

---

# An Examination of Historic NYPD Arrests



By Kenny Jeong

---

# NYPD Arrest Data from 2006–2023 from NYC Open Data

[https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u/about\\_data](https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u/about_data)



# Initial look at the raw data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	ARREST_KEY	ARREST_DATE	PD_CD	PD_DESC	KY_CD	OFNS_DES	LAW_CODE	LAW_CAT	ARREST_BI	ARREST_PI	JURISDICTION	AGE_GROU	PERP_SEX	PERP_RAC	X_COORD	Y_COORD	Latitude	Longitude	Lon_Lat				
2	220756993	11/19/2020	155	RAPE 2	104	RAPE	PL 1303001	F	B	41	0 18-24	M	BLACK	1013232	236725	40.81639	-73.8953	POINT (-73.89529641399997	40.816391847000034)				
3	221995093	12/18/2020	177	SEXUAL AE	116	SEX CRIME	PL 1306503	F	K	84	0 45-64	M	WHITE	989013	192652	40.69547	-73.9828	POINT (-73.98282507899995	40.69546894100006)				
4	221558560	12/9/2020	157	RAPE 1	104	RAPE	PL 1303501	F	S	120	0 <18	M	BLACK	962748	174174	40.64473	-74.0775	POINT (-74.07748315899995	40.64472613100002)				
5	217890363	9/15/2020					PL 2650022	M	B	44	0 25-44	M	BLACK	1009412	245306	40.83996	-73.9091	POINT (-73.90906496999997	40.83995593800007)				
6	219517698	10/21/2020	153	RAPE 3	104	RAPE	PL 1302502	F	Q	112	0 25-44	M	WHITE HIS	1025420	202485	40.72236	-73.8515	POINT (-73.85147389399998	40.72236368700004)				
7	218151681	9/21/2020	157	RAPE 1	104	RAPE	PL 1303501	F	K	77	0 25-44	M	BLACK	1003606	185050	40.67458	-73.9302	POINT (-73.93022154099998	40.67458330800008)				
8	174564403	2/8/2018	440				PL 1553502	F	M	18	0 25-44	M	BLACK	988211	218129	40.7654	-73.9857	POINT (-73.98570226399994	40.765397475000036)				
9	215001943	7/7/2020	153	RAPE 3	104	RAPE	PL 1302503	F	B	41	0 25-44	M	BLACK	1013232	236725	40.81639	-73.8953	POINT (-73.89529641399997	40.816391847000034)				
10	215508944	7/20/2020	263	ARSON 2,3	114	ARSON	PL 1500500	F	S	123	0 <18	M	WHITE	914928	125757	40.51159	-74.2493	POINT (-74.24930372699998	40.51158633800003)				
11	217233706	8/31/2020	153	RAPE 3	104	RAPE	PL 1302503	F	K	77	0 18-24	M	BLACK	1003606	185050	40.67458	-73.9302	POINT (-73.93022154099998	40.67458330800008)				
12	217147644	8/28/2020	157	RAPE 1	104	RAPE	PL 1303503	F	Q	105	0 45-64	M	BLACK	1057636	203954	40.72619	-73.7352	POINT (-73.73523390399998	40.72618713000002)				
13	215620070	7/22/2020	157	RAPE 1	104	RAPE	PL 1303504	F	M	25	0 18-24	M	WHITE	1000555	230994	40.80069	-73.9411	POINT (-73.94110928599997	40.800694331000045)				
14	212465831	4/28/2020	777				PL 1950200	F	K	75	2 45-64	M	BLACK HIS	1016409	178125	40.65554	-73.8840	POINT (-73.88409879299998	40.65553848200005)				
15	216648707	8/16/2020	157	RAPE 1	104	RAPE	PL 1303501	F	K	69	0 25-44	M	BLACK	1010641	175525	40.64842	-73.9049	POINT (-73.90489710899999	40.64842111900003)				
16	192890913	1/29/2019	157	RAPE 1	104	RAPE	PL 1303501	F	Q	112	0 25-44	M	BLACK	1025420	202485	40.72236	-73.8515	POINT (-73.85147389399998	40.72236368700004)				
17	194418197	3/7/2019	594	OBSCENIT	116	SEX CRIME	PL 2350600	F	K	79	0 45-64	M	BLACK	999634	190253	40.68887	-73.9445	POINT (-73.94452903199993	40.688872153000034)				
18	204062162	10/22/2019	153	RAPE 3	104	RAPE	PL 1302501	F	B	49	0 18-24	M	BLACK HIS	1027295	251236	40.85616	-73.8444	POINT (-73.84439712399995	40.85616295700004)				
19	197484672	5/22/2019	157	RAPE 1	104	RAPE	PL 1303501	F	M	25	0 25-44	M	BLACK	1000555	230994	40.80069	-73.9411	POINT (-73.94110928599997	40.800694331000045)				
20	195369731	3/31/2019	177	SEXUAL AE	116	SEX CRIME	PL 1306501	F	K	88	0 45-64	M	BLACK	995188	190691	40.69008	-73.9606	POINT (-73.96055993199997	40.69008097200003)				
21	200070545	7/19/2019	594	OBSCENIT	116	SEX CRIME	PL 2631100	F	M	5	0 45-64	M	WHITE	984946	200203	40.7162	-73.9975	POINT (-73.99749074599998	40.716195914000025)				
22	192501267	1/19/2019	153	RAPE 3	104	RAPE	PL 1302503	F	K	77	0 18-24	M	BLACK	1003606	185050	40.67458	-73.9302	POINT (-73.93022154099998	40.67458330800008)				
23	192015578	1/8/2019	263	ARSON 2,3	114	ARSON	PL 1501500	F	K	73	71 25-44	F	BLACK	1009450	182401	40.6673	-73.9092	POINT (-73.90916337699997	40.667297692000034)				
24	193203736	2/5/2019	777				PL 1950200	F	M	1	0 18-24	F	WHITE	982351	201758	40.72046	-74.0069	POINT (-74.00685220399998	40.720463840000036)				
25	195191904	3/26/2019	440				PL 1553502	F	M	10	0 18-24	M	ASIAN / P	983821	213359	40.75231	-74.0015	POINT (-74.00154983299996	40.75230591800005)				
26	197053144	5/10/2019	157	RAPE 1	104	RAPE	PL 1303502	F	K	75	2 45-64	M	BLACK	1017141	183798	40.67111	-73.8814	POINT (-73.88143295699996	40.67110691100004)				
27	194937955	3/20/2019	157	RAPE 1	104	RAPE	PL 1303501	F	K	77	0 25-44	M	BLACK	1003606	185050	40.67458	-73.9302	POINT (-73.93022154099998	40.67458330800008)				
28	197161753	5/14/2019	263	ARSON 2,3	114	ARSON	PL 1501500	F	K	77	71 45-64	M	BLACK	1005507	185647	40.67622	-73.9234	POINT (-73.92336640699995	40.676217578000035)				
29	193333824	2/8/2019	155	RAPE 2	104	RAPE	PL 1303001	F	S	120	0 18-24	M	BLACK	962748	174174	40.64473	-74.0775	POINT (-74.07748315899994	40.64472613100003)				
30	176090087	3/20/2018	153	RAPE 3	104	RAPE	PL 1302503	F	Q	115	3 25-44	M	BLACK	1018699	215043	40.75686	-73.8757	POINT (-73.87565666399996	40.75686097700003)				
31	192449567	1/17/2019	153	RAPE 3	104	RAPE	PL 1302502	F	K	73	0 18-24	M	BLACK	1008276	183623	40.67066	-73.9134	POINT (-73.91339091999998	40.670655072000045)				
32	206432631	12/13/2019	155	RAPE 2	104	RAPE	PL 1303001	F	K	77	0 18-24	M	BLACK	1003606	185050	40.67458	-73.9302	POINT (-73.93022154099998	40.67458330800008)				
33	193150447	2/4/2019	157	RAPE 1	104	RAPE	PL 1303501	F	K	73	2 25-44	M	BLACK	1008276	183623	40.67066	-73.9134	POINT (-73.91339091999998	40.670655072000045)				
34	1937370513	5/19/2019	155	RAPE 2	104	RAPE	PL 1303001	F	Q	103	0 18-24	M	BLACK	1041749	196938	40.70705	-73.7926	POINT (-73.79261190399995	40.707047475000024)				
35	192174888	1/10/2019	157	RAPE 1	104	RAPE	PL 1303503	F	M	14	0 18-24	M	WHITE	985679	213954	40.75394	-73.9918	POINT (-73.99181355199995	40.753938938000031)				

