

Battle of Neighborhoods in New York and Toronto Cities

Yogish

May 2019

1. Introduction:

- 1.1. Background: New York and Toronto Cities are the two biggest cities in the east coast of USA and Canada. Both cities are considered to be the financial capitals of their respective countries. The cities are diverse in terms of demographics and are bustling with activities all round the clock. This report is an effort in trying to understand the cities in terms of their venues. Crime is an important aspect to understand the perceived safety of any city, we also look to gain an understanding of the crimes documented in Police records in both the cities.
- 1.2. Problem: To get a picture of the cities without actually spending a big amount of money and resources on the ground. Gaining an understanding about a City's make up is very essential for someone from other City, State or country before making a decision to establish a restaurant.
- 1.3. Interest: This analysis will be of specific interest to any client like a future restaurateur interested in setting up a restaurant catering specific cuisine or a multinational café chain looking up to set up a franchise in a specific neighborhood of the two cities.

2. Data acquisition and Cleaning

- 2.1. Data Sources: We need Data for with Latitude and Longitude coordinates for the neighbourhoods in the two cities along with datasets recording the crime details in both New York and Toronto.
 - [New York Neighborhood Dataset](#)
 - [New York Crime Data](#)
 - Toronto Neighbourhood Data: scraped from this [Wikipedia](#) website using Beautiful Soup
 - [Toronto Crime Data](#)
- 2.2. Data Cleaning and Preparation: New York city geojson data is sourced from the cognitive class data sets. We extract the Borough, Neighborhood Names and corresponding Latitude and Longitude co-ordinates into a pandas dataframe for further analysis. Data for Toronto city is scraped from the Wikipedia article on the postal codes of Toronto city using BeautifulSoup package. We get a list of Boroughs and Neighborhoods for Toronto city, and then merge this dataframe with another dataframe with Latitude and Longitude co-ordinates. Crime data is sources fro the open data initiative of both the cities.

- 2.3. Data selection for Analysis: For final analysis of venues with Foursquare API we consider data only for Manhattan Borough from New York City and Neighborhoods with Toronto in their names for Toronto city. Crime analysis is performed on the crimes reported in the year 2018.

