Capstone Project - The Battle of Neighbourhoods (Week 2)

Nikhil

22nd March 2020

Contents

1. Introduction

Problem statement: Toronto is Canada's largest city, the fourth largest in North America and home to people with diverse ethnic and cultural background. Over the years, Canada witnessed increasing in-migration from across the globe especially Asian countries including India and China, which contributes to a largest portion in the pie.

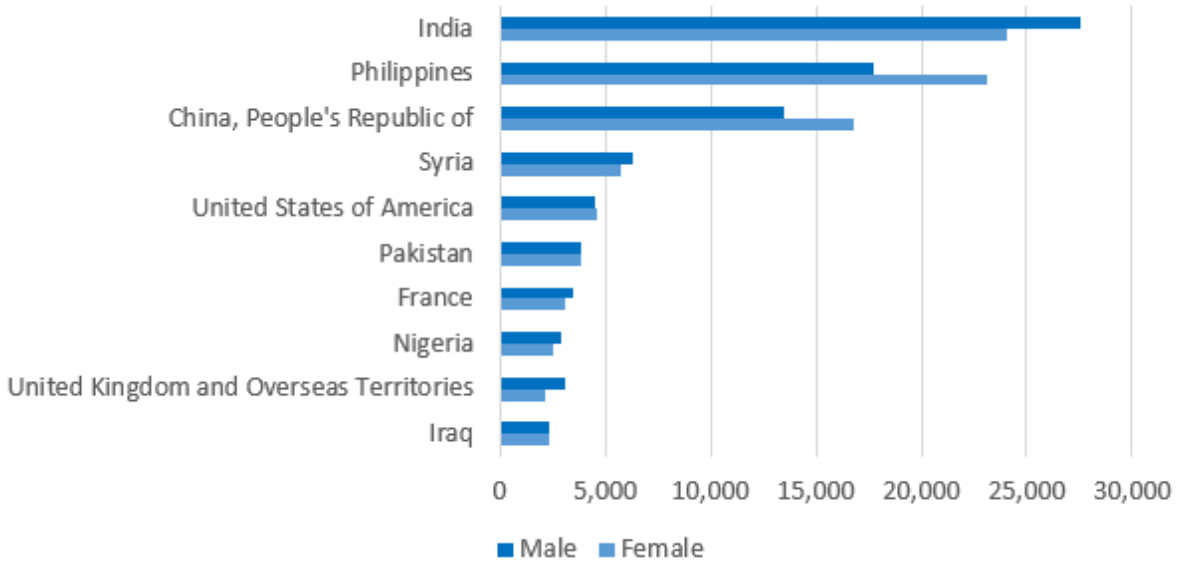

Figure 1: Admission of permanent residents by top 10 countries in 2017 to Canada (Source: Canada.ca

The above diagram represents the admission of permanent residents by top 10 countries in 2017 to Canada (Source: Canada.ca). It is evident that Asian countries (China, India and Pakistan) are amongst the top countries from which large number of people are migrating to Canada. Considering the increasing number of population of Asians in Canada and popularity of Chinese cuisine across Asian nations as well as across the globe, this project will be exploring the right places to start a Chinese restaurant, considering various factors such as proximity to tourist attractions, in Toronto city.

In this project, we will be identifying ideal locations for starting a Chinese restaurant in the Canadian city of Toronto. TripAdvisor ranked **CN Tower** as third amongst the top tourist attractions and first amongst the important landmarks in Canada. This is one of the most important landmark in Canada. Hence, our focus will be to identify suitable locations near CN Tower that can be accessed very easily and fast from this landmark.

We will also try to identify and visualize suitable **neighbourhoods** for starting **targeted marketing and initial promotion activities** of the restaurant launching (near each of the identified locations for starting the restaurant).

Another problem we are trying to explore is that the selection of locations should be considering the **proximity to other places of interest, such as Royal Ontario Museum,** which is another important landmark, thereby increasing maximum footfall of tourists.

2. Data

The above problem statement gives us some indication on the data sets we will be requiring for solving the business problem. For this assignment, we will be requiring following data from various information sources,

a) Data of all the neighbourhoods in Toronto
b) Data of all restaurants near to places of interest such as CN Tower and Royal Ontario Museum
c) Data of all Chinese restaurants near to places of interest such as CN Tower and Royal Ontario Museum

For meeting the above data requirements, we will be using the following data sources,

- Wikipedia for retrieving data of neighbourhoods in Toronto
- Foursquare for retrieving data of all restaurants near to places of interest such as CN Tower and Royal Ontario Museum
- Foursquare for retrieving Data of all Chinese restaurants near to places of interest such as CN Tower and Royal Ontario Museum

2.1 Retrieving data of neighbourhoods in Toronto

For this assignment, we have retrieved the data of all neighbourhoods in Toronto from Wikipedia using the python package Beautiful Soup for HTML parsing. This gave us the following output:

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Queen's Park / Ontario Provincial Government	Queen's Park / Ontario Provincial Government
5	M9A	Etobicoke	Islington Avenue
6	M1B	Scarborough	Malvern, Rouge
7	M3B	North York	Don Mills North
8	M4B	East York	Parkview Hill, Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson
10	M6B	North York	Glencairn

Table 1: Head of neighbourhood data frame obtained using HTML parsing from Wikipedia

Wikipedia source not giving us the geographic co-ordinates hence we retrieved this information from an open source (http://cocl.us/Geospatial_data) and combined both the data frame which gave us the final output as follows.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Queen's Park / Ontario Provincial Government	Queen's Park / Ontario Provincial Government	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills North	43.745906	-79.352188
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073

Table 2: Toronto neighbourhood data frame with geographical co-ordinates

Visualizing this information using Folium library gave us the following output:

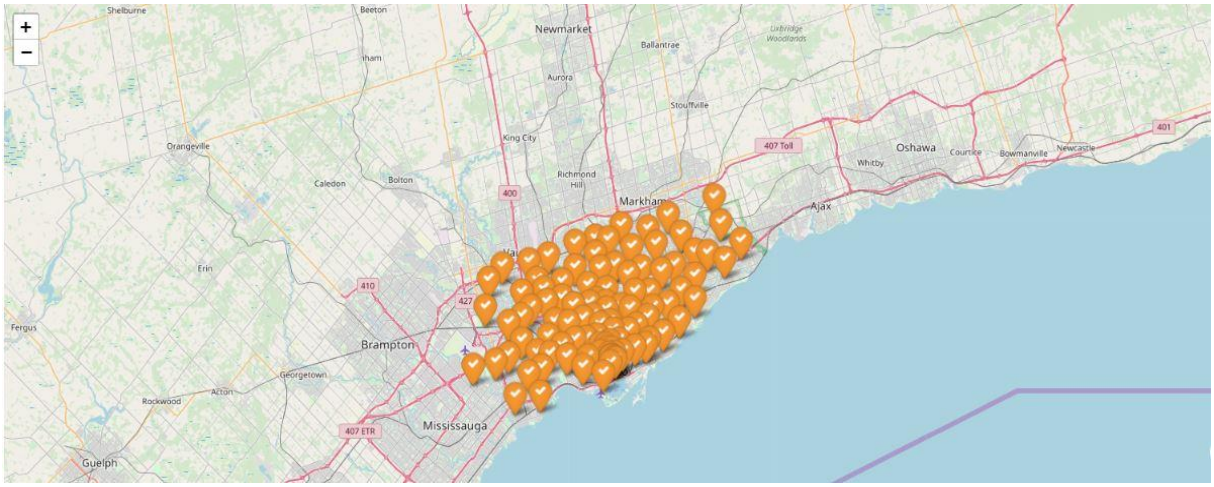

Figure 2: Neighbourhoods in Toronto

2.2 Candidates for suitable locations

We have identified candidate locations (as circular grids) around CN Tower. These locations are at a distance of less than 6 KM from CN Tower and having 300 metres radius each. The visualization of these candidate locations around CN Tower given as follows,