---

# Data Structure

Over 5 million rows

## Relevant Columns:

### ARREST\_KEY

Unique key for each arrest

### ARREST\_DATE

Date of arrest

### OFNS\_DESC

Type of crime

### OFNS\_LEVEL

Level of offense

### BORO

Borough where arrest occurred

### PRECINCT

Precinct where arrest occurred

### AGE\_GROUP

Perpetrator's age within category

### SEX

Perpetrator's sex description

### RACE

Perpetrator's race description

Note: in the data, there is a white-hispanic label and black-hispanic label. There is no hispanic demographic on its own.

# Preparing the Data

(Dataset from NYC Open Data)

Data cleaned in MySQL

- Created table and loaded data

```
USE nypd;
CREATE TABLE nypd_arrests
(
    ARREST_KEY int,
    ARREST_DATE text,
    PD_CD int,
    PD_DESC text,
    KY_CD int,
    OFNS_DESC text,
    LAW_CODE text,
    LAW_CAT_CD text,
    ARREST_BORO text,
    ARREST_PRECINCT int,
    JURISDICTION_CODE int,
    AGE_GROUP text,
    PERP_SEX text,
    PERP_RACE text,
    X_COORD_CD int,
    Y_COORD_CD int,
    Latitude double,
    Longitude double,
    Lon_Lat text
);

LOAD DATA LOCAL INFILE 'C:/Users/Kenny/Desktop/DATA_PORTFOLIO_PROJECTS/NYPD_ARRESTS_HISTORIC/NYPD_Arrests_Data_Historic_2024.csv'
INTO TABLE nypd_arrests
CHARACTER SET utf8
FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"' LINES TERMINATED BY '\n'
IGNORE 1 LINES
SET ARREST_DATE = str_to_date(ARREST_DATE, '%m/%d/%Y');
```

[full SQL script available in github repo](#)

---

# Preparing the Data

(Dataset from NYC Open Data)

Data cleaned in MySQL

- Created table and loaded data
- Dropped unneeded columns

```
ALTER TABLE nypd_arrests  
DROP COLUMN PD_CD,  
DROP COLUMN PD_DESC,  
DROP COLUMN KY_CD,  
DROP COLUMN LAW_CODE,  
DROP COLUMN JURISDICTION_CODE,  
DROP COLUMN X_COORD_CD,  
DROP COLUMN Y_COORD_CD,  
DROP COLUMN Lon_Lat;
```

# Preparing the Data

(Dataset from NYC Open Data)

