Restaurant Recommendation

A. Introduction

A.1. Description and Discussion of the background

Opening his/her own company, shop or hotel has always been a dream for many people. In case of hotel or shop, the location of these plays a vital role. The profit and growth depends upon the selection of the location and the target customers. While selecting the location the main criteria to consider is that the shop should be easily visible for the customer and it should be a place where lot of people come and go in their day to day life. The location of the restaurant, quality and certain other factors determine the profit of the business.

As being said, there are number of things to consider besides location such as

1. Target Customer
2. Locality
3. Operation time
4. Visibility

In this analysis it was assumed that the new restaurant targets customers from entertainment centres such as theatres, mall, park etc..,

A.2. Data Description:

The data need for the analysis includes the location of entertainment centres and the location of the already existing restaurants.

- The location of both restaurant and entertainment centres can be obtained using foursquare API.
- The problem with the foursquare API is that a request is limited to 50 outputs (i.e., we can obtain only 50 restaurants or entertainment zones for a coordinate. Thus a grid of 6 x 6 km with Toronto city hall as centre was constructed.
- Multiple calls are performed for obtaining both restaurant and entertainment datasets and cleaned to removed duplicates caused by overlapping of the location while calling foursquare API.

B. Methodology

Since the calls that can be made with foursquare API per day are limited, the cleaned data is stored in a csv file for easy retrieval of the data. The cleaned data contain name, latitude, longitude, category, id and venue_id of the restaurant or entertainment centre. The id and venue_id can be used if the particular venue details are needed but for analysis it will not be needed.

The sample data of the dataframe with contains entertainment details are shown below.

	name	categories	lat	lng	id	venue_id
0	Ontario Place Cinesphere IMAX	Movie Theater	43.627817	-79.418023	4d544f6c16a6b60c4f5c3ef8	NaN
1	Ontario Place	Theme Park	43.630577	-79.415359	4c10206ae7d295216fbcd7b0	NaN
2	Waterfall Stage	Music Venue	43.628496	-79.420327	4dff4b7b62e1651eb9da6cf1	NaN
3	The Ghost of Exhibition Stadium	Stadium	43.631758	-79.416776	4bac5037f964a520edef3ae3	NaN
4	Echo Beach at Molson Canadian Amphitheatre	Music Venue	43.629332	-79.415072	57ac2e6e498ee8e345a5e96e	NaN
5	Wilderness Adventure	Theme Park	43.628525	-79.421142	4c7d5202d598a0939284c262	NaN

The sample dataframe which contains the restaurant details is shown below.

	name	categories	lat	lng	id	venue_id
0	Bar & Restaurant @ NYC	Bar	43.633246	-79.400224	4d9e80a736818cfadff1f500	NaN
1	Manila Grill Restaurant	Restaurant	43.636230	-79.411454	5dfbdcd482265d0008c050bb	NaN
2	Pearl Restaurant	Asian Restaurant	43.636630	-79.408052	50945a6be4b0ad3a94e01f3b	NaN
3	Rozu Restaurant Lounge	Eastern European Restaurant	43.636899	-79.406064	4d8a998ebc848cfa4c5bde2b	NaN
4	Lusso Restaurant and Bar	Lounge	43.638094	-79.380666	4ad4c05df964a52027f620e3	NaN
5	Gonoe Sushi Japanese Restaurant	Sushi Restaurant	43.639014	-79.385914	501d4d55e4b0825080212f3c	NaN

