

NYC Crime Analysis

Surbhi Thole (sst390) and Divyanshu Singh (ds5387)

Abstract

To perform data analysis on New York City's Crime data.

Dataset: <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Map-Historic-/57mv-nv28>

Introduction and Motivation

NYC being the most popular state of USA also suffers from many criminal activities' year around. In this project we analyzed the historic crime data and found the localities, seasons, days of months etc. that are most vulnerable for such kind of criminal activities. This will help us to take preventive measures and understand the types of crimes taking place throughout the city.

Methodology

Data Pre-processing

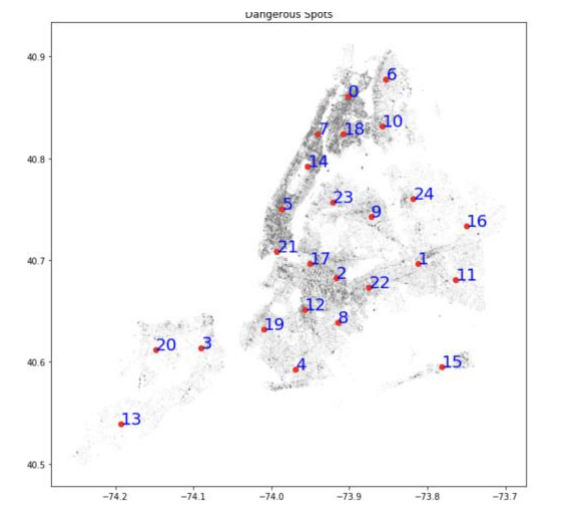
Dataset consist of 35 columns and 6 lakhs rows. For data pre-processing we undertook the following steps:

- Validated the "Complaint_Number" column to check if it contains 7-digit integer values to represent unique complaint id,
- Checked the "Date" column for formatting and range of dates to be 2006 – 2015,
- Validated the "Borough" column to check it the column contains only the 5 boroughs of NYC,
- NYC has a Latitude and Longitude range of (40.477399,40.917577) and (-74.25909, -73.700009) respectively. We validated the Lat – Lon column so that it contains all the values that lie within the mentioned latitude and longitude range,
- We dropped all the columns which contained more than 50% of null values as it won't help us in analysis of the crimes.

Algorithm

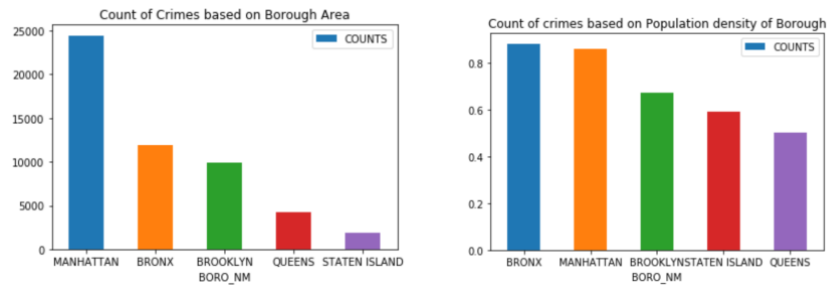
K-means Clustering on Latitude and Longitude of Crime location

- In order to identify the unsafe localities within the boroughs of NYC, we performed K – means clustering so as to find the clusters which helps us to find the density of crimes,
- The cluster centers provide us the information of localities with high crime rate,
- We wanted to identify the most vulnerable locations that should be avoided. Hence, selecting 2 million random samples from the dataset and clustering them provided us with the centers of the clusters giving us the most dangerous locations information. Below graph shows the top 25 vulnerable locations of NYC

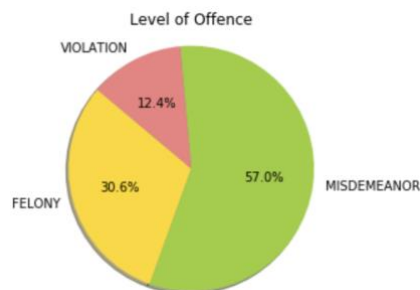


FURTHER ANALYSIS AND RESULTS

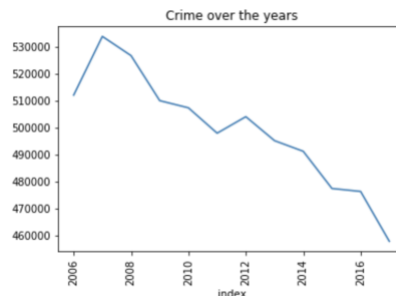
- 1) Validated the most unsafe Borough of NYC based on Area of Borough and Population Density



- Looking at the above graph, we can conclude that Manhattan and Bronx are the most unsafe borough's in NYC.
- 2) We identified the percentage of Felony, Misdemeanor, Violation complaints that are received by NYC Police department.

