Where Two Cities meet and to disagree

(A case study of Application of Foursquare and Machine Learning algorithms)

- The Battle of Neighborhoods -

By

Marcel I. Ngwueke

Submitted to IBM-COURSERA
For the award of IBM Professional Data Science Certification

Table of Content

Part I – In the Beginning

1.1 Introduction

In this project, we apply Foursquare location data to the investigation of two famous cities in the world. These cities are located in the world classified as developed nations, U.S., and Canada. The study is going to be a case study to the concepts of Machine Learning of Data science through the application of Foursquare location data or preferably a foursquare venue database. Foursquare API check-in had been used to redefine the notion of a neighborhood by clustering city venues into aggregate areas depicting the check-in patterns of like-minded peoples.

It's possible, as shown by Cranshaw and Yano[1] to partition cities into grids, and then cluster the output grid in terms of the types of venues found there, which could reveal patterns of land usage across cities.

Moreso, individuals seeking restaurants, bars, movie theatres, coffee shops, malls, spas, colleges, or schools within a city can use the foursquare API to locate any of these places of interest or entertainment. Similarly, corporations that require new office complexes use this application to achieve this purpose. All these are affected by residents, visitors, and the flow of people within this environment or neighborhood. Based on these factors, venues can be recommended for shop complexes or other businesses.

1.2 Study Review Involving the Two Cities

City analysis can be classified into two types, crowdsourcing, and social data system. Using Foursquare location data falls under the latter category, which is a location-based social network. It involves data generation about human activities in city community areas. Foursquare API will facilitate getting venues of important places. And more importantly, we will obtain relevant information for each nearby venue after sending a request.

The city of Chicago is one of the largest cities in the U.S., with a population of close to 3 million. It's located on Lake Michigan in Illinois, and is famous for its magnificent edifices or architecture, with its skyline adorned by skyscrapers such as the John Hancock Center, Willis Tower, and the neo-Gothic Tribune Tower. Like most prominent cities in the world, the City of Chicago is famed for its over 600 fascinating museums, 77 unique community areas, which are grouped into nine distinct sections namely central, far northside, far southeast side, far southwest side, south side, southwest side, and west side; with over 300 informally defined neighborhoods. It has thousands of acclaimed restaurants, bars, and breweries, amphitheaters, colleges, and universities.

Toronto is another city we would study. One thing about Toronto is its diversity and culture, which is reflected in its many neighborhoods. The city of Toronto is composed of 140 neighborhoods grouped into 6 distinct Boroughs. It’s the capital of the province of Ontario and is a major Canadian city along the lake Ontario's Northwestern shore. Toronto, as of 2017, has about 2.93 million inhabitants living in it; and it’s famed with a core of soaring skyscrapers, all dwarfed by the iconic free towering CN Tower.

Some of the neighborhoods have interesting places like the Art Gallery of Ontario, authentic Chinese, Japanese, Thai, and Vietnamese cuisines, Luckee Restaurant and bar, Yuk Yuk’s comedy club. With the application of the Foursquare location data query search, we can identify these places of interest.

In terms of population, New York is the largest and biggest city in the U.S. It has about 8.6 million inhabitants and is composed of five remarkable Boroughs with about 329 unique neighborhoods. New York City, popularly known as ‘the Big Apple’, and ‘the city that never sleeps’, is the home to many of the world’s most iconic neighborhoods, restaurants, museums, and neighborhoods. It is also one of the largest cultural and financial hubs in the world. Manhattan is the core of the city of New York, and it is a densely populated borough that stands among the world's major commercial and financial, and cultural centers.

1.2 Business Problem

The year 2020 has witnessed tremendous events that changed so many things. The foremost is the pandemic that sweeps across the entire universe, that people now live in apprehension of being infected knowing the possible consequence that would follow. So many countries, cities adopted locked-down to curtail the spread within the city neighborhoods, and reduce mortality rates.

A sequel to that is the continuous increase in violence, shootings, and criminal activities as well as social unrest due to police brutality on a certain ethnic group. We augment these into this study to determine its impacts on these cities and their neighborhoods.

(i) This study intends to compare and contrast two prominent cities (Chicago. Toronto) utilizing a clustering technique based on the venue categories of the different cities' neighborhoods. It can help a potential job seeker to match his skills with the location of the neighborhood of his choice.

(ii) We also examined all these cities to ascertain how crime-free their neighborhoods are. This will facilitate someone buying or renting a home to make a firm decision of eschewing crime-infectious neighborhoods.

(iii) Exploring possible relationships between these cities will closely expatiate the similarities and dissimilarities as a result of the finding from the Foursquare API query search.

(iv) Possible benefit for many intra-city developers to compute models based on the city's relationship to other municipalities.

Part II - Data Wrangling

2.1 Introduction of Data Wrangling

In data wrangling, certain processes must be performed to bring the data to a standard that would be able to give the required results needed for analysis. This section would involve the process of manually converting and mapping data from one raw form into another, perhaps format to achieve convenient consumption and presentation of the dataset.

Anomalies occurred due to methods of data acquisition or sources. It may have happened as a result of the instruments utilized in the data collections. Also, human errors may introduce such effects in the data, and if ignored would cause great spurious analytic interpretation leading to erroneous conclusions or inferences.

Most often, acquired data are unstructured ad unorganized which must require organizing and extraction of useful components for easier computation and analysis.

We perform the following operations: data format that converts dates to pandas DateTime. Most of the null values in the Toronto DataFrame, Toronto Crime, and Chicago Crime were formatted. And the Chicago DataFrame that was scraped was complemented with another data frame to enable an analytic process to be executed.

2.2 Scraping from Web pages

2.2.1 Extraction of Toronto Boroughs and Neighbourhoods

The link we used to scrape the city of Toronto, Canada, is https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. The extracted data frame contains over 189 rows and 3 attributes or columns with the index. Observations show that his data frame has cases of null or nan or 'not assigned' values, and as well as not having Latitudes and Longitudes for all the Postal Codes or Boroughs. In this case, the corresponding Latitudes and Longitudes must be derived or determined for all the Postal Codes in the city of Toronto. These known anomalies we shall remove during the preprocessing or cleaning stage in this section.

2.2.2 City of Chicago Neighborhoods

For us to derive the required data frame for the city of Chicago, we need to scrape two links from Wikipedia and Chicago Tribune. https://en.wikipedia.org/wiki/Community_areas_in_Chicago and https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago. The first link enables us to obtain Community Areas, Population, Population density, and areas as columns, while the second link gives the neighborhoods.

An additional resource we use to examine the population, area, and population density is https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods.

2.3 Preprocessing City of Toronto Datasets

We scrape Wikipedia: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, using requests. get, that retrieves the webpage containing the required data, and parse it utilizing the HTML module, which becomes transformed into a pandas data frame by pd.read_html. The *df_list[0]* obtains the first table on the webpage. Looking at the data frame *df_t*, the table output showed that there were three columns and 189 rows. It had several null values that appeared as "not assigned", the equivalent of 'nan'.

We clean the data frame by dropping all the rows that contain 'not assigned' and invoked "reset_index" for the data frame. This dataframe lacks some useful components that would facilitate the analytic processes of the dataframe. We must couple Latitudes and Longitudes to the corresponding Postal Codes of the city of Toronto. So, we used the "geocode" Nominatim library module to generate latitudes and longitudes to the postal codes of the mentioned dataframe.

Some of the columns of this new dataframe tor1 were not required for our overall analysis, therefore we discarded, and using fillna to insert missing longitude and latitude. The resulting dataframe tor1 has 3 attributes (Postal Code, Latitude, and Longitude)

Looking at these two sets of DataFrames (df_tor and tor1) we noticed the unique difference that would make the two accept merging. The columns of the tor1 dataframe were renamed, for instance, postal_code to Postal Code, latitude to Latitude, and longitude to Longitude. The merge was based on terms of their common column, which is the Postal Code. Merging the two dataframes df_tor and tor1 results in a new dataframe known as Toronto, which has 103 rows and 5 columns. This 'Toronto' dataframe we shall use for our analysis to determine the clustering of neighborhoods as well as determining venues and venue categories when we apply the Foursquare API.

2.4 Preprocessing of City of Chicago Dataset

The city of Chicago is another important city in the US; in fact, it is the third largest and prominent city after New York City and Los Angeles. Let's extract or scrape datasets for this city and match them up to obtain the desired goal for our investigation or study.

We scraped the Wikipedia web page mentioned in section 2.2.2 above. The dataframe df_ch has 7 columns and 78 rows. Since the columns did not indicate the names of the columns we modify them to depict and represent appropriate column-names, and at the same time dropping any columns we might not need in our analysis. So, df_ch will now have 4 columns only. Row 31 has a value that looks out of place, we change the value to fit in the proper position. Originally, it was (The) Loop_ and we want it to be in the form "The Loop".

Next, we obtain the Chicago Neighborhood dataset by scraping it from the Wikipedia webpage (https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago). The resulting dataframe dfC was modified by cleaning it up and save to our local machine df_g1.

However, the outcome of this scraping did not meet the required standard. We used it to generate a dataframe which we saved to our local device. The df_g1 is now the constructed dataframe from the above dataframe (df_g2). This has 4 columns or attributes quite different from the original one with Community Area Number, Community Areas, Latitude, and Longitude.

We concated the two dataframes to obtain the actual dataframe we shall use for our further analysis. We used the community area number as the key to achieving the objective, the probability there might be no loss of information as in most cases of if merging. In the above dataframe, we wanted the Community Areas and Neighborhoods to be closed to each other for easy identification and verification; and equally, the Latitude and Longitude remain in their current positions. So, df_Chica is the final dataframe we would use for our analysis, and it has 77 Community Areas and 77 Neighborhoods with 7 columns and 77 rows.

2.5 Other Datasets for This Study

2.5.1 Mapping Crimes with Neighborhoods of Chicago

The dataset we shall use here is Chicago_Crimes_2012_to_2017, a csv file format, which we downloaded from the internet and was saved to our local device. The first step we took was to clean the data by removing unknown columns like Unnamed: 0 and dropping columns that are not relevant to our investigation, such as ID and case number, leaving a total of 19 columns and 1456714 rows. One more thing we did, was to rename all the columns.

Next, we converted the dates of the chic_crimes dataframe to pandas DateTime format and later set it as the index.

2.5.2 Preprocessing Toronto – Crime Dataset

Toronto Crime dataset contains 206435 rows and 16 columns. It was obtained from the Toronto Police service, public safety data portal.

Three steps were taken to clean and mug this dataframe. The first step was to drop some columns that were considered redundant. This process was followed by renaming the remaining columns. The last step after checking and removing null values was to select the range of data observation between two datetimes 2014 to 2018.

Part III – Exploratory Data Analysis (EDA)

3.1 Introduction to EDA

In this section, we would analyze and investigate all the datasets that were preprocessed to generate dataframes in the previous part above, by using the appropriate data science techniques such as data visualization, foursquare API, clustering, folium, exploring Neighborhoods in both cities of Toronto and Chicago.

Subsequently, we shall examine each cluster and determine the discriminating venue categories that distinguish each other. Also, we shall demonstrate the neighborhoods that top the crime rate in both cities.

3.2 Foursquare API Toronto Venue Search

3.2.1 The City of Toronto and Its Neighborhoods

City of Toronto dataset, boro_toronto1 had a total of 103 rows and 5 columns, in which 10 Boroughs were identified. These 10 Boroughs were grouped into 5 groups, namely Toronto, York, Scarborough, Etobicoke, and Mississauga. Figure 1 shows the Folium map of the City of Toronto based on the 10 boroughs.

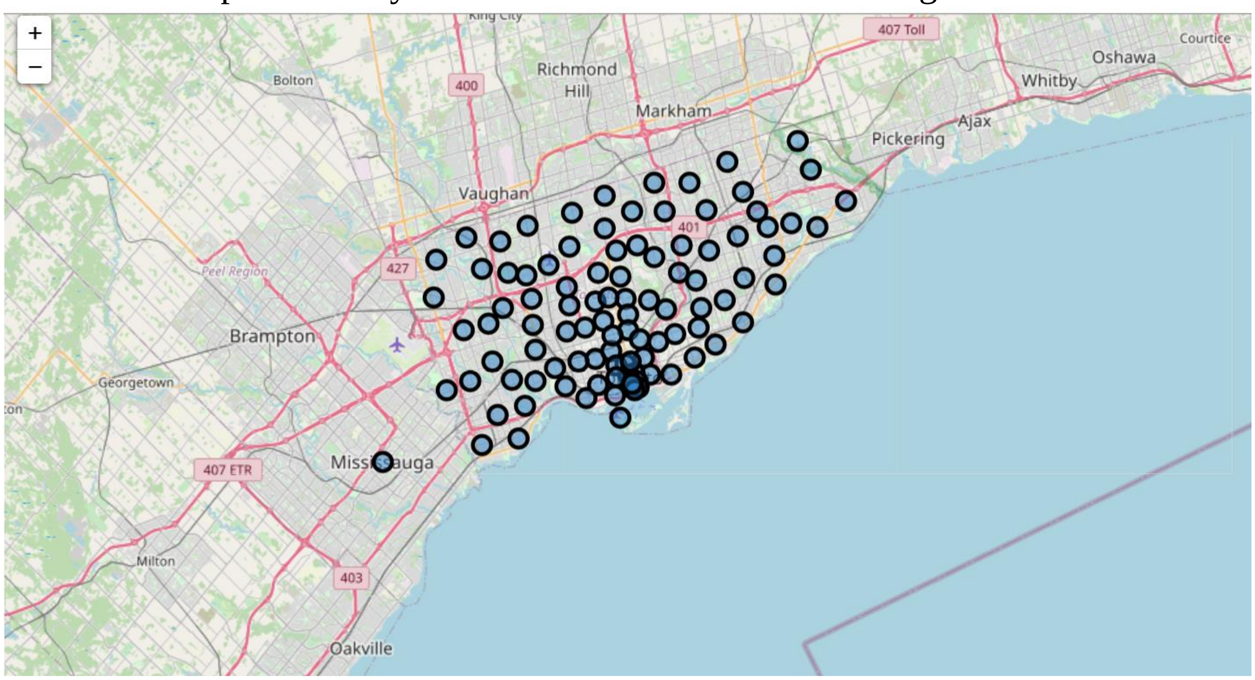

Figure 1: Folium Map of City of Toronto

Each of the boroughs mentioned, the geographical coordinates were obtained using the geolocator (Nominatim) and geocode. There are over 360 neighborhoods, however, we presented a sample of these neighborhoods in Table 1.

Table 1. Sample of Toronto Neighborhoods after Preprocessing.

