# The Battle of Neighborhoods capstone project

## Introduction

New York city and Toronto are very diverse cities, and both are the financial capitals of their respective countries. Each city has different characteristics based on the nature of the city, the culture, the economy and everything.

This project aims at understanding all the aspects we need to compare the two cities to get more understanding of the nature of the city to be able to determine the suitable business use cases and get similarities and dissimilarities for each of them based on the population, culture and running business of different neighborhoods.

To Compare between two cities like Toronto and New York in order to determine the suitable business use cases for each of them, this has a lot of aspects to consider like:

- Economic situation
- crime rates
- Market analysis
- population distribution

As this is a very long research to cover all of these aspects, I decided to take only crime rates with respect to different neighborhoods as my main focal point of the research.

So the proposed project will be a comparison between Toronto and New York cities from crime perspective which neighborhoods has higher crime rates for which type of crimes.

## Data set

the data will be gathered and cleaned from two different resources for new York and Toronto

### ➢ New York data-set

for New York city, the data for crime reporting will be gathered from world data and then will be cleaned and p reprocessed

https://data.world/mithragoesdark/crime-in-new-york-city

Data set consists of 1048575 data element representing different crime incident reporting each has 24 features describing the incident itself

```
CMPLNT_NUM                                101109527
CMPLNT_FR_DT                             12/31/2015
CMPLNT_FR_TM                               23:45:00
CMPLNT_TO_DT                                    NaN
CMPLNT_TO_TM                                    NaN
RPT_DT                                   12/31/2015
KY_CD                                           113
OFNS_DESC                                  FORGERY
PD_CD                                          729
PD_DESC                 FORGERY,ETC.,UNCLASSIFIED-FELO
CRM_ATPT_CPTD_CD                         COMPLETED
LAW_CAT_CD                                 FELONY
JURIS_DESC                          N.Y. POLICE DEPT
BORO_NM                                      BRONX
ADDR_PCT_CD                                     44
LOC_OF_OCCUR_DESC                          INSIDE
PREM_TYP_DESC                        BAR/NIGHT CLUB
PARKS_NM                                        NaN
HADEVELOPT                                      NaN
X_COORD_CD                                1.00731e+06
Y_COORD_CD                                    241257
Latitude                                      40.8288
Longitude                                    -73.9167
Lat_Lon                  (40.828848333, -73.916661142)
Name: 0, dtype: object
```

Example of data element from the data set before cleaning

### ➢ Toronto data-set

Data for Toronto, data for Toronto neighborhood crime rates will be gathered from Toronto police service public data portal

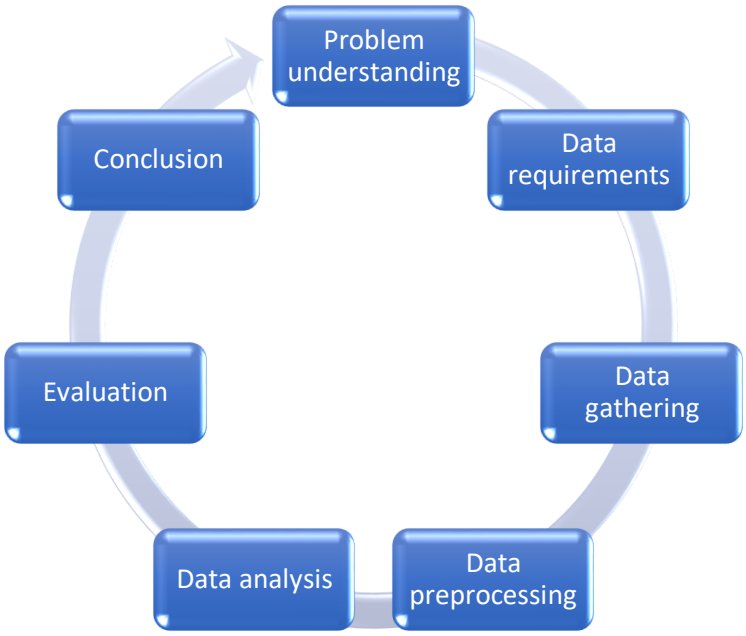
http://data.torontopolice.on.ca/datasets/af500b5abb7240399853b35a2362d0c0_0/data