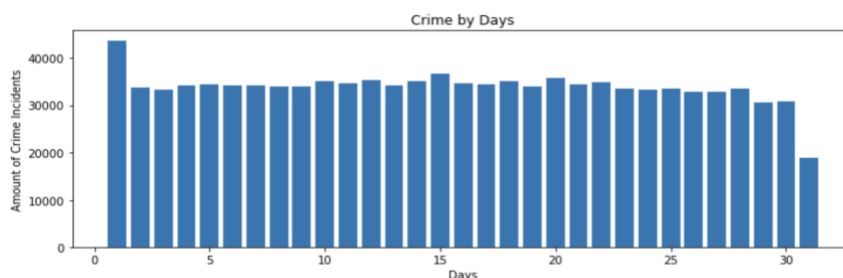


- The above graph shows the most complaints of Misdemeanor being filed in NYC
- 3) To check if the city is getting safe or unsafe to live in, we compared the counts of crimes taking place over the years.

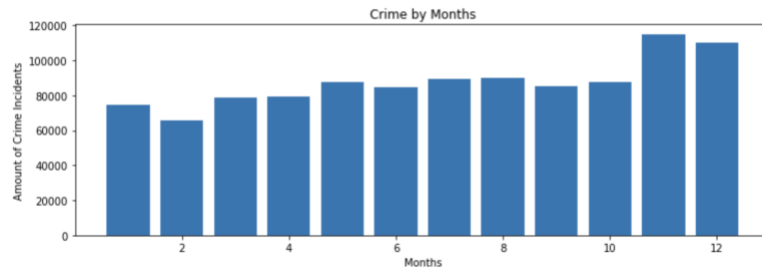


- The above graph shows the crimes rates continues to fall down over the years. We can also see that around the year 2007 there is spike in crime rates. This can be attributed to recession during that period which led to criminal activities.

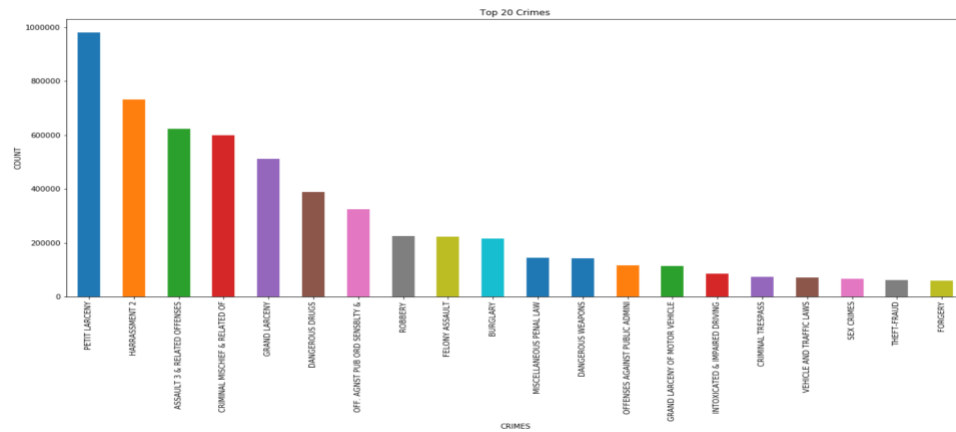
- 4) We also analyzed crimes on Days of month



- The above graph shows that during start of each month there is an increase in criminal activities. This might be because people get their salaries in start of the month which leads to criminal activities
- The below graph shows the most vulnerable months of the year. It can be seen that the criminal activities take place in winter seasons

