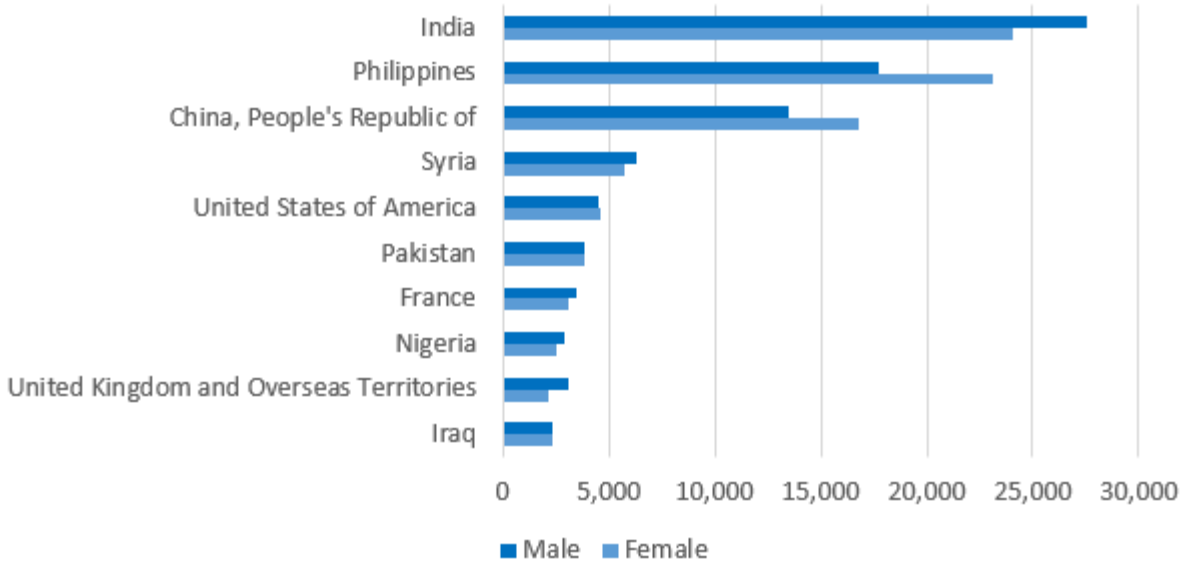
Capstone Project - The Battle of Neighbourhoods (Week 2)

Nikhil

22nd March 2020

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.



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

1. Introduction

- Problem Statement (Contd.)

  - 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.

  - 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

- 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,

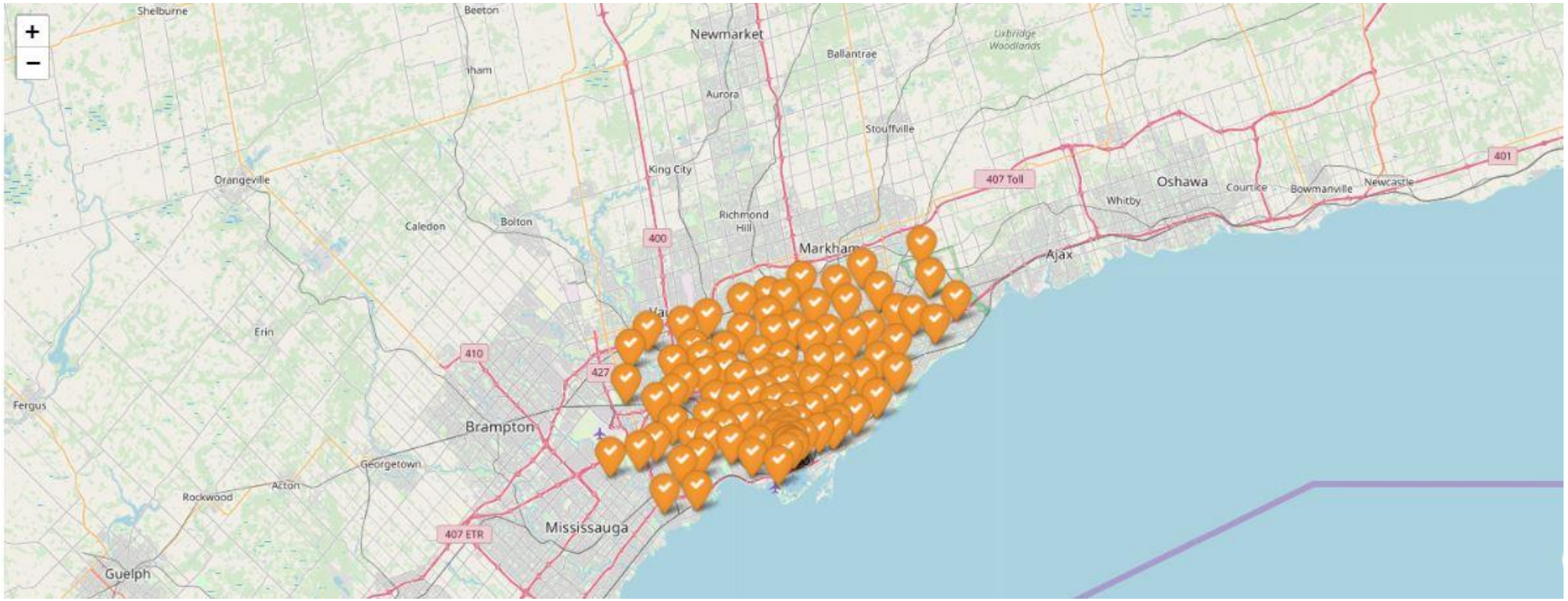
a) Wikipedia for retrieving data of neighbourhoods in Toronto

