



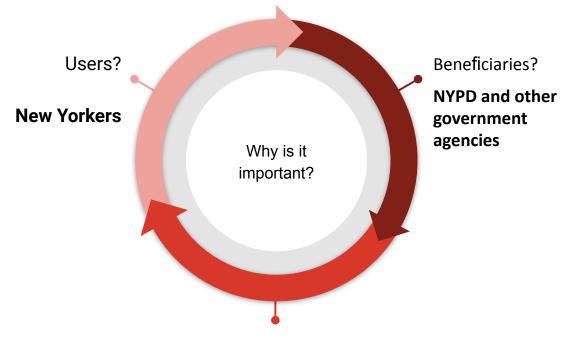
### **Abstract**

- 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**



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

## **Metric Goodness**

- <u>Sense of Safety</u>: Crime rates in a particular area play an important role in the peace and safety index of the entire city
- <u>Trust in Police</u>: 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

6

#### How Safe NYC Really is?

#### **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	1	KY_CD	i	OFNS_DESC :	PD_CD i	PD_DESC :	CRM_ATPT_CPTD_CD	1	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
3/12/2018			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
1/04/2018			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

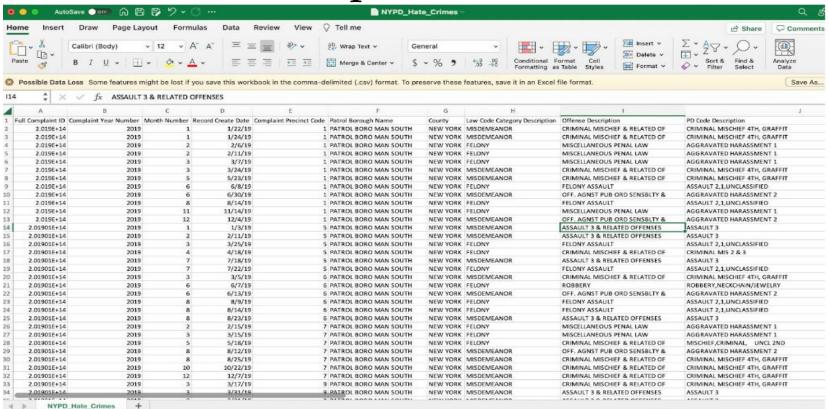
CMPLNT_NUM :	CMPLNT_FR_DT :	CMPLNT_FR_TM :	CMPLNT_TO_DT :	CMPLNT_TO_TM :	ADDR_PCT_CD :	RPT_DT
506547392	03/29/2018	20:30:00			32	03/30/2018
629632833	02/06/2018	23:15:00			52	02/07/2018
787203902	11/21/2018	00:15:00	11/21/2018	00:20:00	75	11/21/2018
280364018	06/09/2018	21:42:00	06/09/2018	21:43:00	10	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	1	OFNS_DESC :	PD_CD i	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