## Data cleaned in MySQL

- Created table and loaded data
- Dropped unneeded columns
- Inspected & cleaned null values across all columns

```
SELECT
SUM(CASE WHEN ARREST_KEY is null THEN 1 ELSE 0 END) AS nullkey,
SUM(CASE WHEN ARREST_DATE is null THEN 1 ELSE 0 END) AS nulldate,
SUM(CASE WHEN OFNS_DESC is null THEN 1 ELSE 0 END) AS nullofns,
SUM(CASE WHEN LAW_CAT_CD is null THEN 1 ELSE 0 END) AS nullcat,
SUM(CASE WHEN ARREST_BORO is null THEN 1 ELSE 0 END) AS nullboro,
SUM(CASE WHEN ARREST_PRECINCT is null THEN 1 ELSE 0 END) AS nullprecinct,
SUM(CASE WHEN AGE_GROUP is null THEN 1 ELSE 0 END) AS nullage,
SUM(CASE WHEN PERP_SEX is null THEN 1 ELSE 0 END) AS nullsex,
SUM(CASE WHEN PERP_RACE is null THEN 1 ELSE 0 END) AS nullrace,
SUM(CASE WHEN Latitude is null THEN 1 ELSE 0 END) AS nulllat,
SUM(CASE WHEN Longitude is null THEN 1 ELSE 0 END) AS nulllong,
( SELECT COUNT(*) from nypd_arrests
WHERE ARREST_KEY IS NULL OR
ARREST_DATE IS NULL OR
OFNS_DESC IS NULL OR
LAW_CAT_CD IS NULL OR
ARREST_BORO IS NULL OR
ARREST_PRECINCT IS NULL OR
AGE_GROUP IS NULL OR
PERP_SEX IS NULL OR
PERP_RACE IS NULL OR
Latitude IS NULL OR
Longitude IS NULL
) AS nulltotal
FROM nypd_arrests;
```



# Preparing the Data

(Dataset from NYC Open Data)

## Data cleaned in MySQL

- Created table and loaded data
- Dropped unneeded columns
- Inspected & cleaned null values across all columns
- Correcting/grouping values across all columns

```
SELECT DISTINCT OFNS_DESC FROM nypd_arrests ORDER BY OFNS_DESC;
-- there are many fields that can be combined, typos, as well as a (nulls) value
UPDATE nypd_arrests SET OFNS_DESC = 'Unknown' WHERE OFNS_DESC = '(null)' OR OFNS_DESC IS NULL;
UPDATE nypd_arrests SET OFNS_DESC = 'ADMINISTRATIVE CODES' WHERE OFNS_DESC = 'ADMINISTRATIVE CODE';
UPDATE nypd_arrests SET OFNS_DESC = 'CHILD ABANDONMENT/NON SUPPORT' WHERE OFNS_DESC = 'CHILD ABANDONMENT/NON SUPPORT 1';
UPDATE nypd_arrests SET OFNS_DESC = 'CRIMINAL MISCHIEF & RELATED OFFENSES' WHERE OFNS_DESC = 'CRIMINAL MISCHIEF & RELATED OF';
UPDATE nypd_arrests SET OFNS_DESC = 'DISRUPTION OF A RELIGIOUS SERVICE' WHERE OFNS_DESC = 'DISRUPTION OF A RELIGIOUS SERV';
UPDATE nypd_arrests SET OFNS_DESC = 'HARRASSMENT' WHERE OFNS_DESC = 'HARRASSMENT 2';
UPDATE nypd_arrests SET OFNS_DESC = 'DISRUPTION OF A RELIGIOUS SERVICE' WHERE OFNS_DESC = 'DISRUPTION OF A RELIGIOUS SERV';
UPDATE nypd_arrests SET OFNS_DESC = 'HOMICIDE-NEGLIGENT, UNCLASSIFIED' WHERE OFNS_DESC = 'HOMICIDE-NEGLIGENT, UNCLASSIFIED';
UPDATE nypd_arrests SET OFNS_DESC = 'INTOXICATED/IMPAIRED DRIVING' WHERE OFNS_DESC = 'INTOXICATED & IMPAIRED DRIVING';
UPDATE nypd_arrests SET OFNS_DESC = 'KIDNAPPING & RELATED OFFENSES' WHERE OFNS_DESC = 'KIDNAPPING' OR OFNS_DESC = 'KIDNAPPING AND RELATED OFFENSES';
UPDATE nypd_arrests SET OFNS_DESC = 'LOITERING/GAMBLING (CARDS, DICE, ETC)' WHERE OFNS_DESC = 'LOITERING/GAMBLING (CARDS, DIC';
UPDATE nypd_arrests SET OFNS_DESC = 'MURDER & NON-NEGL. MANSLAUGHTER' WHERE OFNS_DESC = 'MURDER & NON-NEGL. MANSLAUGHTER';
UPDATE nypd_arrests SET OFNS_DESC = 'MYS LAWS-UNCLASSIFIED' WHERE OFNS_DESC = 'MYS LAWS-UNCLASSIFIED FELONY' OR OFNS_DESC = 'MYS LAWS-UNCLASSIFIED VIOLATION';
UPDATE nypd_arrests SET OFNS_DESC = 'OFF. AGNST PUB ORD SENSBLTY & RIGHTS TO PRIV' WHERE OFNS_DESC = 'OFF. AGNST PUB ORD SENSBLTY &';
UPDATE nypd_arrests SET OFNS_DESC = 'OFFENSES AGAINST PUBLIC ADMINISTRATION' WHERE OFNS_DESC = 'OFFENSES AGAINST PUBLIC ADMINI';
UPDATE nypd_arrests SET OFNS_DESC = 'OTHER OFFENSES RELATED TO THEFT' WHERE OFNS_DESC = 'OTHER OFFENSES RELATED TO THEFT';
UPDATE nypd_arrests SET OFNS_DESC = 'OTHER STATE LAWS (NON PENAL LAW)' WHERE OFNS_DESC = 'OTHER STATE LAWS (NON PENAL LA';
UPDATE nypd_arrests SET OFNS_DESC = 'POSSESSION OF STOLEN PROPERTY' WHERE OFNS_DESC = 'POSSESSION OF STOLEN PROPERTY 5';
UPDATE nypd_arrests SET OFNS_DESC = 'UNAUTHORIZED USE OF A VEHICLE' WHERE OFNS_DESC = 'UNAUTHORIZED USE OF A VEHICLE 3 (UNV)';
UPDATE nypd_arrests SET OFNS_DESC = 'UNLAWFUL POSS. WEAP. ON SCHOOL GROUNDS' WHERE OFNS_DESC = 'UNLAWFUL POSS. WEAP. ON SCHOOL';
```

```
SELECT DISTINCT LAW_CAT_CD FROM nypd_arrests;
-- LAW_CAT_CD normally has F, M, V, I, but also contains nulls, (null), and 9
SELECT * FROM nypd_arrests WHERE LAW_CAT_CD = '9' OR LAW_CAT_CD = '(null)' OR LAW_CAT_CD IS null;
-- 24669 rows but they all still contain location and demographic info
UPDATE nypd_arrests SET LAW_CAT_CD = 'Unknown' WHERE LAW_CAT_CD = '9' OR LAW_CAT_CD = '(null)' OR LAW_CAT_CD IS NULL;
-- replacing the LAW_CAT_CD codes (F, M, V, I) with FELONY, MISDEMEANOR, VIOLATION, INFRACTION
UPDATE nypd_arrests
SET LAW_CAT_CD = REPLACE(REPLACE(REPLACE(REPLACE(LAW_CAT_CD, 'F', 'Felony'), 'M', 'Misdemeanor'), 'V', 'Violation'), 'I', 'Infraction');
```