b) Foursquare for retrieving data of all restaurants near to places of interest such as CN Tower and Royal Ontario Museum

c) Foursquare for retrieving Data of all Chinese restaurants near to places of interest such as CN Tower and Royal Ontario Museum

2.1 Data (All the neighbourhoods in Toronto)

We have retrieved the data of all neighbourhoods in Toronto from Wikipedia using the python package Beautiful Soup for HTML parsing

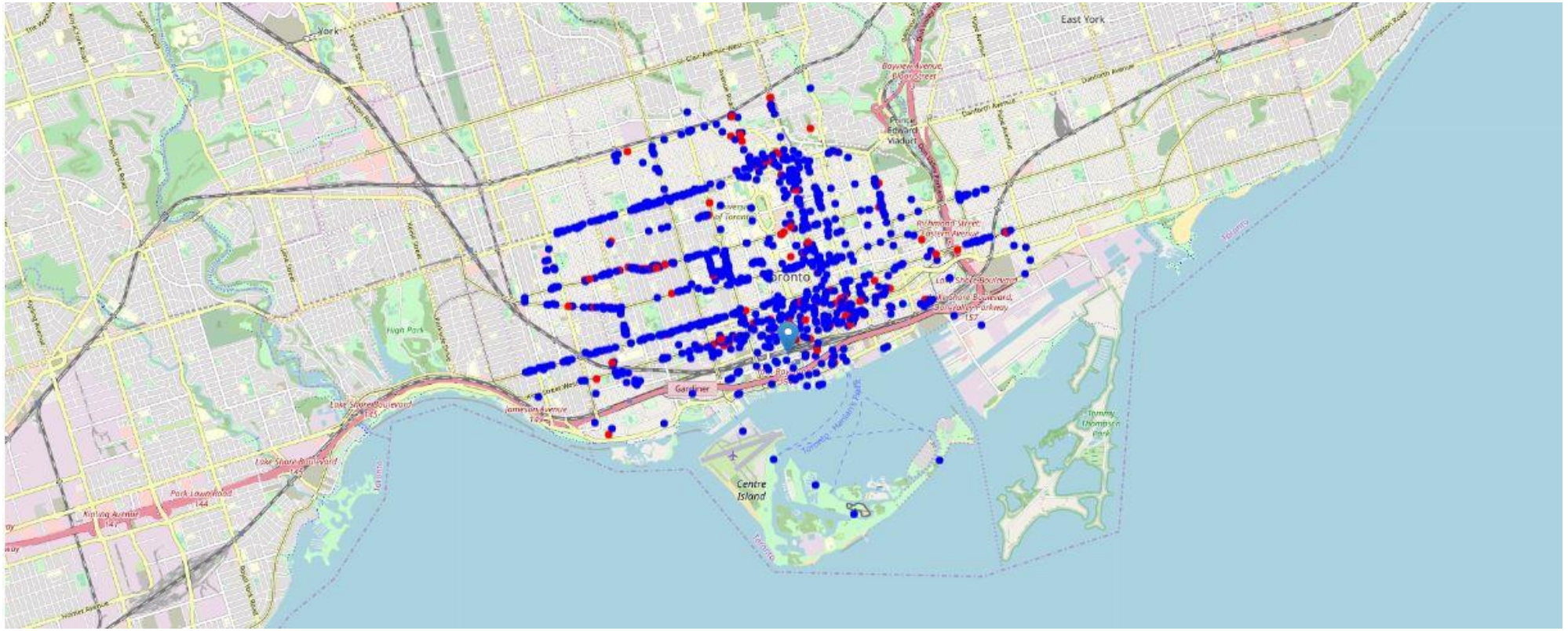


2.2 Data –Candidate locations

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.