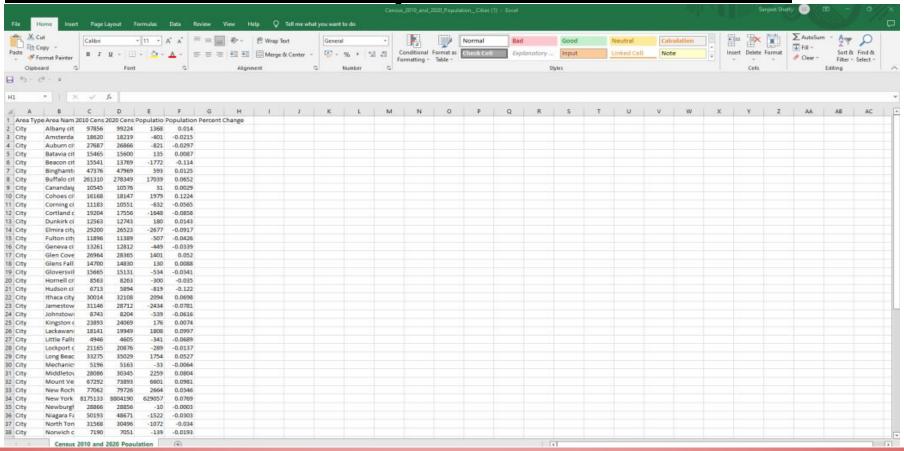
## Data Source 3 Sample -NYPD Arrests

							NYPE, Arrests, Date_7	rance rough						
MINT ARE OF	WEST DATE	PO CD PO DISC	W (0)	DPM DIDC	THE CODE THE D	OR OF APPEAL BOND	AWEST PRICHCT APPEND	TON COOK HOE OF	LP POP I	BE PERF RICE	E,00000,00 Y	COORD CO Lumber	Langitude	Longited
mental territoria	60081	301			PLOSESSEY M	le .		1 45.64	M	BAOL	987407	DESTR. ACTROMOTYPES NO.	-75.852×085×-61888	POHT LTLOGARDANISM
DESIGNATION TO	resoner	100 80/9.0	756	NAME OF THE OWNER	PLINISHED F	# C		1 76-41	M	MHTH HOPPING	101,000	DECK CHISM SCHOOL	-1500000ect extent	PURPLY STANSONATOR
DECREASE T	459/250FF	SET CHEC CHEARGERING ME,FIRE	108	SEX CRIMES	PLIBRISH M	0	111	1.254	W	BACK	VOMENT?	restate acurteroseoscocio	-15/790×768686	PORT LTS TROVIDENSE
MOUNT IN	TROTON	31 °C CONTROLIDO BURGINOS. PORRESENOS	(cm)	DANSSELD DELKIN	PLOSTON N	16	. 87	7 16 60	M	86,401				
	motor	117 SOKUK, KRASE	114	BOLOPAKS -	(FL)1000000 F	100	- 19	9 10-01	M:	BAOK	1000000	20064 40,000040100000		
MARKET	1000000				PL00000 F	160	44	6 25.40	w	siwoschere -	(600.00)	J1575 40/116960960100	-15.6H (11) 11 60000	PRINT 1-10/09/21211000
atterne -	10000	soc .			PL 0000001 F	4	710	1 8141	M	MAIN	1078004	Stein # 457 Commence of	700700-00-000	PUBLIC TO STREET
-	19990	500 ARSON E.S.A		4700H	PLOSEGET F	0	41	7 25-41	M	MALE HOUSE	90/7694	CREET 40004109400000		10841 7 JT 04800 40598
DESKIN T	10000	THE PORQUESTS LINGUISHMENT HALD	103	nanam	PL 1700000 F	- 4	113	8 9996	W	BACK	Years	THE REPORT PROFILE	-0.70796679688	PORC   TE THE MUSIC
140151402 (1)	49096	ISS RAPE I	104	SAPE .	PL100000 F	6	- 0	1 (5.4)	W	BAOL	600000	TORRAL ACAMEDOROROMIC	-15 as another more	PORT 1-73,9590/856899
DITEMBRE TO	PERSONAL PROPERTY.	101 ABASETS	204	MONUTES A PRIABILITY OFFICERS	PL DESSET M	E		1 26-61	M	BACK	1007400	THE SE ACADEMIC PROCESS	- FERT BURGLASSER	PRINT   TERRESONA
COURSE OF THE PARTY OF THE PART	Y110000				PL 2001000 F	88	. 25	0.26-61	M.	MHTS HOPMIC	-	20071 4076767010000		PRINT   TO HELD STORY
BURNOUS -	HERDEN	OE STREEDLINTON IST	198	PEUDHY ASSAULT	PL 1011080 F	to .		8 35-44	M	BAOK	688751	04198 40406TM580010	-75 (A DISSO') MINISTO	PORT   TLORESET + 60
TROOMINE 1	materia	MF RAPE I	194	tors	PL TREBUT   P	F	. 27	1 20-41	W	BLACK	100300	18080 41418ADDMOTT	150000000000	PORT / SHREETHIRK
24 DES 100	P190586	200 TRESPASS & SPARRAN.	900	CRIMINA, TRESPASS	PL HERRY M		. 11	1 68	w	BLACK -	1604960.0	100000.0 4017/054679000	75 887 089	PORT J. TURBET MAIN
