CRIME IN NEW YORK CITY: HOW SAFE NYC REALLY IS?

Yashasvi Choukikar
New York University
(yc4953)
yc4953@nyu.edu

Anmol Ramayanam
New York University
(lr2947)
lr2947@nyu.edu

Sanjeet Shetty
New York University
(ss14431)
ss14431@nyu.edu

December 17, 2022

I. Abstract

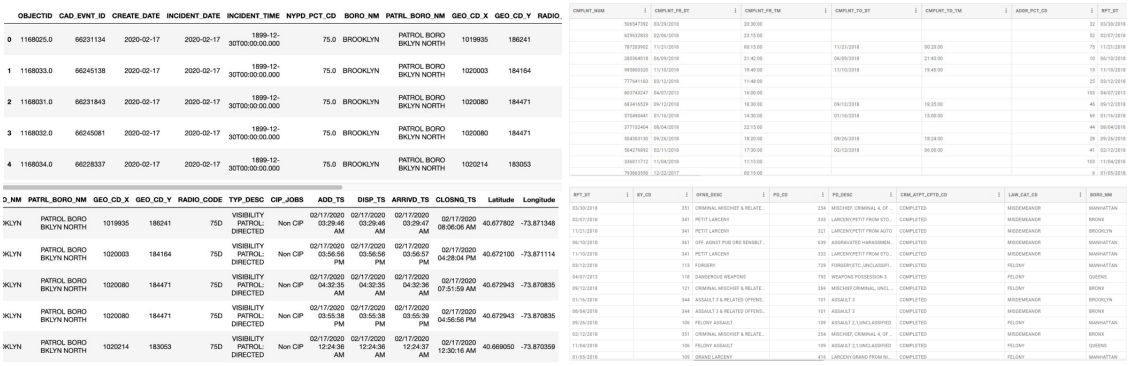
In this project we intend to focus on the crimes committed in New York City. By analyzing data from complaints registered with the NYPD, arrests made by NYPD, hate crimes committed in NYC, population census over the years in NYC, and distressed 911 calls described with various subcategories, we aim to find valuable insights of crime statistics across 5 boroughs of New York City. Our findings we believe shall help diagnosing and prescribing solutions such as delays in 911 dispatch help, which areas to avoid. By discovering the locations of most dangerous crimes , the timing stats can help in establishing better patrolling and prevention policies.

Keywords: *Big Data, HDFS, Apache Spark, Scala, Tableau, Spark SQL*

II. Introduction

Our analysis required data about the 911 calls, NYPD police complaints and NYPD arrests that are made. The datasets are updated and maintained by NYPD and are publicly available. Below are the snapshots of the datasets we have used for the analysis:

- NYC 911 Calls (Source): To communicate with callers and the NYPD, phone takers and dispatchers use this information.
- NYPD Complaints (Source): This dataset comprises where the hate crime is committed and segmented into various categories of crimes like felony, aggravated assault for the years 2019-2022



	OBJECTID	CAD_EVNT_ID	CREATE_DATE	INCIDENT_DATE	INCIDENT_TIME	NYPD_PCT_CD	BORO_NM	PATRL_BORO_NM	GEO_CD_X	GEO_CD_Y	RADIO_
0	1168025.0	66231134	2020-02-17	2020-02-17	1899-12-30T00:00:00.000	75.0	BROOKLYN	PATROL BORO BKLYN NORTH	1019935	186241	
1	1168033.0	66245138	2020-02-17	2020-02-17	1899-12-30T00:00:00.000	75.0	BROOKLYN	PATROL BORO BKLYN NORTH	1020003	184164	
2	1168031.0	66231843	2020-02-17	2020-02-17	1899-12-30T00:00:00.000	75.0	BROOKLYN	PATROL BORO BKLYN NORTH	1020080	184471	
3	1168032.0	66245081	2020-02-17	2020-02-17	1899-12-30T00:00:00.000	75.0	BROOKLYN	PATROL BORO BKLYN NORTH	1020080	184471	
4	1168034.0	66228337	2020-02-17	2020-02-17	1899-12-30T00:00:00.000	75.0	BROOKLYN	PATROL BORO BKLYN NORTH	1020214	183053	

