NYU

Big Data Application Development Symposium - Fall 2022

12.12.22

Crime in NYC

(How Safe NYC Really is?)

Yashasvi Choukikar (yc4953)
Lakshmi Sai Anmol Ramayanam(lr2947)
Sanjeet Vijaya Shetty(ss14431)

Abstract

- Our analysis's primary focus is the **Crimes committed in New York City**,
- The aggregated data for the project comprised of the following datasets: Complaints registered with the NYPD, Arrests made by NYPD, Distressed 911 calls described with various subcategories.
- **AIM: to find valuable insights of crime statistics across the 5 boroughs of New York City**.
- Our findings we believe shall help in diagnosing problems such as **Delays in 911 dispatch help**, **which areas to avoid in NYC.**

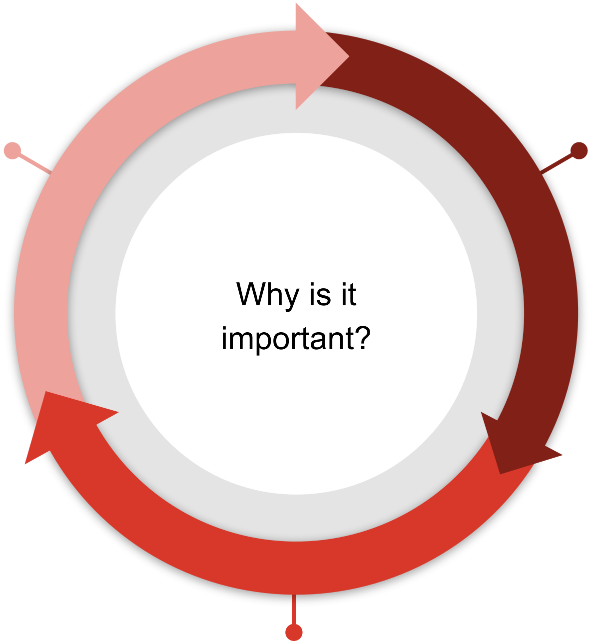
Platforms where this applications run/tools used:

HPC data proc, Tableau, Zeppelin

Motivation

Users?

New Yorkers



Beneficiaries?

NYPD and other government agencies

Active & Successful crime statistics to understand the problematic areas to avoid and reduce crime rate

Metric Goodness

- **Sense of Safety:** Crime rates in a particular area play an important role in the peace and safety index of the entire city

- **Trust in Police :** Features such as severity of the crime,location of the crime,victim age group,suspect age group would be helpful in the goodness analysis.

- **Job Creation :** One of the first things prospective employers look for in establishing a business is the neighborhood environment. Is the area safe? Are robberies and vandalism prevalent? How quickly can police respond when needed?

Data Sources

Data Source 1

- **Name:-** NYC 911 calls
- **Description:-** This dataset is created by the general public and ones started by NYPD employees are included in the data. The information can be used for problems that the NYPD is addressing.
- **Size:-** Dataset Size(6.8 GB)

Data Source 2

- **Name:-** NYPD Complaints
- **Description:-** This dataset comprises every legitimate felony, misdemeanor, and violation crime previous year
- **Size:-** Dataset Size(~2.51GB)

Data Source 3

- **Name:-** NYPD Arrests
- **Description:-** A list of each arrest made in New York City from the beginning of 2006 till the conclusion of the preceding year.
- **Size:-** 1.13 GB

Data Source 4

- **Name:-** NYC Hate Crime
- **Description:-** Dataset of confirmed hate crimes in New York City from the years 2019-2022
- **Size:-** 313 Kb

Data Source 5

- **Name**:- NYC 2010-2020 Census Data
- **Description**:- This dataset comprises different information about the basic housing and demographic information for different NYC neighborhoods.
- **Size:- Dataset Size (5 MB)**

Data Source 6

- **Name:-** NYC Population by Borough
- **Description:-** This dataset comprises every legitimate felony, misdemeanor, and violation crime previous year
- **Size:- Dataset Size (2KB)**

Data Source 1 Sample - NYC 911 Calls

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
629632833	02/06/2018	23:15:00			52	02/07/2018
787203902	11/21/2018	00:15:00	11/21/2018	00:20:00	75	11/21/2018
280364018	06/09/2018	21:42:00	06/09/2018	21:43:00	10	06/10/2018
985800320	11/10/2018	19:40:00	11/10/2018	19:45:00	19	11/10/2018
777641183	03/12/2018	11:48:00			25	03/12/2018
803743247	04/07/2013	16:00:00			103	04/07/2013
683416529	09/12/2018	18:30:00	09/12/2018	18:35:00	46	09/12/2018
570490441	01/16/2018	14:30:00	01/16/2018	15:00:00	69	01/16/2018
377132404	08/04/2018	22:15:00			44	08/04/2018
504303130	09/26/2018	18:20:00	09/26/2018	18:24:00	28	09/26/2018
584276892	02/11/2018	17:30:00	02/12/2018	06:00:00	41	02/12/2018
336011712	11/04/2018	11:15:00			103	11/04/2018
793863550	12/22/2017	00:15:00			6	01/05/2018