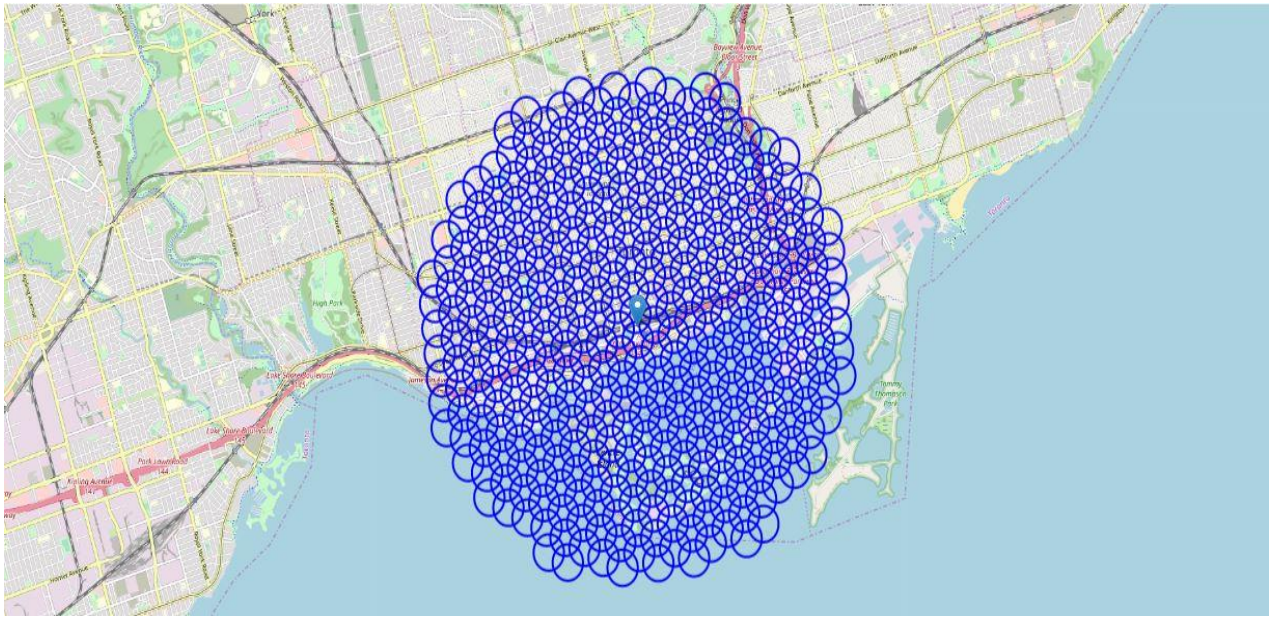

Figure 3: Candidate locations around CN Tower. The candidates are circular grids of 300 metres radius and at a distance of less than six KMs from CN Tower

We have retrieved the address for the centre of each these location candidates using reverse geocoding also the distance from CN Tower and stored in a data frame which looked like,

	Address	Latitude	Longitude	X	Y	Distance from center
0	(CN Tower, 301, Front Street West, Entertainme...	43.635297	-79.336538	-5.314024e+06	1.050238e+07	5992.495307
1	(CN Tower, 301, Front Street West, Entertainme...	43.639021	-79.337100	-5.313424e+06	1.050238e+07	5840.376700
2	(CN Tower, 301, Front Street West, Entertainme...	43.642746	-79.337661	-5.312824e+06	1.050238e+07	5747.173218
3	(CN Tower, 301, Front Street West, Entertainme...	43.646471	-79.338223	-5.312224e+06	1.050238e+07	5715.767665
4	(CN Tower, 301, Front Street West, Entertainme...	43.650197	-79.338785	-5.311624e+06	1.050238e+07	5747.173218
5	(CN Tower, 301, Front Street West, Entertainme...	43.653922	-79.339346	-5.311024e+06	1.050238e+07	5840.376700
6	(CN Tower, 301, Front Street West, Entertainme...	43.657648	-79.339908	-5.310424e+06	1.050238e+07	5992.495307
7	(CN Tower, 301, Front Street West, Entertainme...	43.629357	-79.340136	-5.314924e+06	1.050290e+07	5855.766389
8	(CN Tower, 301, Front Street West, Entertainme...	43.633081	-79.340698	-5.314324e+06	1.050290e+07	5604.462508
9	(CN Tower, 301, Front Street West, Entertainme...	43.636806	-79.341260	-5.313724e+06	1.050290e+07	5408.326913

Figure 4: Address for location candidates and distance from CN Tower

2.3 Restaurant Data

We collected the restaurant data for all and Chinese restaurants using foursquare and a primary analysis gave following results:

a) **Total number of restaurants: 1135**
b) **Total number of Chinese restaurants: 104**
c) **Percentage of Chinese restaurants: 9.16%**
d) **Average number of restaurants in neighbourhood: 5.01**

We stored the restaurant data into another data frame and visualized using folium libraries

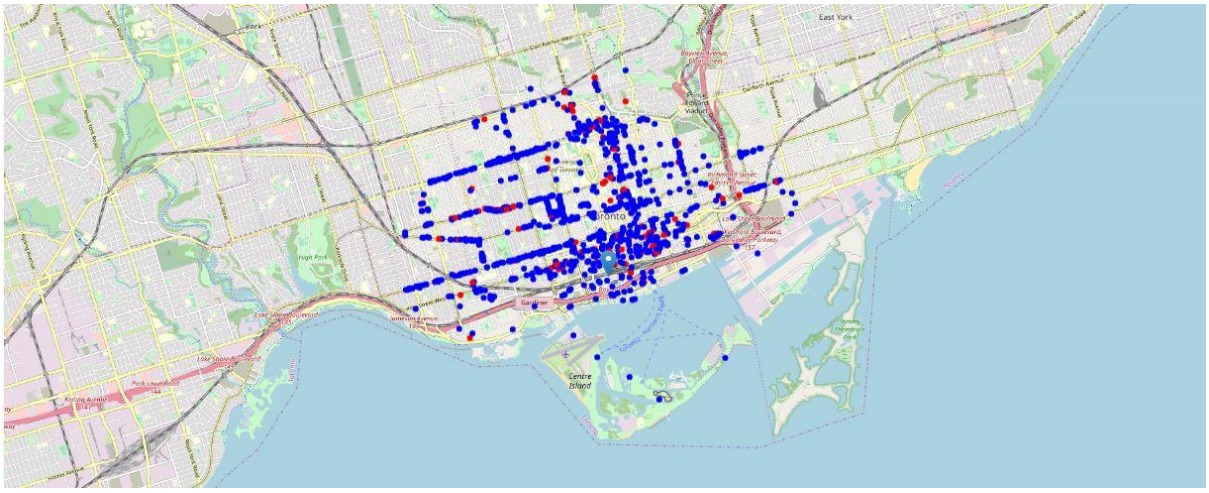

Figure 5: Restaurants around location candidates (red represents Chinese and blue represent other restaurants)