#1001SE 10	rhelmin	200" FROSCHAFER II, CHAMBIA.	_	ANTERPORT INTERNET	PL TORRES M	N.		9 (8) 60		MACH.	1078706	TRACK ALATOTERATORY	TOTAL PROPERTY.	PURT L'ELECTION
100402040 1	100900	ISP RATE:	104	NATE	(45,100000) F	0	10	1 55.40	M	BAOL	1001400	SIGNED ACTORNOMYNONS	-F300-94790H00000	PSRF) 10864 0000
DEFENDED TO	P18/2981	10" ROBERTOPELARCH ONLARGHED		NORSEN	PL 1990900 F	100	4	6 (0) 44	M	seet!	986391	32178 40.7004084080888	-Frankling Street	PUBP 1-14-088602000
-	elektrin.	ISS NATE O	194	NAME:	PL SERVED P	10		1 (0.4)	M	MINERAL PROPERTY.	784458	MARKATAN STATEMENT AND ADDRESS OF	7500000000000	PERC LOS MANAGEMENTS
100000170	remelled	SEY CHLO CHOANGERYG WEJTHIC	- 100	SEX CRIMES	PLEEDING M	0	49	1 (6.4)	M	BAOL	100016	DRYN 418250-D40000	-13.60000+0000HB	PURC ) TLOROGENS
DATEMENT OF	00000	TT BITCH MADE	100	SEE CHARLE	PL 1000001 F		- 4	2 20-61	W	BLACK	1071001	JAMES STREET,	-13.603 (871)	PORT   TURNSTANCE
B790040 12	199981	NY ROBBITATIFEN ARCHIMILATERS	400	ROBBERY	PL 1901000 F	16		1 (5.4)	M.	8,401	982746	08647 4070868808080	-74:005(de) (IIIIIII)	PORT 1-T4:0054280189
MINET IV	03000	SAT RAPE 1	794	NATE .	PE-CORRECT P	6.	. 17	2 846	M	BACK HIPMIC	19048	19600 ASATOROGRADOTOS	-13-0100794700000	PRINT   TURBOTTHER
MITTERS TO	(0.000)	TTS BROWN WALKER S.T.	210	BES CRIMES	PL TORRING M		101	E 26-61	M.	BACK	1071087	SYTTE CONSIDERATION	-11 8000000000000	PRINT   13 600300000
BM176786 **	HIDDEH	SEL CHILD ENGAGERNG METWO	100	SER CRIMES	75,0001001 M		29	8 35-44	W	BLACK	100000	(2009) 40 000004024 00000	-T104-1040980000	PORT L'TUSH + INCHES
SETSMANN TH	remoternia .	TYPE MINISTER, ARRESTS NO.	100	200 CRANIS	PLOSSAT M		78	1 (0.4)	0.1	MACK	(MEDIT)	DESIGNATION OF THE PERSON NAMED IN COLUMN 1	- COMMENSATION	PURPLY STREET
M1909 11	PERSONAL	ION ASSAULTS	200	HOSPILLY STREET, AND CHARGES	75.1000000 M		18	1 75.41	M	SHALL HOUSE CO.	(067591)	2777E 4070454000000	-75 becomes	PORT   13.080 (652)
HT00.001 7-	remen	1011 AMERICA S	ime	ASSESSED SAMELANDS OFFICERS	PL Indicated I M	m	44	9 30-41	M	BACK .	1071004	DETER ALANDESSANDERS	TEMPLEMENT	PRINT   TEACHMAN
scourse -	resesso	ISS RAPE I	104	NATC .	315,1060WH   F	0	10	0.00	M:	BAOK	1001400	350HB 41753668170060	-Follow/Introduction	PRINT ) TURN 4 TRANS
B0884 10	1890400	681 CHLLI DIGASCENIO ASLINEC	108	SCHORAGE.	PL00101 W	E .		8 45-84	M	B6401	966016	160000 40,00040004100010	-153K00/0079W000	PRINT L'ILVESTIONE
-	respect	THE BURGE BRAIN GU	. 100	BED CHARTE	PL TORONAL M	No.	+	3 30 41	M	META HERMAN	1966 1004	STORY ALTOHOLOGICAL	100/00/2009	PORT - NUMBER