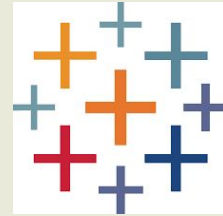
---

# Preparing the Data

(Dataset from NYC Open Data)

Data cleaned in MySQL

- Created table and loaded data
- Dropped unneeded columns
- Inspected & cleaned null values across all columns
- Correcting/grouping values across all columns



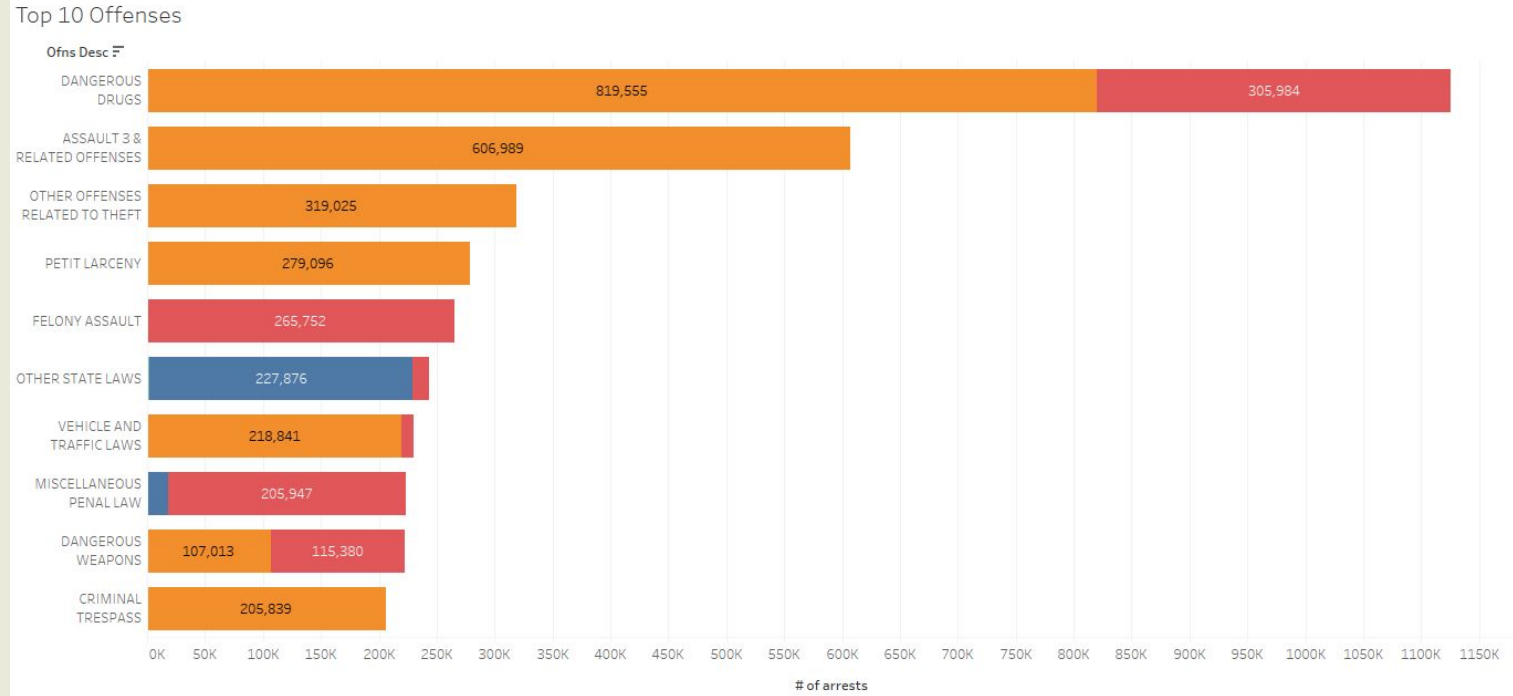
Once data cleansing is completed, imported to Tableau for further transformation & visualization

---

# Key Insights

[Link to dashboard on Tableau Public](#)