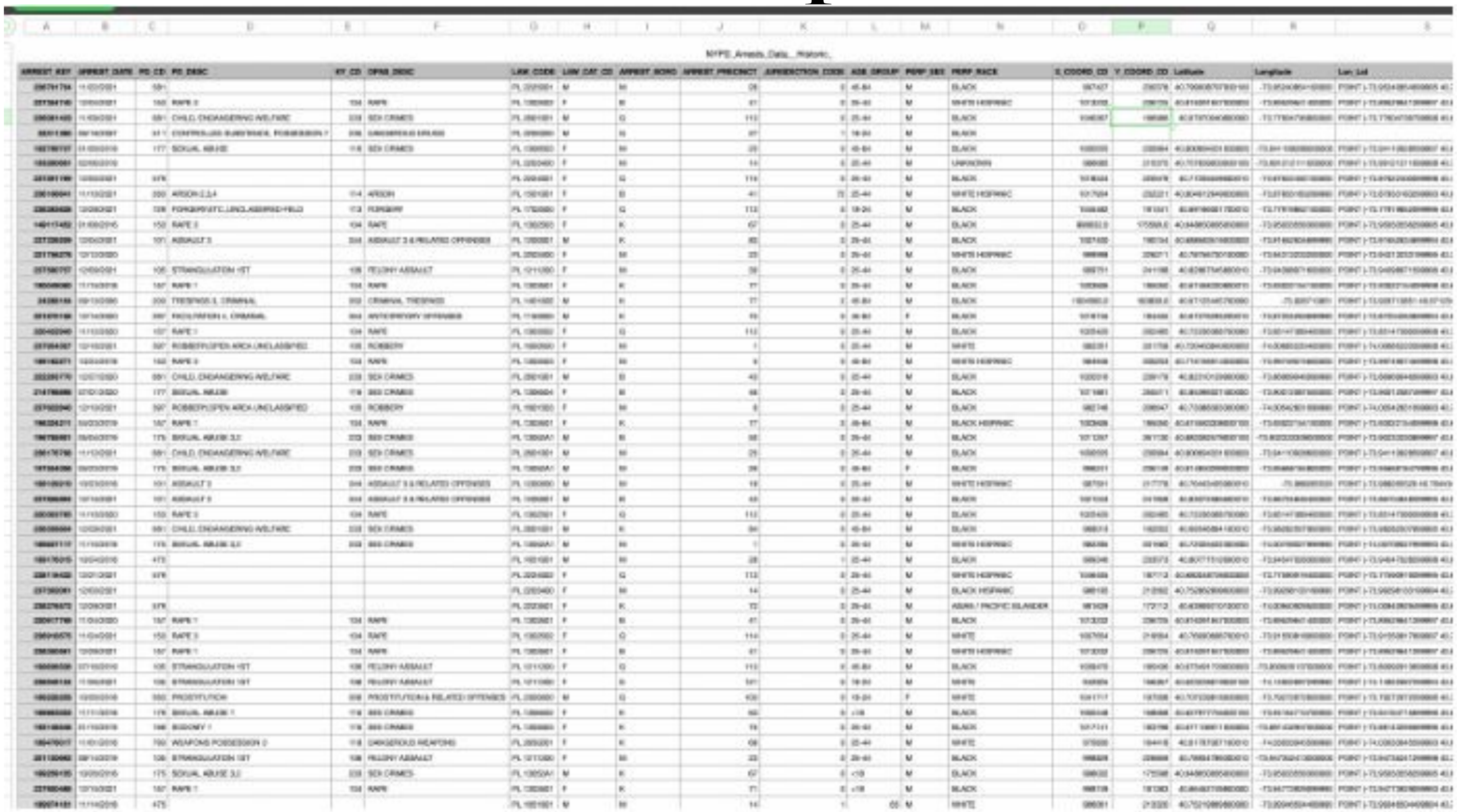
RPT_DT	KY_CD	OFNS_DESC	PD_CD	PD_DESC	CRM_ATPT_CPTD_CD	LAW_CAT_CD	BORO_NM
03/30/2018	351	CRIMINAL MISCHIEF & RELATE...	254	MISCHIEF, CRIMINAL 4, OF ...	COMPLETED	MISDEMEANOR	MANHATTAN
02/07/2018	341	PETIT LARCENY	333	LARCENY,PETIT FROM STO...	COMPLETED	MISDEMEANOR	BRONX
11/21/2018	341	PETIT LARCENY	321	LARCENY,PETIT FROM AUTO	COMPLETED	MISDEMEANOR	BROOKLYN
06/10/2018	361	OFF. AGNST PUB ORD SENSBLT...	639	AGGRAVATED HARASSMEN...	COMPLETED	MISDEMEANOR	MANHATTAN
11/10/2018	341	PETIT LARCENY	333	LARCENY,PETIT FROM STO...	COMPLETED	MISDEMEANOR	MANHATTAN
03/12/2018	113	FORGERY	729	FORGERY,ETC.,UNCLASSIFI...	COMPLETED	FELONY	MANHATTAN
04/07/2013	118	DANGEROUS WEAPONS	793	WEAPONS POSSESSION 3	COMPLETED	FELONY	QUEENS
09/12/2018	121	CRIMINAL MISCHIEF & RELATE...	269	MISCHIEF,CRIMINAL, UNCL ...	COMPLETED	FELONY	BRONX
01/16/2018	344	ASSAULT 3 & RELATED OFFENS...	101	ASSAULT 3	COMPLETED	MISDEMEANOR	BROOKLYN
08/04/2018	344	ASSAULT 3 & RELATED OFFENS...	101	ASSAULT 3	COMPLETED	MISDEMEANOR	BRONX
09/26/2018	106	FELONY ASSAULT	109	ASSAULT 2,1,UNCLASSIFIED	COMPLETED	FELONY	MANHATTAN
02/12/2018	351	CRIMINAL MISCHIEF & RELATE...	254	MISCHIEF, CRIMINAL 4, OF ...	COMPLETED	MISDEMEANOR	BRONX
11/04/2018	106	FELONY ASSAULT	109	ASSAULT 2,1,UNCLASSIFIED	COMPLETED	FELONY	QUEENS
01/05/2018	109	GRAND LARCENY	416	LARCENY,GRAND FROM NI...	COMPLETED	FELONY	MANHATTAN

Data Source 2 Sample -NYPD Complaints

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
629632833	02/06/2018	23:15:00			52	02/07/2018
787203902	11/21/2018	00:15:00	11/21/2018	00:20:00	75	11/21/2018
280364018	06/09/2018	21:42:00	06/09/2018	21:43:00	10	06/10/2018
985800320	11/10/2018	19:40:00	11/10/2018	19:45:00	19	11/10/2018
777641183	03/12/2018	11:48:00			25	03/12/2018
803743247	04/07/2013	16:00:00			103	04/07/2013
683416529	09/12/2018	18:30:00	09/12/2018	18:35:00	46	09/12/2018
570490441	01/16/2018	14:30:00	01/16/2018	15:00:00	69	01/16/2018
377132404	08/04/2018	22:15:00			44	08/04/2018
504303130	09/26/2018	18:20:00	09/26/2018	18:24:00	28	09/26/2018
584276892	02/11/2018	17:30:00	02/12/2018	06:00:00	41	02/12/2018
336011712	11/04/2018	11:15:00			103	11/04/2018
793863550	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	351	CRIMINAL MISCHIEF & RELATE...	254	MISCHIEF, CRIMINAL 4, OF ...	COMPLETED	MISDEMEANOR	MANHATTAN
02/07/2018	341	PETIT LARCENY	333	LARCENY,PETIT FROM STO...	COMPLETED	MISDEMEANOR	BRONX
11/21/2018	341	PETIT LARCENY	321	LARCENY,PETIT FROM AUTO	COMPLETED	MISDEMEANOR	BROOKLYN
06/10/2018	361	OFF. AGNST PUB ORD SENSBLT...	639	AGGRAVATED HARASSMEN...	COMPLETED	MISDEMEANOR	MANHATTAN
11/10/2018	341	PETIT LARCENY	333	LARCENY,PETIT FROM STO...	COMPLETED	MISDEMEANOR	MANHATTAN
03/12/2018	113	FORGERY	729	FORGERY,ETC.,UNCLASSIFI...	COMPLETED	FELONY	MANHATTAN
04/07/2013	118	DANGEROUS WEAPONS	793	WEAPONS POSSESSION 3	COMPLETED	FELONY	QUEENS
09/12/2018	121	CRIMINAL MISCHIEF & RELATE...	269	MISCHIEF,CRIMINAL, UNCL ...	COMPLETED	FELONY	BRONX
01/16/2018	344	ASSAULT 3 & RELATED OFFENS...	101	ASSAULT 3	COMPLETED	MISDEMEANOR	BROOKLYN
08/04/2018	344	ASSAULT 3 & RELATED OFFENS...	101	ASSAULT 3	COMPLETED	MISDEMEANOR	BRONX
09/26/2018	106	FELONY ASSAULT	109	ASSAULT 2,1,UNCLASSIFIED	COMPLETED	FELONY	MANHATTAN
02/12/2018	351	CRIMINAL MISCHIEF & RELATE...	254	MISCHIEF, CRIMINAL 4, OF ...	COMPLETED	MISDEMEANOR	BRONX
11/04/2018	106	FELONY ASSAULT	109	ASSAULT 2,1,UNCLASSIFIED	COMPLETED	FELONY	QUEENS
01/05/2018	109	GRAND LARCENY	416	LARCENY,GRAND FROM NI...	COMPLETED	FELONY	MANHATTAN