Neighborhoods	Latitude	Longitude
Parkwoods	43.7545	-79.33
Victoria Village	43.7276	-79.3148
Regent Park, Harbourfront	43.6555	-79.3626
Lawrence Manor, Lawrence Heights	43.7223	-79.4504
Queen's Park, Ontario Provincial Government	43.6641	-79.3889
Islington Avenue, Humber Valley Village	43.6662	-79.5282
Malvern, Rouge	43.8113	-79.193
Don Mills	43.745	-79.359
Parkview Hill, Woodbine Gardens	43.7063	-79.3094
Garden District, Ryerson	43.6572	-79.3783
Glencairn	43.7081	-79.4479
West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale	43.6505	-79.5517
Rouge Hill, Port Union, Highland Creek	43.7878	-79.1564
Don Mills	43.7334	-79.3329
Woodbine Heights	43.6913	-79.3116
St. James Town	43.6513	-79.3756
Humewood-Cedarvale	43.6915	-79.4307
Eringate, Bloordale Gardens, Old Burnhamthorpe, Markland Wood	43.6437	-79.5767
Guildwood, Morningside, West Hill	43.7678	-79.1866

The city of Toronto Neighborhoods spread towards the North, North-East, North-West of Toronto. The Northside was the York while the central zone is the Downtown Toronto which encompasses north, northeast, and central (see figure 2)

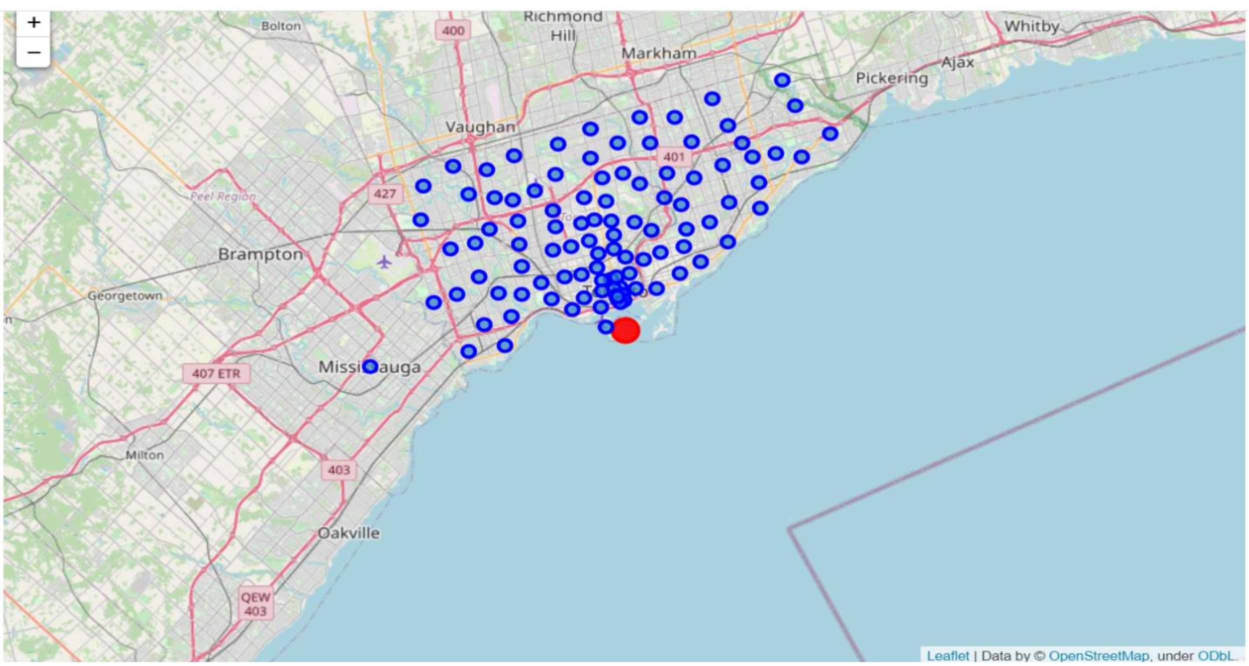

Figure 2: Map_City_Toronto Neighborhoods

3.2.2 Foursquare API to Retrieve Toronto Neighborhoods' Venues

Foursquare API facilitated getting venues of important places of interest near the city of Toronto. Relevant information for each nearby venue was obtained after sending out a request through the foursquare API.

Table 2 shows the results, which were the venue and the venue category. The venue category describes the venues that were returned. For instance, the venue, Brook banks Park found in the Parkwoods neighborhoods, is a Park by venue category convention.

Table2: Location Venues for the city of Toronto Neighborhoods

Neighborhood	Neighborhood Lat.	Neighborhood Long.	Venue	Venue Latitude	Venue Longitude	Venue Category
Parkwoods	43.7545	-79.33	Brookbanks Park	43.751976	-79.332140	Park
Parkwoods	43.7545	-79.33	649 Variety	43.754513	-79.331942	Convenience store
Parkwoods	43.7545	-79.33	Towns on The Ravine	43.754754	-79.332552	Residential Building
Parkwoods	43.7545	-79.33	Subway	43.750334	-79.336906	Sandwich Place
Parkwoods	43.7545	-79.33	Food Basic	43.750549	-79.336045	Supermarket

Further examination shows how many venues would be returned for each Neighborhood. Agincourt neighborhood returned 75, while Alderwood, Long Branch had 77. Similarly, Bathurst manor, Wison Height, Downsview North, and Bayview Village had 83 each.

3.2.3 The City of Toronto Venue Categories

Based on the previous section 3.2.2, we can determine how many unique categories can be curated from all the returned venues. For the city of Toronto, there are a total of 509 unique categories.

Next, we verify how often these unique categories occurred. Table 3 demonstrates a sample of 4 venue categories.

Table 3: Sample of Toronto Venue Categories

VC	N	N Lat.	N Long.	Venue	Venue Lat	Venue Long.
ATM	4	4	4	4	4	4
Accessories Store	15	15	15	15	15	15
Acupuncturist	4	4	4	4	4	4
Adult Boutique	3	3	3	3	3	3

Nota Bene: VC – Venue Category, N - Neighborhood, N Lat – Neighborhood Latitude, N Long – Neighborhood Longitude.

A synopsis shows those venue categories that are very outstanding in terms of frequencies or presences. Such venue categories like American Restaurant (25), Asian Restaurant (15), Bakery (34), Bar (23), Breakfast spot (25), and from Table 3 above, Accessories store (15). These are the most frequently observed venues. The category of that neighborhood is defined by the occurrence of these venues, which shall be used to investigate the clustering of these Neighborhoods in the subsequent section.

Moreover, the mean frequency of occurrence of each category in a neighborhood was determined. And the outputs were used to calculate the top 5 most frequent of its venues for each neighborhood as seen in the main code (because of conciseness we are unable to present in this report). These were used to create the top 10 most common venues per Neighborhood (see Table 4).

Table 4: Top 10 Most Common Venues Per Neighborhood

N	1st MCV	2nd MCV	3rd MCV	4th MCV	5th MCV	6th MCV	7th MCV	8th MCV	9th MCV	10th MCV
Agincourt	Chinese Restaurant	Supermarket	Storage Facility	Hardware Store	Furniture / Home Store	Factory	Food Court	Gas Station	Automotive Shop	Doctor's Office
Alderwood, Long Branch	Convenience Store	Factory	Design Studio	Bank	Medical Center	BBQ Joint	Spa	Asian Restaurant	Lounge	Breakfast Spot
Bathurst Manor, Wilson Heights, Downsview North	Convenience Store	Synagogue	Salon / Barber shop	Residential Building (Apartment / Condo)	Ice Cream Shop	Doctor's Office	Medical Center	Bar	Coffee Shop	Bank
Bayview Village	Dentist's Office	Women's Store	Park	Church	Salon / Barbershop	Clothing Store	Doctor's Office	Kids Store	Kitchen Supply Store	Boutique

Nota Bene: MCV – Most Common Venues
N - Neighborhood

3.2.4 Clustering of Toronto Neighborhoods

For the clustering of the city of Toronto Neighborhoods, we ran the K-means clustering, setting K = 5, and checked the cluster labels generated for each row.

We combined all the DataFrames (*toronto*, and neighborhoods_venues_sorted_A, toronto_grouped_clustering_A as seen in the code) including the Cluster Labels, neighborhoods, and the top 10 venues in each neighborhood and venue name (see Table 5).

Table 5: Sample of Cluster labels and Top 10 Most Common Venues Per Neighborhood

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	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
0	M3A	North York	Parkwoods	43.7545	-79.3300	4	Residential Building (Apartment / Condo)	Convenience Store	Park	School	Elementary School	Coffee Shop	Fast Food Restaurant
1	M4A	North York	Victoria Village	43.7276	-79.3148	2	Automotive Shop	Residential Building (Apartment / Condo)	Park	Government Building	Other Great Outdoors	Grocery Store	Auto Dealership
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.6555	-79.3626	1	Rental Car Location	Art Gallery	Ethiopian Restaurant	Coffee Shop	Automotive Shop	Miscellaneous Shop	Antique Shop
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.7223	-79.4504	3	Clothing Store	Salon / Barbershop	Residential Building (Apartment / Condo)	Shopping Mall	Park	Elementary School	Electronics Store
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial	43.6641	-79.3889	1	Government Building	Coffee Shop	College Academic Building	College Administrative Building	Residential Building (Apartment	College Library	Pharmacy

With Folium library we visualized the clusters of the Neighborhood categories (Figure 3). Five Clusters were identified, and each was marked with a certain color to distinguish each from one another. For instance, Cluster 3 in the figure below is green.

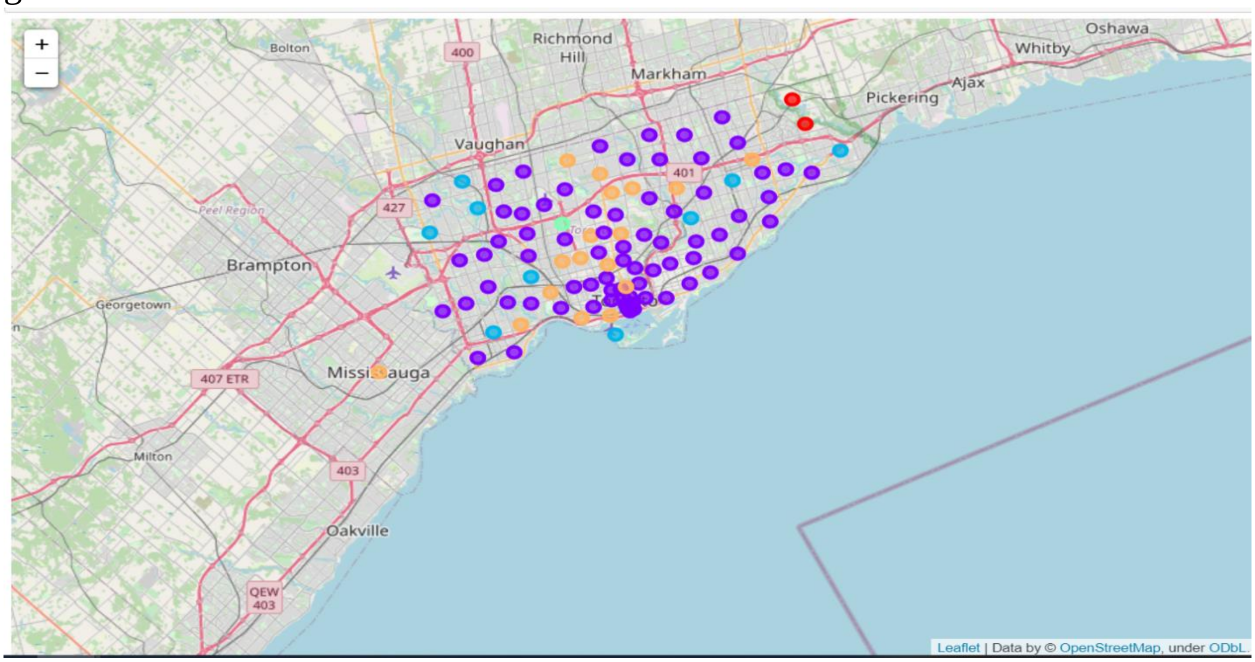

Figure 3: Map_Clusters_Of _Toronto Neighborhoods

The green marker indicates cluster 3. Again, the number of Boroughs in each cluster varies from one cluster to another. For instance, cluster 4 has seventeen Boroughs, while Cluster 3 has only but one Borough. Also, most top venues for cluster label 0 are either Doctor's offices, some Residential Buildings, Medical center, some foreign restaurants such as Chinese, Caribbean, India, etc. And these were designated in red color.

3.2.5 Toronto Crime – Analysis

Ten types of crimes were identified in the city of Toronto from the period between 2014 to 2019. Assault stands as the highest within this time band (see Figure 4 below). Robbery with a weapon had about 3,578 reported cases, however, whether charged or not charged were not mentioned.

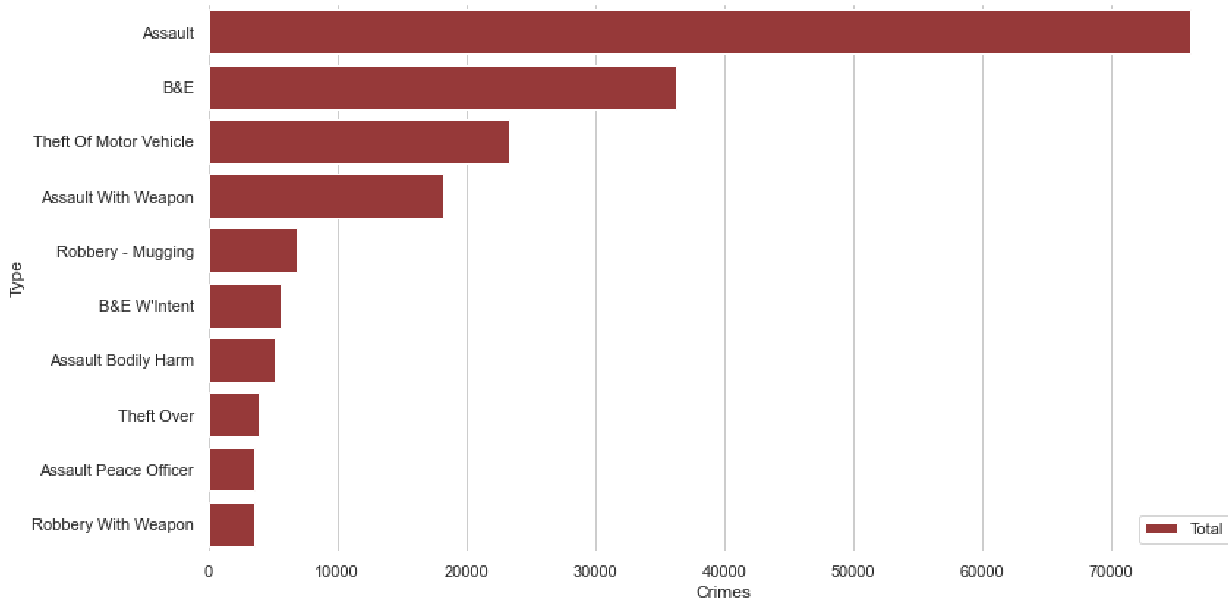

Figure 4: City of Toronto types of crimes with occurrence rate.