# Top 10 Types of Crime

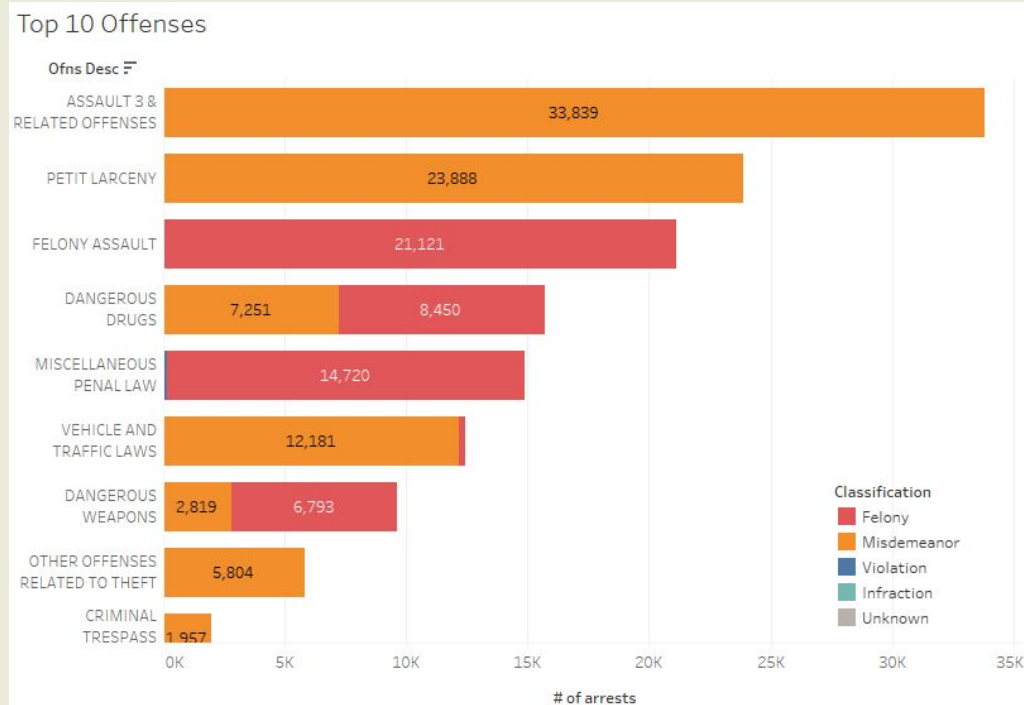


---

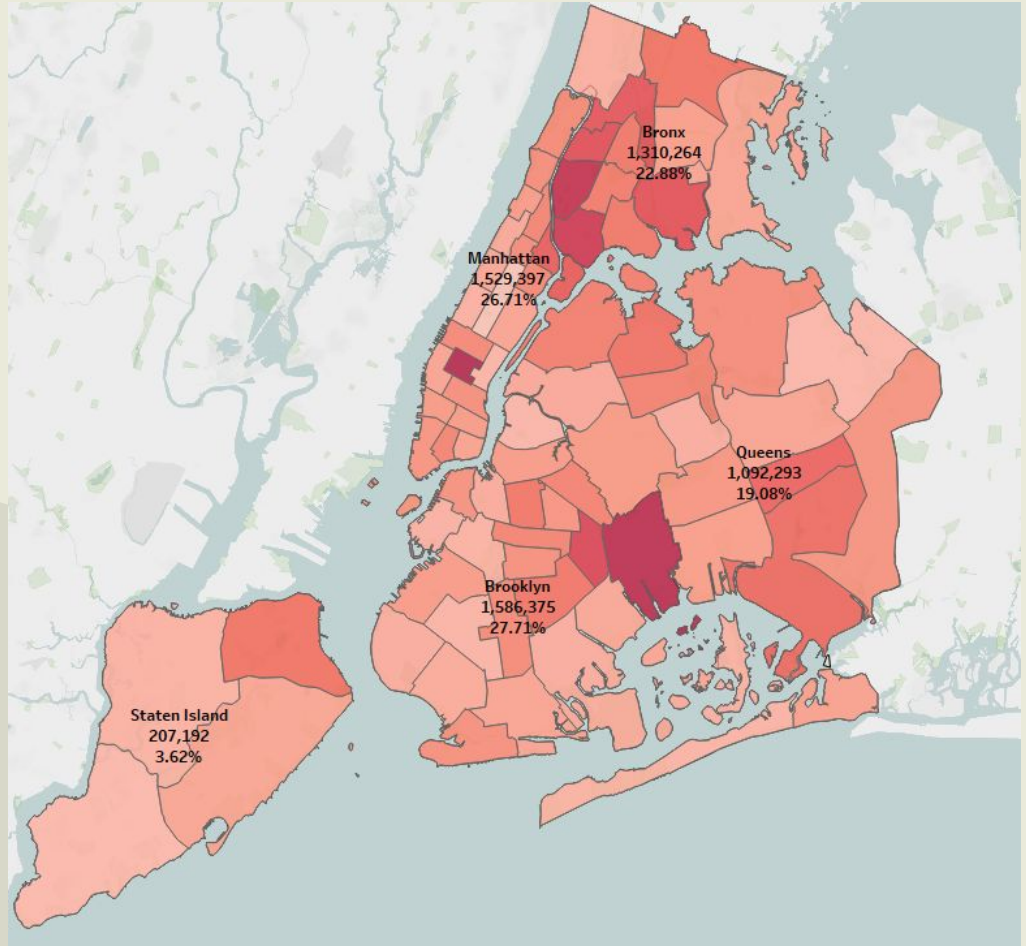
**Dangerous Drugs were the most common offense**