D_NM	PATRL_BORO_NM	GEO_CD_X	GEO_CD_Y	RADIO_CODE	TYP_DESC	CIP_JOBS	ADD_TS	DISP_TS	ARRIVD_TS	CLOSNG_TS	Latitude	Longitude
KLYN	PATROL BORO BKLYN NORTH	1019935	186241	75D	VISIBILITY PATROL: DIRECTED	Non CIP	02/17/2020 03:29:46 AM	02/17/2020 03:29:46 AM	02/17/2020 03:29:47 AM	02/17/2020 08:06:06 AM	40.677802	-73.871348
KLYN	PATROL BORO BKLYN NORTH	1020003	184164	75D	VISIBILITY PATROL: DIRECTED	Non CIP	02/17/2020 03:56:56 PM	02/17/2020 03:56:56 PM	02/17/2020 03:56:57 PM	02/17/2020 04:28:04 PM	40.672100	-73.871114
KLYN	PATROL BORO BKLYN NORTH	1020080	184471	75D	VISIBILITY PATROL: DIRECTED	Non CIP	02/17/2020 04:32:35 AM	02/17/2020 04:32:35 AM	02/17/2020 04:32:36 AM	02/17/2020 07:51:59 AM	40.672943	-73.870835
KLYN	PATROL BORO BKLYN NORTH	1020080	184471	75D	VISIBILITY PATROL: DIRECTED	Non CIP	02/17/2020 03:55:38 PM	02/17/2020 03:55:38 PM	02/17/2020 03:55:39 PM	02/17/2020 04:56:56 PM	40.672943	-73.870835
KLYN	PATROL BORO BKLYN NORTH	1020214	183053	75D	VISIBILITY PATROL: DIRECTED	Non CIP	02/17/2020 12:24:36 AM	02/17/2020 12:24:36 AM	02/17/2020 12:24:37 AM	02/17/2020 12:30:16 AM	40.669050	-73.870359

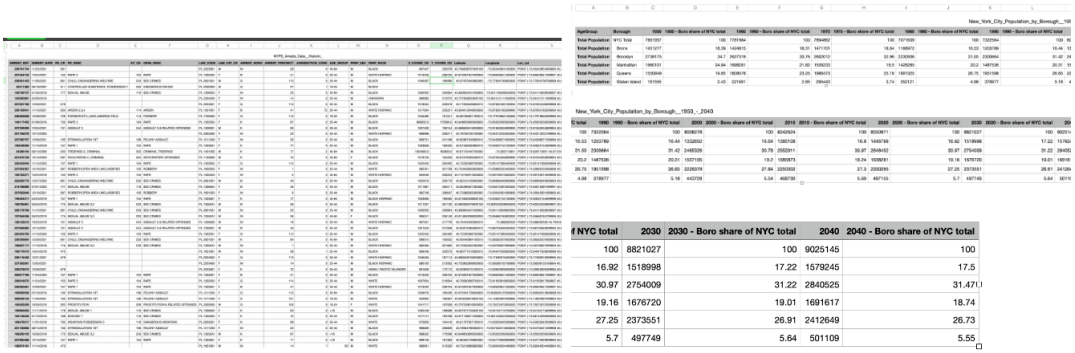
CMPLNT_NUM	CMPLNT_FR_DT	CMPLNT_FR_TM	CMPLNT_TO_DT	CMPLNT_TO_TM	ADDR_PCT_CD	RPT_DT
506547392	03/29/2018	20:30:00			32	03/30/2018
629632933	02/06/2018	23:15:00			52	02/07/2018
787283902	11/21/2018	00:15:00	11/21/2018	00:20:00	75	11/21/2018
283364618	05/09/2018	21:42:00	05/09/2018	21:43:00	19	05/10/2018
165865530	11/10/2018	19:40:00	11/10/2018	19:48:00	19	11/10/2018
777441183	03/12/2018	11:46:00			23	03/12/2018
603743247	04/07/2013	16:00:00			103	04/07/2013
683415529	09/12/2018	18:30:00	09/12/2018	18:35:00	40	09/12/2018
570495441	01/16/2018	14:30:00	01/16/2018	15:00:00	69	01/16/2018
377132404	06/04/2018	22:15:00			44	06/04/2018
164831189	09/26/2018	18:20:00	09/26/2018	18:24:00	28	09/26/2018
584276892	02/11/2019	17:30:00	02/12/2019	06:00:00	41	02/12/2019
336017712	11/04/2018	11:15:00			103	11/04/2018
793863350	12/22/2017	00:15:00			6	01/05/2018