2.3 Data – Restaurant Data

We collected the restaurant data for all and Chinese restaurants using foursquare

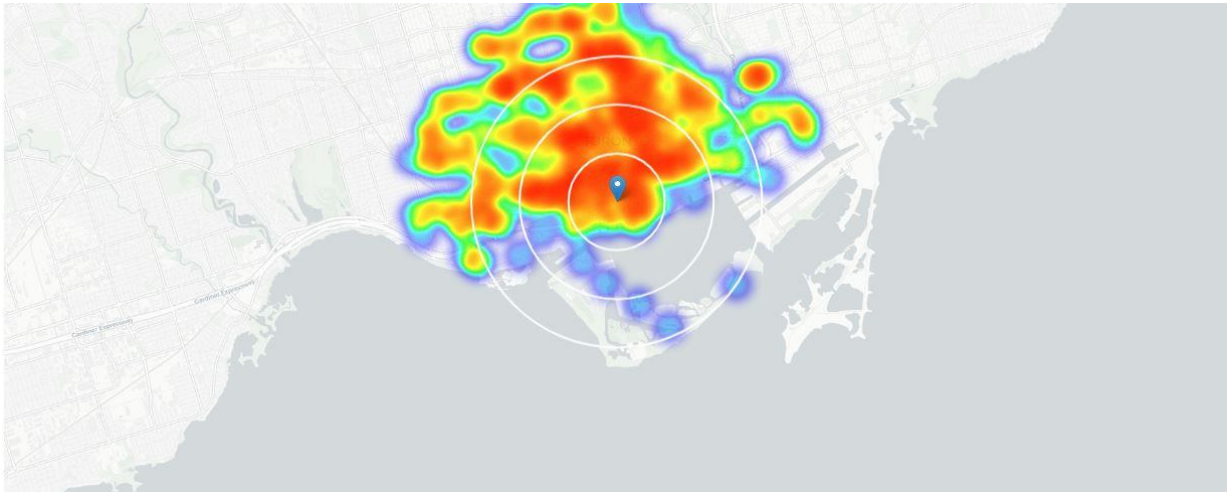


3. Methodology

1. we have **retrieved all the relevant data** required for our analysis
2. 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**
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

3.1 Exploratory Data Analysis

- Average number of restaurants in every area with radius=300m: 5.01
- We can find one Chinese restaurant at a distance of nearly 1.5 KM from every candidate centre