To articulate the number of crimes committed, we used the count method in terms of offense and premise-type, which were grouped for every occurrence year, occurrence month as per the event's unique id (as shown in Figure 5). We saw a sharp drop in 2015 and a rapid spike increase in 2017 and 2018.

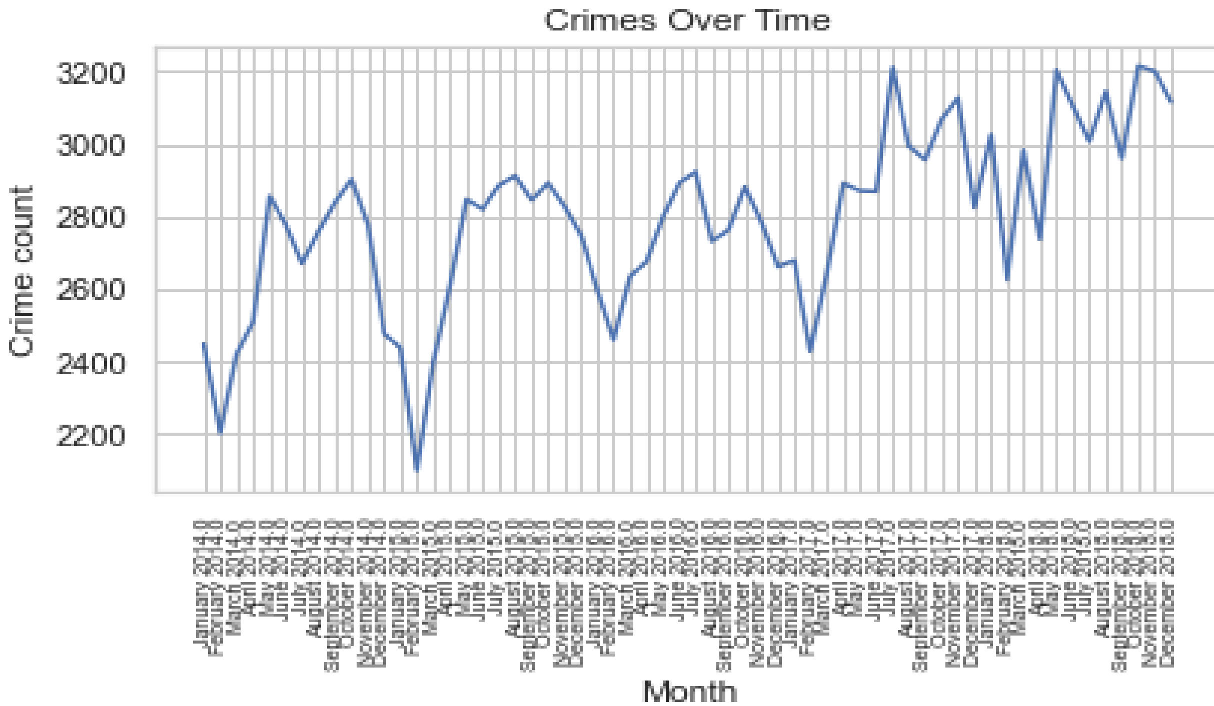

Figure 5: Toronto Crimes Over Time from 2014 to 2019

Evaluating top crimes that have trended over the past years in the city of Toronto were shown in Figure 5. Assault was outstanding with a continuous rise in its rate, as compared to Theft-Over, which had been very low and flattened over the years from 2014 to 2018. BreakandEnter was also high in comparison to AutoTheft and Robbery crimes as shown in Figure 6.

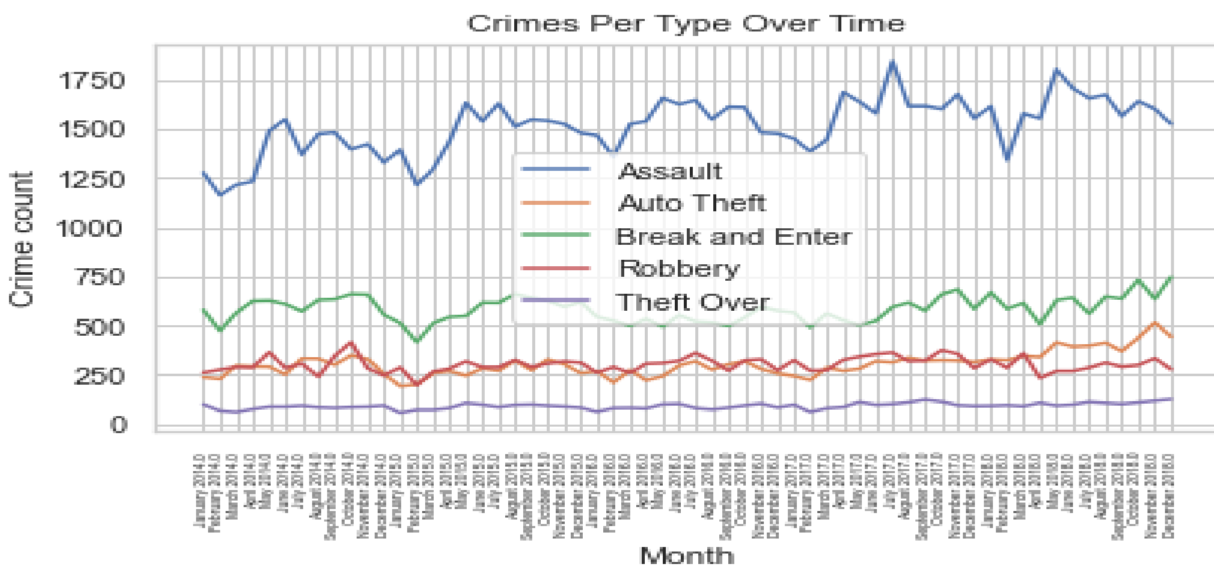

Figure 6: City of Toronto Crimes Per Type Over Time.

Figure 7 shows the crimes overtime per neighborhood. Waterfront Communities – The Island has the highest rate of crimes and was followed closely by the Bay Street Corridor and Church-Yonge Corridor. Moss Park (23) has the least in terms of crimes overtime per neighborhood.

We further analyzed individual crime types such as Robbery and Assault. These were performed based on frequencies of occurrence per month and year for the period between 2014 and 2019. Robbery cases were mostly committed outside relative to House (see Figure 8). Also, in 2017, we saw the highest number of Robbery cases committed as shown in Figure 9. Assault cases reported shown to be on increase from 2014 to 2018 (see Figure 10).

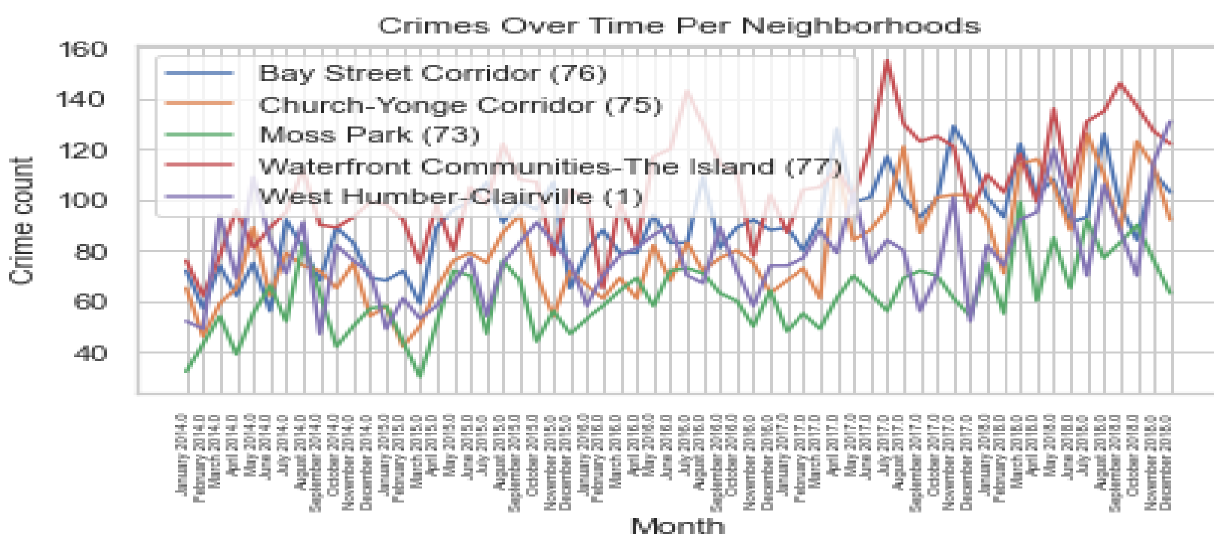

Figure 7: City of Toronto Crimes Over Time Per Neighborhood

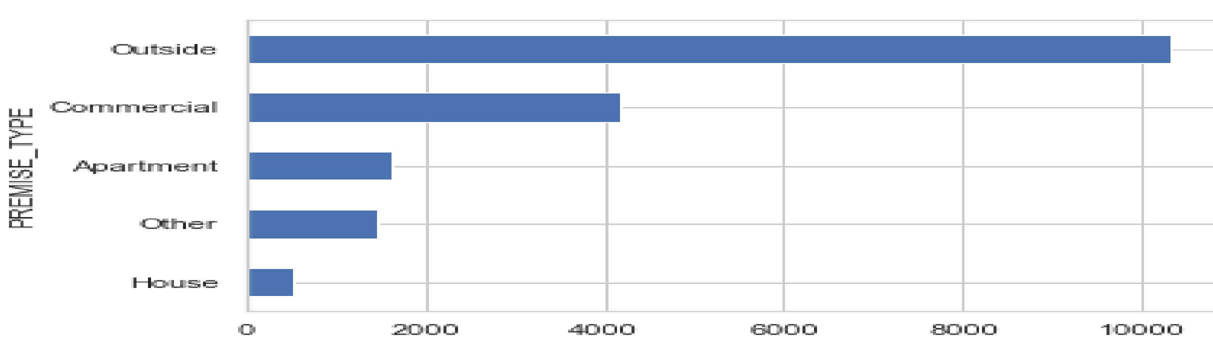

Figure 8: Committed Robbery Cases and Locations

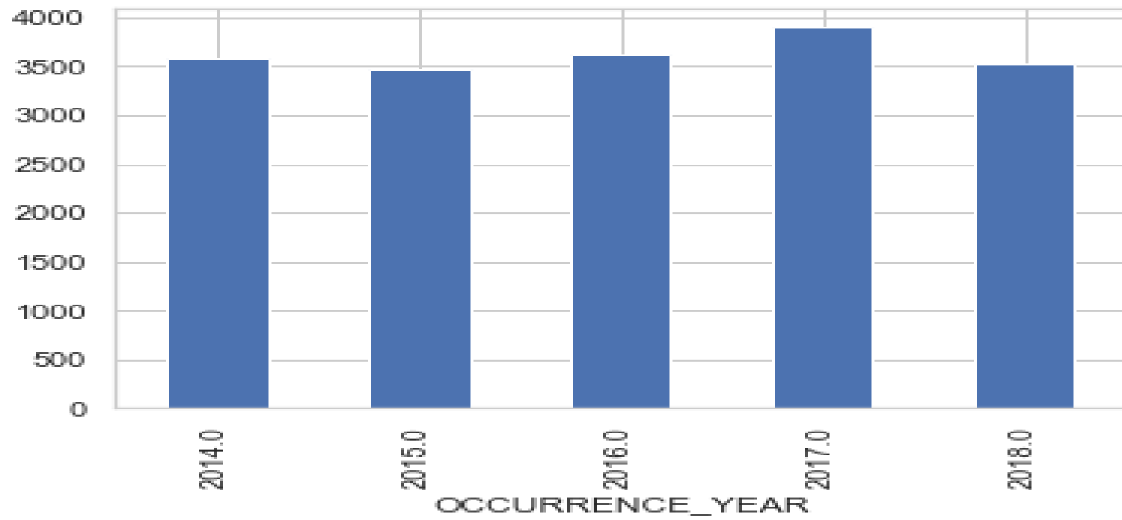

Figure 9: Robbery Cases over the occurrence year.

We also considered the top number of neighborhoods associated with crimes. And the neighborhoods where different crimes are menacingly more compare to other neighborhoods.

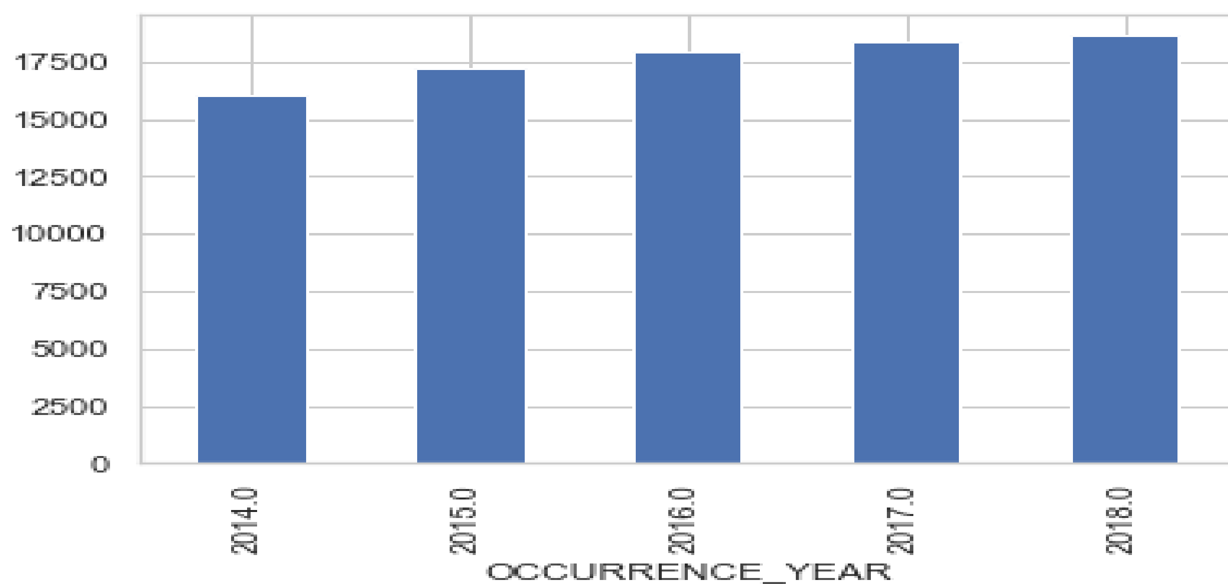

Figure 10: Rising Cases of Assault from 2014 to 2018

As shown in Figure 11, a high number is an indication of the number of occurrences in that Neighborhood.