RPT_DT	KY_CD	OFNS_DESC	PD_CD	PD_DESC	CRM_ATPT_CPTD_CD	LAW_CAT_CD	BORO_NM
03/30/2018		CRIMINAL MISCHIEF & RELATE...	254	MISCHIEF, CRIMINAL 4 OF ...	COMPLETED	MISDEMEANOR	MANHATTAN
02/07/2018		PETIT LARCENY	340	LARCENY,PETIT FROM STO...	COMPLETED	MISDEMEANOR	BRONX
11/21/2018		PETIT LARCENY	341	LARCENY,PETIT FROM AUTO	COMPLETED	MISDEMEANOR	BROOKLYN
05/10/2018		OFF AGNST PUB ORD SENSBLT...	639	AGGRAVATED HARASSMEN...	COMPLETED	MISDEMEANOR	MANHATTAN
11/10/2018		PETIT LARCENY	340	LARCENY,PETIT FROM STO...	COMPLETED	MISDEMEANOR	MANHATTAN
03/12/2018		FORGERY	729	FORGERY,ETC.,UNCLASSIFI...	COMPLETED	FELONY	MANHATTAN
04/07/2013		DANGEROUS WEAPONS	793	WEAPONS POSSESSION 3	COMPLETED	FELONY	QUEENS
09/12/2018		CRIMINAL MISCHIEF & RELATE...	269	MISCHIEF,CRIMINAL, UNCL...	COMPLETED	FELONY	BRONX
01/16/2018		ASSAULT 3 & RELATED OFFENS...	101	ASSAULT 3	COMPLETED	MISDEMEANOR	BROOKLYN
06/04/2018		ASSAULT 3 & RELATED OFFENS...	101	ASSAULT 3	COMPLETED	MISDEMEANOR	BRONX
09/26/2018		FELONY ASSAULT	109	ASSAULT 2,1,UNCLASSIFIED	COMPLETED	FELONY	MANHATTAN
02/12/2018		CRIMINAL MISCHIEF & RELATE...	254	MISCHIEF, CRIMINAL 4 OF ...	COMPLETED	MISDEMEANOR	BRONX
11/04/2018		FELONY ASSAULT	109	ASSAULT 2,1,UNCLASSIFIED	COMPLETED	FELONY	QUEENS
01/05/2018		GRAND LARCENY	439	LARCENY,GRAND FROM NI...	COMPLETED	FELONY	MANHATTAN

Figure 1: *911 Calls and NYPD Complaints Dataset Snapshot*

- NYPD Arrests (Source): It includes a breakdown of each arrest that the NYPD made in NYC from 2006 through the end of the previous calendar year

- NYC Population by Borough Source: Summary table of New York City population numbers and percentage shares by Borough, including school-age (5 to 17), 65 and Over, and total population for over the years 1950-2040.



f NYC total	2030	2030 - Boro share of NYC total	2040	2040 - Boro share of NYC total
100	8821027	100	9025145	100
16.92	1518998	17.22	1579245	17.5
30.97	2754009	31.22	2840525	31.47
19.16	1676720	19.01	1691617	18.74
27.25	2373551	26.91	2412649	26.73
5.7	497749	5.64	501109	5.55

Figure 2: *NYPD Arrests and Population By Borough Dataset Snapshot*

- NYC Hate Crimes Source: This dataset comprises where the hate crime is committed and segmented into various categories of crimes like felony, aggravated assault for the years 2019-2022
- NYC 2010-2020 Census Data Source: This dataset comprises different information about the basic housing and demographic information for different NYC neighborhoods between the years 2010 and 2020.