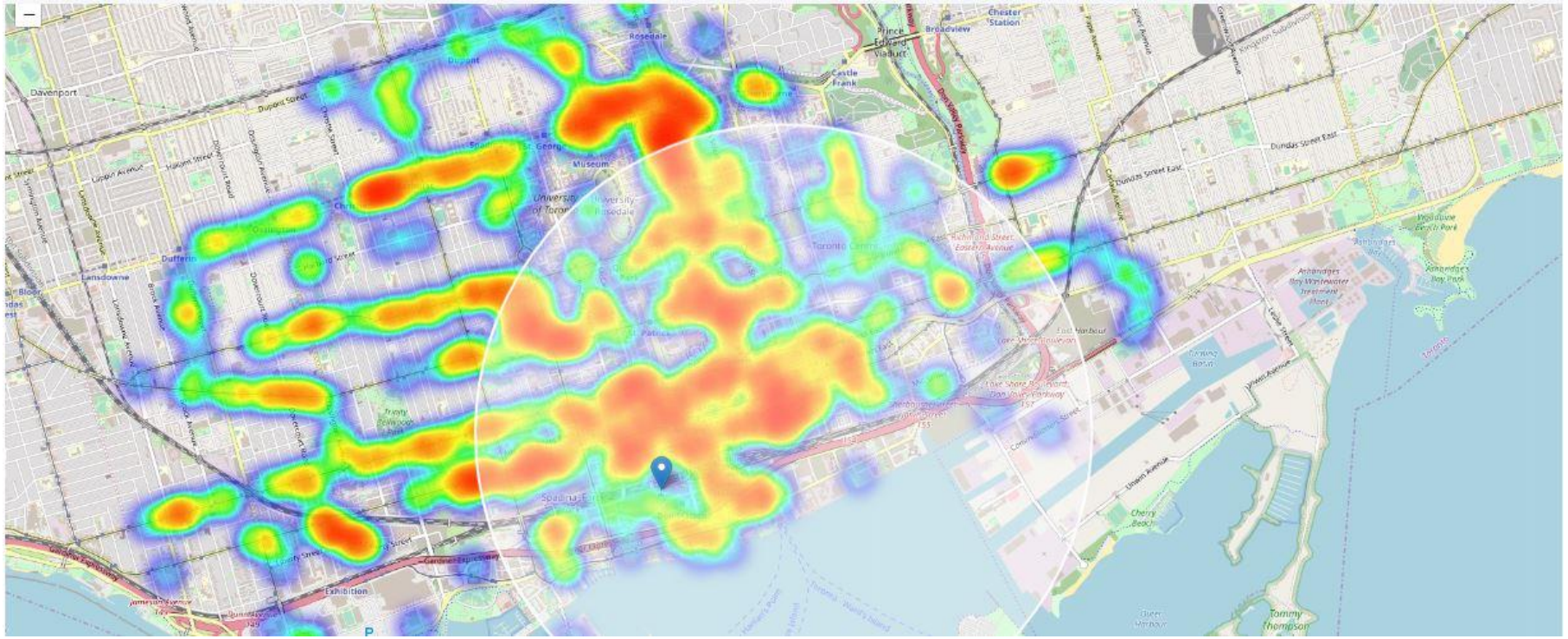


Heat map of all restaurant around CN Tower

Heat map of Chinese restaurant around CN Tower

3.1 Exploratory Data Analysis (Narrowing the region of scope)

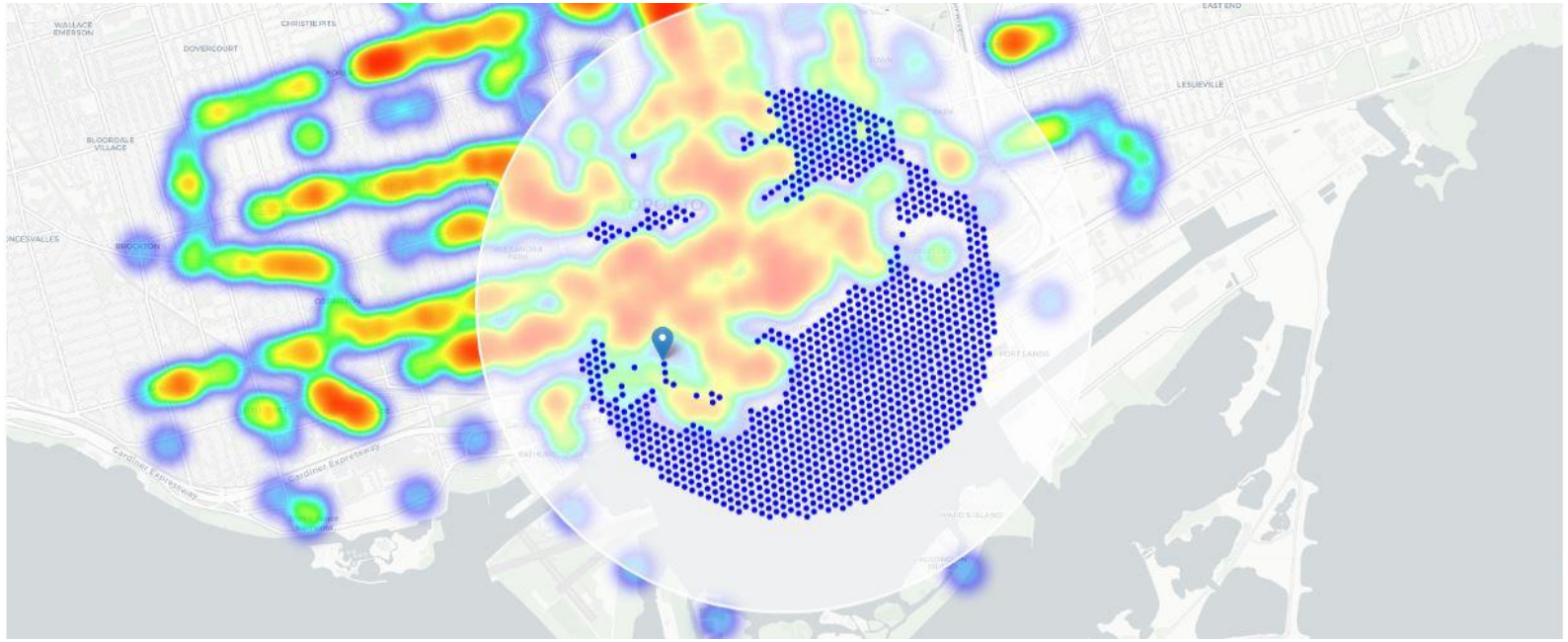
- Now let us focus on a narrow region that covers our areas of interest such Royal Ontario Museum, Toronto University etc. Let us move the center of interest and reduce our radius to 2.5 KMs. Now our focus will be within the white circled location,



