



VISUALIZE YOUR CITY

Data Science Project



Contents

1. Introduction	2
2. Data Acquisition	2
New York city crime data acquisition:	2
Toronto city crime data acquisition:.....	2
3. Data Pre-processing	3
New York city crime data:	3
Toronto city crime data:	3
4. Visualization	4
New York crime data Visualization:	4
Visualization through maps.....	6
Toronto city crime data visualization:	7
Visualization through maps.....	9
5. Conclusion	9

1. Introduction

Considering the type of neighbourhood people presently living in, they would prefer to move to a similar neighbourhood or move to another neighbourhood with better living conditions. It is always better to be prepared before making a big move, by researching more about the city you would like to move. This report with statistical and visual representation of crimes in New York and Toronto help those people.

Comparing two cities is based on different categories like cost of living, connectivity through transport, crime rate, employment opportunities, education institutions, Immigration rate, population, ethnically and culturally diversity, cleanliness, population, Tax rates, weather and climate and local amenities.

In this project, I have compared the neighbourhoods of New York and Toronto to find out how similar or dissimilar they are based on crime statistics.

This report represents statistical and visual representation of crimes in New York and Toronto using matplotlib and Folium API. Geo spatial Analysis of the crime rate is shown for New York and Toronto.

2. Data Acquisition

The Data required for the cities of New York and Toronto is extracted and Processed.

New York city crime data acquisition:

- **NYPD Complaint Data Historic** which is available on NYC open data (<https://data.cityofnewyork.us/browse?q=crime&sortBy=relevance>). This dataset includes all valid felony, misdemeanour, and violation crimes reported to the New York City Police Department (NYPD) from 2006 to the end of last year (2019). As the dataset has near to 6M rows, to reduce computational costs filtered version of this data which is available at Kaggle (<https://www.kaggle.com>) is used for this analysis.
The New York City Police Department (NYPD) provides statistics that are categorized by police borough and precinct. For statistical presentation purposes the numerous law categories and subsections are summarized by law class: felony, misdemeanour and violation. These legal categories are then subdivided into broad crime and offense categories, e.g., Felonious Assault, Grand Larceny, Misdemeanour Criminal Mischief, etc.
- **New York city Borough boundaries, New York city Police Precinct boundaries** in the form geojson format which are available at NYC open data. (<https://data.cityofnewyork.us/City-Government/Borough-Boundaries-Water-Areas-Included-/tv64-9x69>)
- **Neighbourhood location data**-To get the location data of New York city precincts from Foursquare, we need this dataset that contains all the 5 boroughs and the neighbourhoods that exist in each borough and latitude and longitude coordinates of each neighbourhood. The data is downloaded from (https://geo.nyu.edu/catalog/nyu_2451_34572)

Toronto city crime data acquisition:

- **Major Crime Indicators(MCI) 2014 to 2019** occurrences by reported date is used for the Visualization. Toronto crime data is available at Toronto Public Safety Data Portal. (<https://data.torontopolice.on.ca/search?q=crime>)

The Major Crime Indicators categories are Assault, Break and Enter, Auto Theft, Robbery and Theft Over (Excludes Sexual Assaults). The location of crime occurrences has been approximated to nearest road intersection node.

- **Toronto city divisions boundaries** geojson file is downloaded. (https://data.torontopolice.on.ca/datasets/43ef8c93684f44a78eade66b3350ce9f_0)

3. Data Pre-processing

New York city crime data:

- **NYPD Complaint Data Historic** This dataset has all the crimes reported in the years 2013 to 2015 and has 24 features and 1048575 samples.

Data fields of interest in this data set are as follows. (to compare with Toronto dataset)

- ADDR_PCT_CD: The precinct in which the incident occurred
- BORO: The name of the borough in which the incident occurred
- LAW_CAT_CD: Level of offense: felony, misdemeanour, violation
- RPT_DT: Date event was reported to police
- X_COORD, Y_COORD_CD: coordinates for New York State Plane Coordinate System
- Longitude, Longitude: coordinates for Global Coordinate System

The dataset is checked for any missing data fields of interest and missing rows are dropped. To compare both cities for the same reporting year, the RPT_DT is converted to Date time format and ADDR_PCT_CD which is in float is converted to integer.

new columns RPT_year, RPT_month are added by extracting only year and month from RPT_DT and 2015 data is filtered.