Gathered data set consists of 140 entry each has 56 features, each entry describing the crime rates at each neighborhood for different types of crimes

```
OBJECTID                                1
Neighbourhood_Crime_Rates_Neigh    Yonge-St.Clair
Neighbourhood_Crime_Rates_Hood_             97
Hood_ID                                    97
Neighbourhood                      Yonge-St.Clair
Assault_2014                               58
Assault_2015                               38
Assault_2016                               51
Assault_2017                               46
Assault_2018                               61
Assault_AVG                              50.8
Assault_CHG                               33%
Assault_Rate_2018                      1912.8
Auto_Theft_2014                            28
Auto_Theft_2015                            32
Auto_Theft_2016                            22
Auto_Theft_2017                            46
Auto_Theft_2018                            69
AutoTheft_AVG                            39.4
AutoTheft_CHG                             50%
AutoTheft_Rate_2018                    2163.7
BreakandEnter_2014                         29
BreakandEnter_2015                         16
BreakandEnter_2016                         28
BreakandEnter_2017                         32
BreakandEnter_2018                         23
BreakandEnter_AVG                        25.6
BreakandEnter_CHG                        -28%
BreakandEnter_Rate_2018                721.2
Robbery_2014                               12
Robbery_2015                               25
Robbery_2016                               14
Robbery_2017                               21
Robbery_2018                               19
Robbery_AVG                              18.2
Robbery_CHG                              -10%
Robbery_Rate_2018                      595.8
Theft_Over_2014                             3
Theft_Over_2015                             6
Theft_Over_2016                             4
Theft_Over_2017                             6
Theft_Over_2018                             3
TheftOver_AVG                             4.4
TheftOver_CHG                            -50%
TheftOver_Rate_2018                      94.1
Homicide_2014                               0
Homicide_2015                               0
Homicide_2016                               0
Homicide_2017                               0
Homicide_2018                               0
Homicide_AVG                              NaN
Homicide_CHG                              NaN
Homicide_Rate_2018                          0
Population                               3189
Shape__Area                      1.16131e+06
Shape__Length                       5873.27
Name: 0, dtype: object
```

Example of Toronto data-set before cleaning

## Methodology




## Data  preprocessing and analysis

in this section, I am describing all the data preprocessing and analysis has been done for the two cities

- ### New York

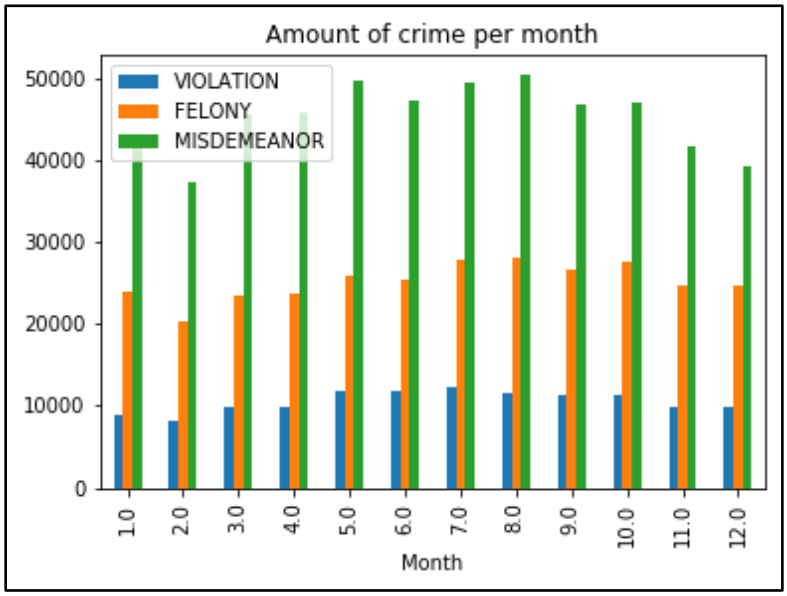
  Data preprocessing has been done to preform the following