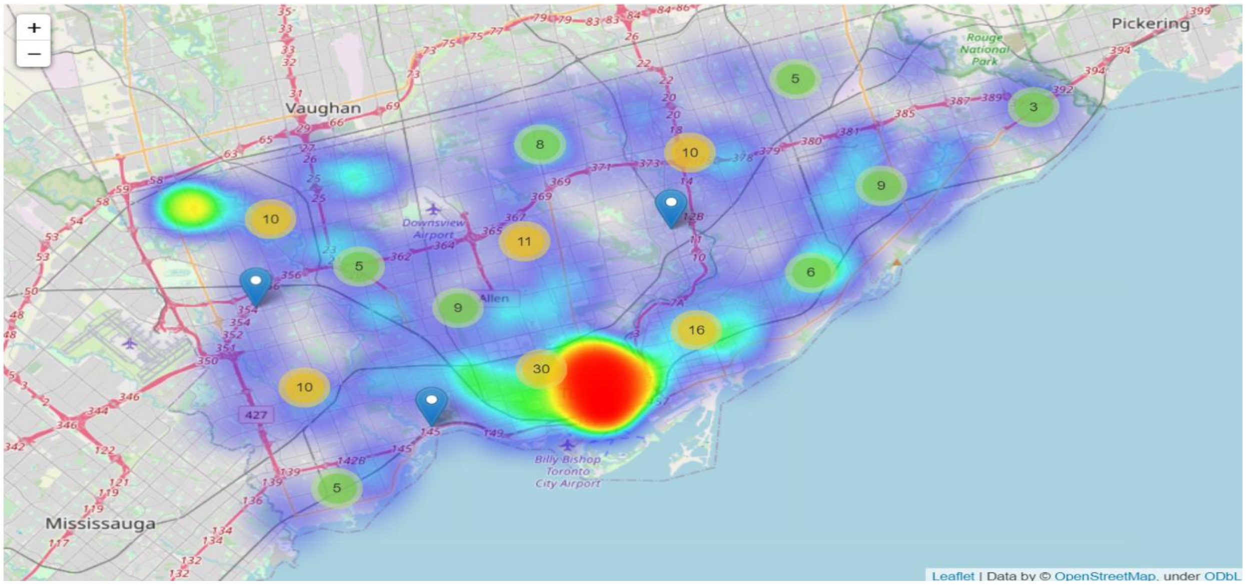

Figure 11: Toronto Neighborhoods with a high number of crime rates.

3.3 Foursquare API City of Chicago Venue Search.

3.3.1 The City of Chicago and Its Neighborhoods

Unlike the city of Toronto, the city of Chicago uses Community Areas instead of Boroughs. There are 77 Community Areas and more than 200 Neighborhoods in the city of Chicago. From the dataframe, df_Chica (see the main code), three key geographical sections surrounded Downtown Chicago, which is The Loop. These are the North Side, the South Side, and West Side as divided by the Chicago River. Table 6 shows a sample of the neighborhood surrounding with latitude and longitude.

For each of the Community Areas, the geographical coordinates were derived using the geolocator (Nominatim) and geocode. Figure 12 shows the Folium map of the city of Chicago

More so, we created a map of Chicago and its surrounding Neighborhoods, where Chicago (The Loop) is shown as Red marker on the map (Figure 13)

Table 6: Sample of Chicago Neighborhoods after Preprocessing.

Neighborhoods	Latitude	Longitude
Loyola, Rogers Park	42.007	-87.67799
LincolnPark, Old Town Triangle, ParkWest, Ranch	41.926	-87.6488
Cabrini–Green, Gold Coast, Goose Island, Magnif.	41.903999999999996	-87.6315
Edison Park	42.0035	-87.8171
Big Oaks, Norwood Park East, Norwood ParkWest,	41.9858	-87.8069
Gladstone Park, Jefferson Park	41.9825	-87.7704
Hollywood Park, North Park, Rivers Edge	41.9828	-87.7284
Albany Park, Mayfair, North Mayfair, Ravenswood	41.9683	-87.728
Portage Park	41.9532	-87.7646
Avondale Gardens, Irving Park, Kilbourn Park	41.9538	-87.7143
Humboldt Park, West Humboldt Park	41.8991	-87.7218
West Garfield Park	41.8779	-87.7305
East Garfield Park, Fifth City	41.881	-87.7012
Fulton River District, Greektown, Illinois Med.	41.8668	-87.6664
East Pilsen, Heart of Chicago, Lower West Side	41.8523	-87.666
The Loop, New Eastside, Printer s Row, South	41.8786	-87.6251
Central Station, Dearborn Park, Museum Campus,	41.8608	-87.6257
Fuller Park	41.8091	-87.6334
Washington Park	41.7945	-87.616
East Hyde Park, Hyde Park	41.7948	-87.5917
Avalon Park, Marynook, Stony Island Park	41.7442	-87.5856
East Side	41.708	-87.5352
Brighton Park	41.8171	-87.6994
McKinley Park	41.8316	-87.6729
Gage Park	41.7935	-87.7036
Beverly Woods, Kennedy Park, Morgan Park, West.	41.687	-87.6692

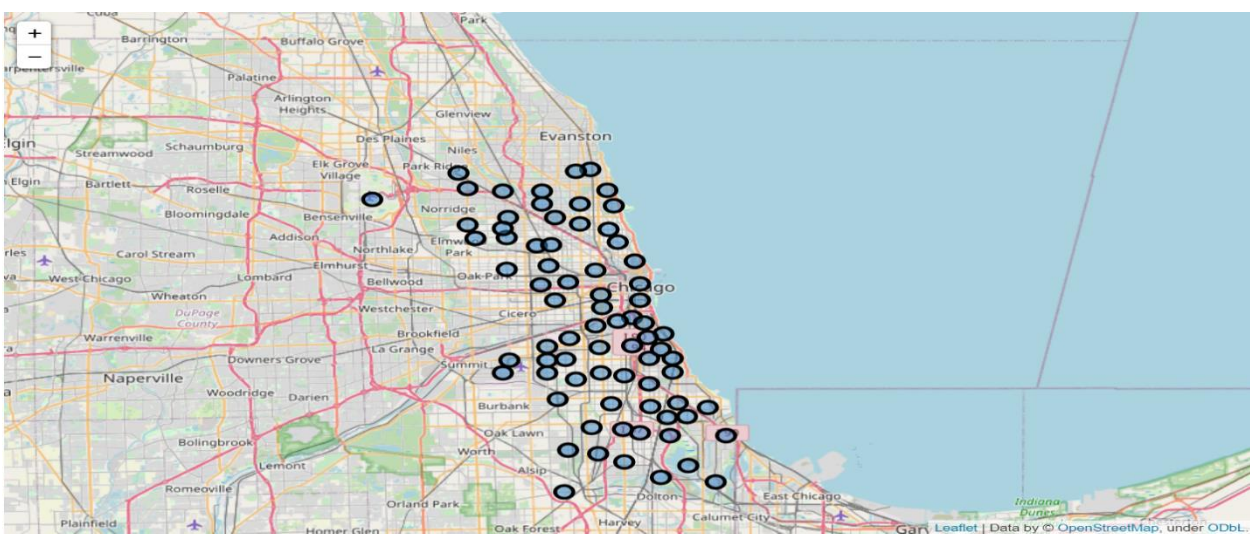

Figure 12: Folium Map of City of Chicago with its surrounding Neighborhoods

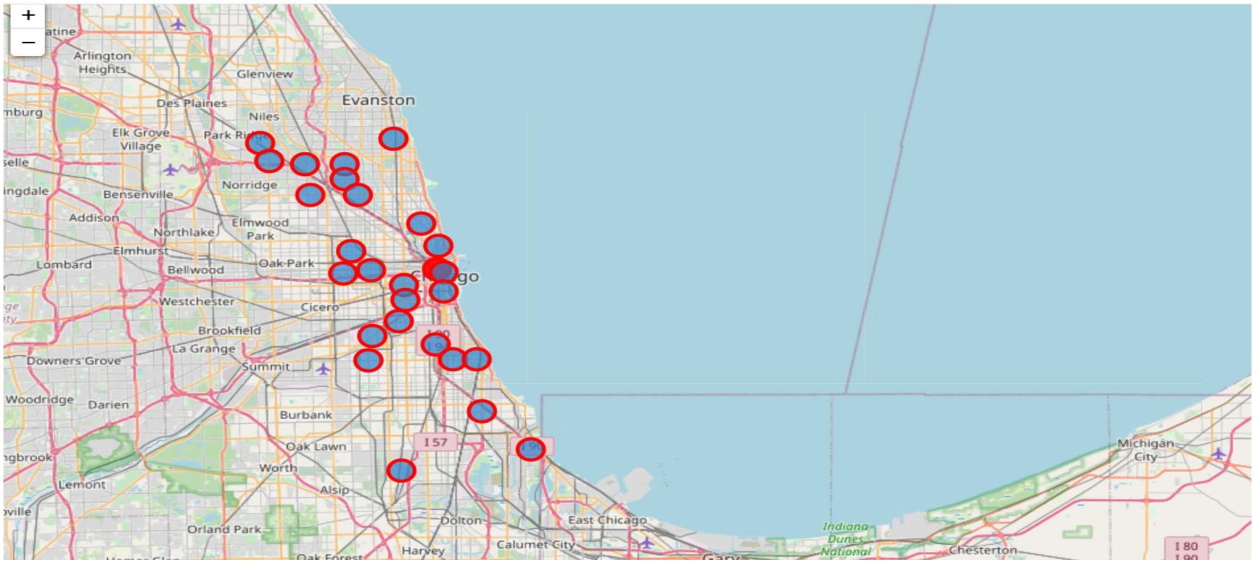

Figure 13: Chicago and its surrounding Neighborhoods with The Loop as red

3.3.2 Foursquare API To Retrieve Neighborhoods' Venues for Chicago

For every venue search request, the Foursquare API credentials and current version must be provided. Foursquare will then retrieve venues of important places in the city of Chicago. It will be also possible to obtain relevant information for each nearby venue from the foursquare API. After dropping records where the venue was building or office, we had Table 6, which showed the neighborhood, neighborhood latitude and longitude, venue, venue latitude, and longitude and venue category. This table is a sample of the neighborhood venue generated with categories. These were retrieved from different Community Areas and neighborhoods just like the Loyola, Rogers Park.

For every neighborhood, we checked how many venues were returned. For instance, the neighborhood called Albany Park, Marynook, and Stony Island Park had a total of 87 venues while Avondale Gardens, Irving Park, and Kilbourn Park returned about 80 venues. Similarly, other neighborhoods returned a certain number of venues, which the Python code could generate the entire results for the city of Chicago neighborhoods.

Table6: Location Venues for City of Chicago Neighborhoods

N	N. Lat.	N. Long.	Venue	Venue Latitude	Venue Long	Venue Category
Loyola, Rogers Park	42.007	-87.677996	St. Jerome Parish	42.008652	-87.672294	Church
Loyola, Rogers Park	42.007	-87.677996	The Wooly Mammoth	42.006889	-87.678305	Antique Shop
Loyola, Rogers Park	42.007	-87.677996	Olshansky's Estate	42.006835	-87.678124	Assisted Living
Loyola, Rogers Park	42.007	-87.677996	West Ridge Chicago	42.007141	-87.683101	Professional & Other Places
Loyola, Rogers Park	42.007	-87.677996	**Casa Roman**	42.008693	-87.676377	Convenience Store
Loyola, Rogers Park	42.007	-87.677996	**Metro by T-Mobile**	42.008618	-87.674356	Mobile Phone Shop

N – Neighborhood, Lat. – Latitude, Long. - Longitude

3.3.3 City of Chicago Venue Categories

There was a total of 331unique categories returned for the city of Chicago. It is possible to determine how many possible unique categories can be curated from all the returned venues. Furthermore, we verified how often these unique categories occurred. As shown in Table 7, which is a sample table showing the first four rows, namely ATM (3), Accessories store (1), Adult Boutique (2), and American Restaurant (17). Others that are found in this table include Bakery (1), Café (11), Coffeeshop (16), Church (85), Pubs, Hotels, Italian, Japanese, and American Restaurant are the most frequently observed venues.

Table 7: Location Venues for the city of Chicago Neighborhoods

N	N. Lat	N. Long	Venue	Venue Lat	Venue Long	Venue Category
Loyola, Rogers Park	42.007	-87.677996	St. Jerome Parish	42.008652	-87.672294	Church
Loyola, Rogers Park	42.007	-87.677996	The Wooly Mammoth	42.006889	-87.678305	Antique Shop
Loyola, Rogers Park	42.007	-87.677996	Olshansky's Estate	42.006835	-87.678124	Assisted Living
Loyola, Rogers Park	42.007	-87.677996	West Ridge Chicago	42.007141	-87.683101	Professional & Other Places
Loyola, Rogers Park	42.007	-87.677996	Casa Roman	42.008693	-87.676377	Convenience Store
Loyola, Rogers Park	42.007	-87.677996	Metro by T-Mobile	42.008618	-87.674356	Mobile Phone Shop

N – Neighborhood, N. Lat – Neighborhood Latitude, N. Long – Neighborhood Longitude, Venue Lat – Venue Latitude, Venue Long – Venue Longitude

For us to investigate the clustering of the city of Chicago neighborhoods in the subsequent section, we shall make use of the category. Furthermore, the mean frequency of occurrence for each category in a neighborhood must be determined and the outputs used to derive the top 5 most frequent of its venues as presented in the Python code. We then used them to generate the top 10 most common venues per Neighborhoods as shown in Table 8.

Table 8: Top 10 Most Common Venues of the city of Chicago Per Neighborhood

	Neighbourhood	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	Albany Park, Mayfair, North Mayfair, Ravenswoo..	Hookah Bar	Automotive Shop	Salon / Barbershop	Coffee Shop	Dentist's Office	Cosmetics Shop	Discount Store	Mobile Phone Shop	Miscellaneous Shop	Mexican Restaurant
1	Avalon Park, Marynook, Stony Island Park	Salon / Barbershop	Dentist's Office	Church	Miscellaneous Shop	Sandwich Place	Burger Joint	Boutique	Fast Food Restaurant	Nail Salon	Car Wash
2	Avondale Gardens, Irving Park, Kilbourn Park	Speakeasy	Convenience Store	Financial or Legal Service	Smoke Shop	Strip Club	Automotive Shop	Salon / Barbershop	Filipino Restaurant	Dentist's Office	Art Gallery
3	Beverly Woods, Kennedy Park, Morgan Park, West.	Doctor's Office	Salon / Barbershop	Church	Park	Medical Center	High School	Shoe Store	BBQ Joint	Dentist's Office	School
4	Big Oaks, Norwood Park East, Norwood ParkWest,..	Taxi	Hospital	Doctor's Office	Medical Center	Emergency Room	Furniture / Home Store	Park	Laundry Service	Church	Café

3.3.4 Clustering of the city of Chicago Neighborhoods

Clustering of the city of Chicago Neighborhoods requires running the K-means clustering which we set at K = 5. Also, cluster labels generated for each row were checked. To perform the above operation requires a combination of all the dataframes (“New_df_Chicago and chica_neighborhoods_venues_sorted, chicago_grouped_clustering”, see the main Python code) which would append the cluster labels to the “New_df_Chicago” columns and the top 10 venues (see Table 9).