- **New York city Borough boundaries, Precinct boundaries** data is loaded into IBM Streaming Body object.

The Object is read into memory. The Object is then converted to BytesIO to be able to read it using json.

- **The neighbourhood location data** which is in Json format is loaded into server. All the data is in features key. The json format is converted to a data frame. The original data frame is sliced for each borough.

To get the exact location of Police precincts using Foursquare API, a search query is formed to return all the venues having "NYPD precinct" within 1000 meters' radius of the lat, lon coordinates of neighbourhood. The response from the query is carefully observed and relevant information is appended to the precinct data frame for each borough.

The 5 borough data frames are cleaned by removing repeated, unwanted rows and merged to NY precincts which has 71 precinct rows with the exact locations, borough and neighbourhoods they serve.

Toronto city crime data:

- **Toronto crime Dataset Major Crime Indicators(MCI)**

Data fields of interest in this data set are as follows.

- Reported Date: Date the incident reported to Police
- MCI: The Major Crime Indicators categories are Assault, Break and Enter, Auto Theft, Robbery and Theft Over
- Division: Police Division in which the incident occurred
- Neighbourhood: Neighbourhood in which the incident occurred

- Lat, Long: coordinates for Global Coordinate System

The data is checked for missing rows for the fields and missing rows are deleted. Unwanted column fields are deleted.

- **Toronto division boundaries** data is loaded into IBM Streaming Body object.

The Object is read into memory. The Object is then converted to BytesIO to be able to read it using json.

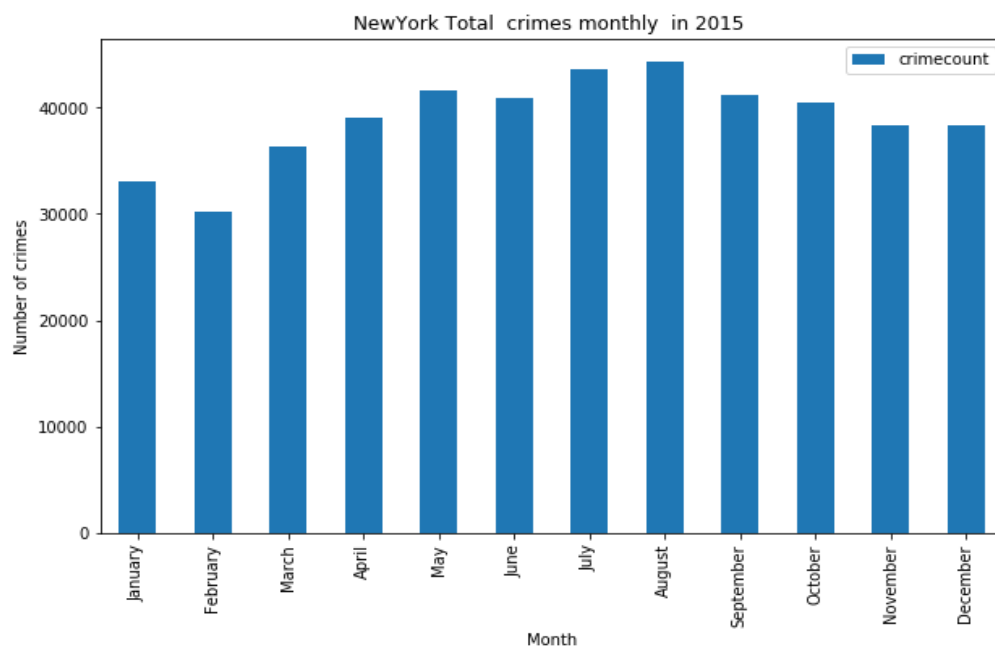
4. Visualization

The data is loaded, Analysed and processed for the visualization. Visualization is performed using matplotlib and chlopleth maps. The crime statistics based on year, month and Division are visualized using matplotlib

New York crime data Visualization:

- **Total crimes monthly** -2015 crime data is grouped by reported month and visualized using matplotlib bar plots.