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

3.1 Exploratory Data Analysis (Filtering the locations)

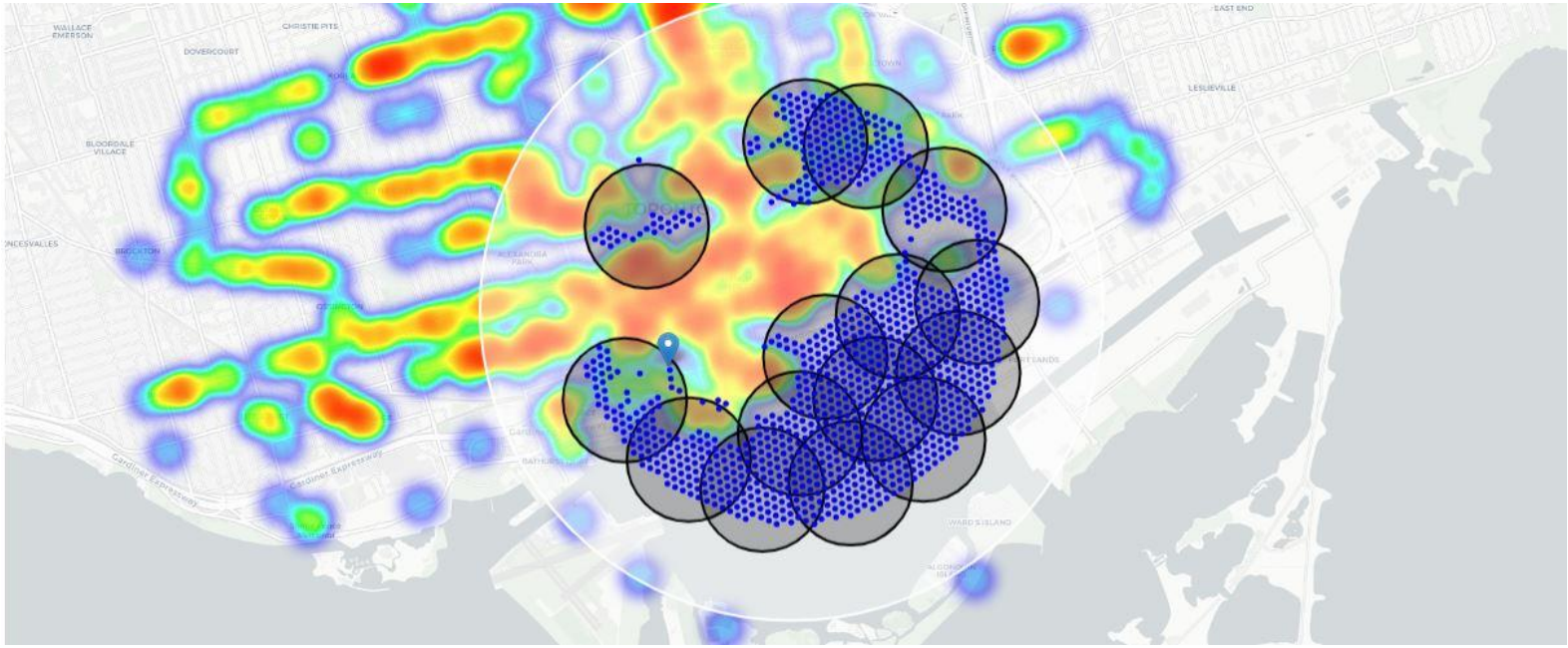
- Now we filtered the above data frame using the candidate selection criteria outlined in the methodology section. Let us recall the criteria we mentioned in the methodology section
- a) Locations with no more than two restaurants in radius of 250 meters
- b) Locations without Chinese restaurants in radius of 400 meters