At this stage of development, we have the following set of data for our future analysis

1. Toronto neighbourhood data
2. Candidate locations around CN Tower
3. Restaurant data of all and Chinese restaurants

3. Methodology

Figure 6: three-step methodology adopted for data analysis

For this project, we will follow a three-step methodology and the main stages described as follows,

Stage 1: In this step, we have **retrieved all the relevant data** required for our analysis, which are

a) Toronto neighbourhood data
b) Candidate locations data around CN Tower
c) Restaurant data of all and Chinese restaurants in the candidate locations

We used multiple secondary sources for data including Wikipedia (Using parser), Geopy libraries and foursquare data of restaurants

Stage 2: The second stage involves analysis of the data we have collected. In this stage, we will explore different location candidates around CN Tower and calculate **restaurant density** in each of the candidate locations. We will use **heat maps** and other visualizations to explore the density

in each area and limit our analysis to areas where **low number of restaurants are located and few chinese restaurants**

Stage 3: In the final stage of analysis, we will filter the candidate locations based on following criteria

a) Locations with **no more than two restaurants in radius of 250 meters**
b) Locations **without Chinese restaurants in radius of 400 meters**
c) Proximity to other areas of interest such as Royal Ontario museum

Once we filtered the candidate areas we will be visualizing the suitable locations for starting Chinese restaurants around CN tower and superimpose the neighbourhoods in Toronto on the map (for identifying suitable locations for targeted marketing)

3.1 Exploratory data analysis

In this section, we will be doing some relevant exploratory analysis of data, which are essential for identifying suitable locations

1. Count of number of restaurants in every area candidate

We obtained the following header for data frame while analysing the number of restaurants in every area candidate,

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
0	(CN Tower, 301, Front Street West, Entertainme...	43.635297	-79.336538	-5.314024e+06	1.050238e+07	5992.495307	0
1	(CN Tower, 301, Front Street West, Entertainme...	43.639021	-79.337100	-5.313424e+06	1.050238e+07	5840.376700	0
2	(CN Tower, 301, Front Street West, Entertainme...	43.642746	-79.337661	-5.312824e+06	1.050238e+07	5747.173218	0
3	(CN Tower, 301, Front Street West, Entertainme...	43.646471	-79.338223	-5.312224e+06	1.050238e+07	5715.767665	0
4	(CN Tower, 301, Front Street West, Entertainme...	43.650197	-79.338785	-5.311624e+06	1.050238e+07	5747.173218	0
5	(CN Tower, 301, Front Street West, Entertainme...	43.653922	-79.339346	-5.311024e+06	1.050238e+07	5840.376700	2

6	(CN Tower, 301, Front Street West, Entertainme...	43.657648	-79.339908	-5.310424e+06	1.050238e+07	5992.495307	2
7	(CN Tower, 301, Front Street West, Entertainme...	43.629357	-79.340136	-5.314924e+06	1.050290e+07	5855.766389	0
8	(CN Tower, 301, Front Street West, Entertainme...	43.633081	-79.340698	-5.314324e+06	1.050290e+07	5604.462508	0
9	(CN Tower, 301, Front Street West, Entertainme...	43.636806	-79.341260	-5.313724e+06	1.050290e+07	5408.326913	0

Table 3: Number of restaurants in every candidate area

Average number of restaurants in every area with radius=300m: 5.01

2. Distance to nearest Chinese restaurants from all candidate centre,

We have also calculated the distance to nearest Chinese restaurant from each of the candidate centre,

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to chinese restaurant
0	(CN Tower, 301, Front Street West, Entertainme...	43.635297	-79.336538	-5.314024e+06	1.050238e+07	5992.495307	0	3625.257418
1	(CN Tower, 301, Front Street West, Entertainme...	43.639021	-79.337100	-5.313424e+06	1.050238e+07	5840.376700	0	3196.618019
2	(CN Tower, 301, Front Street West, Entertainme...	43.642746	-79.337661	-5.312824e+06	1.050238e+07	5747.173218	0	2830.943688

3	(CN Tower, 301, Front Street West, Entertainme...	43.646471	-79.338223	-5.312224e+06	1.050238e+07	5715.767665	0	2302.391039
4	(CN Tower, 301, Front Street West, Entertainme...	43.650197	-79.338785	-5.311624e+06	1.050238e+07	5747.173218	0	1710.258433
5	(CN Tower, 301, Front Street West, Entertainme...	43.653922	-79.339346	-5.311024e+06	1.050238e+07	5840.376700	2	1126.482718
6	(CN Tower, 301, Front Street West, Entertainme...	43.657648	-79.339908	-5.310424e+06	1.050238e+07	5992.495307	2	577.011892
7	(CN Tower, 301, Front Street West, Entertainme...	43.629357	-79.340136	-5.314924e+06	1.050290e+07	5855.766389	0	4083.857483
8	(CN Tower, 301, Front Street West, Entertainme...	43.633081	-79.340698	-5.314324e+06	1.050290e+07	5604.462508	0	3560.023504
9	(CN Tower, 301, Front Street West, Entertainme...	43.636806	-79.341260	-5.313724e+06	1.050290e+07	5408.326913	0	3064.252398