III. Design and Architecture

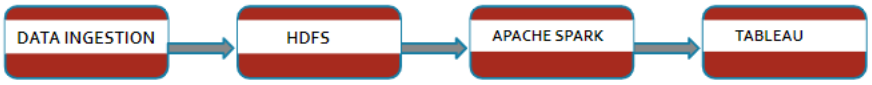


Figure 3: *Architecture*

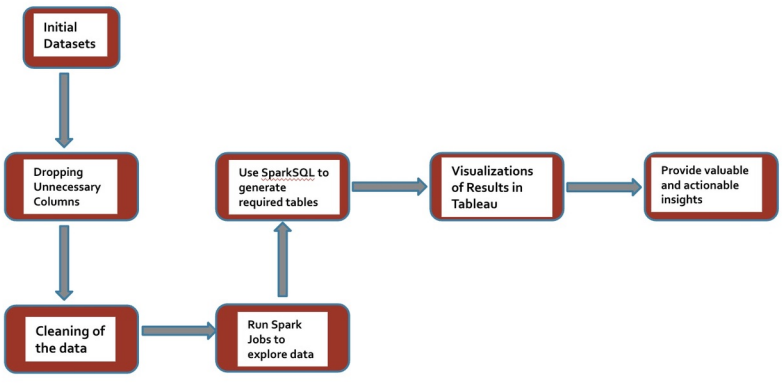


Figure 4: *Flowchart*

By integrating Scala with Hadoop, using Apache Spark [10]we were able to successfully store data for longer periods of time , making it easiest to use APIs to search against it. Having done data profiling leveraging Scala, creating Spark sessions and with the help of UDF to transform dates in required format and dropping unnecessary null values from our datasets, we have stored the cleaned datasets (10GBs in size) in HDFS [11]. Aggregated data results of the combined datasets were done in Zeppelin [12] by activating the enableHiveInput() option in client mode. ML modeling to predict the number of arrests were helped by using Apache Spark's MLLib feature [13] in Zeppelin. Owing to the size of the datasets, all our resulting queries were saved in cvs format(obtained by the spark job submission), stitched together using Python [14] and visualized in Tableau [15]

IV. Analysis and Findings

Initial idea was to track the 911 calls which translated into complaints at NYPD and the complaints to arrests, but doing so posed a challenge as we did not have case IDs information or a single unique identifier to join all three tables. Since our datasets have varying dates (data accumulated through 2006 to present), our aggregated data was found for years 2018, 2019, 2020 and 2021.

Ratio of calls to complaints to arrests

By comparing the number of calls, complaints and arrests, we observed the trend remained same with a consistent ratio between 911 Calls, complaints registered and arrests made. We also observed that every year most number of calls are made in the month of May, with maximum calls in the period 2018-2021 happening in the year 2020.

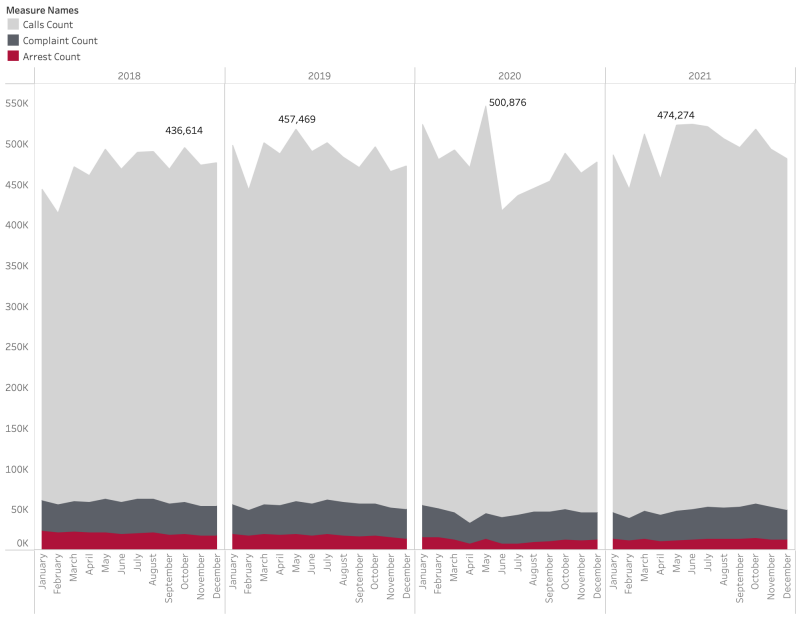


Figure 5: *Ratio of Calls, Complaints and Arrests over the period 2018-2021*

Offense

When we categorized the complaints registered to arrests made by NYPD, we see that maximum number of complaints are registered as Petit Larceny and maximum arrests made are of the type Assault. The conversion of a assault complaint to an arrest is highest among all the crime categories in the past 4 years.