  1- remove all unnecessary elements

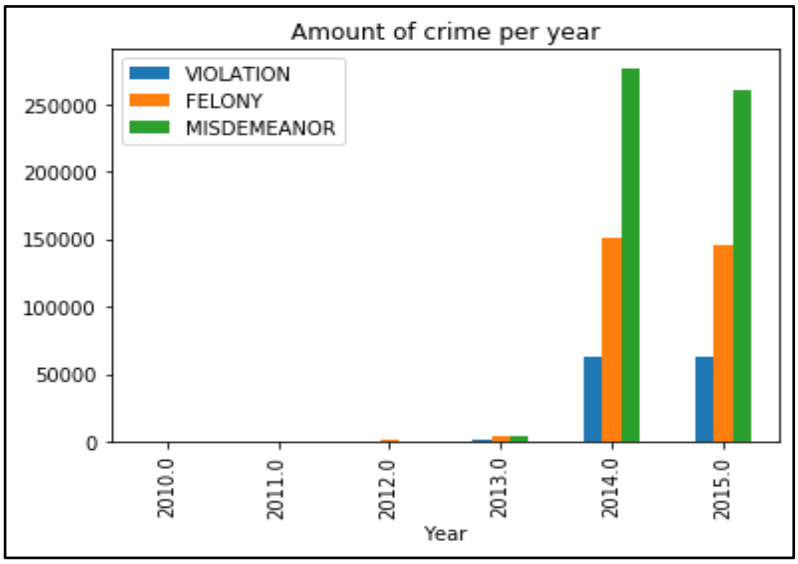
  2- remove unnecessary features

  3- remove Nan values

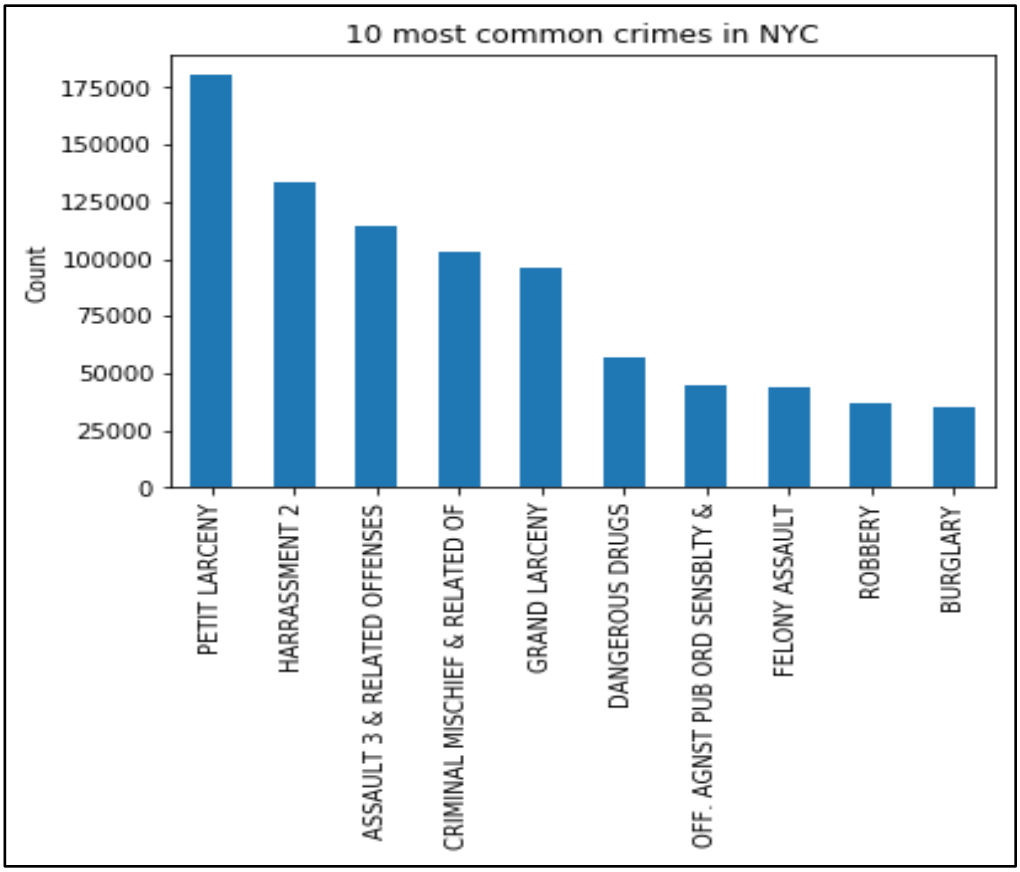
  4- make analysis to occurrence data and time to conclude the following