Table 4: Distance to the nearest Chinese restaurant

We can find one Chinese restaurant at a distance of nearly 1.5 KM from every candidate centre. In order to get a clear picture of the scenario we visualized all and Chinese restaurants in locations candidates using heat maps (We also included white circles on the map to show 1km, 2km, 3km radius from CN Tower), which gave us the following outputs:

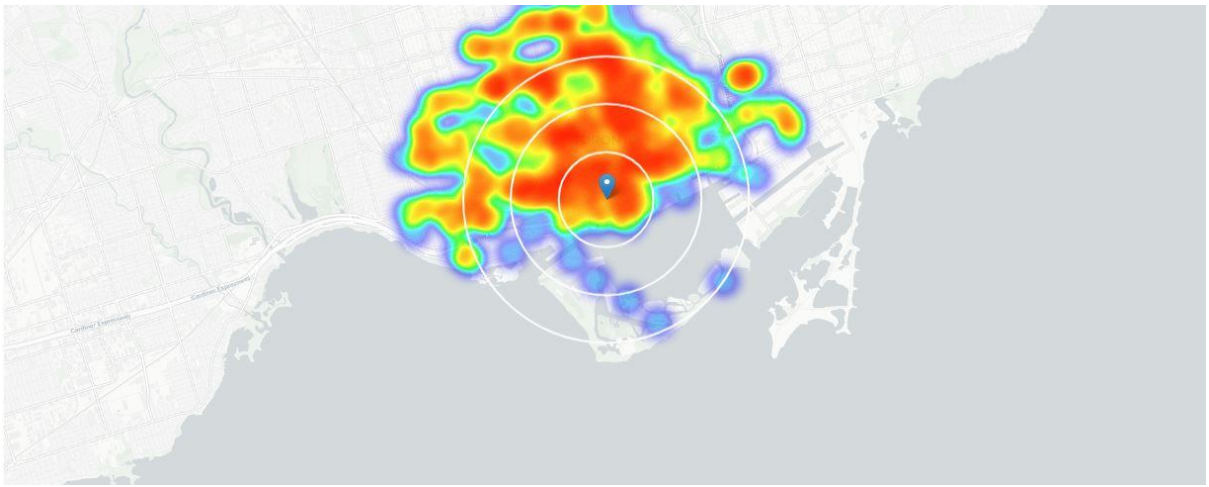

Figure 7: Heat map of all restaurant around CN Tower

Figure 8: Heat map of Chinese restaurants around CN Tower

The above map gives us some clear indication on the areas with low density of Chinese restaurants. However, as evident, the southern side in the map largely covered by Ontario river and hence we can ignore that location. Now let us focus on a narrow region that covers our areas of interest such Royal Ontario Museum, Toronto University etc. Let us move the centre of interest and reduce our radius to 2.5 KMs. Now our focus will be within the white circled location,

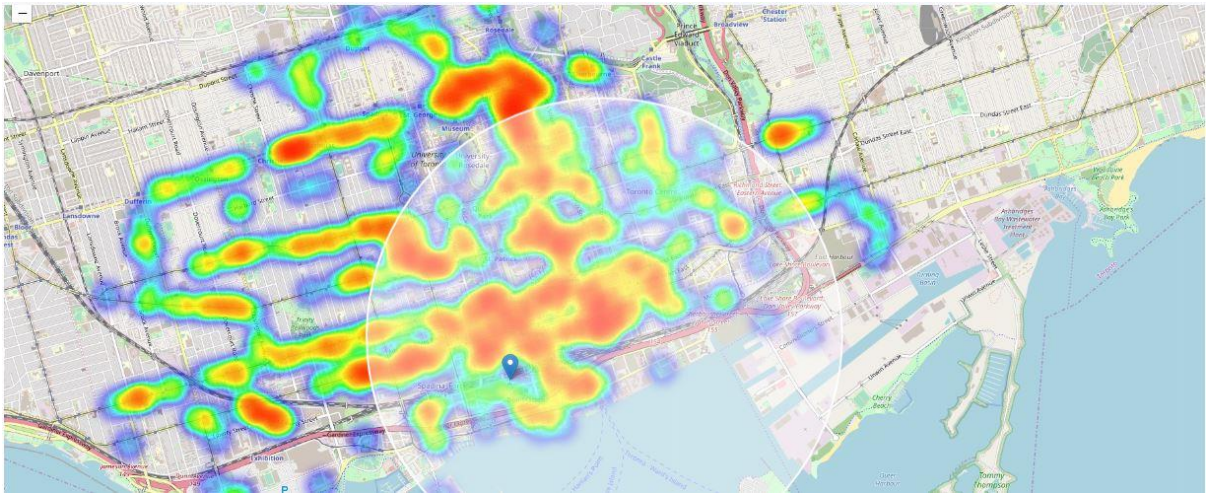

Figure 9: New focus region obtained by narrowing the regions and reducing the radius to 2.5 KMs

We created new location candidates focusing on the above region in the white circle. We obtained the data for these candidates and obtained following data frame.