**However, in 2018, Assault overtook it as the most common offense**

# Top 10 Types of Crime in 2023



# Borough and Precinct Arrest Heatmap



### Boroughs by # of arrests:

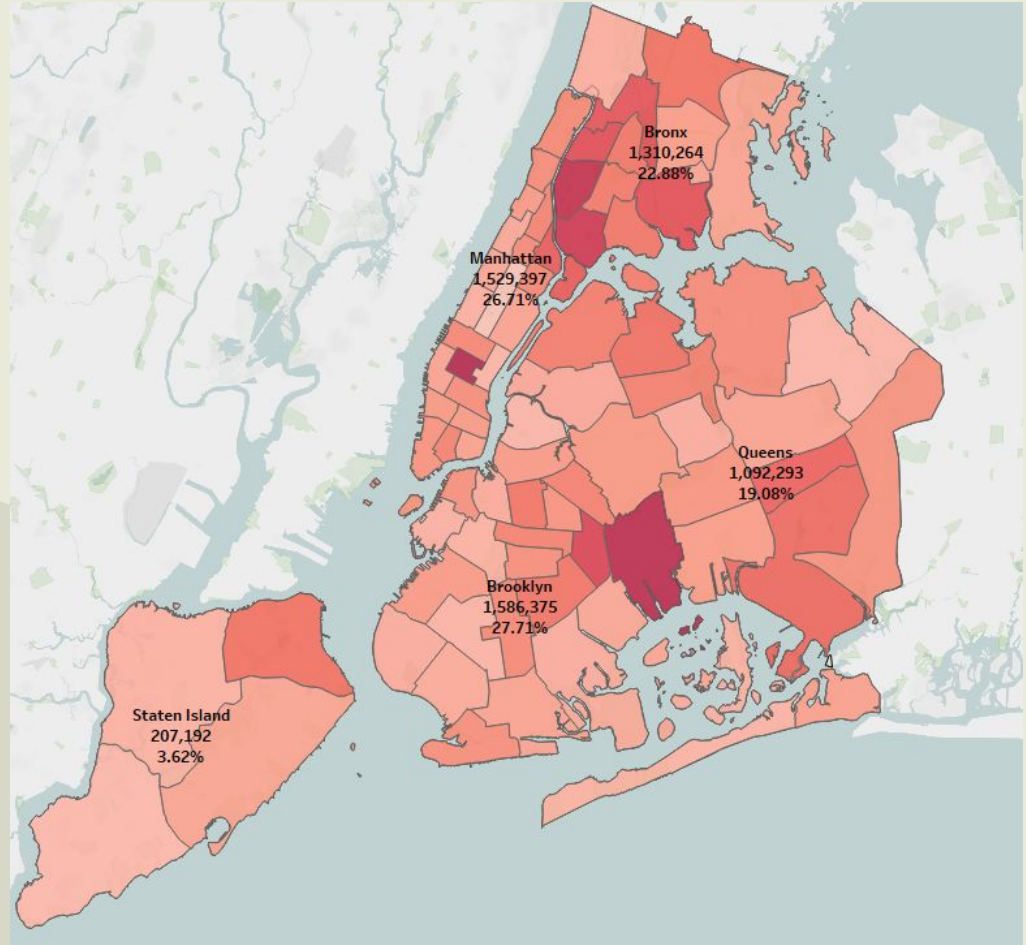
Brooklyn

Manhattan

Bronx

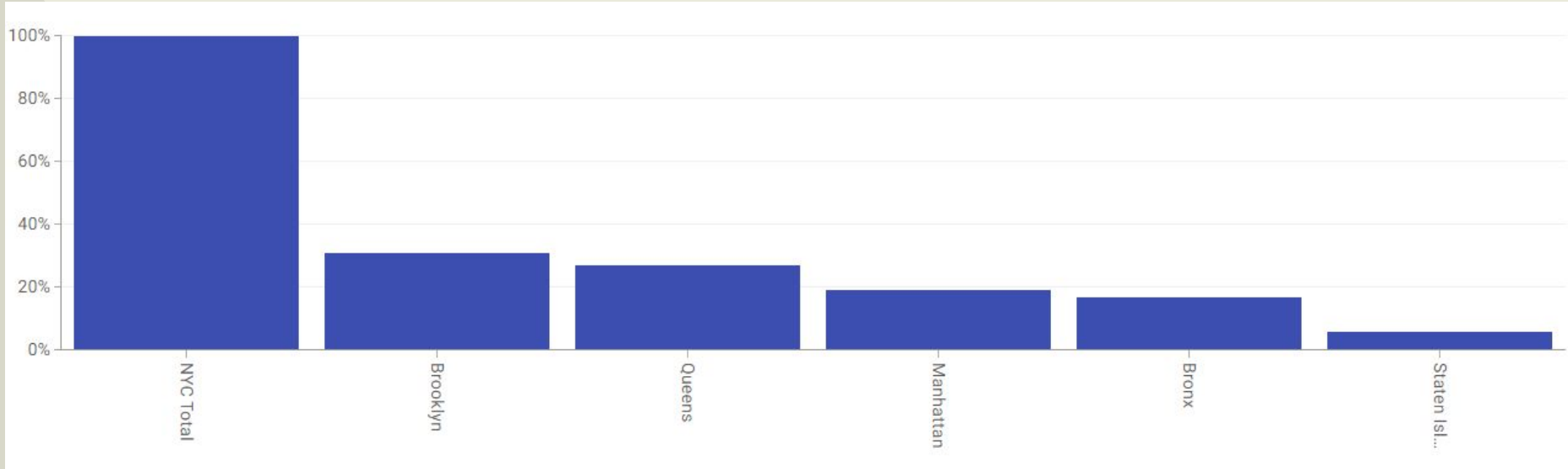
Queens

Staten Island





## Boroughs by population (in 2020):



---