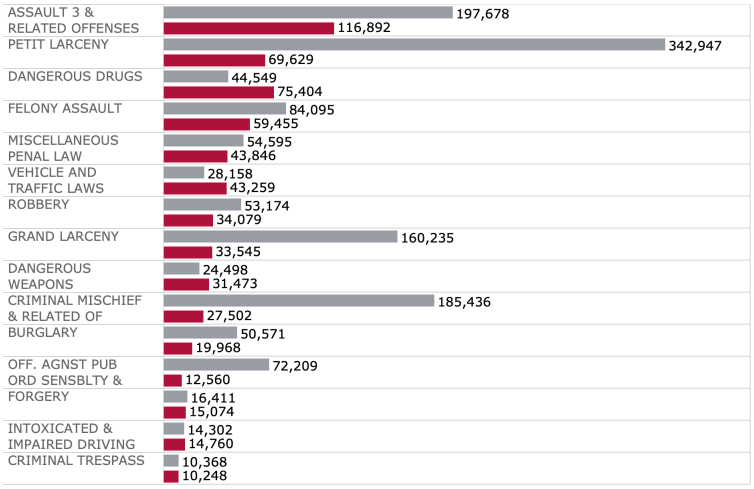


Figure 6: *Top Offenses*

Dispatch time

When we categorized the 911 calls into their respective boroughs to see the time it takes for the police to reach the crime scene, we noticed that out the 5 boroughs of New York City, Bronx has the slowest average dispatch time of about 10.262 minutes.

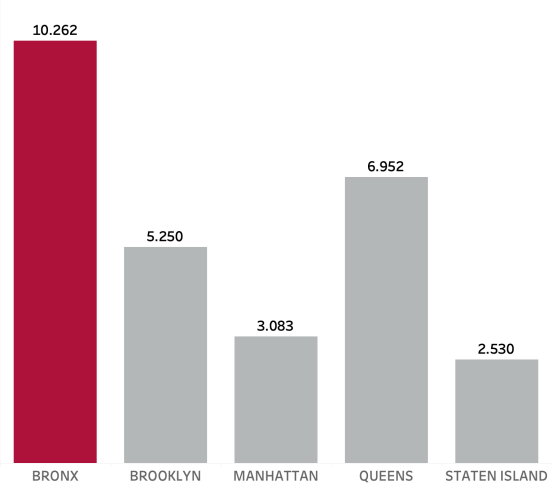


Figure 7: *Dispatch Delays*

Borough wise gender

Our data has kept with the modern times, having included Diverse and Inclusive/Neutral in its category. We observed for genders, victims were predominantly females across all five boroughs. [3]

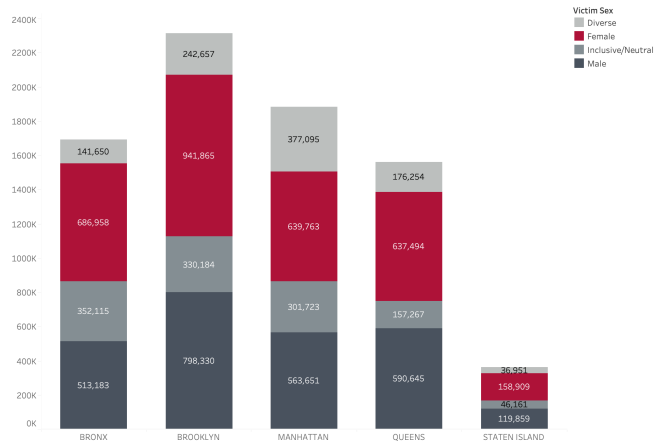


Figure 8: *Gender Distribution in Boroughs*

Borough wise race

Further diving into the victim subset formed from aggregated data set, on analysing in terms of race, African-Americans are most susceptible to be victims in Brooklyn and Bronx while Caucasians in Brooklyn and Manhattan. This is justified based on our population consensus data of 2020.