MICHAEL TO	review.	470			PL HERSET M	10	- 28	1 35-0	M	SUACH	599340	20075 40,007111090010	-13:3+0+15:0000000	PORT 1-73.944-750800
MINE T	10010	100			PL2004000 F	9	113	8 20-61	W	SHITCH HOPING	100600	18772 ASMISSIONED	-12/1000 HARRISON	PERC   TETROPISE
B796001 <	1000001				PL000400 F	10	140	1 25-44	W	BLACK HISPINEC	9818	21882 40.7528088888888	-11:00tentineses	PDRF 1-71.9659810019
MINNEY TO	1090001	un.			PL000901 F	6.	10	1 26-61	M	AURI/PICFC BLADER	187429	177172 60.63888110100010	- FLODRICHONNISS	PRINT   TURNISHING
SSORTTEN TO	01000	107 88901	108	AMPS:	PS.1000001 F	8.	41	8 29-41	M.	BACK	10300	DETER CONTRACTORS	-15 Rether (888)	PRINT   TURBURATOR
2969-0676 ***	91991	150 TAPE 3	194	NAME.	Pt. 00000 F	0	114	0.25-44	M	10475	1007004	21494 4176806870010	-75.04 EXC#1000000	PORT 1-TL0+9508+760
IMBORT T	osomi	ter News	104	ANTE:	PL DESIGN F		41	1 25-61	M	MATERIAL PROPERTY.	10730100	DESCRIPTION OF THE PERSONS NAMED IN	-11890961 6089	PURPLY THROUGH THE
100000000	restore	IOS STRINGULATION IST	100	PELISHY ARBAGET	75,1011000 F	0	718	10.46	M	BACK	1098476	190406 ACETHOR TORONO	75.800kW 10705000	PORT   13 80000H 960
-	miner:	DE STREEDGLIGTER OF	100	PRINCIPLY AMBAGAT	PLATONE F		ter	9 1990	M	10-610	manuse.	TAKEN ALAKSKANINGSTAN	1010000000000	PORT I'VE TREAMINE
MISSION IN	PERSONAL PROPERTY.	SSC PROSTUTION	100	PROSTITUTION & RELATED INTENSICS	PLUMONO: N	0	100	0 19-01		WEST	694707	19768 40.70Y2081886888	13.50(1915)00000	PRINTED TO THE PATTERN
-	rmisine	THE BUILDS MAJOR !	79	BED CROSSES	PL SERVICE   F		66	0.70		NACK.	Transition .	TABLE ASSESSMENT OF THE PARTY NAMED IN	1949/04/10/08	PRINT PRINTED IN
-	recere	THE BUDGHY I	78	BEE CRAKES	PLORES F	6		1 8141	44	MAIN	38230	19299 40471000100000	TRANSCOME VALUE	PORT / TARRASON
MARKET III	1010010	790. WSAPONS PORSESSION 0		DANSONO REALTHS	PL00000 F	E	- 04	1 (5.4)	M	WEST	37969	19415 418197971000	1400000400000	P08473-T4:0003084889
#11000 III	PHODDR	too arrenguscron ter	188	PRESENT ASSAULT	PLOTORO F	10	n	8 20-60	W	BLACK	(MARK)	2000 41/494760000	-0.87904100000	PERCHIDING
1909/05 10	remone.	175 SONIAL HOUSE SU	.00	SCHORMS-	(FL1000A1 M	6	67	8 <8	W	BAOL	100000	1750M 403400000000000	-15/85/800000000000	PDMT  -11,9580000000
227900480	recen	10° NAS 1	754	N/S	PLOSSET F	8	n	8 19	w	BACK	498.10	19700 - 41 AVAILABRE SEED	-DHODESON	PERCHIPMENT NUMBER
GOOTA LEE	esudina.	476			PL HERRY W		14		05 M	MATE .	1000001	2/202 4070108080000	TO SOCIAL PROPERTY AND INC.	PORT L'ELOGIA GENAL