Data Source 3 Sample -NYPD Arrests



Data Source 4 Sample -NYPD Hate Crime

NYPD_Hate_Crimes

Possible Data Loss Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

I14 ASSAULT 3 & RELATED OFFENSES

	A	B	C	D	E	F	G	H	I	J
1	Full Complaint ID	Complaint Year Number	Month Number	Record Create Date	Complaint Precinct Code	Patrol Borough Name	County	Law Code Category Description	Offense Description	PD Code Description
2	2.019E+14	2019	1	1/22/19	1	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	CRIMINAL MISCHIEF & RELATED OF	CRIMINAL MISCHIEF 4TH, GRAFFIT
3	2.019E+14	2019	1	1/24/19	1	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	CRIMINAL MISCHIEF & RELATED OF	CRIMINAL MISCHIEF 4TH, GRAFFIT
4	2.019E+14	2019	2	2/6/19	1	PATROL BORO MAN SOUTH	NEW YORK	FELONY	MISCELLANEOUS PENAL LAW	AGGRAVATED HARASSMENT 1
5	2.019E+14	2019	2	2/11/19	1	PATROL BORO MAN SOUTH	NEW YORK	FELONY	MISCELLANEOUS PENAL LAW	AGGRAVATED HARASSMENT 1
6	2.019E+14	2019	3	3/7/19	1	PATROL BORO MAN SOUTH	NEW YORK	FELONY	MISCELLANEOUS PENAL LAW	AGGRAVATED HARASSMENT 1
7	2.019E+14	2019	3	3/24/19	1	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	CRIMINAL MISCHIEF & RELATED OF	CRIMINAL MISCHIEF 4TH, GRAFFIT
8	2.019E+14	2019	5	5/23/19	1	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	CRIMINAL MISCHIEF & RELATED OF	CRIMINAL MISCHIEF 4TH, GRAFFIT
9	2.019E+14	2019	6	6/8/19	1	PATROL BORO MAN SOUTH	NEW YORK	FELONY	FELONY ASSAULT	ASSAULT 2,1,UNCLASSIFIED
10	2.019E+14	2019	6	6/30/19	1	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	OFF. AGNST PUB ORD SENSBLTY &	AGGRAVATED HARASSMENT 2
11	2.019E+14	2019	8	8/14/19	1	PATROL BORO MAN SOUTH	NEW YORK	FELONY	FELONY ASSAULT	ASSAULT 2,1,UNCLASSIFIED
12	2.019E+14	2019	11	11/14/19	1	PATROL BORO MAN SOUTH	NEW YORK	FELONY	MISCELLANEOUS PENAL LAW	AGGRAVATED HARASSMENT 1
13	2.019E+14	2019	12	12/4/19	1	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	OFF. AGNST PUB ORD SENSBLTY &	AGGRAVATED HARASSMENT 2
14	2.01901E+14	2019	1	1/3/19	5	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	ASSAULT 3 & RELATED OFFENSES	ASSAULT 3
15	2.01901E+14	2019	2	2/11/19	5	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	ASSAULT 3 & RELATED OFFENSES	ASSAULT 3
16	2.01901E+14	2019	3	3/25/19	5	PATROL BORO MAN SOUTH	NEW YORK	FELONY	FELONY ASSAULT	ASSAULT 2,1,UNCLASSIFIED
17	2.01901E+14	2019	4	4/18/19	5	PATROL BORO MAN SOUTH	NEW YORK	FELONY	CRIMINAL MISCHIEF & RELATED OF	CRIMINAL MIS 2 & 3
18	2.01901E+14	2019	7	7/18/19	5	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	ASSAULT 3 & RELATED OFFENSES	ASSAULT 3
19	2.01901E+14	2019	7	7/22/19	5	PATROL BORO MAN SOUTH	NEW YORK	FELONY	FELONY ASSAULT	ASSAULT 2,1,UNCLASSIFIED
20	2.01901E+14	2019	3	3/5/19	6	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	CRIMINAL MISCHIEF & RELATED OF	CRIMINAL MISCHIEF 4TH, GRAFFIT
21	2.01901E+14	2019	6	6/7/19	6	PATROL BORO MAN SOUTH	NEW YORK	FELONY	ROBBERY	ROBBERY,NECKCHAIN/JEWELRY
22	2.01901E+14	2019	6	6/13/19	6	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	OFF. AGNST PUB ORD SENSBLTY &	AGGRAVATED HARASSMENT 2
23	2.01901E+14	2019	8	8/9/19	6	PATROL BORO MAN SOUTH	NEW YORK	FELONY	FELONY ASSAULT	ASSAULT 2,1,UNCLASSIFIED
24	2.01901E+14	2019	8	8/14/19	6	PATROL BORO MAN SOUTH	NEW YORK	FELONY	FELONY ASSAULT	ASSAULT 2,1,UNCLASSIFIED
25	2.01901E+14	2019	8	8/23/19	6	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	ASSAULT 3 & RELATED OFFENSES	ASSAULT 3
26	2.01901E+14	2019	2	2/15/19	7	PATROL BORO MAN SOUTH	NEW YORK	FELONY	MISCELLANEOUS PENAL LAW	AGGRAVATED HARASSMENT 1
27	2.01901E+14	2019	3	3/15/19	7	PATROL BORO MAN SOUTH	NEW YORK	FELONY	MISCELLANEOUS PENAL LAW	AGGRAVATED HARASSMENT 1
28	2.01901E+14	2019	5	5/18/19	7	PATROL BORO MAN SOUTH	NEW YORK	FELONY	CRIMINAL MISCHIEF & RELATED OF	MISCHIEF,CRIMINAL, UNCL 2ND
29	2.01901E+14	2019	8	8/12/19	7	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	OFF. AGNST PUB ORD SENSBLTY &	AGGRAVATED HARASSMENT 2
30	2.01901E+14	2019	8	8/25/19	7	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	CRIMINAL MISCHIEF & RELATED OF	CRIMINAL MISCHIEF 4TH, GRAFFIT
31	2.01901E+14	2019	10	10/22/19	7	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	CRIMINAL MISCHIEF & RELATED OF	CRIMINAL MISCHIEF 4TH, GRAFFIT
32	2.01901E+14	2019	12	12/7/19	7	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	CRIMINAL MISCHIEF & RELATED OF	CRIMINAL MISCHIEF 4TH, GRAFFIT
33	2.01901E+14	2019	3	3/17/19	9	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	CRIMINAL MISCHIEF & RELATED OF	CRIMINAL MISCHIEF 4TH, GRAFFIT
34	2.01901E+14	2019	3	3/31/19	9	PATROL BORO MAN SOUTH	NEW YORK	MISDEMEANOR	ASSAULT 3 & RELATED OFFENSES	ASSAULT 3