Applying Folium library we visualized the clusters of the Neighborhood categories (Figure 13). And five clusters were identified and each was labeled with a certain color to distinguish each from one another. For example, cluster 0 is identified with Red color, and it has only two Community Areas (Fuller Park and Washington Park). Moreso, most top venues for cluster label 0 were either salon/Barbershop, some

Residential Buildings, Medical Center, some foreign restaurants such as Chinese, Caribbean, Indian, etc.

Table 9: **Sample of Cluster labels, Top 10 Most Common Venues Per Neighborhood(Chicago)**

Community Area Number	Community Areas	Neighbourhood	2017 Population	Area(sq mi.)	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Commor Venue
01	Rogers Park	Loyola, Rogers Park	55062	1.84	42.0070	-87.677996	1	Mexican Restaurant	Train	Doctor's Office	Residential Building (Apartment / Condo)	Salon / Barbershop	Professic & Ot Pla
07	Lincoln Park	LincolnPark, OldTownTriangle, ParkWest, Ranch	67710	3.16	41.9260	-87.648800	2	Taxi	Residential Building (Apartment / Condo)	Thai Restaurant	Dentist's Office	Sandwich Place	B
08	Near North Side	Cabrini–Green, Gold Coast, Goose Island, Magnif.	88893	2.74	41.9040	-87.631500	2	Residential Building (Apartment / Condo)	Bank	Moving Target	Gym	Taxi	T Star
09	Edison Park	Edison Park	11605	1.13	42.0035	-87.817100	1	Salon / Barbershop	Bar	Spa	Italian Restaurant	Speakeasy	Seaf Restaur
10	Norwood Park	Big Oaks, Norwood Park East, Norwood ParkWest,..	37089	4.37	41.9858	-87.806900	3	Taxi	Hospital	Doctor's Office	Medical Center	Emergency Room	Furnitu Home St

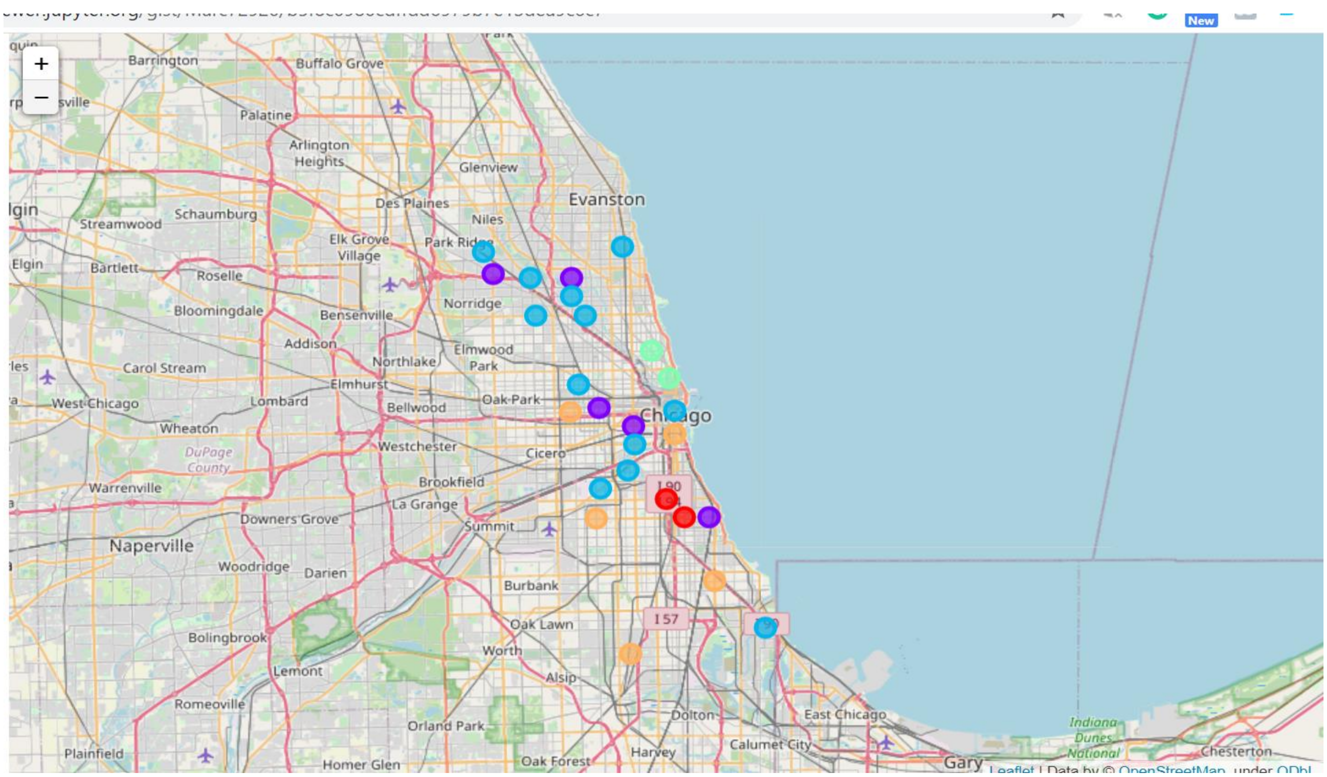

Figure 13: **Map_Clusters_Of _Chicago Neighborhoods**

Cluster 1 was marked as purple, and it has 5 identifiable community areas and plenty of venues, of which Bank, Park, Residential Building (Apartment/ Condo), and Church were among the most common venues. Also, cluster label 2 was marked as blue with 12 neighborhoods. It could be seen that the other cluster labels had color identification.

3.3.5 City of Chicago Crime Analysis

As in the case of the city of Toronto, 10 primary types of crimes were identified in the city of Chicago. The theft was the highest (329,460) followed by Battery and Criminal Damage (263,700 and 155,455 respectively). Narcotics and Assault recorded 135,240 and 91,289 respectively (see Figure 14 for these crime types). From the period, 2012 to 2017, we consider the arrests that took place due to some of the crimes that were committed. Both yearlies, weekly and daily arrests were seen declining from 2012 to 2017, probably as Police efforts to clamp down on crime increased as shown in Figure 15. Similarly, we equally examined the domestic violence in the city of Chicago, and we saw a trend that declined from 2012 to 2014, and rose in 2016 and dropped sharply in 2017 (see Figure 16).

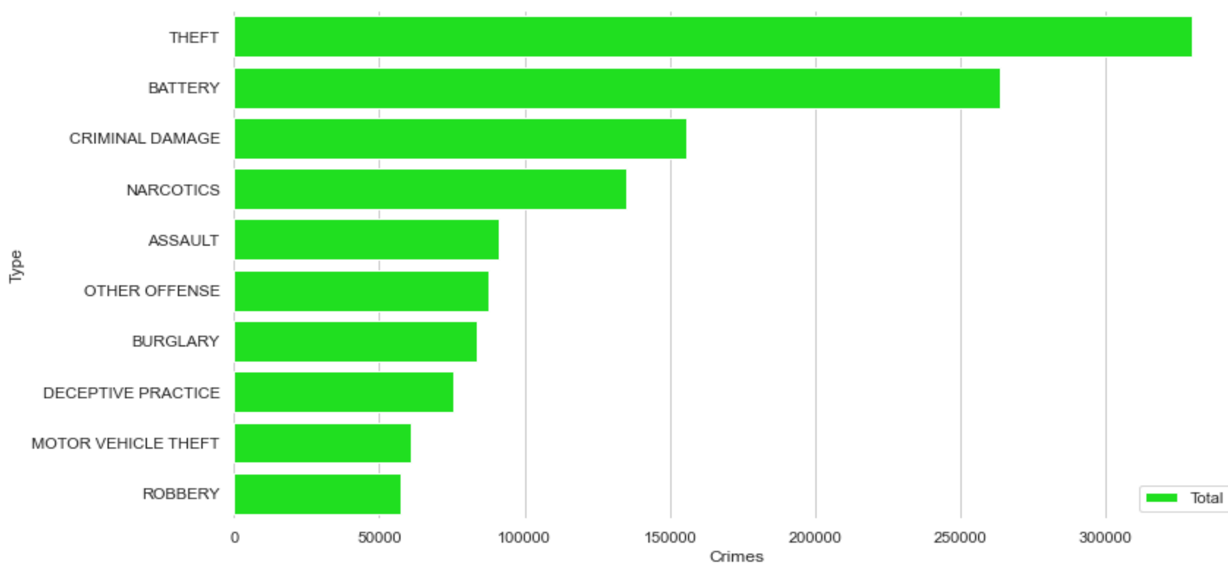

Figure 14: City of Chicago types of crimes with occurrence rate.

We evaluate five top crimes that have trended over the past years in the city of Chicago. These are Assault, Battery, Criminal damage, Narcotics, and Theft. S shown in Figure 17, primary-type crime appeared in that order of trending and peaked between July and August for every year (2012 to 2016). Theft cases were on the low side until 2016 when Narcotics cases nosed down. However, in December 2015, and both Narcotics and Theft cases had the same low values.

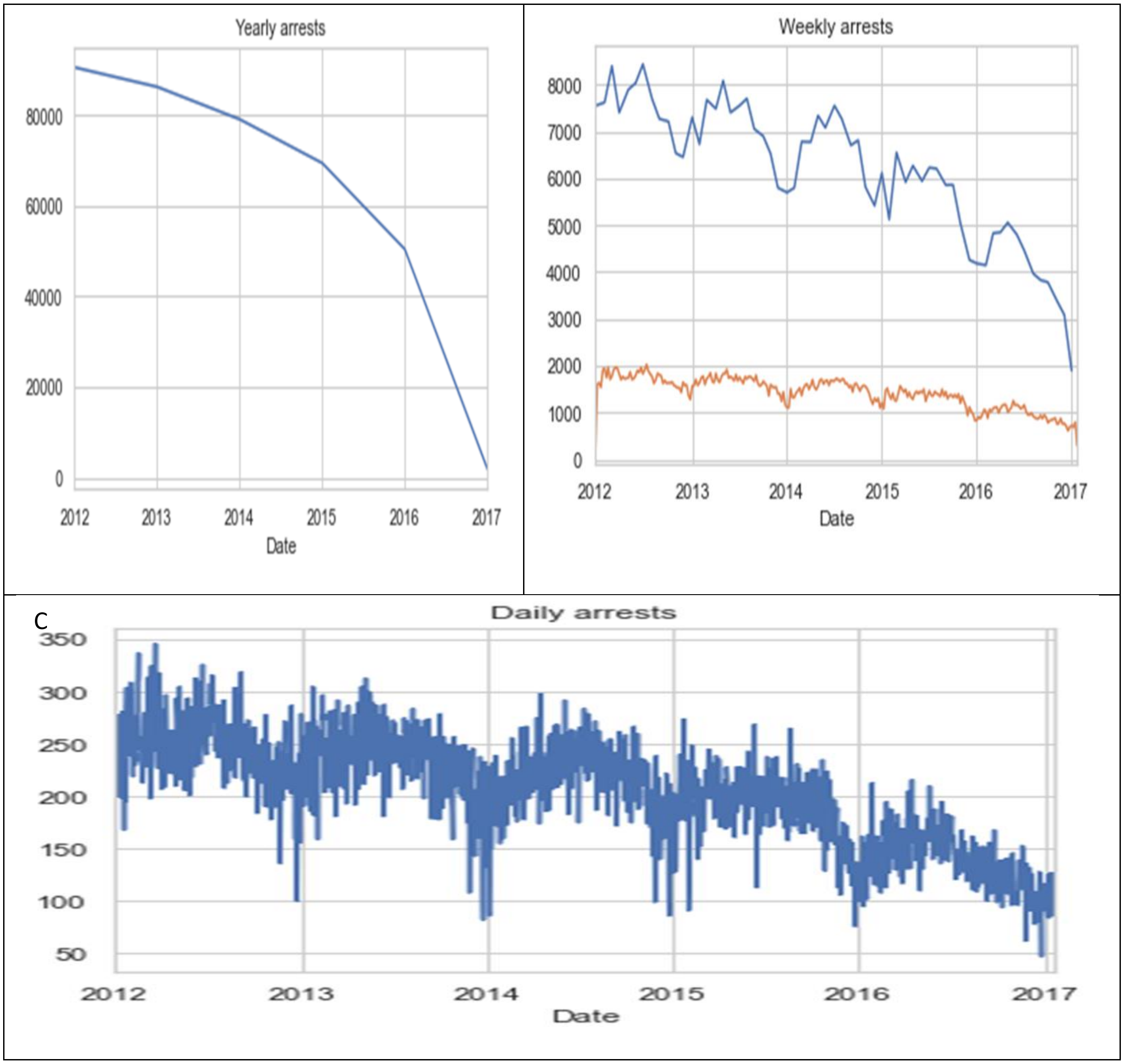

Figure 15: Yearly, Weekly and Daily Arrests due to Crime types in Chicago

Figure 18 shows a similar display to that of figure 17, only that these were weekly displays. Using the primary-type crimes we could show the top number of criminal neighborhoods in Chicago. The primary-types were accumulated for each of the

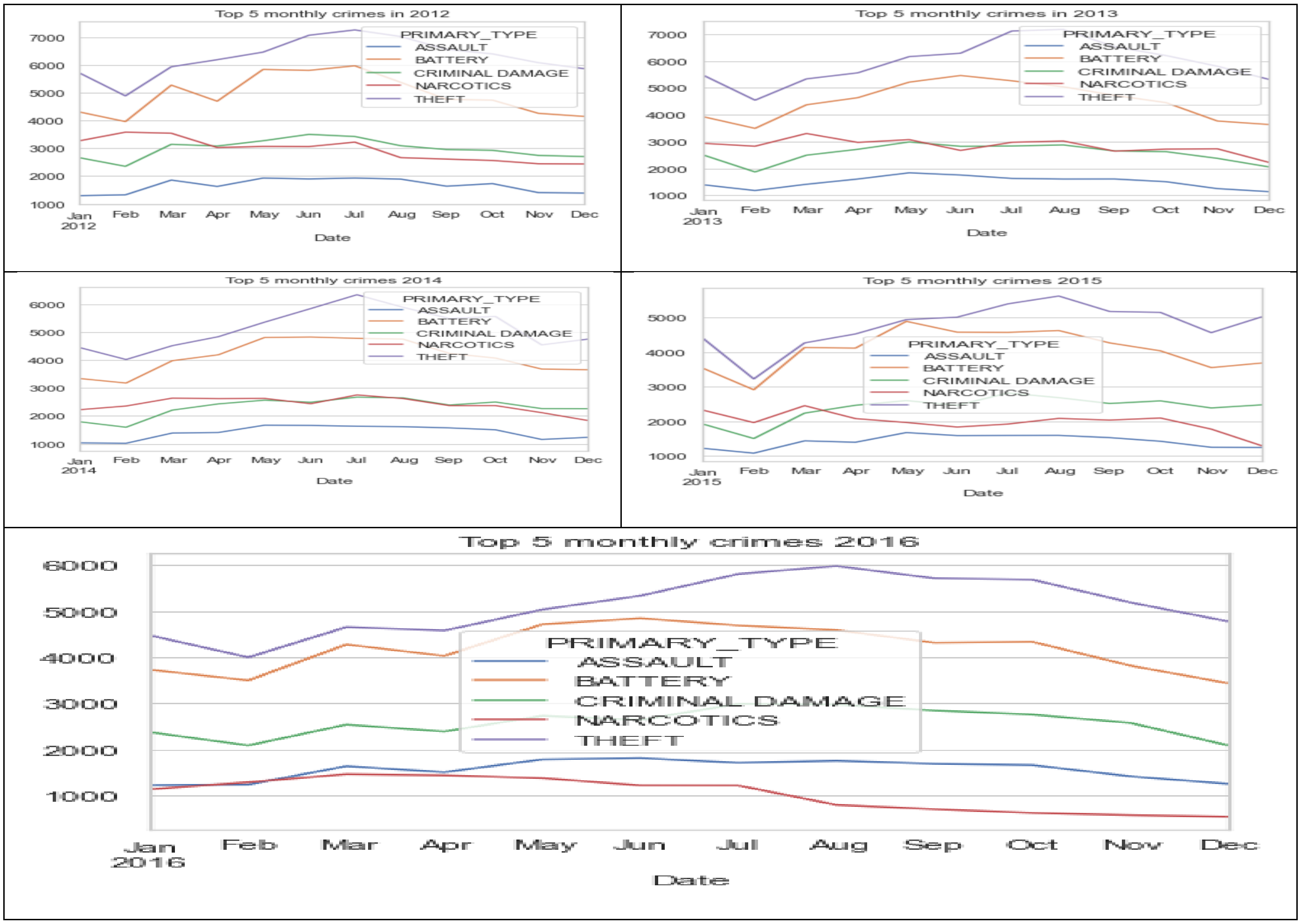

Figure 17: City of Chicago 5 Top crimes trending from 2012 to 2016

neighborhood identified by the Community Area Number. Table 10 shows sample data with community area number, latitude longitude, and primary-type as columns. Figure 19 represents a folium map of the top number of criminal neighborhoods in the city of Chicago.

