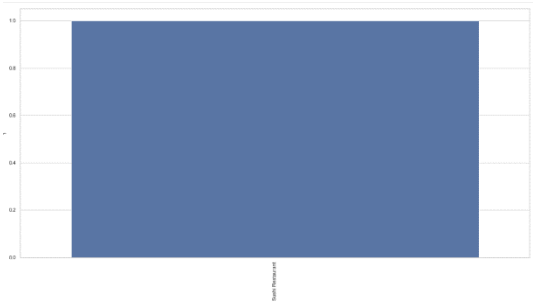
3.5 Analyzing the result

Our most suit location contains 3 criteria
1 high density of restaurant number
2 low density of Korean restaurant number
3.high density of similar cuisine such as Japanese, thai , etc.
3.5.1 plot bar chart to analyze which cluster match the criteria

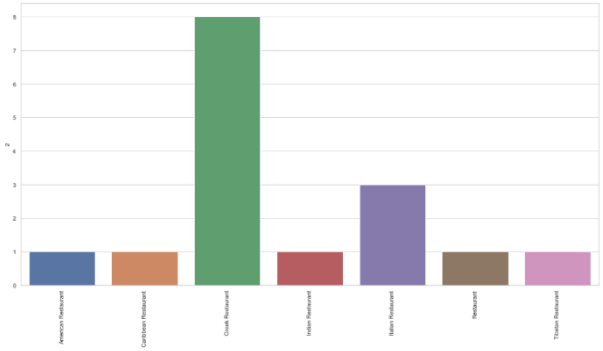
3.5.1.1 cluster 0



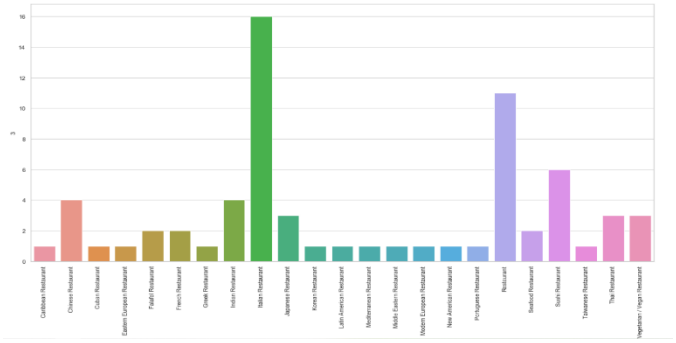
3.5.1.2 cluster 1



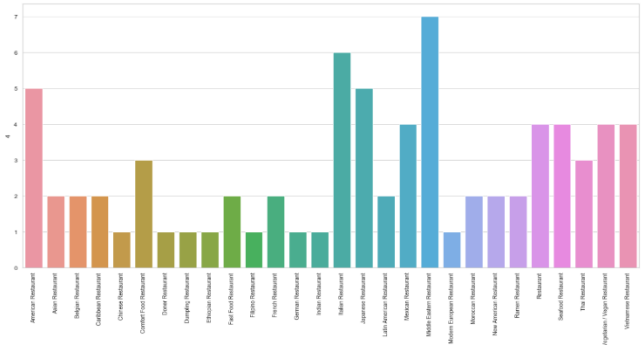
3.5.1.3 cluster 2



3.5.1.4 cluster 3



3.5.1.5 cluster 4



4. Conclusion

As shown in the report the most suit cluster for Korean restaurant based on given criteria is cluster 0 because it contains total of 235 restaurants with 25 of Japanese restaurants and only 1 korean restaurant. In the area shown below is the best location for open Korean restaurant