NYPD_Hate_Crimes

Data Source 5 Sample- NYC 2010-2020

	A	B	C	D	E	F
1	Area Type	Area Nam	2010 Cens	2020 Cens	Populatio	Population Percent Change
2	City	Albany cit	97856	99224	1368	0.014
3	City	Amsterda	18620	18219	-401	-0.0215
4	City	Auburn ci	27687	26866	-821	-0.0297
5	City	Batavia cit	15465	15600	135	0.0087
6	City	Beacon cit	15541	13769	-1772	-0.114
7	City	Binghamt	47376	47969	593	0.0125
8	City	Buffalo cit	261310	278349	17039	0.0652
9	City	Canandaig	10545	10576	31	0.0029
10	City	Cohoes ci	16168	18147	1979	0.1224
11	City	Corning ci	11183	10551	-632	-0.0565
12	City	Cortland c	19204	17556	-1648	-0.0858
13	City	Dunkirk ci	12563	12743	180	0.0143
14	City	Elmira city	29200	26523	-2677	-0.0917
15	City	Fulton city	11896	11389	-507	-0.0426
16	City	Geneva ci	13261	12812	-449	-0.0339
17	City	Glen Cove	26964	28365	1401	0.052
18	City	Glens Fall	14700	14830	130	0.0088
19	City	Gloversvil	15665	15131	-534	-0.0341
20	City	Hornell ci	8563	8263	-300	-0.035
21	City	Hudson ci	6713	5894	-819	-0.122
22	City	Ithaca city	30014	32108	2094	0.0698
23	City	Jamestow	31146	28712	-2434	-0.0781
24	City	Johnstow	8743	8204	-539	-0.0616
25	City	Kingston c	23893	24069	176	0.0074
26	City	Lackawan	18141	19949	1808	0.0997
27	City	Little Falls	4946	4605	-341	-0.0689
28	City	Lockport c	21165	20876	-289	-0.0137
29	City	Long Beac	33275	35029	1754	0.0527
30	City	Mechanic	5196	5163	-33	-0.0064
31	City	Middletov	28086	30345	2259	0.0804
32	City	Mount Ve	67292	73893	6601	0.0981
33	City	New Roch	77062	79726	2664	0.0346
34	City	New York	8175133	8804190	629057	0.0769
35	City	Newburgh	28866	28856	-10	-0.0003
36	City	Niagara Fa	50193	48671	-1522	-0.0303
37	City	North Ton	31568	30496	-1072	-0.034
38	City	Norwich c	7190	7051	-139	-0.0193

Census 2010 and 2020 Population

Data Source 6 Sample- NYC Population by Borough

New_York_City_Population_by_Borough__195

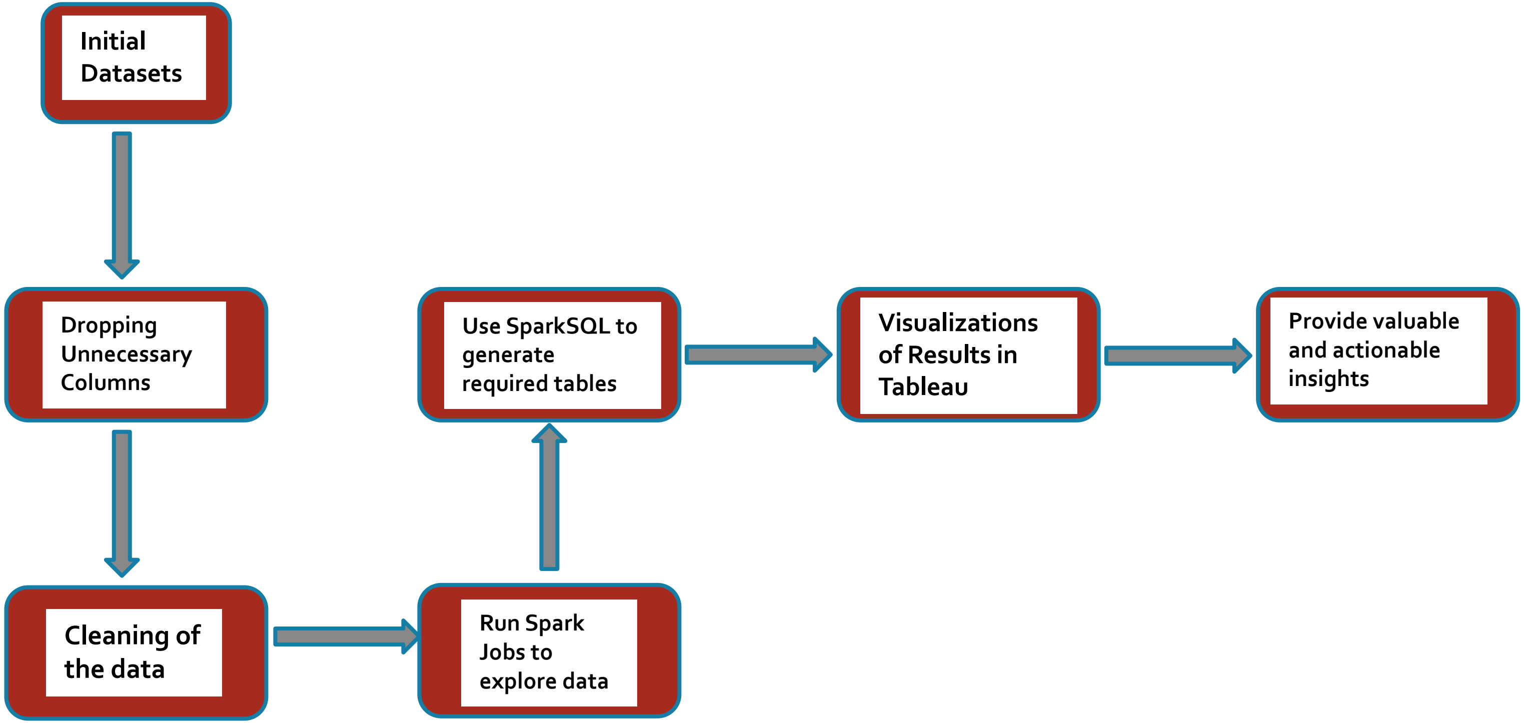
AgeGroup	Borough	1950	1950 - Boro share of NYC total	1960	1960 - Boro share of NYC total	1970	1970 - Boro share of NYC total	1980	1980 - Boro share of NYC total	1990	1990 - Boro share of NYC total	
Total Population	NYC Total	7891957	100	7781984	100	7894862	100	7071639	100	7322564	100	80
Total Population	Bronx	1451277	18.39	1424815	18.31	1471701	18.64	1168972	16.53	1203789	16.44	13
Total Population	Brooklyn	2738175	34.7	2627319	33.76	2602012	32.96	2230936	31.55	2300664	31.42	24
Total Population	Manhattan	1960101	24.84	1698281	21.82	1539233	19.5	1428285	20.2	1487536	20.31	15
Total Population	Queens	1550849	19.65	1809578	23.25	1986473	25.16	1891325	26.75	1951598	26.65	22
Total Population	Staten Island	191555	2.43	221991	2.85	295443	3.74	352121	4.98	378977	5.18	4