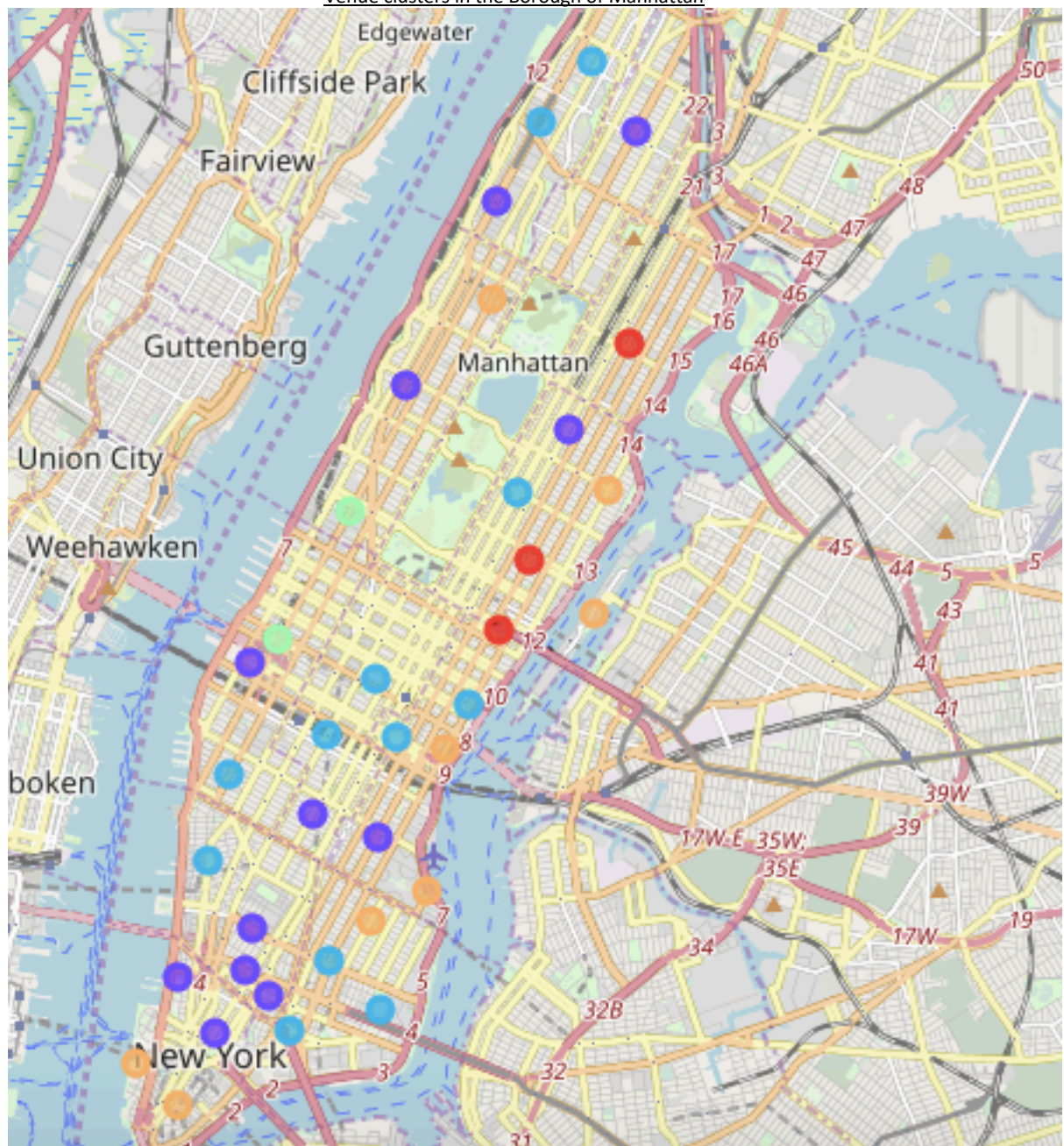
3. Analysis

- 3.1. Foursquare API: Foursquare is a social location service that allows users to explore the world around them. Users can download the Foursquare application to their iPhone, Blackberry, or Android phone and sign up for free, then connect their Foursquare accounts to their other social media accounts. After users download the free application and connect on Facebook or Twitter, they can connect with their friends who are also active on Foursquare. Whenever they or their friends *check in* to a place (that is, they're at a location and they tell others that they are there), the message is broadcast to their friends via Twitter or Facebook. When a user checks in enough times, that user becomes the *mayor* of a location, which may or may not give the user access to special offers, depending on the business running a location. For example, a coffee shop might extend a free drink to anyone who becomes a mayor. Users can also earn badges as they explore and check in to locations, leave tips for other users (for instance, "The margaritas here are great!"), and may even have the opportunity to create locations that haven't appeared yet on the service.
- 3.2. Venue Analysis: We use the venues group but we use the explore endpoint instead of the search endpoint, and we pass the latitude and the longitude coordinates of the neighbourhoods along with our credentials. Then we make the call to the database and in return we get a list of the popular spots around the endpoint Just like with the search endpoint, for each venue we get mostly its name, unique ID, location, and category. With the Foursquare API we get trending venues nearby. These are venues that have the highest foot traffic when the call to the database is made. Therefore the results vary depending on when the call is made. And to get the trending venues, we use the trending endpoint along with the credentials and the latitude and the longitude coordinates of the place of interest. We make the call to the database and get a JSON file of the trending venues that are nearby. In the JSON file, for each trending venue, we get mostly its name, unique ID, location, and category.
- 3.3. Where to setup a Restaurant or a Café: The table below provides a percentage of how many eating places are found trending in each of the clusters in the two cities.