January	33080
February	30238
March	36402
April	38985
May	41554
June	40897
July	43560
August	44250
September	41186
October	40530
November	38385
December	38394



- **Precinct crimes in 2015** -crime data is grouped based on ADD_PCT_CD and LAW_CAT_CD to find out major crimes in each precinct and plotted using bar plots. The highest number of

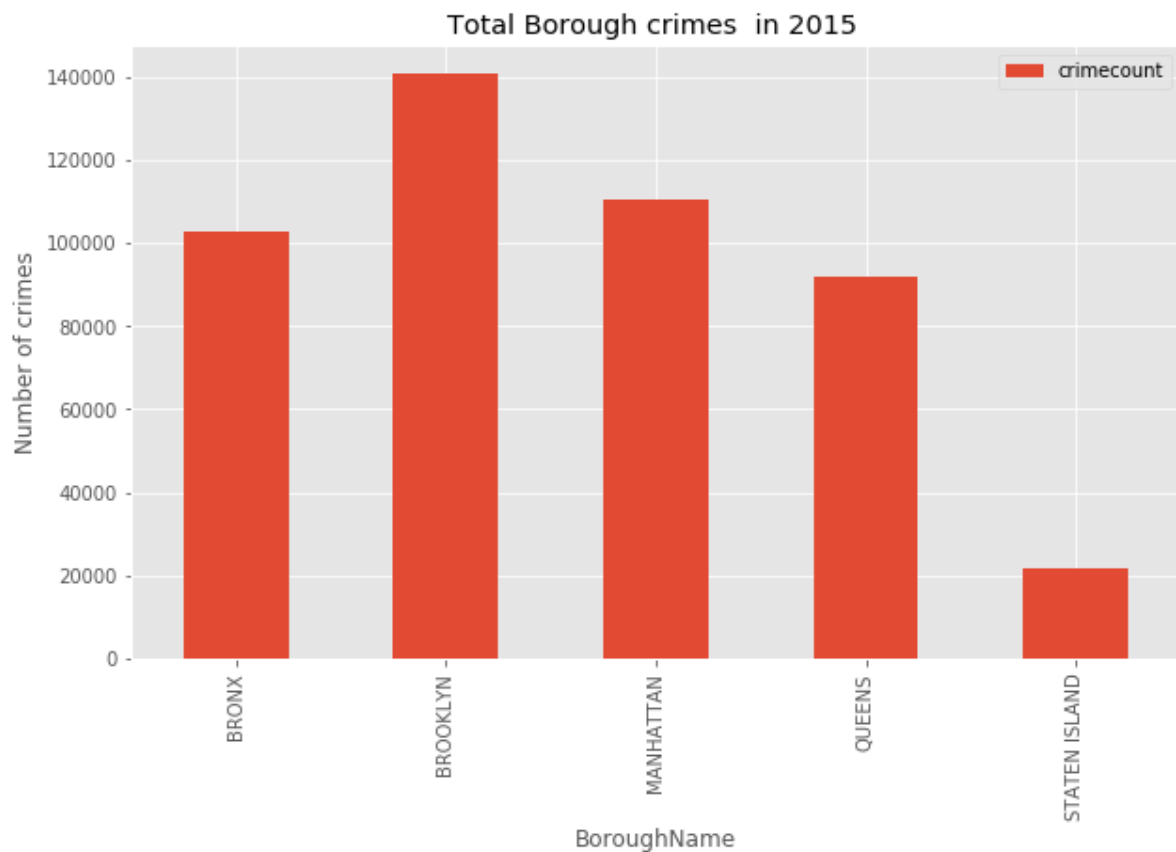
crimes in precinct 75 which serves Easternmost portion of Brooklyn with highest felony, misdemeanour and violation crimes compared with other precincts.