New_York_City_Population_by_Borough__1950_-_2040

C total	1990	1990 - Boro share of NYC total	2000	2000 - Boro share of NYC total	2010	2010 - Boro share of NYC total	2020	2020 - Boro share of NYC total	2030	2030 - Boro share of NYC total	20
100	7322564	100	8008278	100	8242624	100	8550971	100	8821027	100	90251
16.53	1203789	16.44	1332650	16.64	1385108	16.8	1446788	16.92	1518998	17.22	15792
31.55	2300664	31.42	2465326	30.78	2552911	30.97	2648452	30.97	2754009	31.22	28405
20.2	1487536	20.31	1537195	19.2	1585873	19.24	1638281	19.16	1676720	19.01	16916
26.75	1951598	26.65	2229379	27.84	2250002	27.3	2330295	27.25	2373551	26.91	24126
4.98	378977	5.18	443728	5.54	468730	5.69	487155	5.7	497749	5.64	5011

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

Design Diagram



Code Challenges

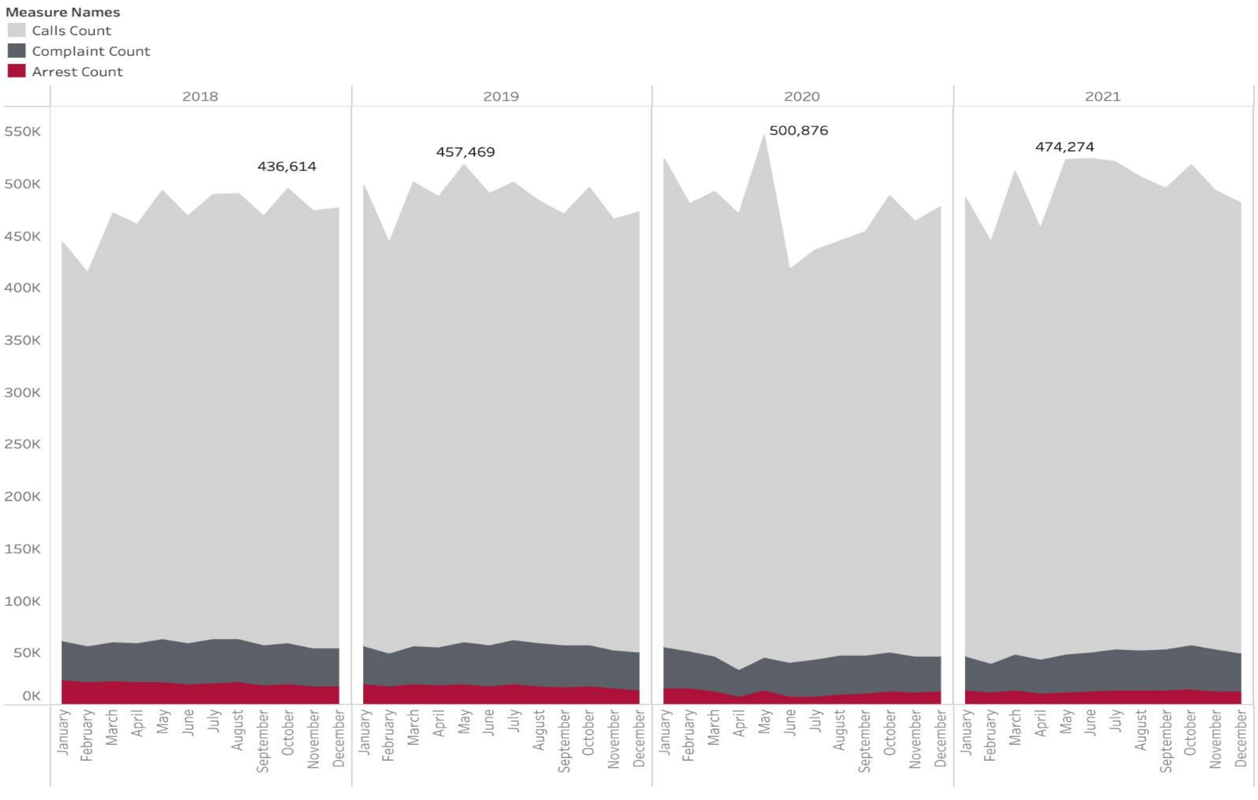
01	**Null Types**	• NA values also include - “(null)”, “” . String Matching query helps removes it
02	**System Resources**	• 26 million rows often clogged the kernels • Overloaded clusters
03	**Cohesive yet covariance**	• Common columns such as offense types, counties but highly varied data values