Good locations after filtering represent in blue color

3.1 Exploratory Data Analysis (Clustering)

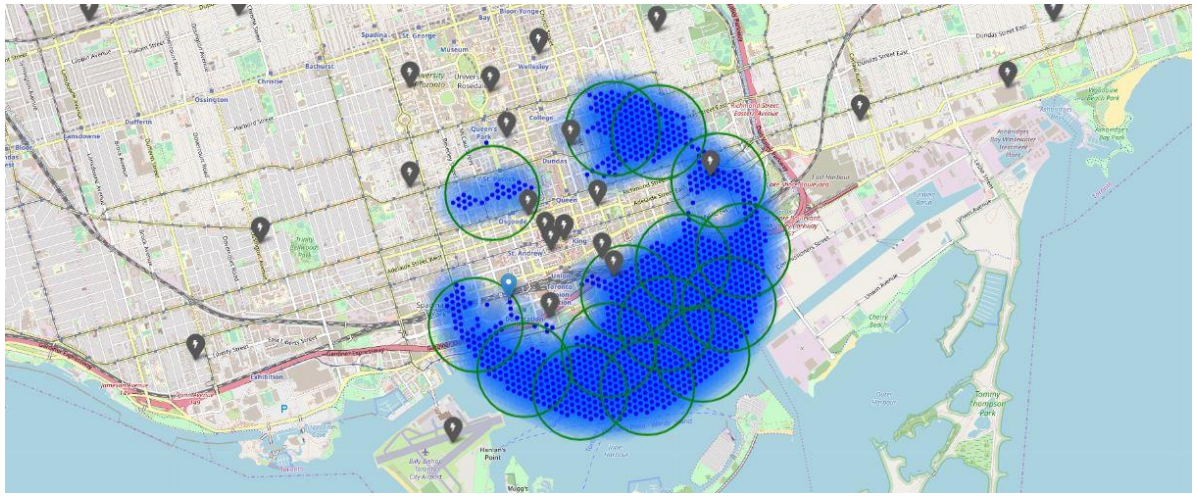
We cluster these location points using K-means clustering. We will make 15 clusters using these points and the output displayed as follows,

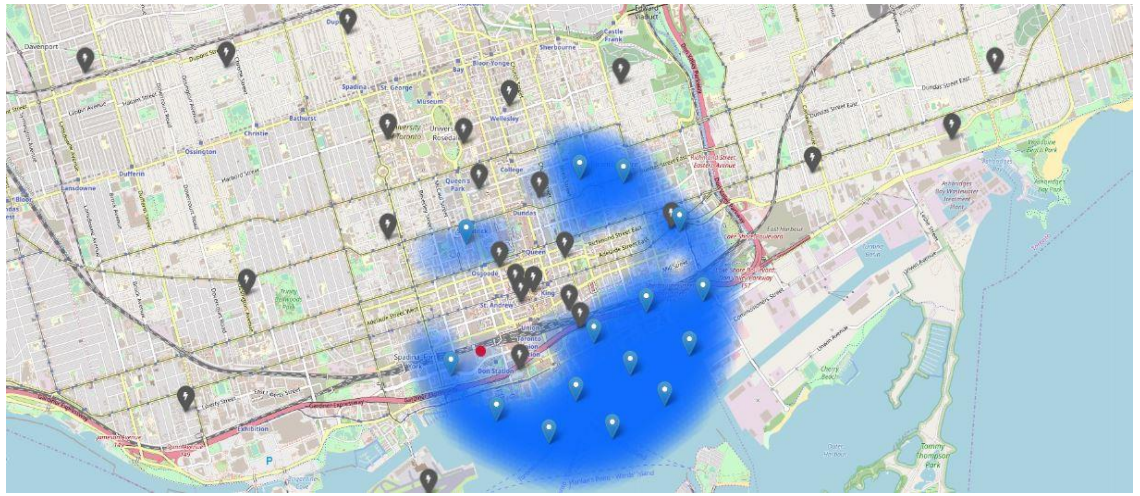


Clustering good locations into 15 clusters using K-Means

3.1 Exploratory Data Analysis (Final results)

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





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 discussion

- The analysis found that, every candidate areas of 300m radius around CN Tower comprises of nearly five restaurants on an average
- 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
- By narrowing the focus region and filtering using criteria we found,
    a) Locations with no more than two restaurants nearby: 1292
    b) Locations with no Chinese restaurants within 400m: 1342
    c) 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
- 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

5. Conclusion

- 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