Moreso, we created a dataframe to show crimes associated with gang violence in the city of Chicago neighborhoods. These crimes include homicide, concealed carry license violation, Narcotics, weapons violation within the period 2012 to 2017. And Figure 20 shows the folium map of the criminal neighborhoods in the city of Chicago (Tables 11a, 11b, and 11c). These tables are one in a horizontal continuation from a to c. We determined the correlation coefficient between the pairs of attributes, which gave the measures of the interplay between different attributes. Using the above data, we were able to predict hierarchy within the data, which allowed ordering the data in clusters, Figure 21.

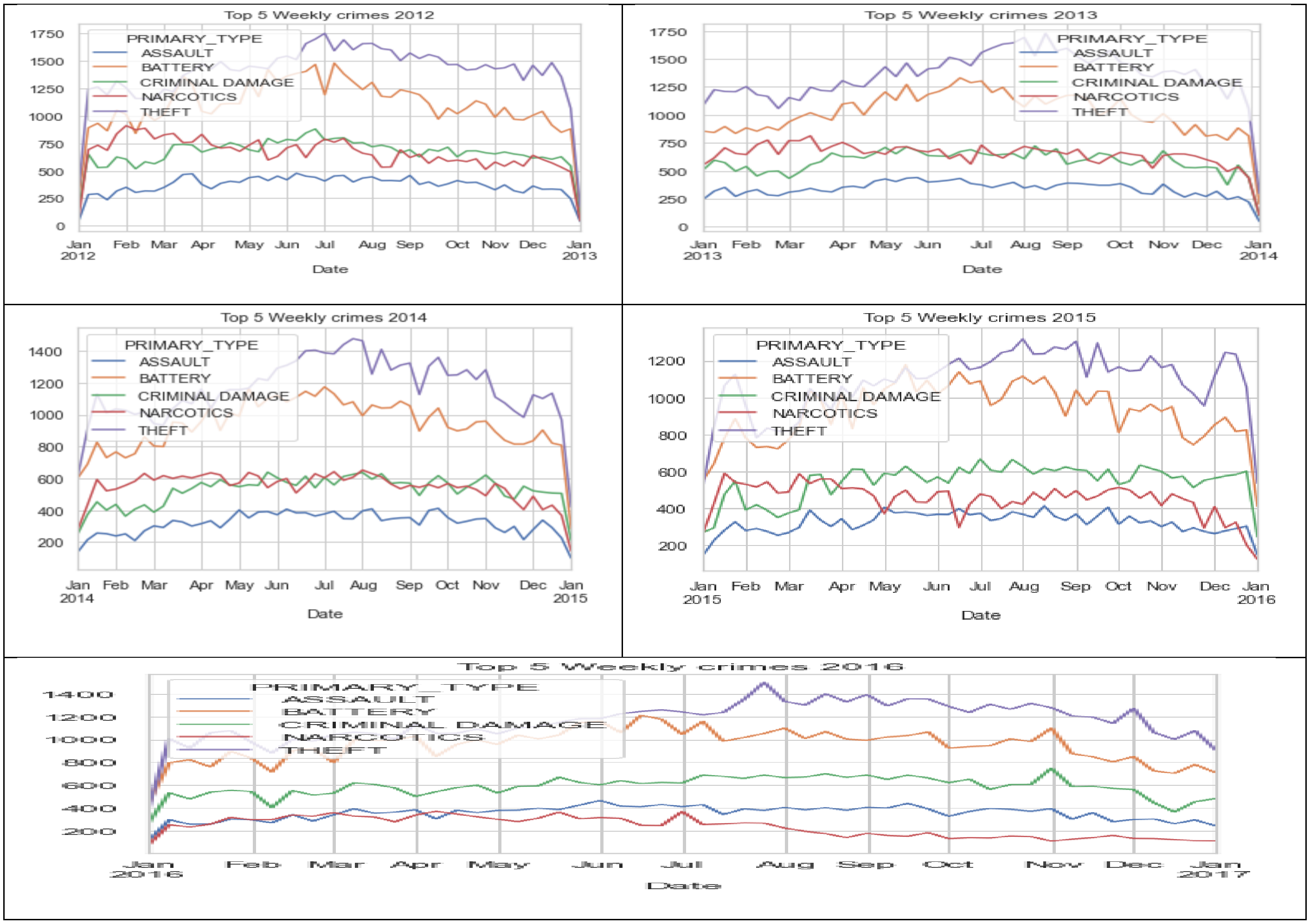

Figure 18: City of Chicago Top 5 crimes trending weekly from 2012 to 2016

Table 10: Top Number Criminal Neighborhoods in City of Chicago

Community Area Number	Latitude	Longitude	Primary Type
29.0	**41.8641**	**87.7068**	**46151**
42.0	**41.7829**	**87.6044**	**21765**
25.0	**41.8949**	**87.7584**	**94730**
44.0	**41.7454**	**87.6038**	**33152**
35.0	**41.844**	**87.6269**	**13325**

The threshold used was 0.8, and we were able to generate a correlation heatmap which resulted from the combination of correlation coefficient and dendrogram (hierarchical clustering, see Figure 22). It showed the heat patterns, the correlation coefficient, and the hierarchical ordering within the features or attributes. Deep blue and blue colors indicated how related these features are.

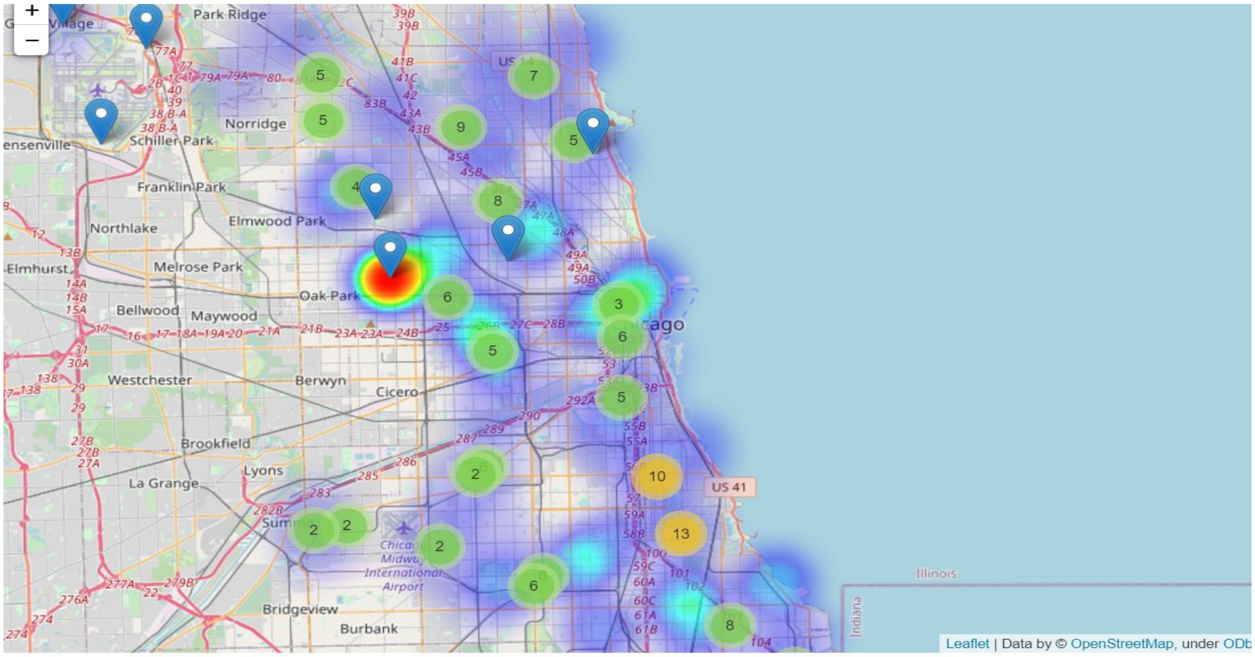

Figure 19: Top Number Criminal Neighborhoods in City of Chicago.

Table 10: Top Number Criminal Neighborhoods in City of Chicago

Community Area Number	Latitude	Longitude	Primary Type
29.0	41.8641	87.7068	46151
42.0	41.7829	87.6044	21765
25.0	41.8949	87.7584	94730
44.0	41.7454	87.6038	33152
35.0	41.844	87.6269	13325

Table 11a: crimes associated with gang violence in the city of Chicago neighborhoods

	YEAR	ARSON	ASSAULT	BATTERY	BURGLARY	CONCEALED CARRY LICENSE VIOLATION	CRIM SEXUAL ASSAULT	CRIMINAL DAMAGE	CRIMINAL TRESPASS	DECEPTIVE PRACTICE	GAMBLING	HOMICIDE	HUMAN TRAFFICKING
0	2012	469	19898	59132	22844	0	1406	35853	8215	13426	652	503	0
1	2013	364	17971	54003	17894	0	1264	30853	8135	13461	596	422	2
2	2014	399	16896	49444	17894	15	1305	27797	7538	15300	393	424	2
3	2015	451	17040	48904	17894	34	1339	28669	6401	15365	310	497	13
4	2016	521	18710	50243	17894	36	1445	30978	6306	17291	189	772	11
5	2017	13	774	1974	17894	5	64	1305	317	652	0	31	0

Table 11b: Crimes associated with gang violence in the city of Chicago neighborhoods

INTERFERENCE WITH PUBLIC OFFICER	INTIMIDATION	KIDNAPPING	LIQUOR LAW VIOLATION	MOTOR VEHICLE THEFT	NARCOTICS	NON-CRIMINAL	NON-CRIMINAL (SUBJECT SPECIFIED)	OBSCENITY	OFFENSE INVOLVING CHILDREN	OTHER NARCOTIC VIOLATION	OTHER OFFENSE	PRO
1228	156	236	573	16492	35486	6	2	26	2183	6	17480	
1281	134	243	465	12582	34144	4	0	24	2317	5	17984	
1398	117	220	395	9913	28953	16	1	37	2344	10	16962	
1308	121	192	292	10076	23833	115	0	46	2200	5	17534	
933	128	202	225	11363	12413	49	1	54	2258	4	17040	
47	6	6	3	712	411	3	0	0	96	0	874	

Table 11c: Crimes associated with gang violence in the city of Chicago neighborhoods

OTHER OFFENSE	PROSTITUTION	PUBLIC INDECENCY	PUBLIC PEACE VIOLATION	ROBBERY	SEX OFFENSE	STALKING	THEFT	WEAPONS VIOLATION
17480	2203	17	3007	13484	1058	207	75454	3904
17984	1652	10	3135	11819	1013	153	71524	3246
16962	1625	10	2903	9798	949	140	61530	3114
17534	1322	14	2421	9639	943	154	57292	3362
17040	799	10	1602	11960	910	167	61167	3430
874	32	1	54	613	20	7	2493	177