Code Challenge - Aggregated data

01	Merging Criteria	• Offense Description have various levels to it differing in each data set • Based on Year, able to aggregate data for 2018, 2019, 2020, 2021 • 77 NYPD Precincts common for all • 26 million rows with multiple columns.
02	Using Spark SQL Queries	• One View for all data insights led to kernel overboard issue • Created multiple views focusing one key sight • This helped avoid "long running joins", no additional memory for cluster manager and system resources
02	Choice of File Format Storage	• Chosen platform for visualization: Tableau, works best with csv (used for focused insights) • Apache Parquet generates lower storage costs for data file, **very efficient compression and encoding schemes,** useful for 6 GB Data

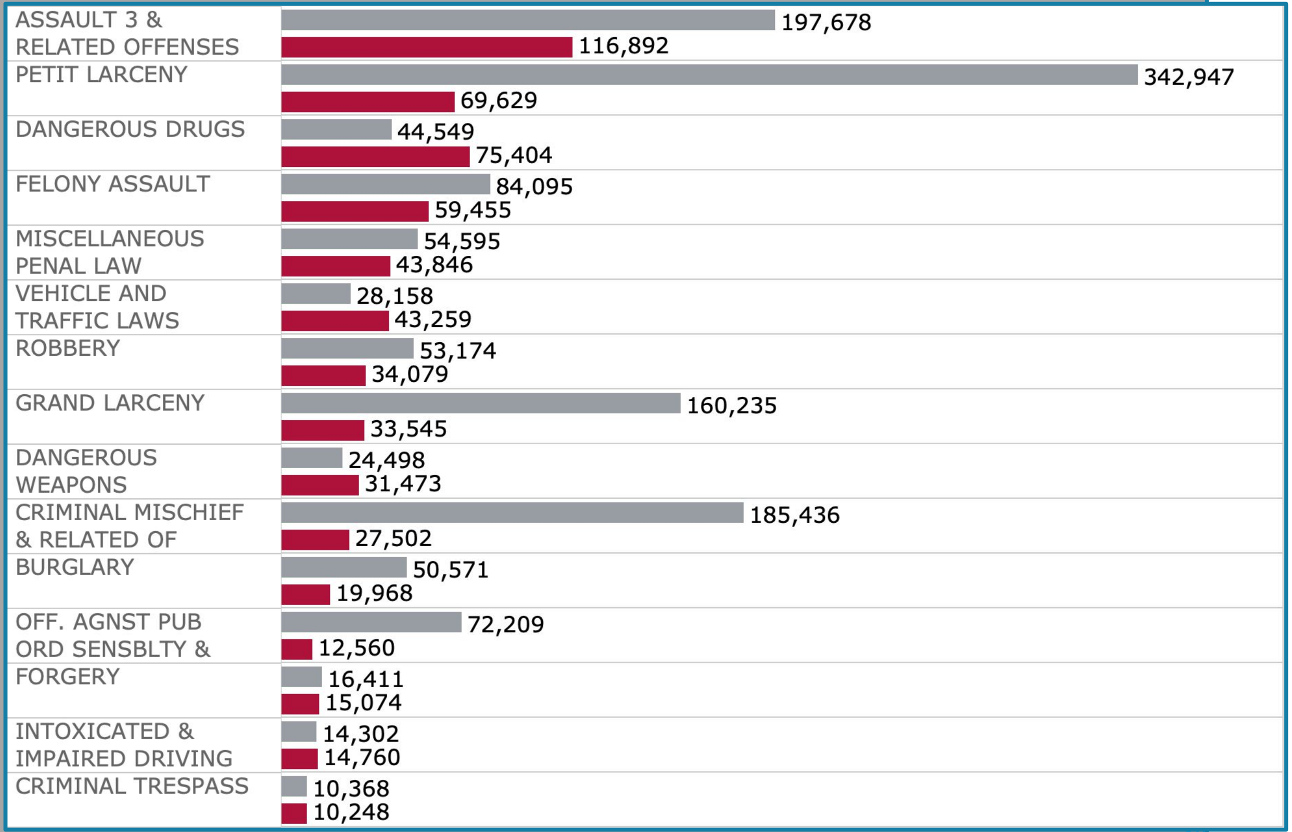
Result 1

2020 was the year where **most number of 911 calls** took place .It reached a numerical peak of **500,876(number of calls)** in May just after the advent of the Covid Pandemic. And another interesting thing we noticed is that every year(2018-2021) in the month of May most number of 911 calls are reported.



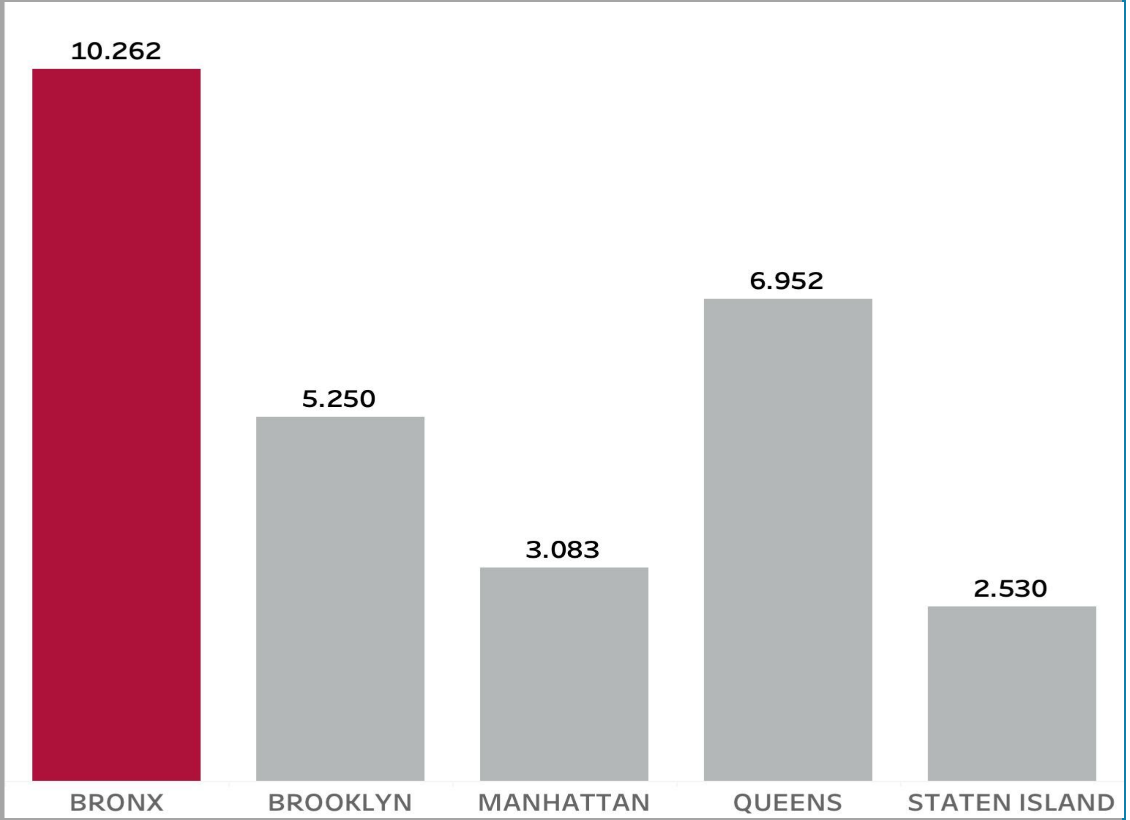
Result 2

Top offense people in NYC complaint and get arrested for are **assault 3, dangerous drugs, petit larceny and Felony assault**