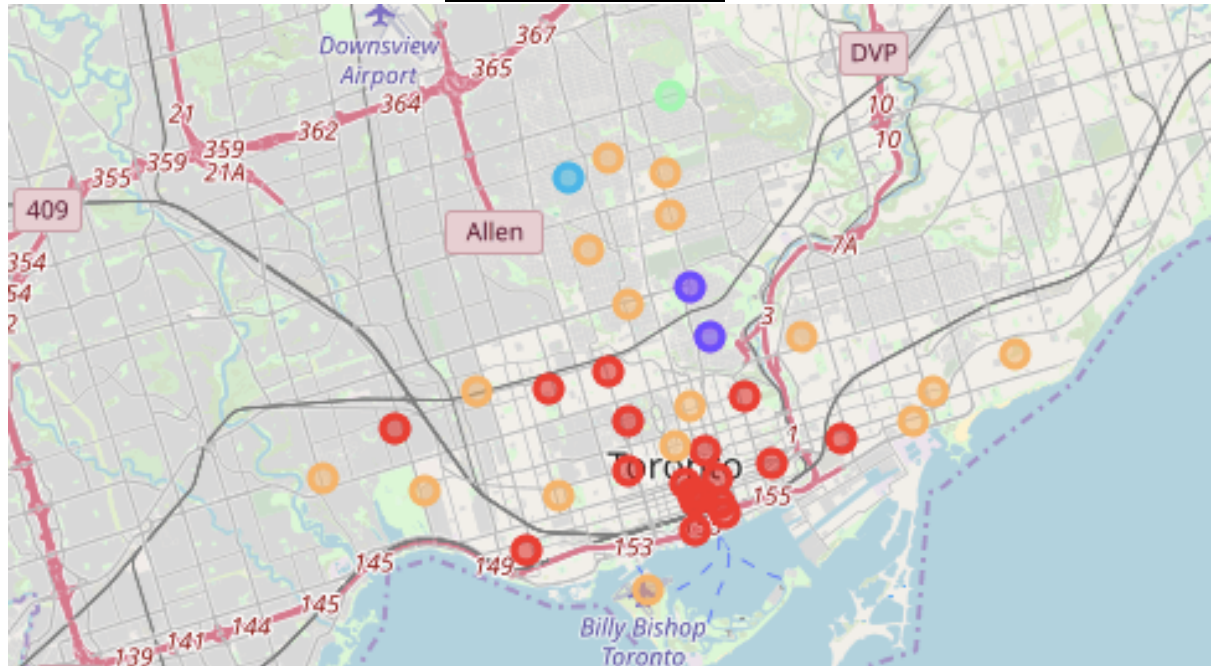
Borough	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Manhattan	67%	71%	67%	0%	73%
Toronto	81%	0%	0%	0%	56%

4. Results: **Battle of Neighbourhoods:** New York is geographically a larger city with an area on 783.8 Square KMs compared to Toronto City with an area of 630.2 Square KMs. New York city is divided into 5 Boroughs containing 306 Neighbourhoods in total compared to Toronto's 103 Neighbourhoods in 10 Boroughs. Analysis of top venues using Foursquare API is limited to Manhattan Borough in NYC and Neighbourhoods with "Toronto" in their names for Toronto City. Looking at 20 venues within a radius of 500 meters from the centre of each neighbourhood Manhattan has 203 unique venue categories. Cluster analysis indicate that there are specific clusters in both Manhattan and Toronto with an eating place (restaurants/Cafes/Pub-Bar-Lounges) as the top two most common venues in them, and both the cities have clusters without any eating place at all. Parks, Theatres, Galleries are the common venues in the clusters without any eating places within them.