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to Chinese restaurant
0	43.648097	-79.353315	-5.311774e+06	1.050410e+07	1	821.495793
1	43.648718	-79.353409	-5.311674e+06	1.050410e+07	1	766.486294
2	43.644623	-79.353538	-5.312324e+06	1.050418e+07	0	1185.615807
3	43.645244	-79.353632	-5.312224e+06	1.050418e+07	0	1099.240985
4	43.645865	-79.353726	-5.312124e+06	1.050418e+07	0	1015.370201
5	43.646485	-79.353820	-5.312024e+06	1.050418e+07	0	934.677777
6	43.647106	-79.353914	-5.311924e+06	1.050418e+07	1	858.060865
7	43.647727	-79.354008	-5.311824e+06	1.050418e+07	1	786.711096
8	43.648348	-79.354102	-5.311724e+06	1.050418e+07	1	722.191284
9	43.648969	-79.354196	-5.311624e+06	1.050418e+07	1	666.487923

Table 5: New location candidates on focussed region and distances to Chinese restaurants

Now we filtered the above data frame using the candidate selection criteria outlined in the methodology section. Let us recall the criteria we mentioned in the methodology section

a) Locations with no more than two restaurants in radius of 250 meters
b) Locations without Chinese restaurants in radius of 400 meters

Filtering these candidates helped us in identifying good locations candidates, which represented in the following map (Blue colour indicates the locations that meet our criteria)

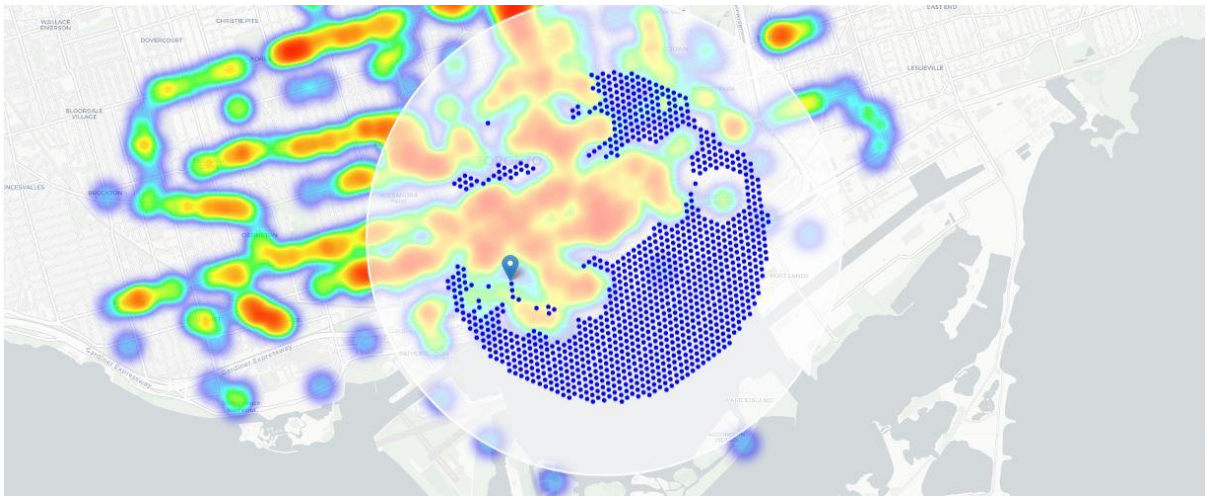

Figure 10: Good locations (represented in blue) for starting restaurant that meet our criteria requirements

Clustering the location candidates:

Let us cluster these location points using K-means clustering. We will make 15 clusters using these points and the output displayed as follows,