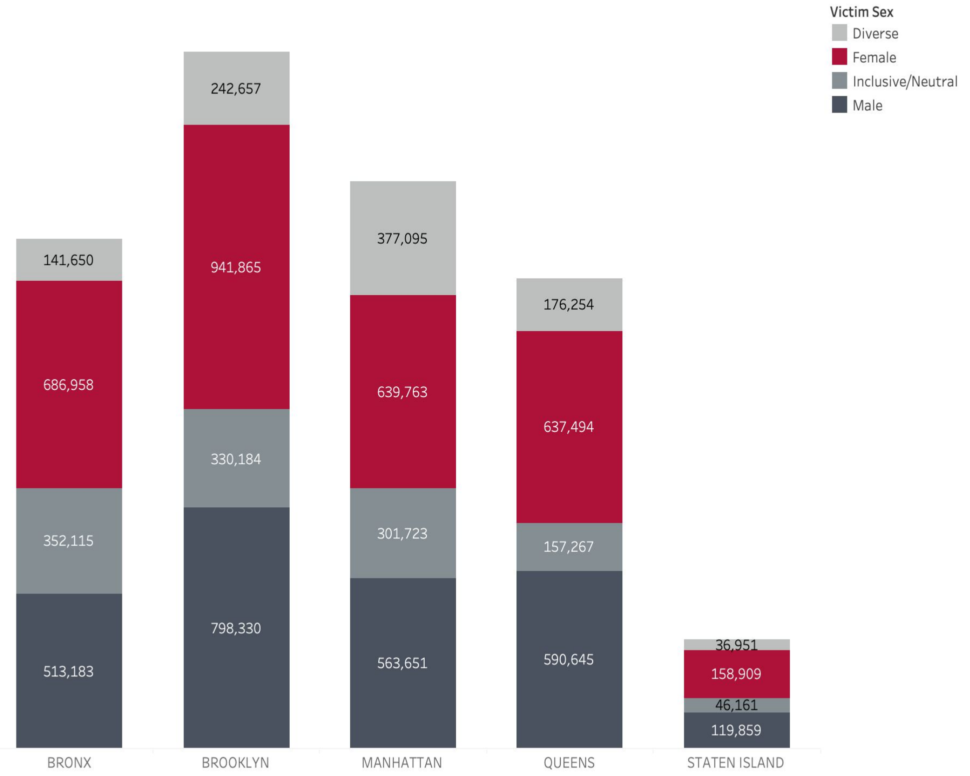
Result 3

Out of the 5 borough in NYC-the Bronx, Brooklyn, Manhattan, Queens and Staten Island.**Bronx** has **slowest dispatch time** once a 911 call is made.



Result 4

In all the 5 boroughs,the victims were primarily **“Female”** across the years -2018,2019,2020.



Result 5

Further diving into the victim dataset and location dataset we could conclude that most black people were targeted in **Brooklyn**-an alarming **785,214!**

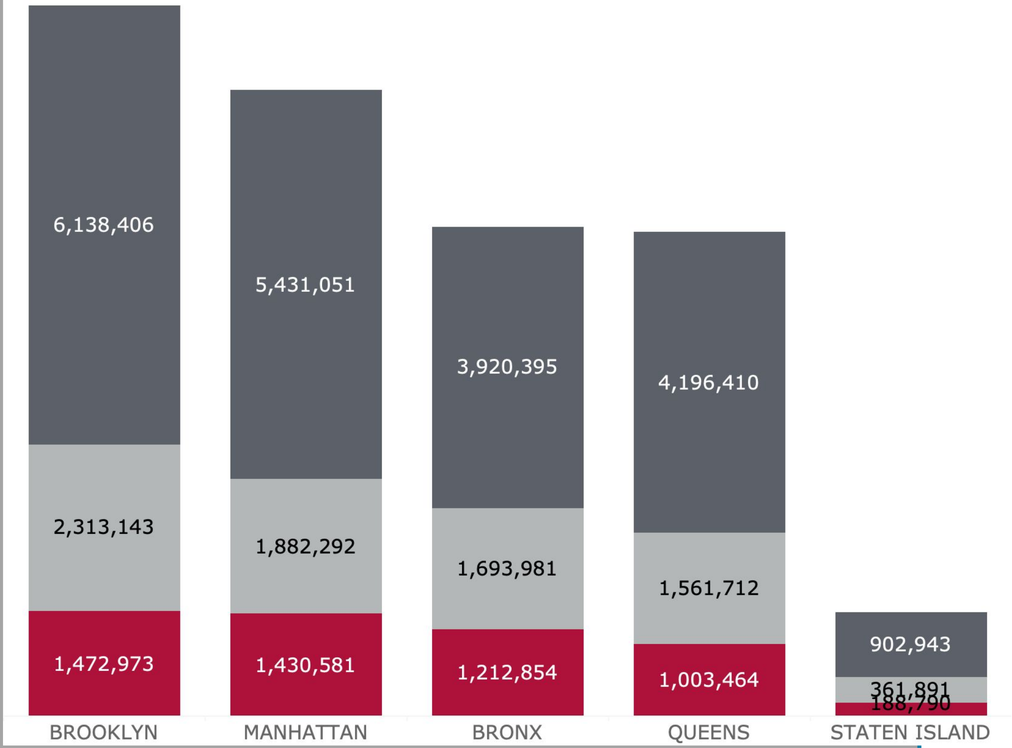
	BLACK	WHITE	WHITE HISPANIC	ASIAN / PACIFIC ISLANDER	BLACK HISPANIC	AMERICAN INDIAN/ALASKAN ..
BROOKLYN	785,214	423,858	265,821	105,334	45,833	8,929
MANHATTAN	306,093	394,803	242,160	97,524	64,679	6,883
QUEENS	319,534	275,906	297,396	222,052	26,319	12,020
BRONX	418,268	96,033	425,632	29,828	128,077	5,466
STATEN ISLAND	54,066	143,161	39,923	8,895	4,541	1,598