Venue clusters in the Borough of Manhattan



Cluster of Venues in Toronto

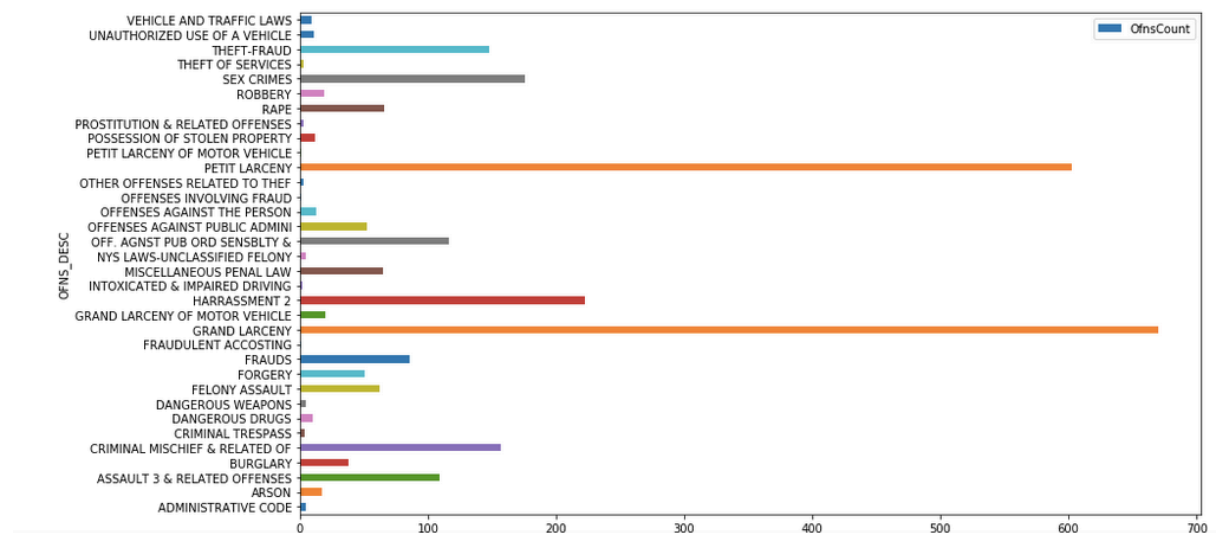


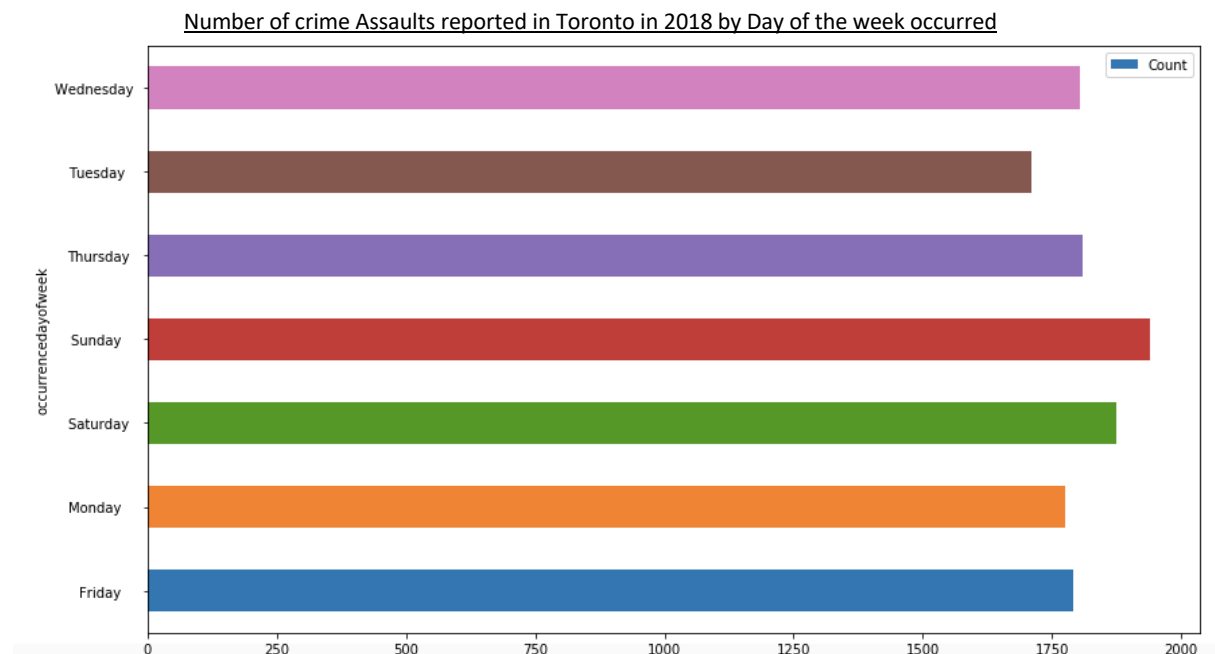
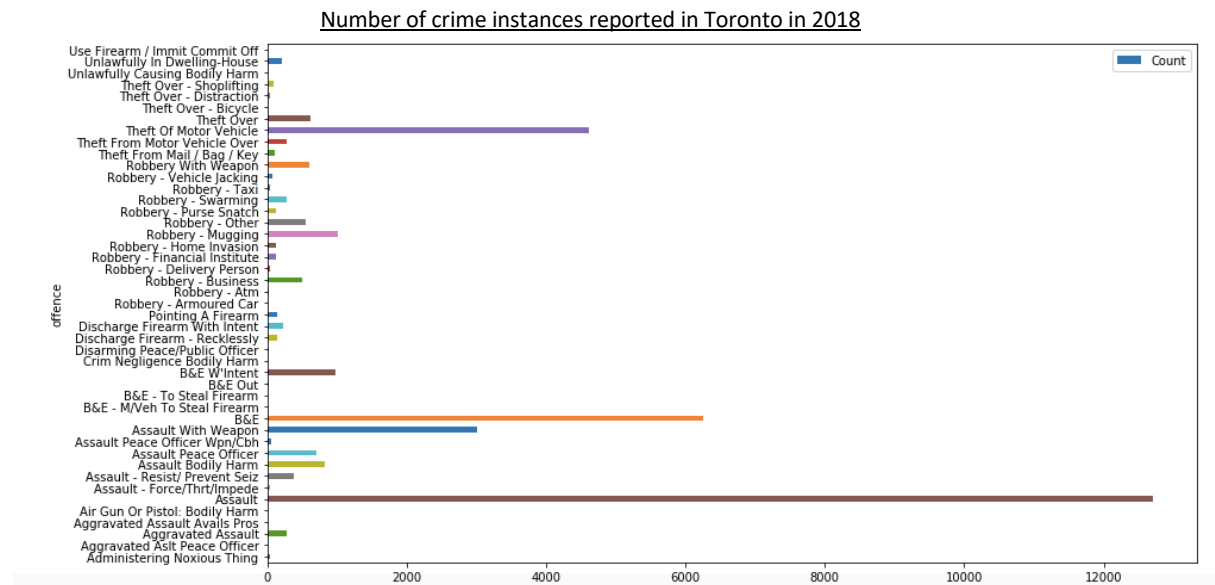
Crime in NYC and Toronto