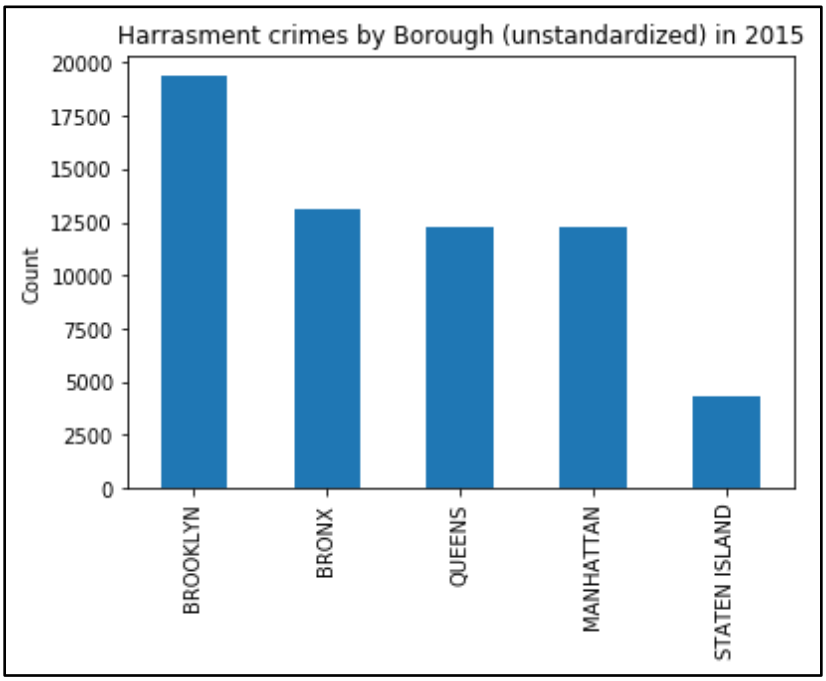

Crime rate per month in New York city



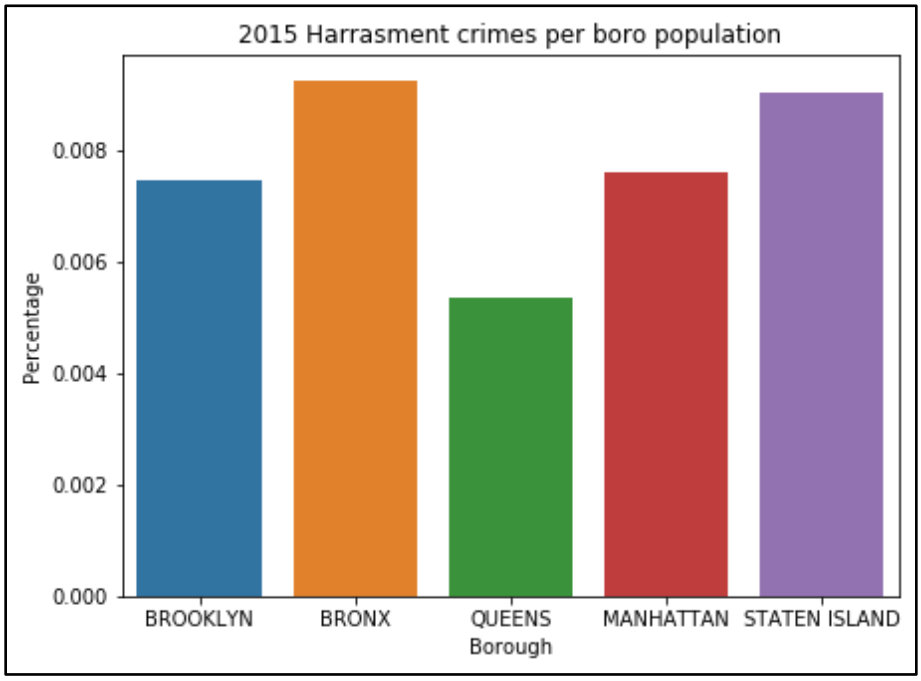

Crime rate per year in New York city

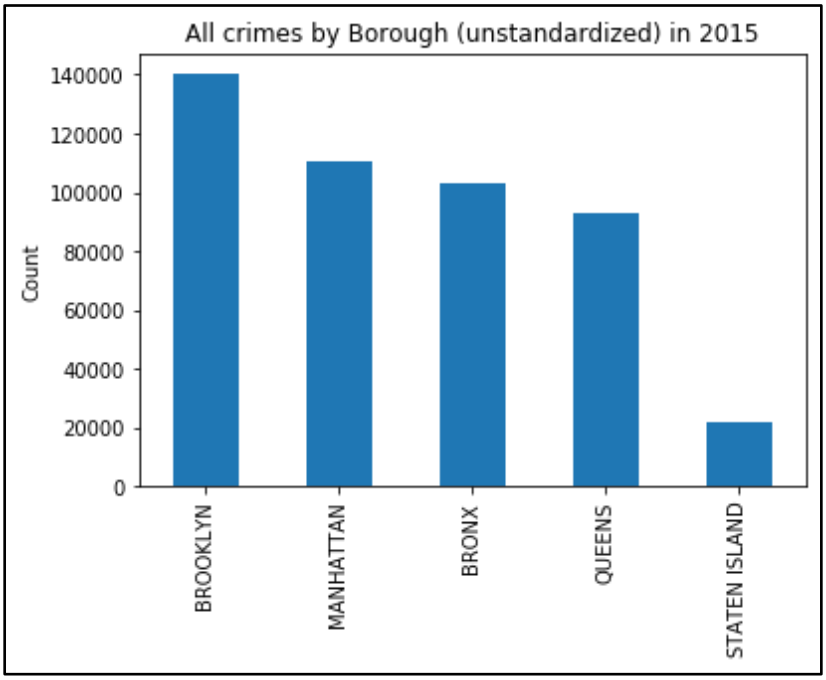



10 most common crimes in NY