### Data Source 4 Sample -NYPD Hate Crime



### Data Source 5 Sample- NYC 2010-2020

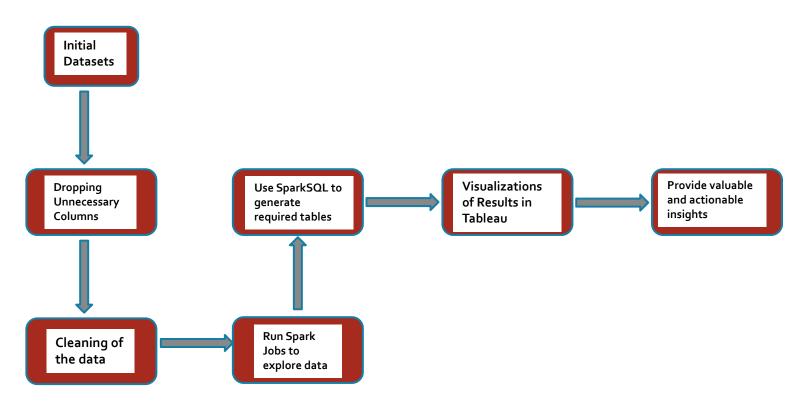


### Data Source 6 Sample- NYC Population by Borough

	Α.	- 8	- C		0		E		j.	- 0			1	- 3		Κ.	4	
															New	York_Cit	y_Population_by_Boroup	p_1
AgeiGr	eup	Borough	1950	1950 - Bor	ro share of t	NYC total	1960 1	1900 - Bora st	have of NYC to	tel 1970	1970 - Baro shere	of NYC total	1980	1960 - Boro share of N	PYC total	1990	1990 - Bors share of NYC	total
Total I	Population	NYG Torse	7891957			100	7781984		1	00 7884862		100	7971639		100	7322564		100
Yotal F	Population	Bronx	1451277			16.39	1424915		18.	21 1471701		18,04	1168972		16.53	1203769	1	5.46
Total I	Population	Brooklyn	2738175			34.7	9167585		33	76 2602012		32,96	6 2230936		21.66	2303664		1,42
Total I	Population	Manhattan	1960101			24.84	1698261		21	62 1539233		19.5	1428285		20.2	1487536	2	9.31
Total i	Population	Queens	1530949			19.65	1809518		23	25 1989473		35.16	1891325		26.75	1901098		5.05
****	Population	Staten teland	191555			2.45	221991			85 295443		9:24	352121		4.00	378977		5.18
							221991			0 20-0		***			4.30	270017		2.10
vew_		y_Population	_by_Box		10002000	140		NVC total				750		State 2VV to seade a		J-15000		
New,	York_Cit		_by_Box	rC total	2000 2	140	o share of		2010 20		are of NYC total	2020		share of NYC total	2030	2030 -	Boro share of NYC total	
New_total	York_Cit; 1960 7322564	y_Population	_by_Box	100 I	2000 S 8008278	140		100	2010 20 8242924		are of NYC total	2020 S		100	2030 8821027	2000 -	Boro share of NYC total	902
New_ total 100 16.53	York_Cit 1960 7322564 1203789	y_Population	_by_Box	100 / 16.44 1	2000 1 8008278 1332680	140		100 16.64	2010 20 8242924 1385108		are of NYC total 100 16.8	2029 1 8550971 1445788		100 16.92	2030 8621027 1518996	2000 -	Boro share of NYC total 100 17.22	902
New_ total 100 16.53	York_City 1990 7322564 1203789 2300664	y_Population	_by_Box	100 / 16.44 1 31.42 2	2000 1 8008278 1332650 2465326	140		100 16.64 30.78	2010 29 8242924 1385108 2552911		are of NYC tetal 100 16.8 50.97	2026 1 8550971 1445788 2649452		100 16.92 30.97	2030 8821027 1518996 2754008	2000 -	Boro share of NYC total 100 17 22 31 22	902 157 284
New_ total 100 16.53 91.55	York_City 1990 7322564 1203789 2300664 1487526	y_Population	_by_Box	100 / 16.44 1 31.42 2 20.31 1	2000 4 8008278 1332650 2466326 1537185	140		100 16.64 30.78 19.2	2019 29 8242924 1385108 2552911 1560873		are of NYC state 100 16.8 30.97 19.24	2029 1 8500971 1445788 2648452 1635281		100 16.92 30.97 19.16	2030 8821027 1518986 2754008 1676720	2600 -	Boro share of NYC total 100 17.22 31.22 13.01	902 157 284 169
New_ 100 16.53 H.65	York_City 1990 7322564 1203789 2300664	y_Population	_by_Box	100 8 16.44 1 31.42 2 20.31 1 26.66 2	2000 1 8008278 1332650 2465326	140		100 16.64 30.78 19.2	2010 29 8242924 1385108 2552911		are of NYC state 100 16.8 30.97 19.24	2026 1 8550971 1445788 2649452		100 16.92 30.97 19.16	2030 8821027 1518986 2754000 1676770 2373551	2000 -	Boro share of NYC total 100 17 22 31 22	902 157 284 169 241