By placing the precincts in descending order based on total crimes and considering top 10 precincts.

ADDR_PCT_CD	Crimes
75	15825
40	12280
44	11412
43	11411
14	10218
52	9953
67	9703
73	9555
46	9280
47	8973

- **Borough crimes in 2015**-crime data is grouped based on BORO_NUM and plotted. Highest number of crimes in Brooklyn borough.

BORO_NM	crime count
BRONX	102629
BROOKLYN	140600
MANHATTAN	110383
QUEENS	92113
STATEN ISLAND	21736



Visualization through maps

- The above grouped data frame is used to visualize the total crimes of each borough using choropleth maps. Highlight functionality and popup are added to map.
- The precinct location data which is extracted is overlapped on to New York city map and Borough boundaries map which gives a clear understanding of number of precincts in each borough.

A feature Group, precincts is created and circle markers are created for each precinct location and added to the FeatureGroup. This feature Group is added to map. A marker for label is created and added to map which displays the name of the precinct.

- 2015 crime dataset is grouped only on ADD_PCT_CD to find total number of crimes in each precinct and displayed using choropleth maps. Threshold scale is defined with 6 spacing using minimum and maximum crimes as range. Threshold scale is passed as bins which gives colour shading. A popup instance is created and added to map to display the precinct name. Highlight functionality is enabled when hovering over a GeoJSON area.
- For interactive Layer display of the major type of crimes in New York city in 2015, a random dataset is created considering 1000 samples.

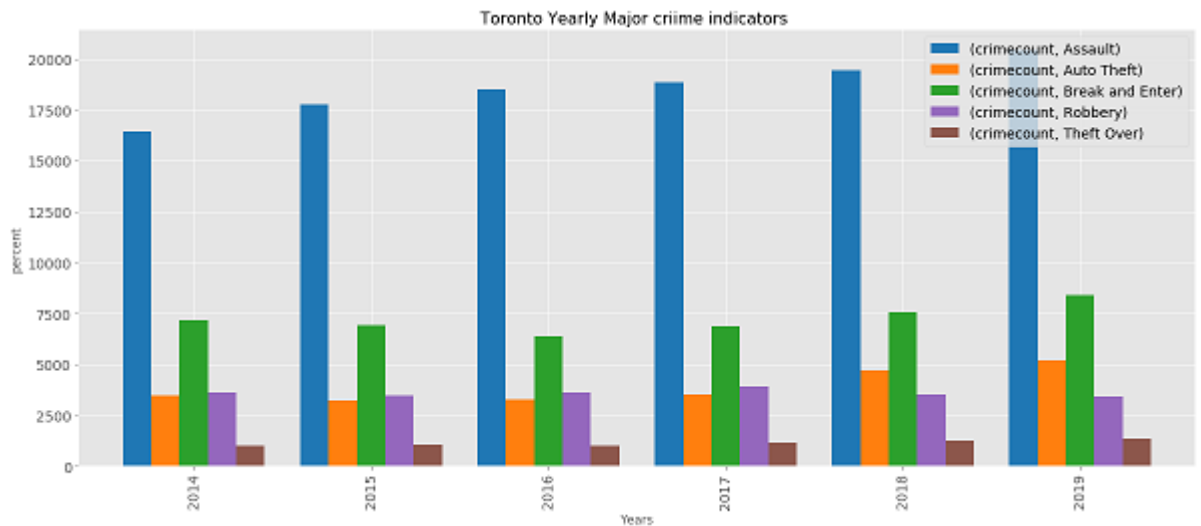
A dictionary of Labels is created and a new column Labels is added which specifies the crime type. A colour List is provided depending on the intensity of the crime to compare with the Toronto crimes. FELONY-red, MISDEMEANOR-Green, VIOLATION-Orange

Adding coloured map markers to each type of crime type -Using Folium Circle Marker and Popup displays the type of crime when user click on the marker and the marker is added to the Feature Group and the Feature Group is added to map. The layer controls for the overlay layers can be checked and unchecked.