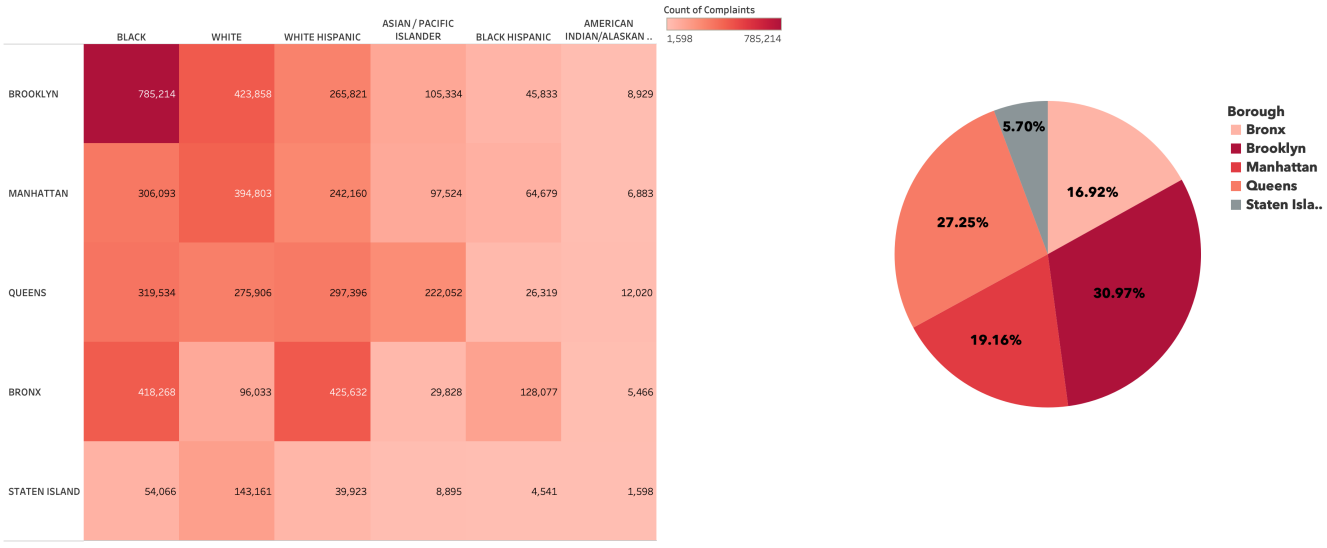


Figure 9: *Race Distribution and Population Density in Boroughs*

Borough wise calls, complaints and arrests

When we compared the volumes of calls, complaints and arrests that happened in NYC's 5 boroughs, we find Brooklyn has the highest crime rate that gets reported while Staten Island has the lowest crime rate. The ratio of complaints to arrests is higher in Manhattan than in Brooklyn.

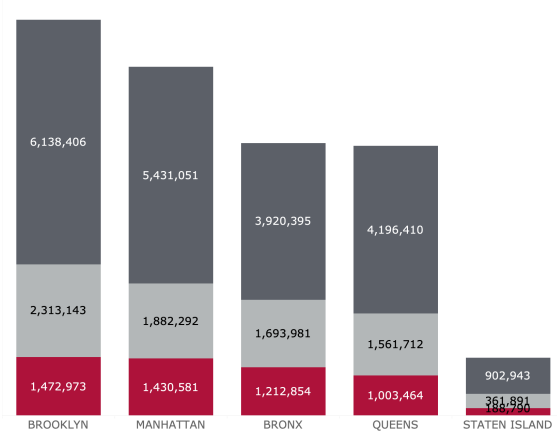


Figure 10: *Call Distribution in Boroughs*

Hate crimes monthly

Over the years we saw an increase in Hate crime rate in the month of March, and overall the number of hate crime committed remain consistent and we do not see any decline.

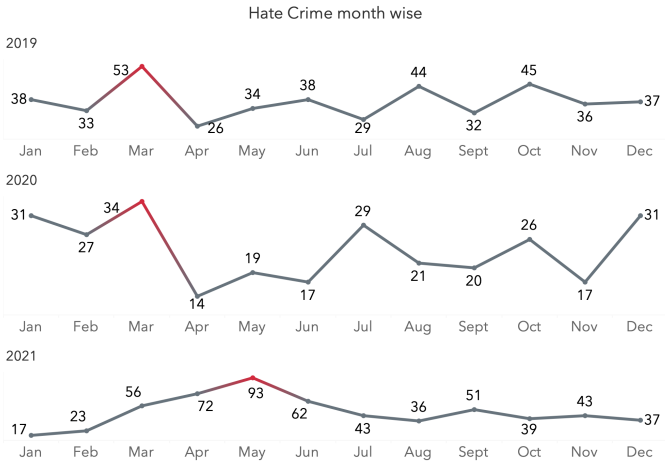


Figure 11: *Hate crimes Month Wise*

ML Modeling

We also attempted to apply ML Modeling to our dataset by mapping the number of 911 calls along with official NYPD complaints to assess number of successful arrests. Given the number of 911 calls, complaints registered per day per NYPD precinct, month and day of week, we predicted the number of arrests by comparing the models leveraging Linear Regression, Random Forest Regressor and Gradient Boosted Tree Regression.