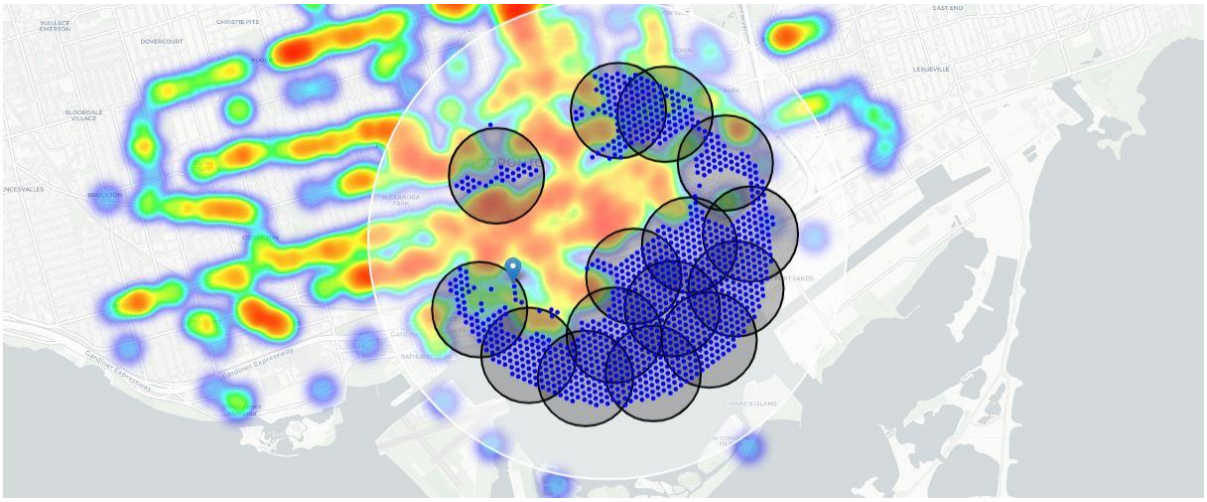

Figure 11: Clustering using K-means

We will make the map little more comprehensible by removing the heat maps. We will also include all the neighbourhoods in Toronto for targeted marketing activities of restaurant.

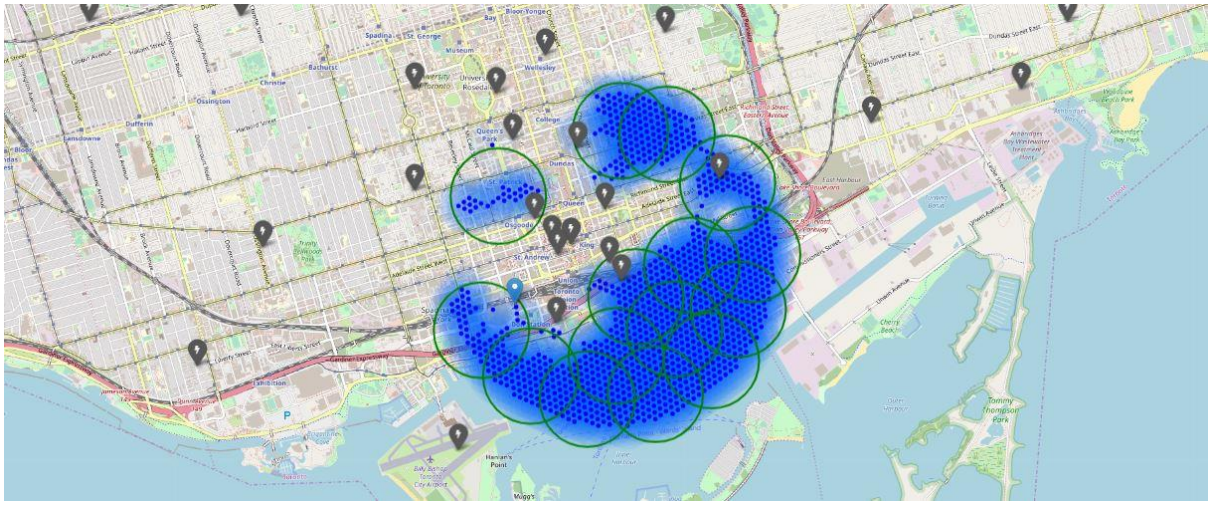

Figure 12: Good location candidates around CN tower and neighbourhoods near each candidates

We obtained the address of each of the centres of good locations, and also distance from CN Tower and got following results:

==========

Addresses of centers of areas recommended for further analysis

==========

D 173, Tasque, Mirande, Gers => 2.5km from CN Tower

Tasque, Mirande, Gers => 2.7km from CN Tower

D 3, Termes-d'Armagnac, Mirande => 3.0km from CN Tower

D 173, Tasque, Mirande => 1.1km from CN Tower

Route de Tasque, Tasque, Mirande => 3.1km from CN Tower

Route de Tasque, Tasque, Mirande => 2.4km from CN Tower

Route de Tasque, Tasque => 0.6km from CN Tower

D 3, Termes-d'Armagnac, Mirande => 1.6km from CN Tower

Tasque, Mirande, Gers => 3.7km from CN Tower

D 173, Tasque, Mirande, Gers => 1.8km from CN Tower

Route de Tasque, Tasque => 1.7km from CN Tower

D 3, Termes-d'Armagnac => 3.7km from CN Tower

D 3, Termes-d'Armagnac => 3.4km from CN Tower

Route de Tasque, Tasque, Mirande, Gers, Occitanie, France métropolitaine, 32160, France => 3.4km from CN Tower

D 173, Izotges, Mirande => 1.8km from CN Tower

Final output:

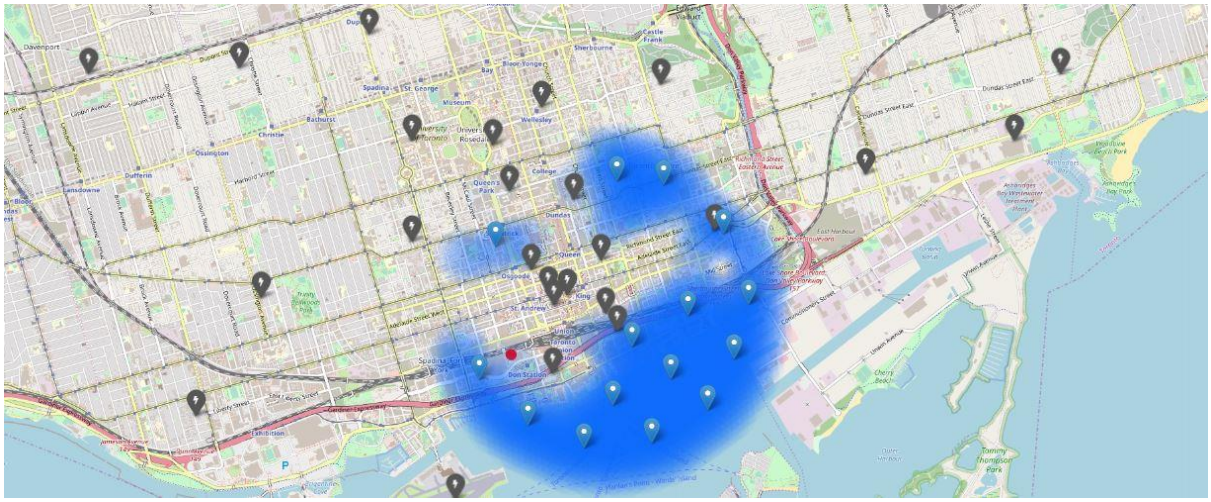

Figure 13: Final output of analysis

The pop-up marker in blue represents centre of ideal locations for starting Chinese restaurant and grey represents the neighbourhoods in Toronto. The green marker represent CN tower, which is our area of interest.

4. Results and discussions

Results: We have started our analysis by looking at all the restaurants and Chinese restaurants by keeping CN Tower (One of the most important landmark and place of interest in Toronto). The analysis found that, every candidate areas of 300m radius around CN Tower comprises of nearly five restaurants on an average. Thus, there was a requirement of some additional detailed analysis to identify suitable locations, as it was not giving us any clear indication on possible locations, which meets our criteria as outlined in the methodology section. At this stage, we also found that we could find one Chinese restaurant at a distance of nearly 1.5 KM from every candidate centre and hence there was a need of narrowing the focus region.

Our further analysis was to focus on narrow region that covers our areas of interest such Royal Ontario Museum, Toronto University etc. Let us move the centre of interest and reduce our radius to 2.5 KMs. This transformation of centre and reduction in radius was crucial at coming to some important findings. We applied our criteria for filtering the candidate centre, which are,

a) Locations with no more than two restaurants in radius of 250 meters
b) Locations without Chinese restaurants in radius of 400 meters

And found the following results on locations,

Locations with no more than two restaurants nearby: 1292
Locations with no Chinese restaurants within 400m: 1342
Locations with both conditions met: 1100

We applied K-Means clustering on the locations that satisfies our criteria to transform the areas points into 15 clusters and found the address of centre of these clusters and distance from CN Tower as follows,

1. D 173, Tasque, Mirande => 2.5km from CN Tower
2. Tasque, Mirande, Gers => 2.7km from CN Tower

3. D 3, Termes-d'Armagnac, Mirande => 3.0km from CN Tower
4. D 173 => 1.1km from CN Tower
5. Route de Tasque, Tasque, Mirande => 3.1km from CN Tower
6. Route de Tasque, Tasque, Mirande => 2.4km from CN Tower
7. Route de Tasque, Tasque, Mirande => 0.6km from CN Tower
8. D 3, Termes-d'Armagnac, Mirande => 1.6km from CN Tower
9. Tasque, Mirande, Gers => 3.7km from CN Tower
10. D 173, Tasque, Mirande => 1.8km from CN Tower
11. Route de Tasque, Tasque, Mirande => 1.7km from CN Tower
12. D 3, Termes-d'Armagnac, Mirande => 3.7km from CN Tower
13. D 3, Termes-d'Armagnac, Mirande => 3.4km from CN Tower
14. Route de Tasque, Tasque, Mirande => 3.4km from CN Tower
15. D 173, Izotges => 1.8km from CN Tower

Now it is up to the decision makers/stakeholders to decide their interest area for choosing a location amongst these 15 clusters. We also visualized the neighbourhoods in Toronto for decision makers for their targeted marketing campaigns near any of these clusters.

Note: There is some error in the address due to the complexity of data sets. However, the locations visualized on the map, distance and other information are accurate.

5. Conclusion

The objective of our project was to identify ideal locations for starting Chinese restaurants around CN Tower and other places of interest such as Royal Ontario Museum given the increasing number of immigrants from Asian nations and tourist footfall in these locations. Our analysis was able to find 15 clusters with location points having no more than two restaurants in radius of 250 meters from each point and locations without Chinese restaurants in radius 400 meters. This implies, the identified locations are suitable due to less competition and increasing demand. However, a further filtering of these clusters using another set of criteria such as tourist's traffic may help in identifying most suitable clusters amongst these 15 clusters but such filtering is beyond the scope of this project and involves higher complexity.