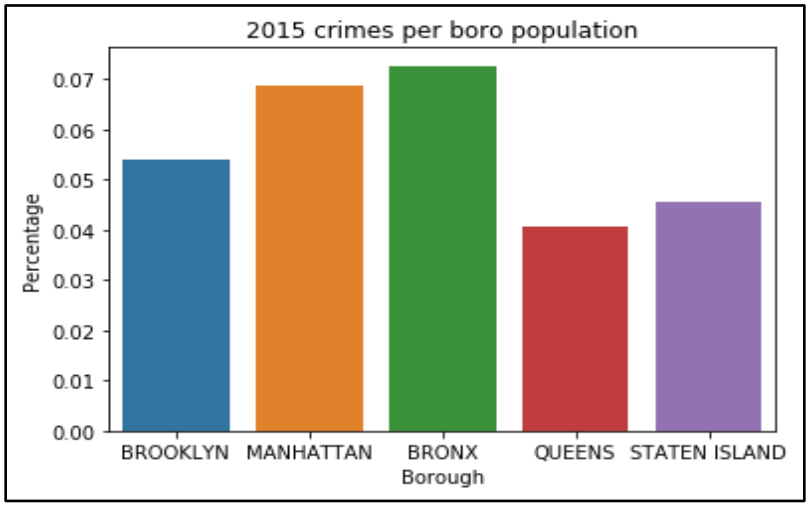



Harassment crime by Borough


2015 Harrasment crimes per boro population
Percentage
0.000
0.002
0.004
0.006
0.008
BROOKLYN
BRONX
QUEENS
MANHATTAN
STATEN ISLAND
Borough