Toronto city crime data visualization:

- **Toronto yearly major crime Indicators:**

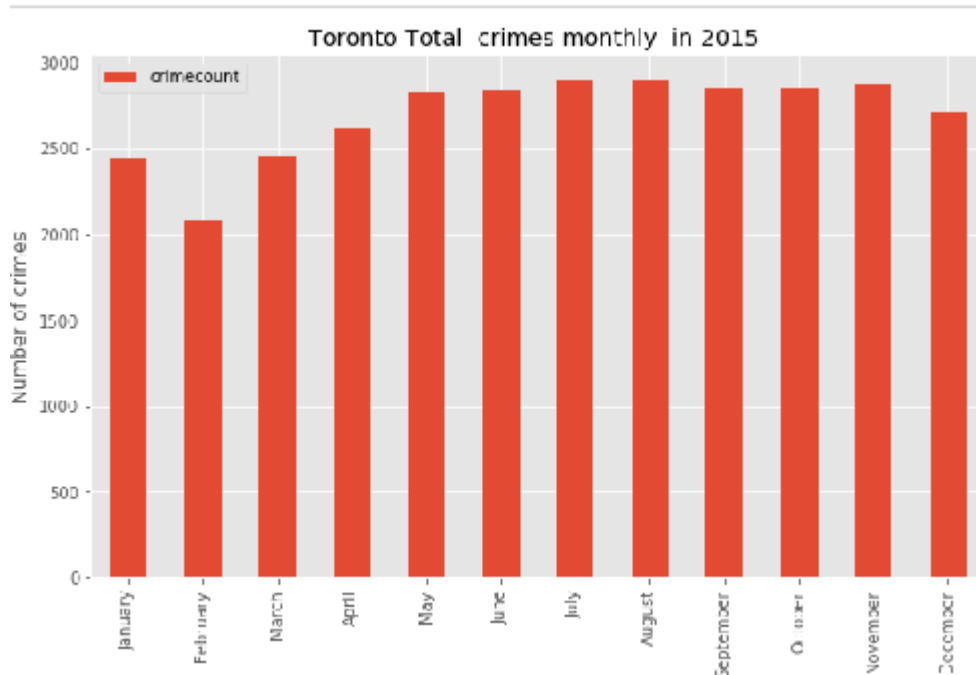
The dataset is grouped by reported year and MCI and converted to pivot. And plotted using area and bar plots. From the plot it is clear that Major number of crimes are in the year 2019.



- **Toronto major crimes monthly in the year 2015:**

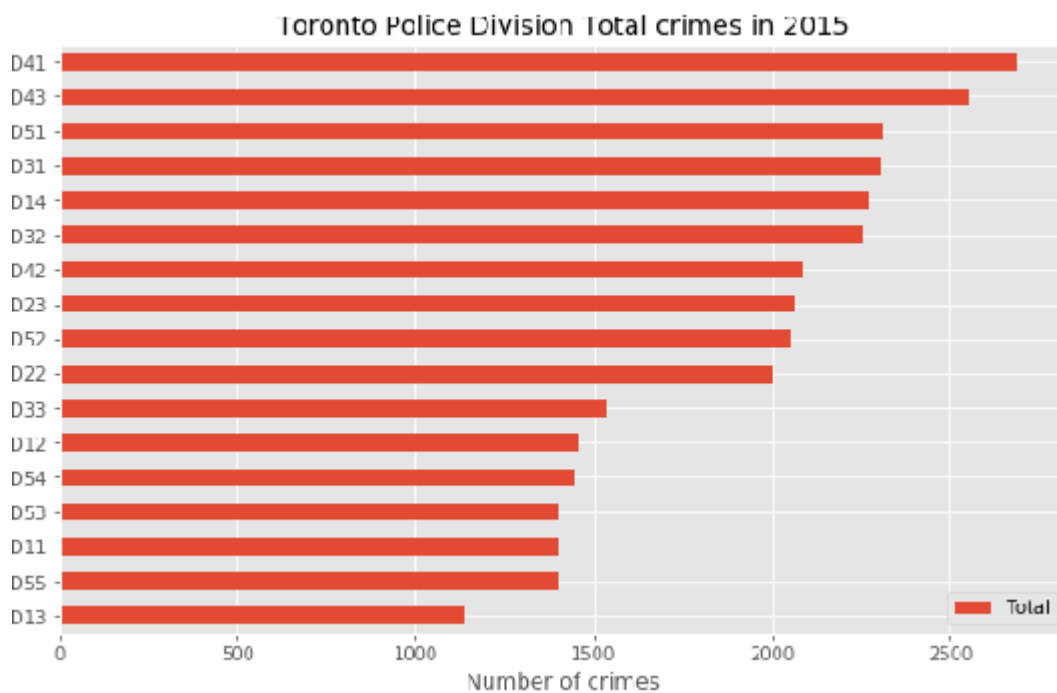
To compare with New York city reported year 2015 is filtered and grouped based on reported month. And plotted using bar plots. it is clear that highest number of crimes in the months of August and July.