Folium, which can be used to build interactive maps, is used to visualize the location of the restaurants and the entertainment centres. The interactive map provides information about the distribution of the data. An image of the map is shown below.
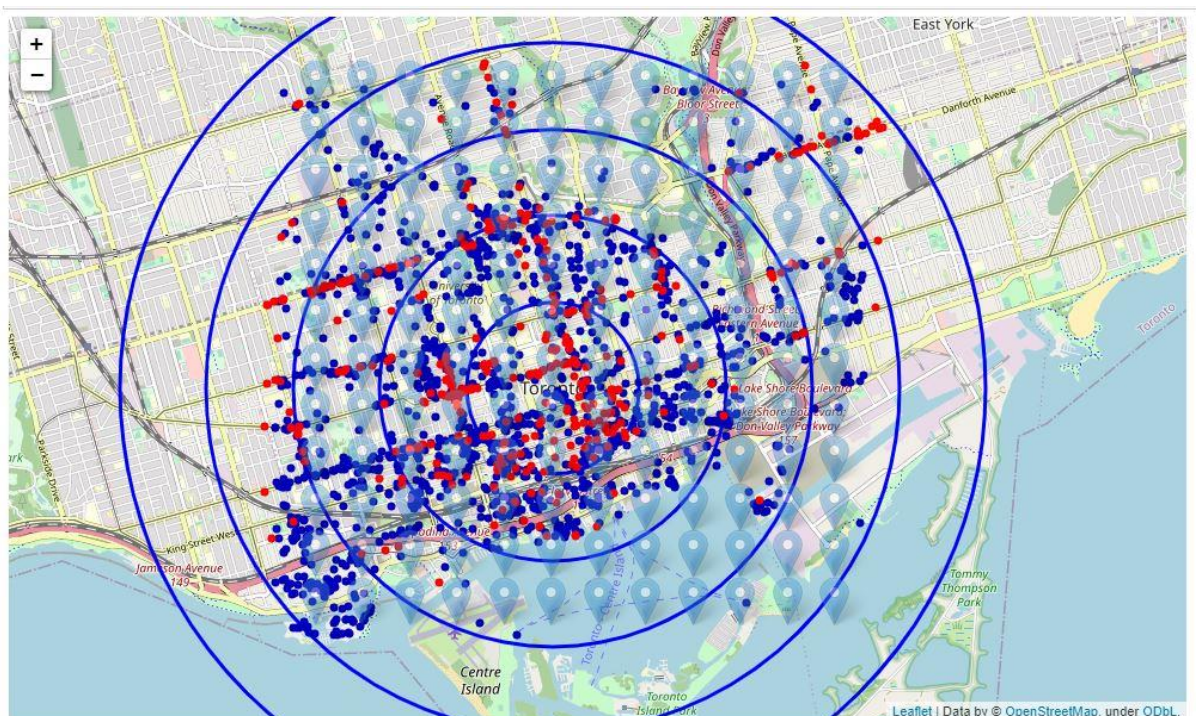

The red dot indicates the position of the restaurant and the blue dot indicates the presence of the entertainment centres. The circles indicate the 1km, 2km to 5 km from the centre which facilitates easy measure of distance from the centre. The Blue Marker indicates the location where the foursquare API was called for the collection of data.