All crimes by Borough (unstandardized) in 2015
Count
0
20000
40000
60000
80000
100000
120000
140000
BROOKLYN
MANHATTAN
BRONX
QUEENS
STATEN ISLAND

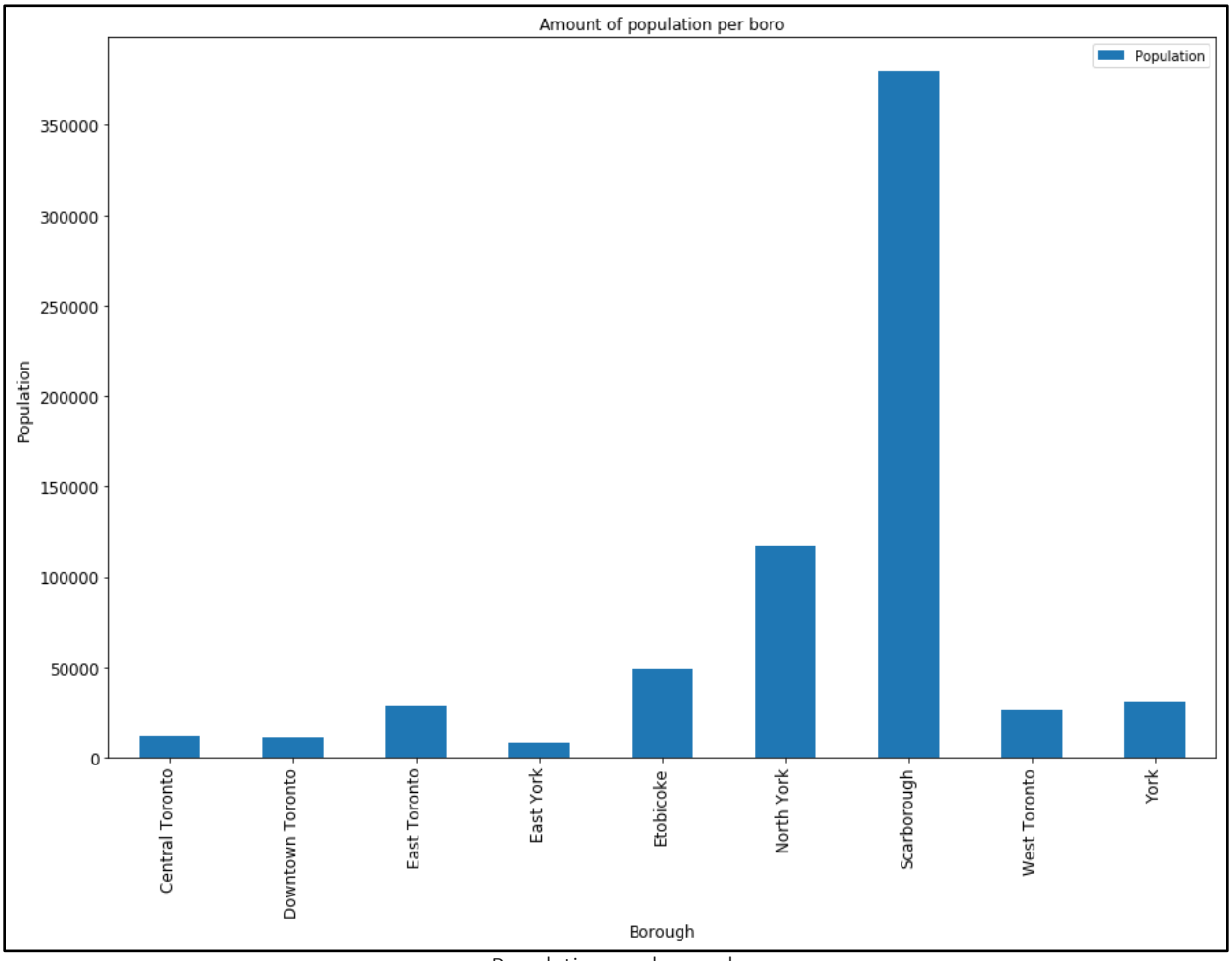
ALL crimes by Borough



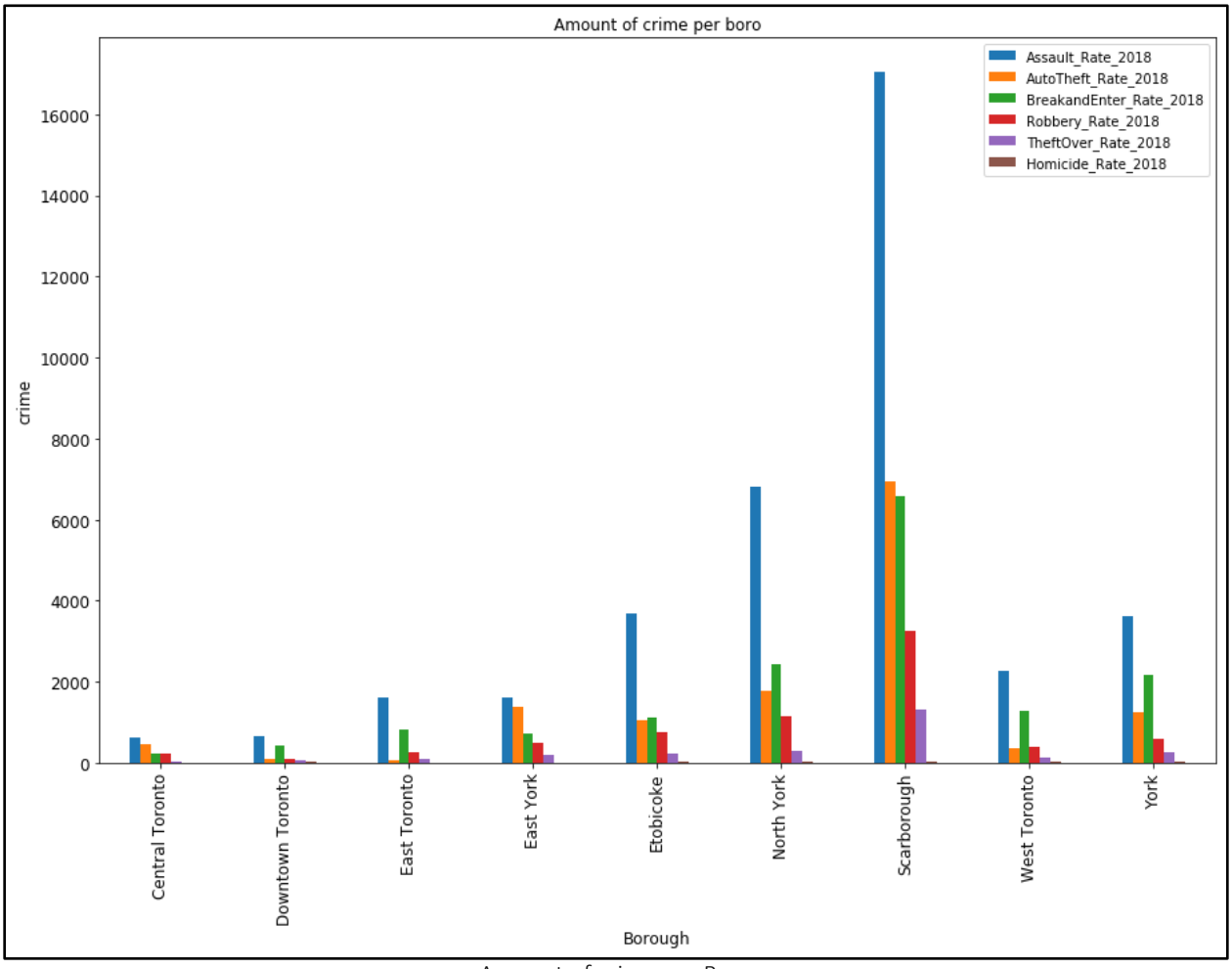

crime per boro population

## ➢ Toronto city

The same preprocessing procedure has been applied to Toronto city



Population per borough



Amount of crime per Boro

## Conclusion

In New York city maximum crime rates occurs in Broklyn while in Toronto it occurs in Scarborough, so these districts are not the best option for investment.