Based on the available Larceny (Grand and Petit), Harrasment 2 are top 3 frequently reported crimes in New York City, while Assault, Breaking & Entering, and Motor Vehicle Theft are the most commonly reported crime in Toronto city. The borough of Brooklyn had the highest number of Larceny reported at 424 (more than one incident a day), followed by Manhattan that had 364 incidents reported.

Day of the week distribution of Assault in Toronto shows the number of assaults on Sunday is the most followed by Saturday and Wednesday

Number of crime instances reported in NYC in 2018





5. **Conclusions:** The objective of the project was to look at venues in the cities of New York and Toronto, the data suggest that the two cities are not only similar to each other in some aspects but also are different to each other in certain aspects. Both the cities have neighbourhoods which have Theaters, Parks, and Other public places as the most common venues. Certain neighbourhoods in both the cities cater to specific cuisine. E.g. Mexican Restaurant is the most frequently found venue in East Harlem, Manhattan. While Breakfast Spots are the most common venue in Parkdale Toronto. We also found that Cluster 3 in Manhattan and Clusters 1, 2, and 3 in Toronto might be the places one can explore to set up a restaurant or a cafe.
- Larceny and Harassment are the frequently reported crime in New York city whereas Assault, B&E, and MV Theft are the most commonly reported incidents in Toronto City

6. Future Directions: Further data and analysis is required to explore the relationship between the type of Venues and the demographics of any particular neighbourhood in both cities and also the relationship between Crime frequency, nature etc to the demographics of a particular location in the two cities. Reverse geocoding is another step that we would like to explore visualizing crimes on an interactive map using Nominatim package. Perform this venue analysis again to observe if the venues and cluster membership remain consistent.