For preliminary enquiry of the distribution of the data more clearly, a heatmap of entertainment centre is drawn using folium. The map is shown below.
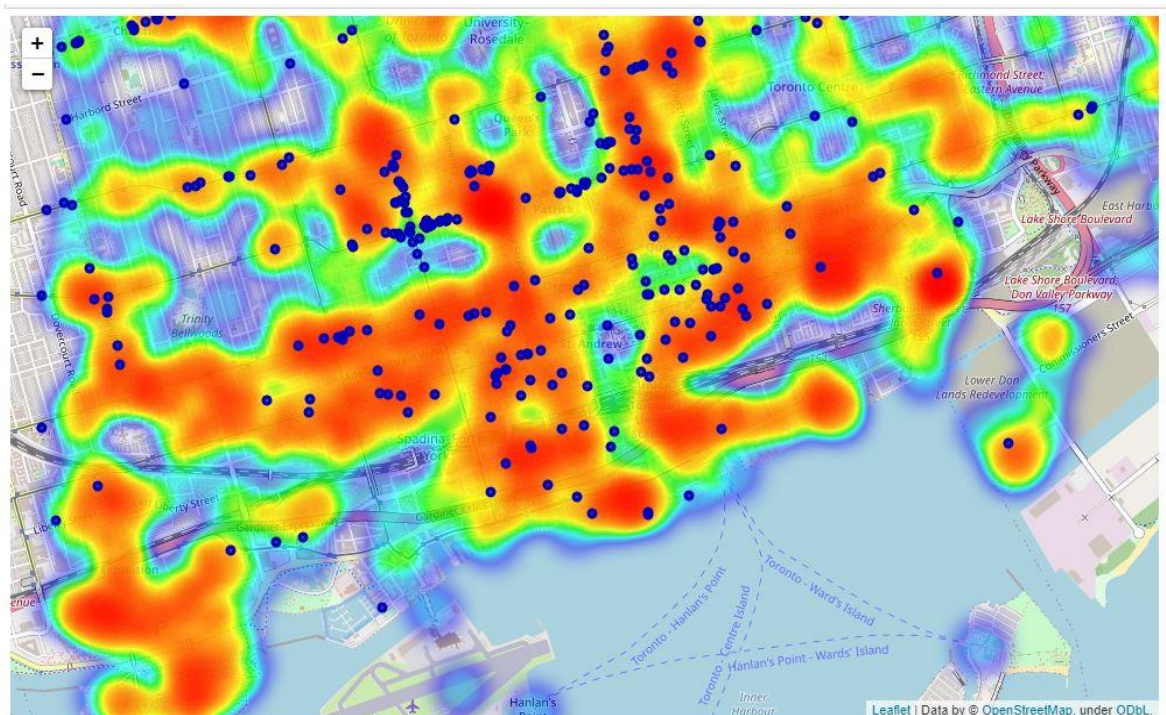
The density based clustering is used to detect the cluster of entertainment zones. The main advantage with the density based clustering is that there is no need to specify the number of clusters explicitly, thus it can effectively capture the areas with large number of entertainment centres. The clusters after running the density based algorithm on our data is shown below in the map
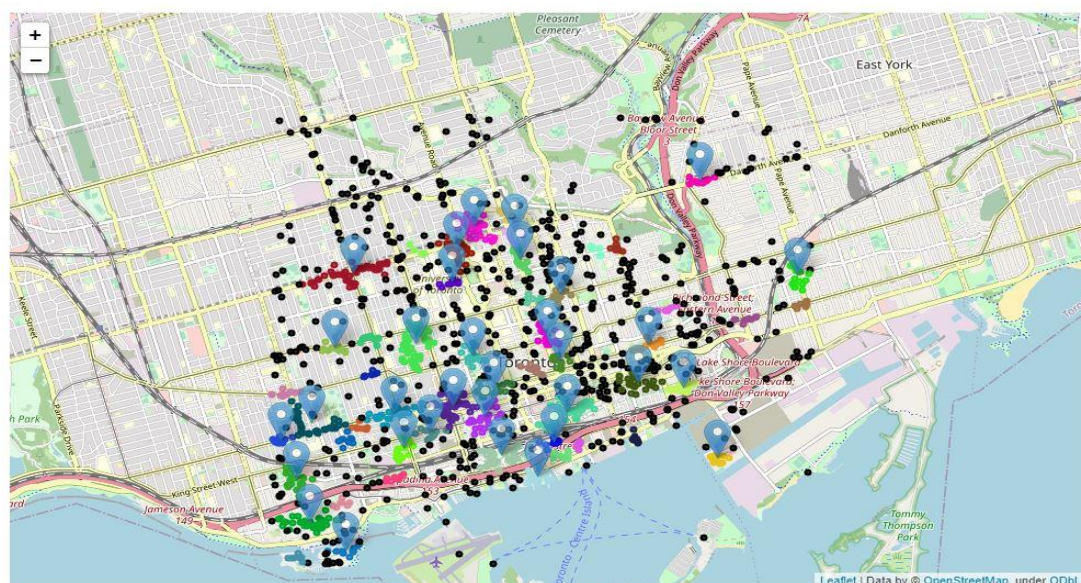
Each marker in the map represent a cluster and each colour represent the member of the clusters except the black dots which does not belongs to any cluster indicating that the entertainment centres near that place is sparse.

Now that we have found the areas with high entertainment centres let's shift our focus to the number of restaurants. The next objective is to find the entertainment areas with less number of restaurants (less than 4) around the 300 m of the vicinity. This can be found by calculating the distance of the restaurant and the mean location of the cluster areas (entertainment centres) and calculating the number of restaurants within 300 m.

C. Results

The result is the pandas dataframe with cluster number, no of entertainment centres in the cluster, cluster mean latitude and longitude and the number of restaurants within 300 m from the mean latitude and longitude. The dataframe is shown below

	clusternumber	entertainment count	clustlat	clustlng	Count
10	1	32	43.632987	-79.417993	0
19	45	16	43.664428	-79.341468	0
26	42	12	43.663597	-79.395592	0
28	2	12	43.629343	-79.412288	0
7	27	39	43.650132	-79.359367	1
15	3	24	43.638441	-79.420073	1
17	26	22	43.650738	-79.366431	1
21	11	14	43.643453	-79.422463	1
29	10	12	43.641156	-79.354036	1
3	12	42	43.645593	-79.417615	2
18	6	17	43.642359	-79.403238	2
20	35	15	43.655577	-79.364957	2
31	52	11	43.676592	-79.356755	2
14	8	25	43.638814	-79.382153	3
24	20	13	43.643216	-79.379714	3