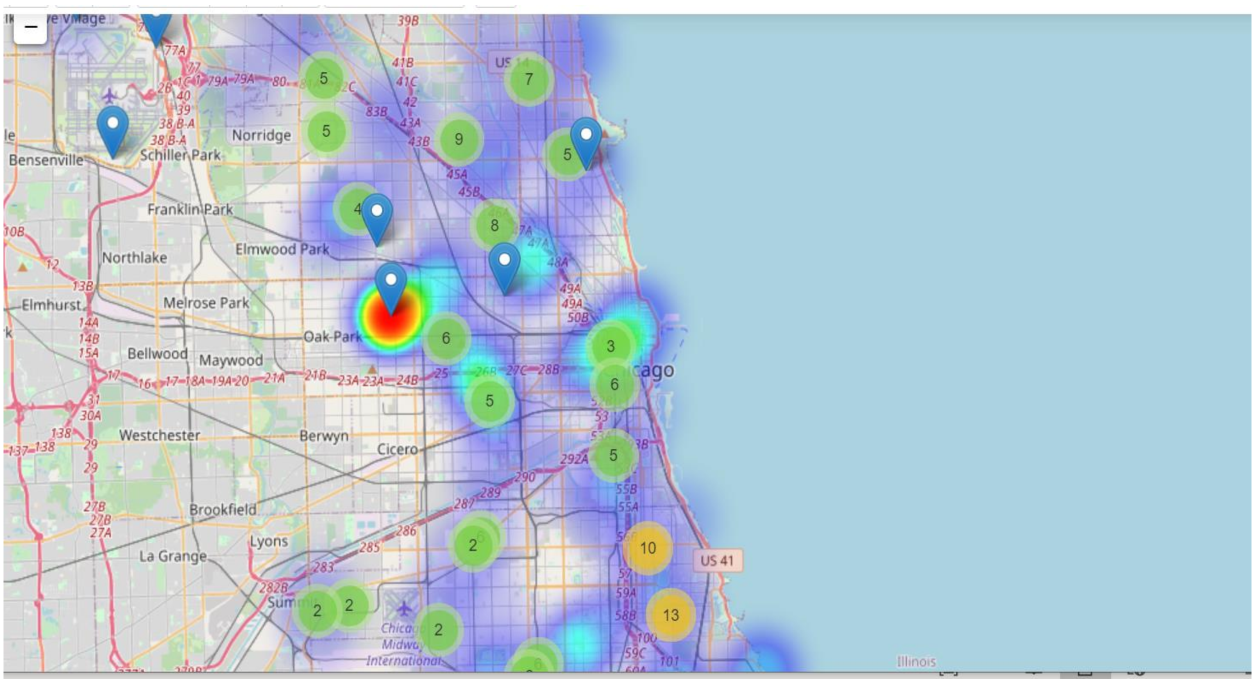

Figure 20: Folium Map of Criminal Neighborhoods in the city of Chicago

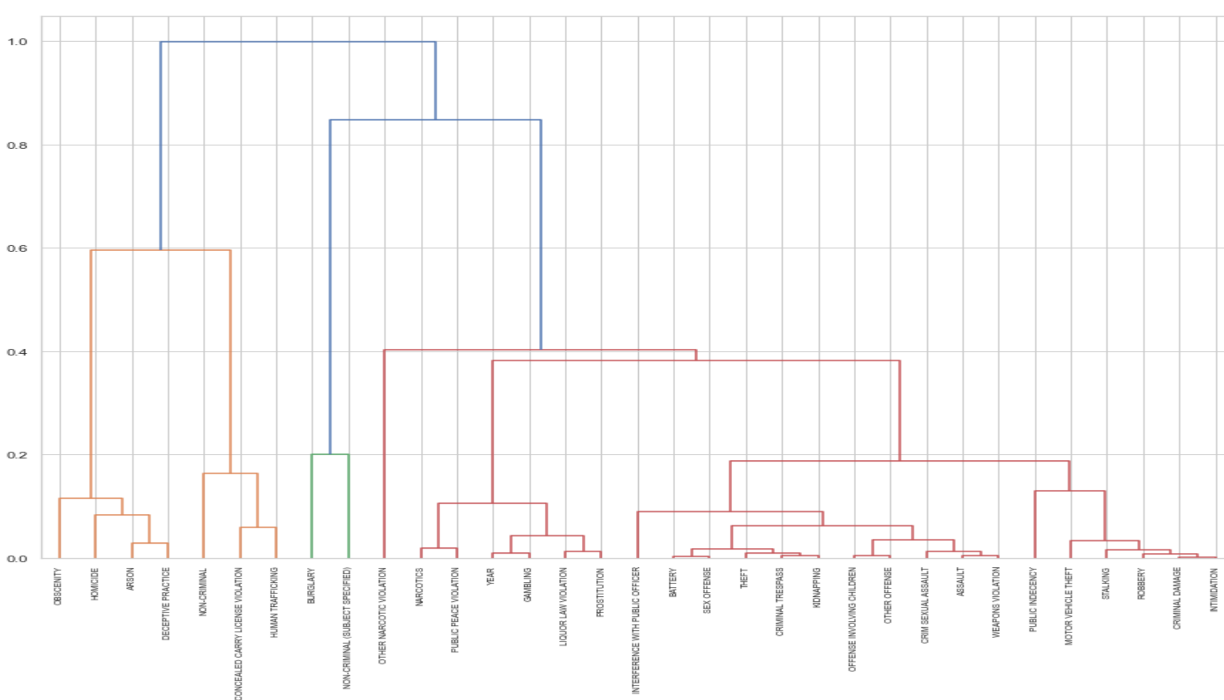

Figure 21: Dendrogram of Chicago Crime rate associated with Gang violence

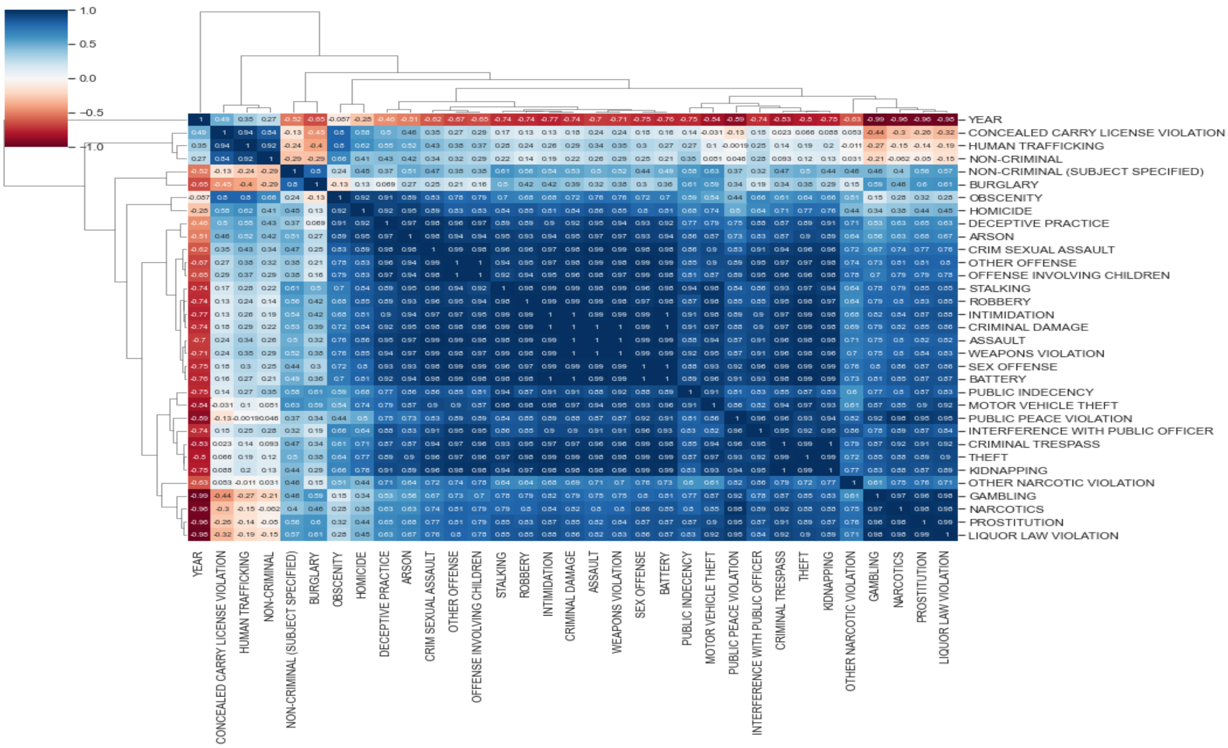

Figure 22: Correlation Heatmap with hierarchical clustering of Chicago gang violence crime

Part IV - Discussion and Conclusion

4.1 Discussion and Conclusion

4.1.1 Cities of Chicago and Toronto Neighborhood Clustering

In this section, we interchange Boroughs with Community Areas as they bore the same meaning. The city of Toronto used Boroughs and the city of Chicago used Community Areas. Again, the city of Toronto is characterized as having 103 Boroughs with over 360 neighborhoods, while the city of Chicago has 77 Community Areas with more than 200 neighborhoods. After removing records where the venue is building or office, bus line, bus station, and road, foursquare API returned 7607 and 2052 venues for the cities of Toronto and Chicago respectively.

The maximum number of venues returned for each neighborhood is 336 and the least was 6 for the city of Toronto, while in the city of Chicago the maximum number of venues for the neighborhood was 75 and the least of venues was 1. Also, the total number of unique categories that was returned for the city of Toronto is 511, while the city of Chicago is 332, which is low compare to that of Toronto.

The frequency of occurrence of these unique categories lies in the range of 1 to 304. Those unique categories that occurred once include Airport, Airport Food couch, Airport lounge, Aquarium, Arepa Restaurant, Baseball stadium, Beach Bar, Blood Donation center, Zoo, etc. The following unique categories with the number they appeared in the neighborhoods: Automotive shop(139), Bank(119), Church(137), Coffeeshop(174), Convenience store(120, Dentist's office(188), Doctor's office(165), Medical center(145), Park(171), Pizza Place(96), Residential Building(Apartment/Condo)(304), salon/Barbershop(141), school(106). Similarly, venue categories found in the city of Chicago Neighborhoods range between 1 and 87. Those with the least occurrence of frequency include a wine shop, winery, warehouse store, weight loss center, Theme Restaurant, Accessories store, Animal shelter, Antique shop, etc. America Restaurant(18), Art Gallery(14), Automotive shop(44) in contrast to the 139 found in the city of Toronto neighborhoods, Bank(32), Church(75), Coffeeshop(18), Convenience store(19), Dentist's office(45), Doctor's office(67), Medical Center(18), Mexican Restaurant(55), Park(32), Pizza Place(22), Residential Building(Apartment/Condo)(62), Sal0n/Barbershop(87), School(23). These venue categories found in the city of Chicago were far less than those identified in the city of Toronto.

In the city of Toronto, most neighborhoods have the top 5 most frequent of its venues. Agincourt has 11percent Chinese Restaurants and 4 percent of churches, Coffeeshop, storage facility, and supermarket. In the neighborhoods of CN Tower, King and Spadina, Railway lands, Harbourfront West, Bathurst Quay, Residential Building(Apartment/Condo) take 17 percent, and Coffeeshop takes 2 percent. Central By street records 12 percent for Doctor's office, 9 percent Medical Center and 6 percent for Japanese Restaurant, Dentist's office and Residential Building, Church and Wellesly neighborhoods record 16 percent for Residential Building(Apartment/Condo), and Bank and Bubble Tea shop have 5 percent each. The automotive shop has 22 percent occurrence in Dorset Park, WaxFood Heights, Scarborough Tom Center, and 3 percent for Factory. HarbourFront East, union State, Toronto Islands Neighborhoods have 29 percent of the Plane venue. In Hillcrest village and Humber summit, there are 17 percent college classroom, and 28 percent Automotive shop respectively. A seemed commercial Neighborhood is the Lawrence Manor and Lawrence Heights that has 33 percent clothing store, 17 percent for parks, Residential Building (Apartment/Condo), Salon/Barbershop and shopping mall.

Table 12 shows the number of Boroughs/Community Areas in each city for the Cluster Labels indicating the clustering of neighborhoods. The first cluster label, Toronto has 15 neighborhoods while Chicago has only 3. Cluster label 1 indicates Toronto to have 1 while Chicago has 9.

Table 12: Number of Boroughs/Community Areas in Each City for the Cluster Labels

Cluster Labels	Toronto	Chicago
0	15	3
1	1	9
2	36	7
3	17	3
4	14	4

Figure 23 is a vertical bar chart showing these cluster labels for the two cities. For the city of Toronto, Figure 24 demonstrates venue categories with Residential Building(Apartment/condo) getting a total of over 300.

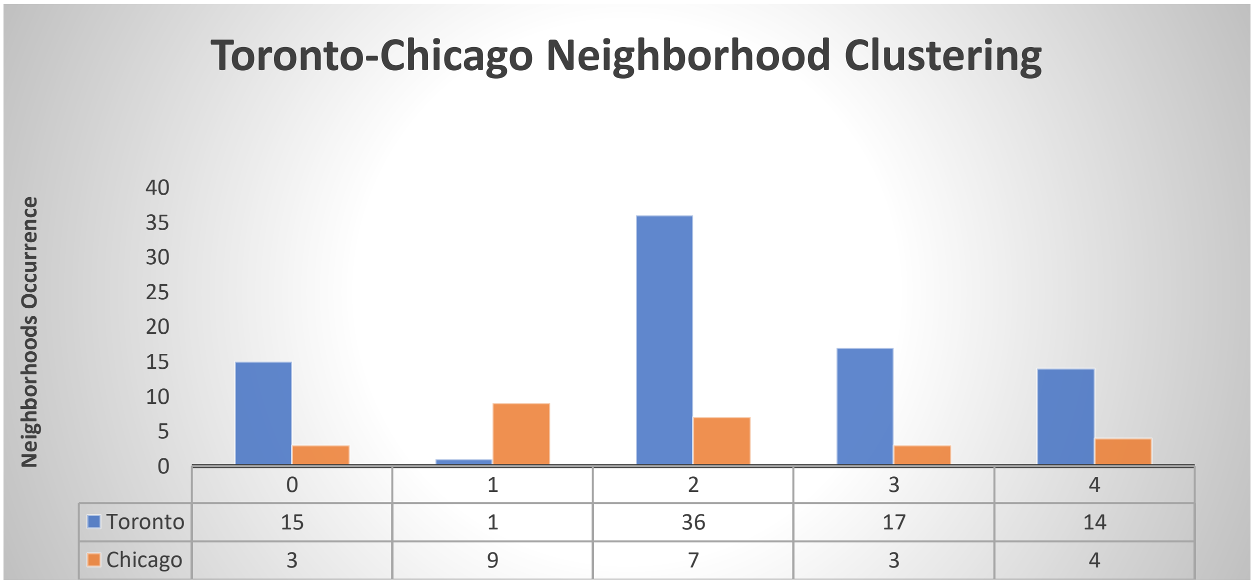

Figure 23: Comparison of Toronto and Chicago Clustering of Neighborhoods.

The dentist's office and Coffeeshop were next in rank, while Café and Gas Station have the least of less than 80. From Figure 25, we presented cases where venue categories exist more in most of the city of Toronto Neighborhoods.

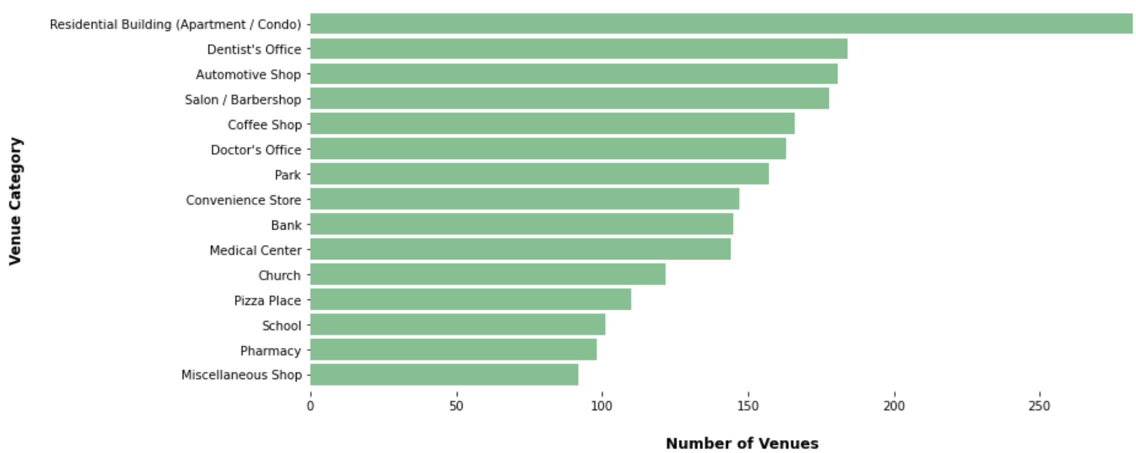

Figure 24: Number of venues for venue categories in the City of Toronto

This shows the number of neighborhoods that contain the venue categories. About 78 neighborhoods in the city of Toronto have Coffeeshops. 69 of the neighborhoods have Park, Church, Residential Building(Apartment/Condo), and Medical Center. 52 neighborhoods get Pharmacy, Café, and School.