features	COUNT_ARREST	prediction
(1470,[1,1464,1466,1467,1468,1469],[1.0,1.0,122.0,187.0,6.0,1.0])	8	4.878614125290836
(1470,[1,1464,1466,1467,1468,1469],[1.0,1.0,123.0,97.0,6.0,1.0])	4	3.696190126976548
(1470,[1,1464,1466,1467,1468,1469],[1.0,1.0,19.0,182.0,20.0,1.0])	8	9.435968387697114
(1470,[1,1464,1466,1467,1468,1469],[1.0,1.0,24.0,129.0,13.0,1.0])	1	6.528575484608318
(1470,[1,1464,1466,1467,1468,1469],[1.0,1.0,26.0,85.0,11.0,1.0])	1	5.318230718562197

ML Modeling	RMSE Score	R2 Score
Linear Regression	4.18	0.44
Random Forest Regression	4.61	0.32
GBT Modelling	4.45	0.36

Figure 12: *Model Feature Selection and Performance Metrics*

V. Conclusion

1. Based on our findings, 7, Dispatch timings for 911 Calls needs to be improved for Bronx significantly in comparison to other boroughs. [1], [2]
2. Our findings for females being most susceptible to crime 8 is in line with the current scenario as cited by City Journal News outlet [3]
3. Hate Crimes 11have been on the rise [4],[5] and [6]. Raising social awareness and using resources such as [7] and [8] will help in mitigating this phenomenon.
4. In recent years, NYPD has been more pro-active in arresting law-defying citizens for petty crimes. [9]

VI. Future Scope

1. For ML modeling, these are average results 12 from our preliminary analysis. This could be improved by adding more features to our model such as offense, age
2. Unique identifier to be added in complaints and arrests, so that better analytics and focus areas can be identified by NYPD, and NYPD could use these analytics to see in what areas they could improve and efficiency for the same can be improved.

VII. Acknowledgements

1. For graphs and visualizations, we utilized Tableau Desktop with a student license (associated with NYU, Stern, Data Visualization Class).
2. We thank NYU HPC[10] Team for providing us their platform as a computational and storage resource.
3. We appreciate Prof. Yang Tang's [16]fast responses to our Discord questions.

References

[1] https://ibo.nyc.ny.us/cgi-park2/2019/05/theres-a-call-reporting-a-crime-in-progress-how-long-does-it-

[2] https://www.vera.org/downloads/publications/911-call-processing-system-review-of-policing-literature.pdf

[3] https://www.city-journal.org/surge-in-violent-crime-against-nyc-women

[4] https://ag.ny.gov/press-release/2020/ag-james-launches-hotline-combat-coronavirus-hate-crimes-and-xen

[5] https://www.thecity.nyc/2022/1/4/22867479/2021-saw-another-surge-of-anti-asian-hate-crimes-in-subway

[6] https://www.cnn.com/2020/01/07/us/new-york-crime-stats-anti-semitic-hate-crimes

[7] https://www.nyc.gov/site/cchr/about/report-discrimination.page

[8] https://www.nyc.gov/site/cchr/help/i-need-help.page

[9] https://www.bloomberg.com/news/articles/2022-08-30/nyc-s-rise-of-low-level-arrests-worry-critics-of-b

[10] NYU HPC Team https://sites.google.com/nyu.edu/nyu-hpc/home

[11] Apache Hadoop https://hadoop.apache.org/

[12] Apache Zeppelin https://zeppelin.apache.org/

[13] Apache MlLib https://spark.apache.org/docs/latest/ml-guide.html

[14] Python 3.8 - Anaconda https://www.anaconda.com/products/distribution

[15] Tableau - licensed with NYU https://www.tableau.com/

[16] Professor Yang Tang https://ytang.com