The Locations are depicted on the map and shown below
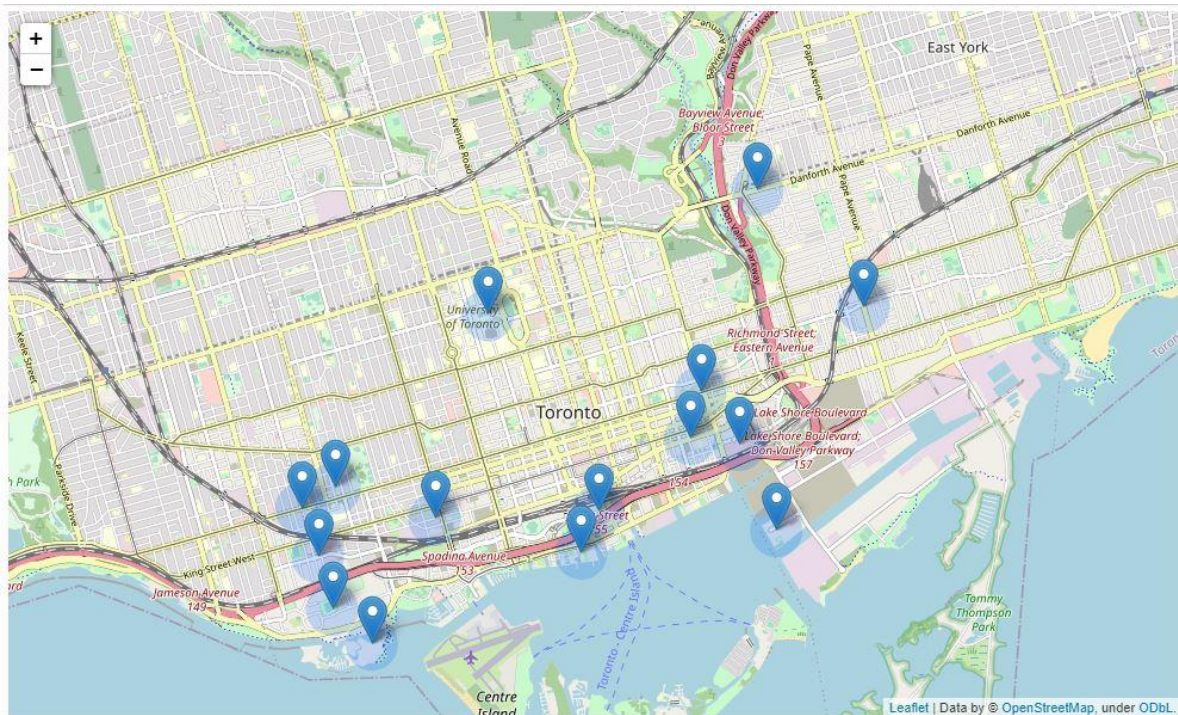
D. Discussion

Our analysis shows that there around ~1550 entertainment zones around the city hall(6x6 km). The data showed that with the exception of few areas the entertainment zones are everywhere with some areas having high density of entertainment centres. The restaurants data as collected from the foursquare api indicates around 389 different kinds of restaurant. We tried to find the pockets of entertainment zones by clustering algorithm and found 52 such spots. Ensuring a minimum number of 10 entertainment centres further reduction was performed on the selected spots.

The number of the restaurant around 400m of the each spots was calculated by measuring the distance from the mean location of every spots to every restaurant. Then further deduction was done based on the number of restaurants in the particular zone. The spots with less than or equal to 3 restaurants were selected. The spots are reduces to 14 at the end of the analysis.

Most interesting fact is that nearly 9 out of 14 spots were located near the coast and another place to consider is the area near University of Toronto which may attract the attention of college students.

E. Conclusion

The purpose of the project is to find the best possible locations for the opening of new restaurant which may help the stackholders to decide the location of the restaurant. The best possible location was found from the analysis and plotted on the interactive map. Final decision on the optimal location on the restaurant depends on the other factors such as neighbourhood, availability of resources and manpower, price etc..,