reported month	Crime count
January	2449
February	2083
March	2452
April	2619
May	2832
June	2834
July	2893
August	2899
September	2856
October	2858
November	2872
December	2716



- Division crimes in 2015:** Data is grouped based on Division and is plotted using bar plots. It is clear that Division 41 has the highest number of crimes in the year 2015. Also, major crime indicators in each division are plotted using bar plots which show the highest robberies in Division 43 followed by 41.

Division	Total
D41	2691
D43	2552
D51	2309
D31	2307
D14	2274



Visualization through maps

- The grouped data frame on division is used to visualize the total crimes of each division using Choropleth maps. Highlight functionality and popups are added to map. Threshold scale is defined with 6 spacing using minimum and maximum crimes as range. Threshold scale is passed as bins which gives colour shading. A popup instance is created and added to map to display the division name. Highlight functionality is enabled when hovering over a GeoJSON area.
- For interactive Layer display of the major crime indicators in Toronto city in 2015, a random dataset is created considering only 1000 samples. A dictionary of Labels is created and a new column Labels is added which specify the crime type. Label dictionary= {'Robbery':4,'Assault':3,'Break and Enter':2,'Theft Over':1,'Auto Theft':0} A colour List is provided depending on the intensity of the crime to compare with the New York crimes. Orange-Auto Theft, purple- Theft Over, green- Break and Enter, blue- Assault, Red-Robbery. Adding coloured map markers to each type of crime type -Using Folium Circle Marker and Popup displays the type of crime when user click on the marker and the marker is added to the Feature Group and the Feature Group is added to map. The layer controls for the overlay layers can be checked and unchecked.

5. Conclusion

The above analysis can be extended by finding the crime rate and ranking the place on a scale of 1 (low crime rate) to 10 (high crime rate).

CRIME RATE – A crime rate describes the number of crimes reported to law enforcement agencies per 100,000 total populations. A crime rate is calculated by dividing the number of reported crimes by the total population; the result is multiplied by 100,000.

With the use of New York Borough population in 2020(projected which is near to actual 2015 population) and crimes reported, the crime rate is calculated.

Borough	crime rate
Bronx	7093.576
Brooklyn	5308.761
Manhattan	6737.733
Queens	3952.847
Staten Island	4461.824
NYC Total	5466.759

The total crime rate is highest in Bronx followed with Manhattan.

If the neighbourhood crime data and population of New York city can be extracted it can be compared with Toronto neighbourhood crime rate and visualized with plots and maps. Unfortunately, I could not find the source for this.

With the use of Toronto neighbourhood_crime_rates file, the top 10 neighbourhoods based on total crime rate are as follows with Bay Street Corridor having the highest crime rate.

Neighbourhood	crime rate
Bay Street Corridor	4093.499244
University	3260.15512
Moss Park	3257.583146
Kensington-Chinatown	3053.775425
Church-Yonge Corridor	2613.273772
West Humber-Clareville	2512.608069
York University Heights	2410.03153
Oakridge	2166.847237
Kennedy Park	2119.955615