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



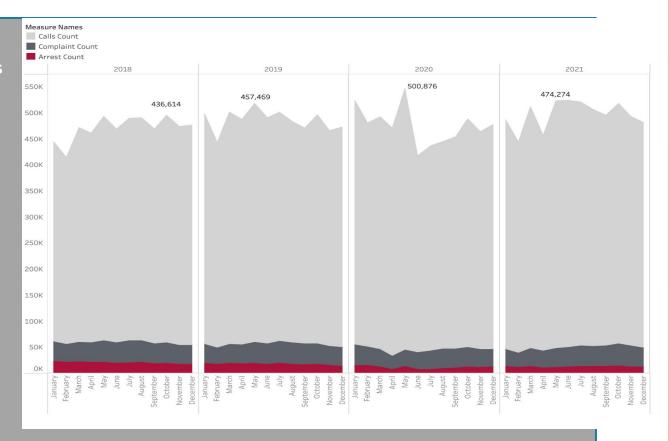
### **Code Challenges**

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

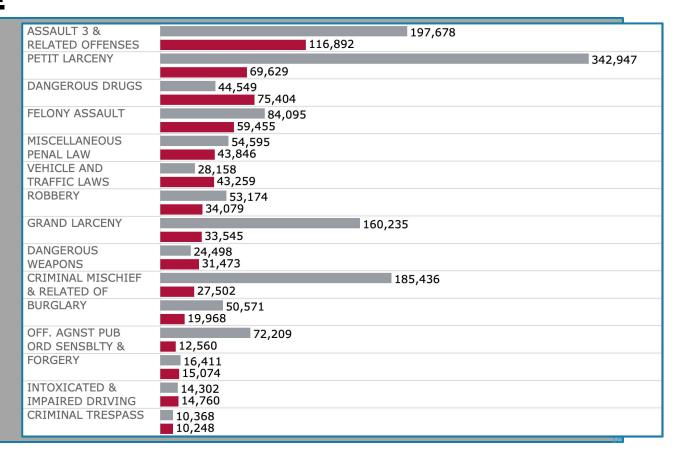
# Code Challenge - Aggregated data

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

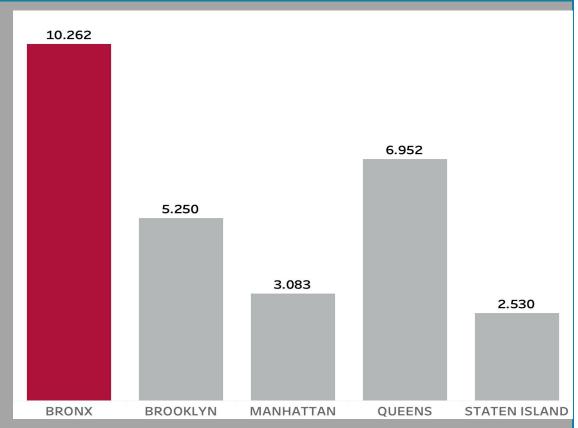
**2020** was the year where most number of 911 calls took place .It reached a numerical peak of 500,876(number of calls) in May just after the advent of the Covid Pandemic. And another interesting thing we noticed is that every year(2018-2021) in the month of May most number of 911 calls are reported.



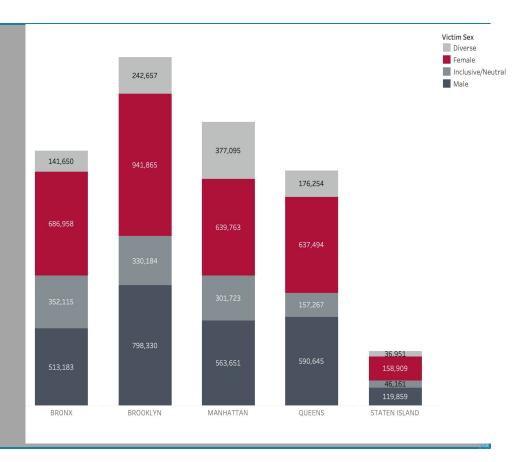
Top offense people in NYC complaint and get arrested for are assault 3, dangerous drugs, petit larceny and Felony assault



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.