Count of Complaints

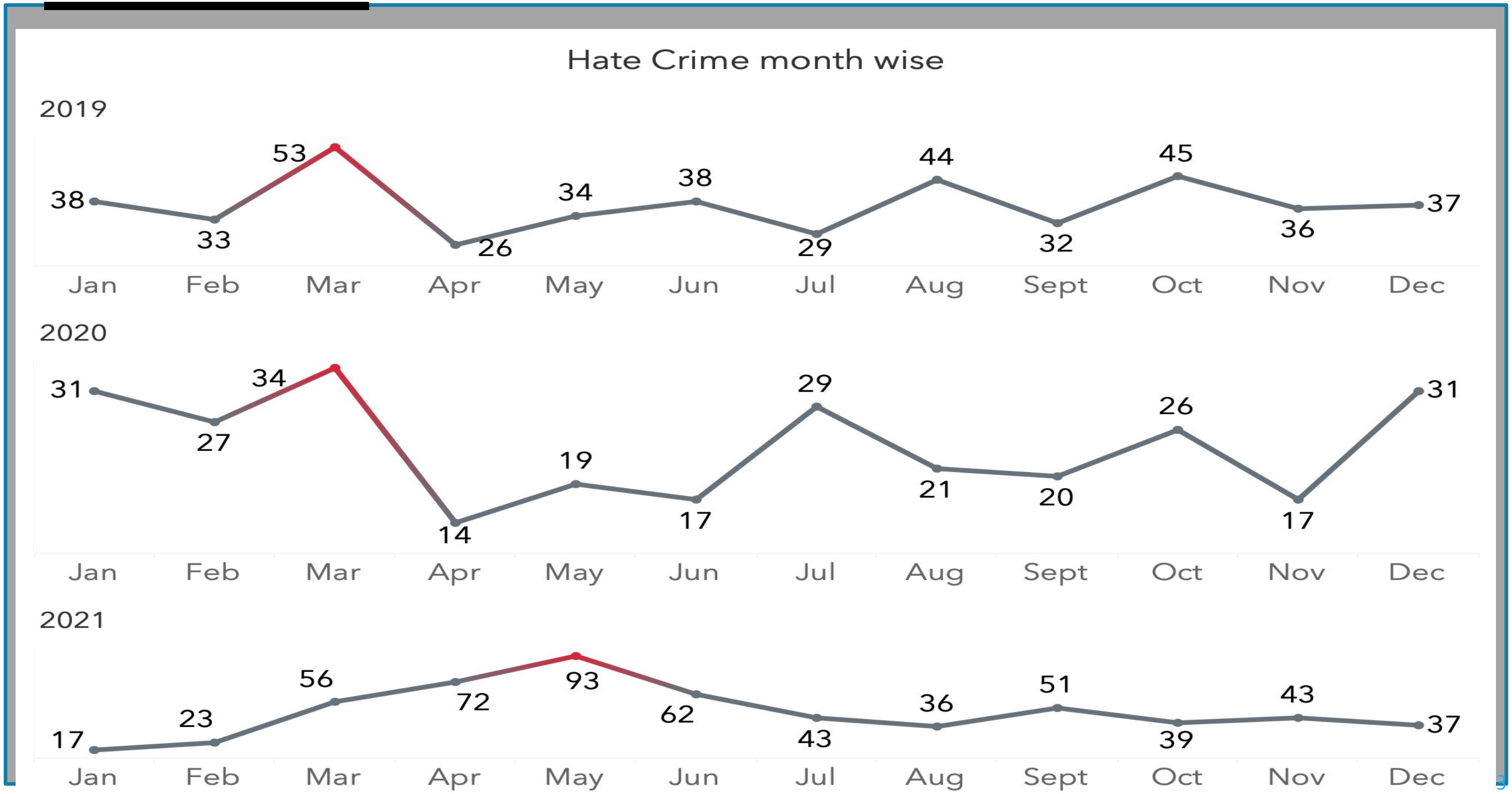
1,598 785,214

Result 6

When we checked at the volumes of calls, complaints and arrests that happened in NYC 5 boroughs, we can conclude that **Brooklyn** has the **highest crime rate** that gets reported while **Staten Island** has the **lowest crime rate** that is reported. Also another interesting thing to notice is that **ratio of complaints to arrests is higher in Manhattan than in Brooklyn.**



Result 7


Hate Crime month wise
2019
38 33 53 26 34 38 29 44 32 45 36 37
Jan Feb Mar Apr May Jun Jul Aug Sept Oct Nov Dec
2020
31 27 34 14 19 17 29 21 20 26 17 31
Jan Feb Mar Apr May Jun Jul Aug Sept Oct Nov Dec
2021
17 23 56 72 93 62 43 36 51 39 43 37
Jan Feb Mar Apr May Jun Jul Aug Sept Oct Nov Dec

ML Modeling

We intend to map 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**, predict the **number of arrests.**

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

Model	R2 score
Linear Regression	0.44
Random Forest Regressor	0.32
Gradient Boosted Tree Regression	0.36

Linear Regression model performs better in comparisons . These are average results from our preliminary analysis. This could be improved by better feature extraction (more cohesive dataset features).

Obstacles

- Tableau integration with HPC data proc
- Choosing an appropriate Machine Learning algorithm to implement with the given datasets
- Down time of HPC data proc
- Data Stitching (file creation (csv) done for key insights by joining datasets)

Learnings

Integration of Datasets and analysis of the key insights using Spark,SparkSQL.

All the **transformations** are lazily evaluated.This gives Spark the ability to make optimization decisions, as all the transformations become visible to the Spark engine before performing any action.

Tableau helped us in analysis of large amounts of complex data through a visual interface.

Summary

- Worked on the the integration of data set for further analysis focusing on the 5 boroughs in NYC.
- Pre-processing and data cleaning was done as required .
- Exploratory data analysis of the data sets by running Spark Jobs
- Merging of datasets using a common attribute using Spark SQL and implementation of Linear Regression
- Analyzed visualizations on Tableau to come up with trends and statistics.

Acknowledgements

- We’d like to thank **Prof. Yang Tang** for giving us the opportunity to work on this project.
- We’d also like to thank the **NYU HPC team** for giving us the necessary dev environment to perform our experiments.
- We’re also grateful for the DataViz stern course by Professor **Kristen Soulski**, where we got the tableau licenses.

References

- https://www.tableau.com/resources/reference-materials
- https://spark.apache.org/docs/latest/ml-classification-regression.html
- https://spark.apache.org/docs/latest/sql-ref.html
- https://data.cityofnewyork.us/browse?tags=911

NYU

Thank You!