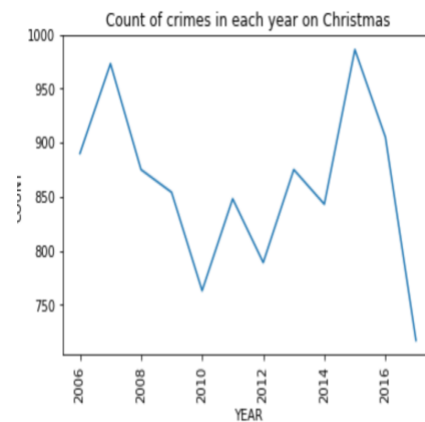
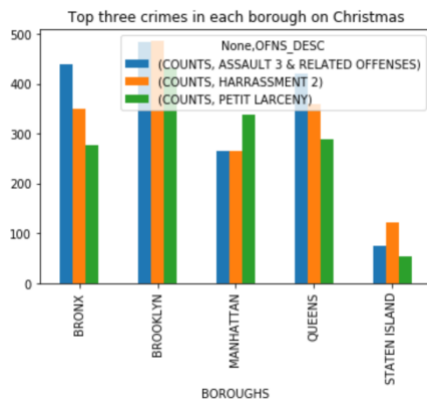


5) The top 20 crimes that takes place in NYC

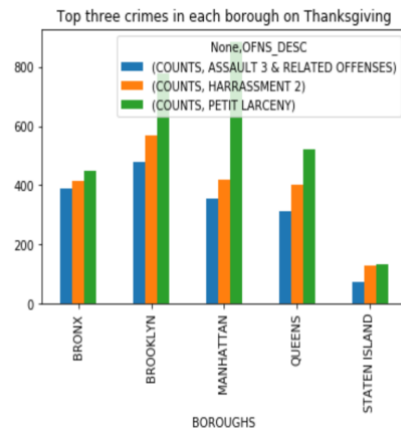
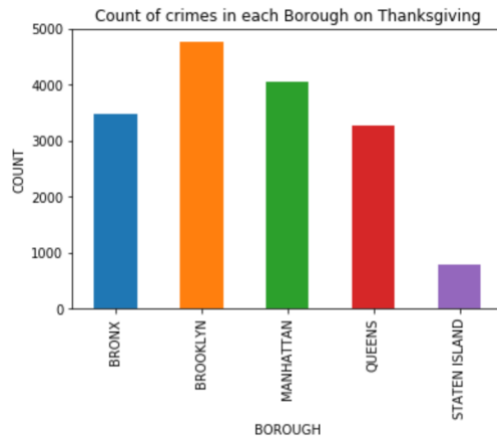


6) Seasonal Crimes – 25th December

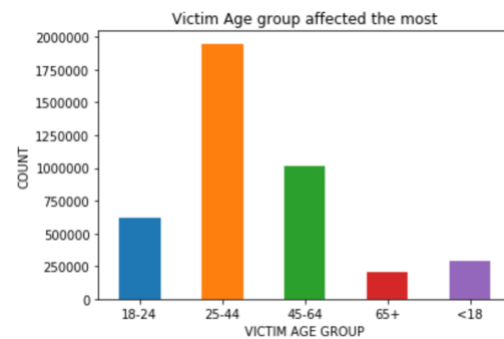
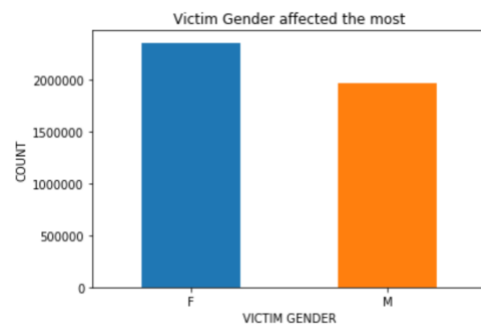
- We analyzed the crimes taking place in **Christmas**. The results are as follows



- We analyzed the crimes taking place on **Thanksgiving**. The results are as follows



7) The victim age group, gender that is affected the most



DISCUSSION

We analyzed various crimes taking place in NYC in each borough. We analyzed that crimes that place in each month, and around the year range of 2006 – 20015. The analysis of seasonal crimes helped us to take a deeper insights into the activities that takes place on Christmas, thanksgiving etc

CONCLUSION

To conclude with our analysis, we found out that the most unsafe place to live in NYC is Manhattan followed by Bronx. We analyzed that people should take care at the start of each month and during the holiday seasons and winters as the crime activities are at its peak during these days and seasons. To be safe, girls should be more careful as they are the victims of most of the crimes. To further add the age group of 25-44 years is the victim of most criminal activities. Our analysis also showed that the people are not safe inside the parks, streets during late nights in certain localities highlighted by our algorithm and hence should be aware of the surroundings.

REFERENCES

- <https://www.kaggle.com/adamschroeder/crimes-new-york-city>
- <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i>