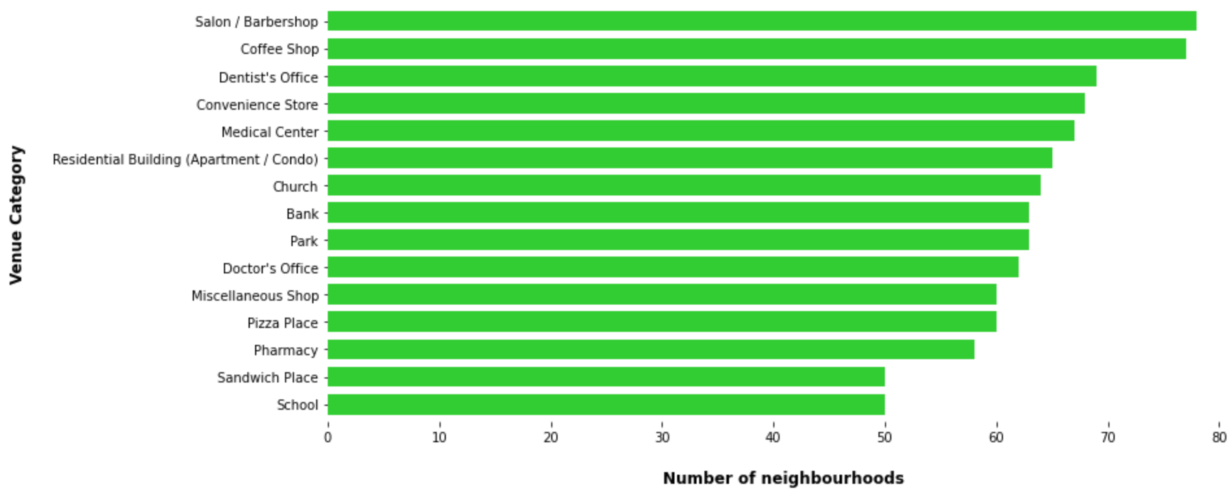

Figure 25: Number of Neighborhoods Having different venue categories (City of Toronto)

For the city of Chicago, we presented the top 5 most frequent of its venues, noting that 13 percent was the highest and the least was 2 percent. A breakdown showed that Albany Park, Mayfair, North Mayfair, and Ravenswood neighborhood has Automotive shop(5%), Hookah Bar(5%), Salon/Barbershop(5%), Coffeeshop(4%), and Discount store(3%). In Avalon Park, Marynook, Stony Island Park Neighborhood, there was 8 percent of salon/barbershop, 6 percent of Dentist's office, 5 percent of Miscellaneous shop and Church, 4 percent of Nail salon. All the venues in Avondale Gardens, Irving Park, Kilbourn Park neighborhood were 4 percent for Strip club, speakeasy, smoke shop, convenience store, and Salon/Barbershop. In Big Oaks, Norwood Park East, Norwood Park West neighborhood, Taxi was 10 percent in occurrence, Doctor's office and Hospital were 6 percent, while Emergency room and Medical center were 4 percent. We saw Residential Building(Apartment/Condo) be 9 percent, Tech startup, Taxi, Gym, Bank to be 4 percent in the neighborhood of Cabrini-Green, gold Coast, Goose Island. These increased in the Central Station, Dearborn Park, Museum Campus neighborhood, where Residential Building(Apartment/Condo) was 11 percent, Salon/Barbershop 10 percent, Gym 7 percent, Dentist's office is 6 percent, and Miscellaneous shop is 4 percent. However, in the neighborhood of Gladstone Park, Jefferson Park, the Dentist's office was prominent to about 12 percent, Doctor's office was 11 percent, Salon/Barbershop(5%), and Playground and Automotive shop were about 4 percent. In the Edison Park neighborhood, speakeasy venue has the frequency of 2 percent, the least as seen in the study while 13 percent maximum was found in the West Garfield Park neighborhood. These venues were the church.

Figure 26 is a bar chart for the city of Chicago venue category, which shows the number of venues. Salon/Barbershop venue category tops the list, followed by Church which is 80 in number. The doctor’s office is 70 while Residential Building(Apartment/Condo) is 60. The number of venues for the following venue categories, Mexican Restaurant, and Dentist’s office was 50 and 45 respectively, while Automotive shop was 44.

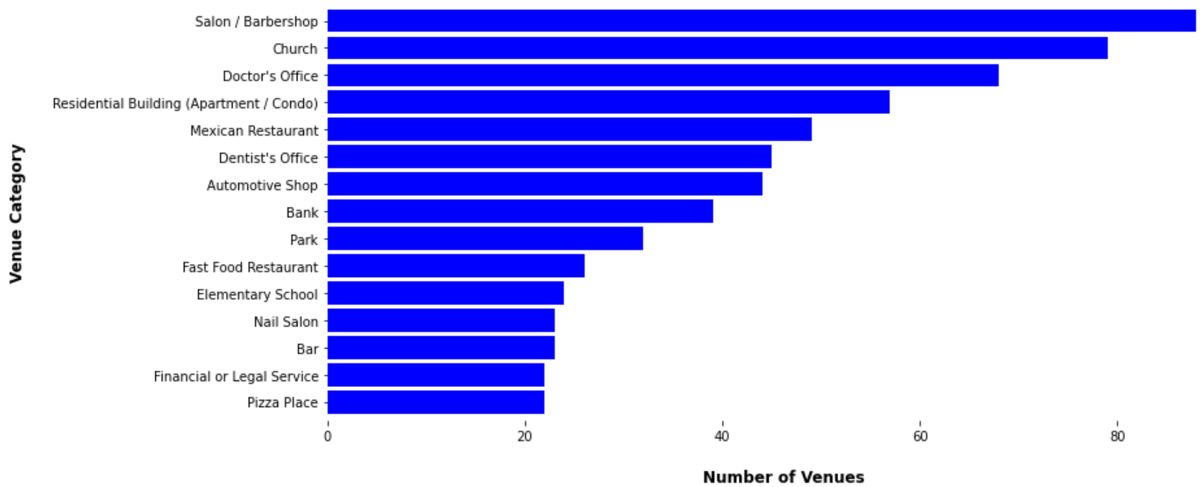

Figure 26: Number of venues for venue categories in the City of Chicago

For the venue categories that exist more in most of the city of Chicago neighborhoods, Figure 27 shows that about 30 neighborhoods have Salon/Barbershops and Doctor’s office. 23 neighborhoods have a Church venue category. 18 neighborhoods have Residential Building(Apartment/Condo). 16 neighborhoods have both Dentist’s office, Park and Bank. 15 neighborhoods have a Nail salon, financial or legal service, mobile phone shop, Bar, school, Automotive shop, and Mexican Restaurant while 14 neighborhoods have a Miscellaneous shop. Most of the venue categories that appear once are considered to be rare categories for both cities of Toronto and Chicago. Like college Arts Building, Creole Restaurant(Chicago), Language school Locksmith(Toronto).

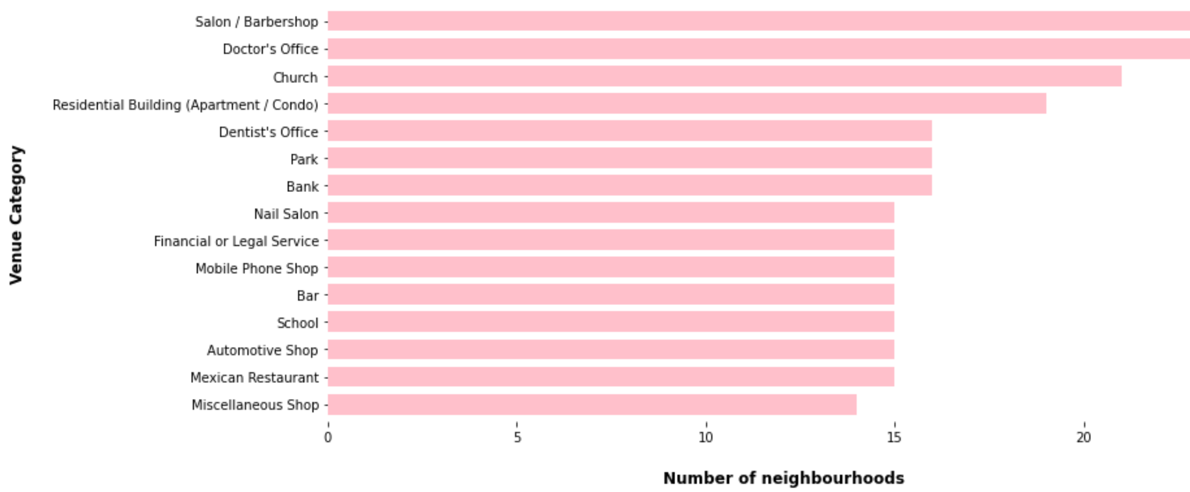

Figure 27: Number of Neighborhoods Having different venue categories (City of Chicago)

For the two cities under study, the most common venue categories have been identified such as Residential Building(Apartment/Condo) which has 35 in the number of venues for both Chicago and Toronto (see Figure 28). Similarly, Salon/Barbershop has 27 in the number of venues. Miscellaneous shops and pharmacies are the least.

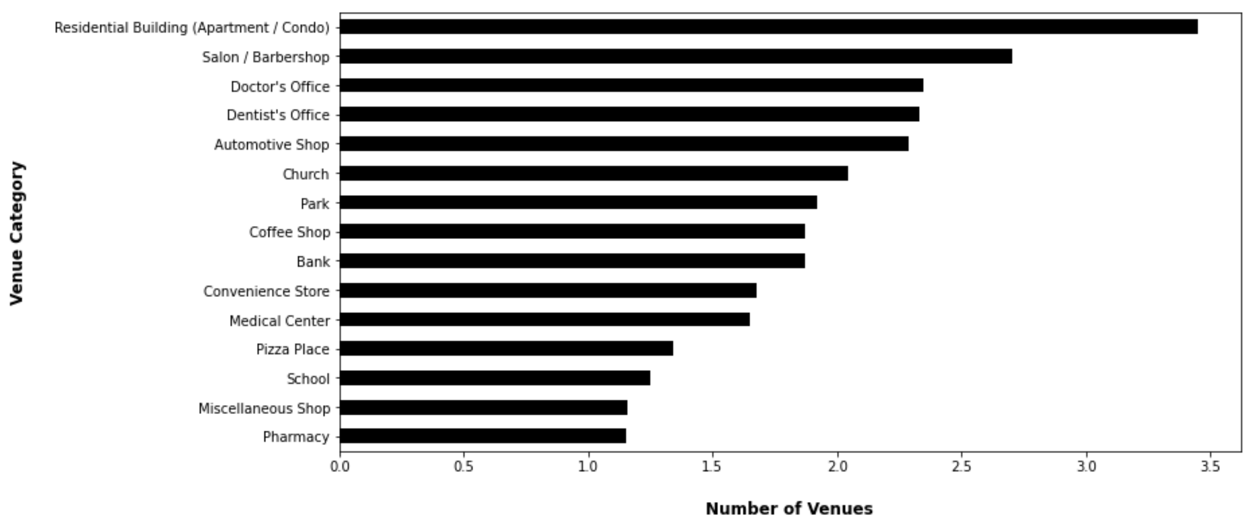

Figure 28: Most Common Venue Categories in both Toronto and Chicago.

4.1.2 Discussion on the two cities Crime rate Analysis

As shown in Figures 3 and 14 in the previous sections, the city of Toronto had the following types of crimes based on-premise type and offense, which were far less than those recorded in the city of Chicago. The Theft, for instance, had been more than 84 times in the city of Chicago compare to in the city of Toronto. Burglary (Chicago) and B&E(Toronto) were in the ratio of 2:1, an indication that more of this crime was committed in the Chicago neighborhoods. In the city of Toronto, every single Theft of motor vehicle committed would be three times more likely to happen in the city of Chicago. Similarly, we observed that all types of Assaults combined in the city of Toronto were slightly above that recorded in the city of Chicago. It's one of the crimes that surpassed the usual Chicago trend in the crime wave. Robbery cases recorded in Chicago were five times more than in the city of Toronto, a clear indication that this city(Toronto) is almost crime-free in comparison to the city of Chicago, where there were high rates of gangs, Narcotics, other offense(87,874) and criminal damage (155,455) from 2012 to 2018.

These, were reflected in the daily, weekly monthly, and yearly arrests shown in Figures 4,5, and 6 for the city of Toronto and Figures 15, 16, and 17 for the city of Chicago. However, most cases in the city of Chicago seem to be on the decline, unlike in the city of Toronto where cases were not declining. There were so many crimes quite prominent in the city of Chicago neighborhoods, ranging from Arson, Assault, Concealed carry license violation to homicide to human trafficking, prostitution, sex offense, etc. As shown in Figure 19 and 20, the city of Chicago neighborhoods had recorded cases of gang violence and other crimes making most of these neighborhoods not safe to live and operate businesses, quite unlike in the city of Toronto neighborhoods that established a haven for living and operating corporate, small and medium scale businesses of all kinds.

4.2 Conclusion

The cities of Toronto and Chicago were analyzed using the system of the battle of the neighborhood, which involves the use of Foursquare API applications to determine the venues and venue categories. Later, the results were used to find the clustering of the city neighborhoods.

Foursquare API facilitated getting venues of important places of interest. The clusters show the neighborhoods in each of the five clusters both for the city of Toronto and the city of Chicago. The top common venues mapped show that

Residential Building(Apartment/Condo) topped the list in the city of Toronto, while the most common venue category ie top is salon/barbershop and coffeeshop.

In the city of Chicago, the results obtained in terms of top common venues and most common venue categories were not prominent like in that of the city of Toronto. The clustering of the neighborhoods shows five clusters and indicating that the neighborhoods that form these were few in comparison to the Toronto neighborhood clustering.

In terms of similarities, there are close form- venue categories that exist in both cities, such as the salon/barbershop, church, Doctor's office, etc.

Finally, the crime rate analysis of the two cities, Toronto and Chicago demonstrates great remarkable differences indicating that the city of Chicago is immersed in crime wave of all kinds and these have been on increase, making the city not safe; while Toronto city shows a completely different trend indicating haven neighborhoods.

References

Akira Ito, et al: Region – to – Region Similarity Analysis based on Foursquare Venue database.

Amy X. Zhang, et al, (2013): Hoodsquare: Modeling and Recommending Neighborhoods in location-based social networks. arXiv; 130836657v1

Hierarchical Clustering with Python & SciKit Learn

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

https://en.wikipedia.org/wiki/Community_areas_in_Chicago

https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago

https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods

Preotiuc-Pietro, et al, (2013): Exploring Venue based city – city similarity measures, UrbComp 13.

Scipy Hierarchical Clustering