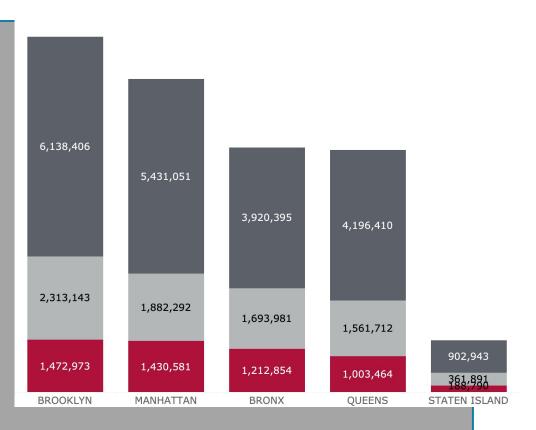
In all the 5 boroughs, the victims were primarily "Female" across the years -2018, 2019, 2020.

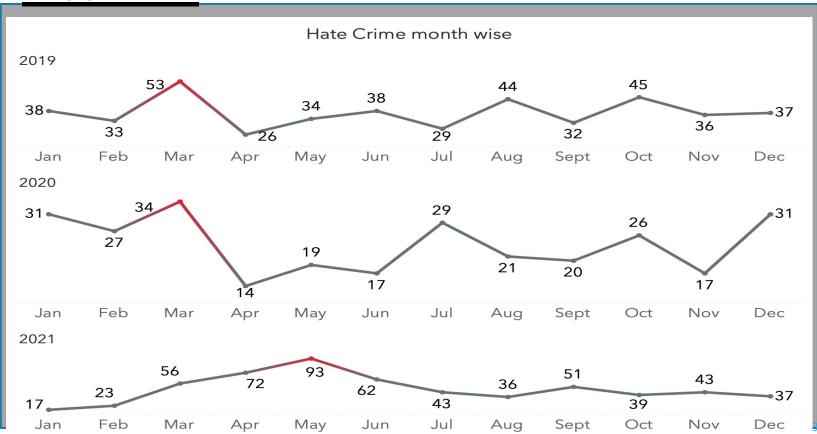


Further diving into the victim dataset and location dataset we could conclude that most black people were targeted in **Brooklyn**-an alarming **785,214!** 

	BLACK	WHITE	WHITE HISPANIC	ASIAN / PACIFIC ISLANDER	BLACK HISPANIC	AMERICAN INDIAN/ALASKAN	Count of Complaints 1,598 785,214
BROOKLYN	785,214	423,858		105,334			1,000
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	

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.





# 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.

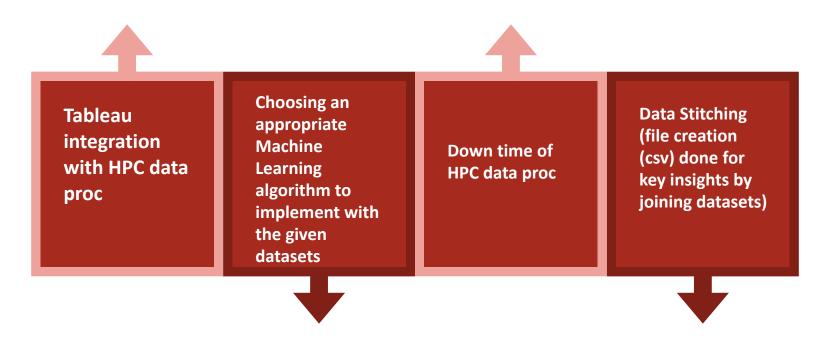
features •	COUNT_ARREST	v prediction	~	≡
(1470,[1,1464,1466,1467,1468,1469], [1.0,1.0,122.0,187.0,6.0,1.0])	8	4.878614125290836		_
(1470,[1,1464,1466,1467,1468,1469], [1.0,1.0,123.0,97.0,6.0,1.0])	4	3.696190126976548		ı
(1470,[1,1464,1466,1467,1468,1469], [1.0,1.0,19.0,182.0,20.0,1.0])	8	9.435968387697114		ı
(1470,[1,1464,1466,1467,1468,1469], [1.0,1.0,24.0,129.0,13.0,1.0])	1	6.528575484608318		ı
(1470,[1,1464,1466,1467,1468,1469], [1 0 1 0 26 0 85 0 11 0 1 0]\	1	5.318230718562197	ı	<b>+</b>

# 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**



# **Learnings**

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

the **transformations** are lazily evaluated. This gives Spark the ability to make optimization decisions, as

ΑII

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.

# <u>Summary</u>

- 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