# Differences in Boroughs

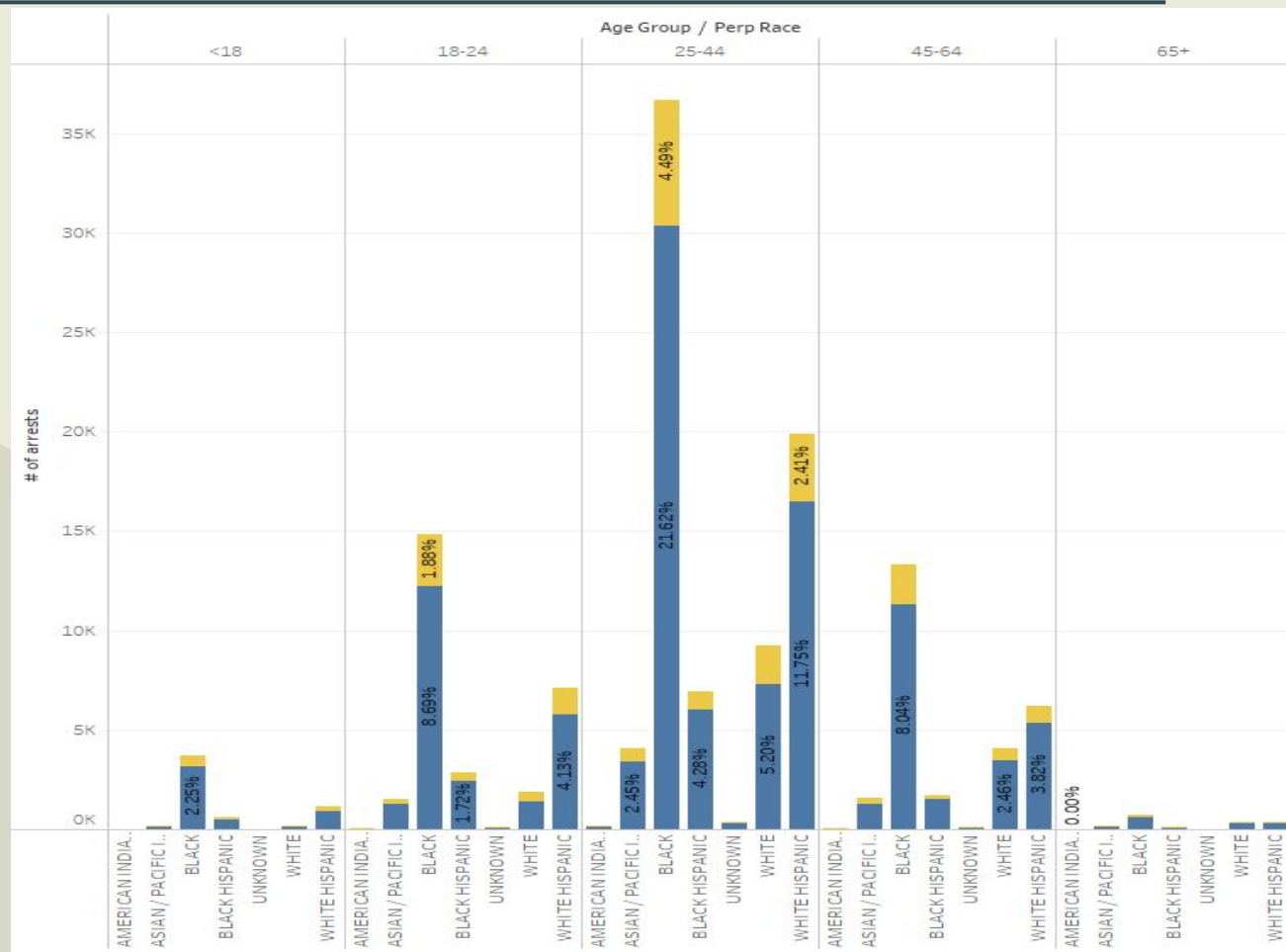
- Despite its reputation for high crime, The Bronx came in 3rd for number of arrests
  - This is probably mainly a result of its lower total population size
- Staten Island has a significantly smaller population and number of arrests



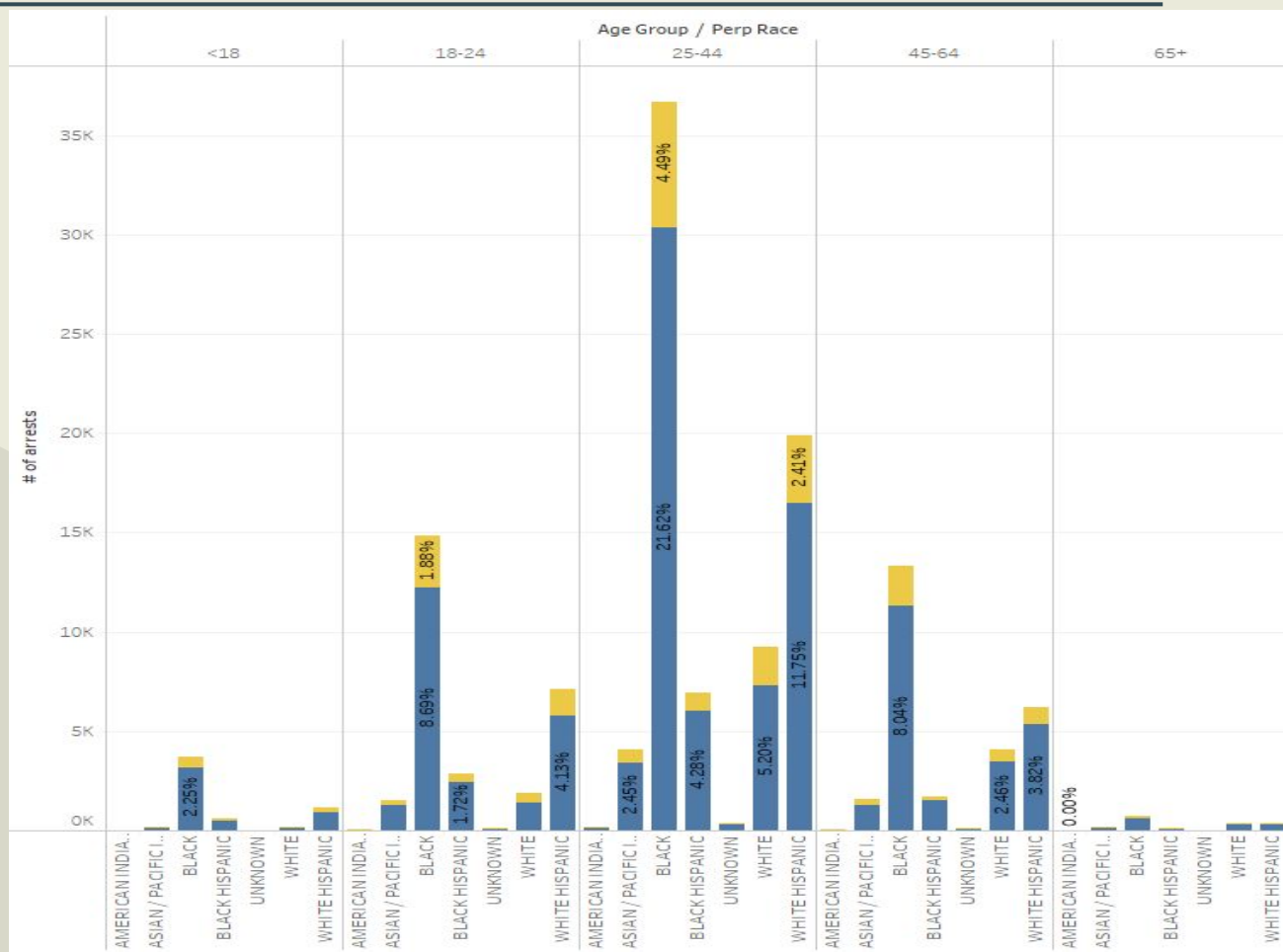
Further research into each borough that would be useful:

- Income/Poverty levels
- Housing/Homelessness
- Racial demographics

# Arrests by Age Group, Race, and Sex



- Black males age 25-44 are arrested at the highest rates by far
  - Followed by White-Hispanic males age 25-44
- In every age group, the black male demographic has the highest number of arrests

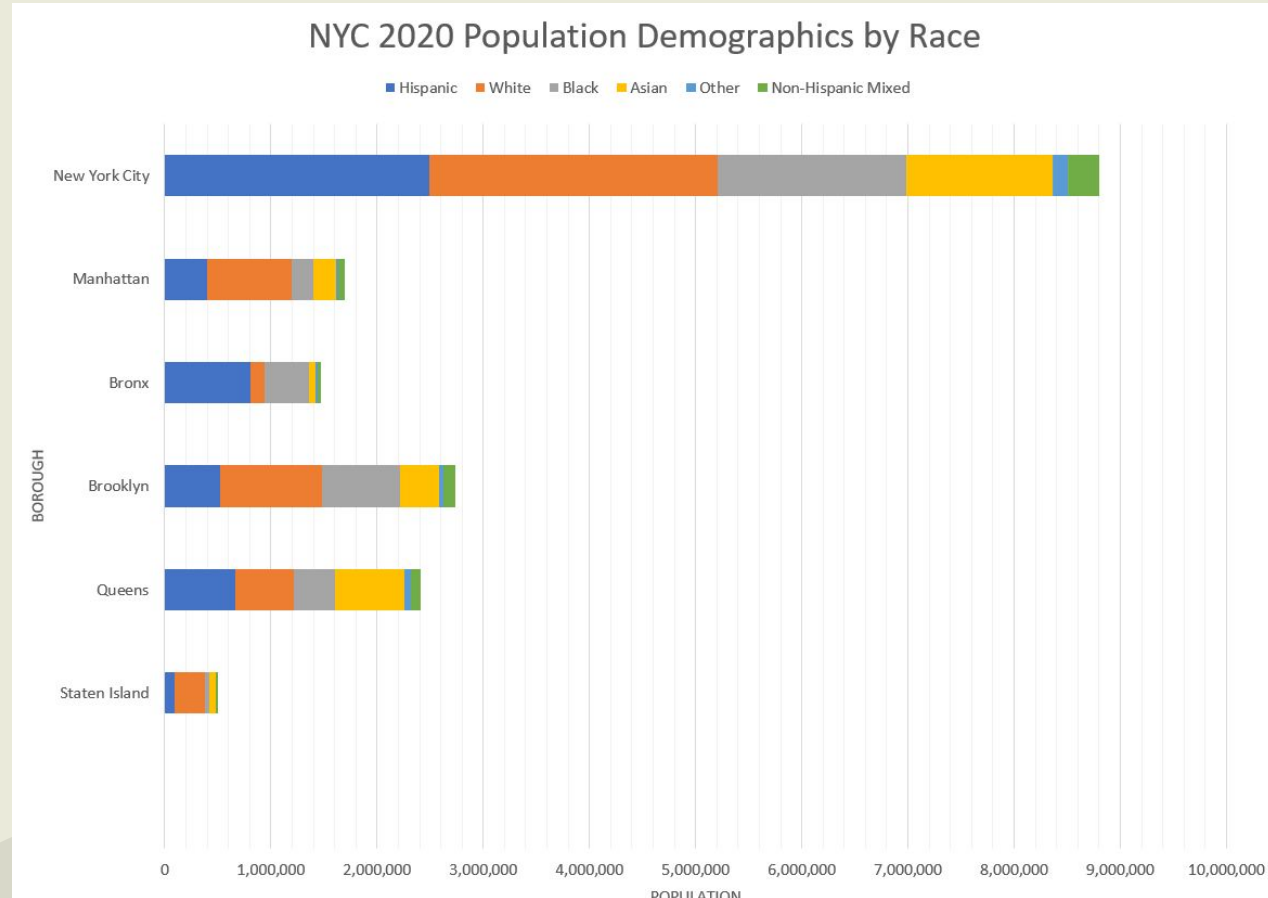


# NYC 2020 Census

Highest population by race:

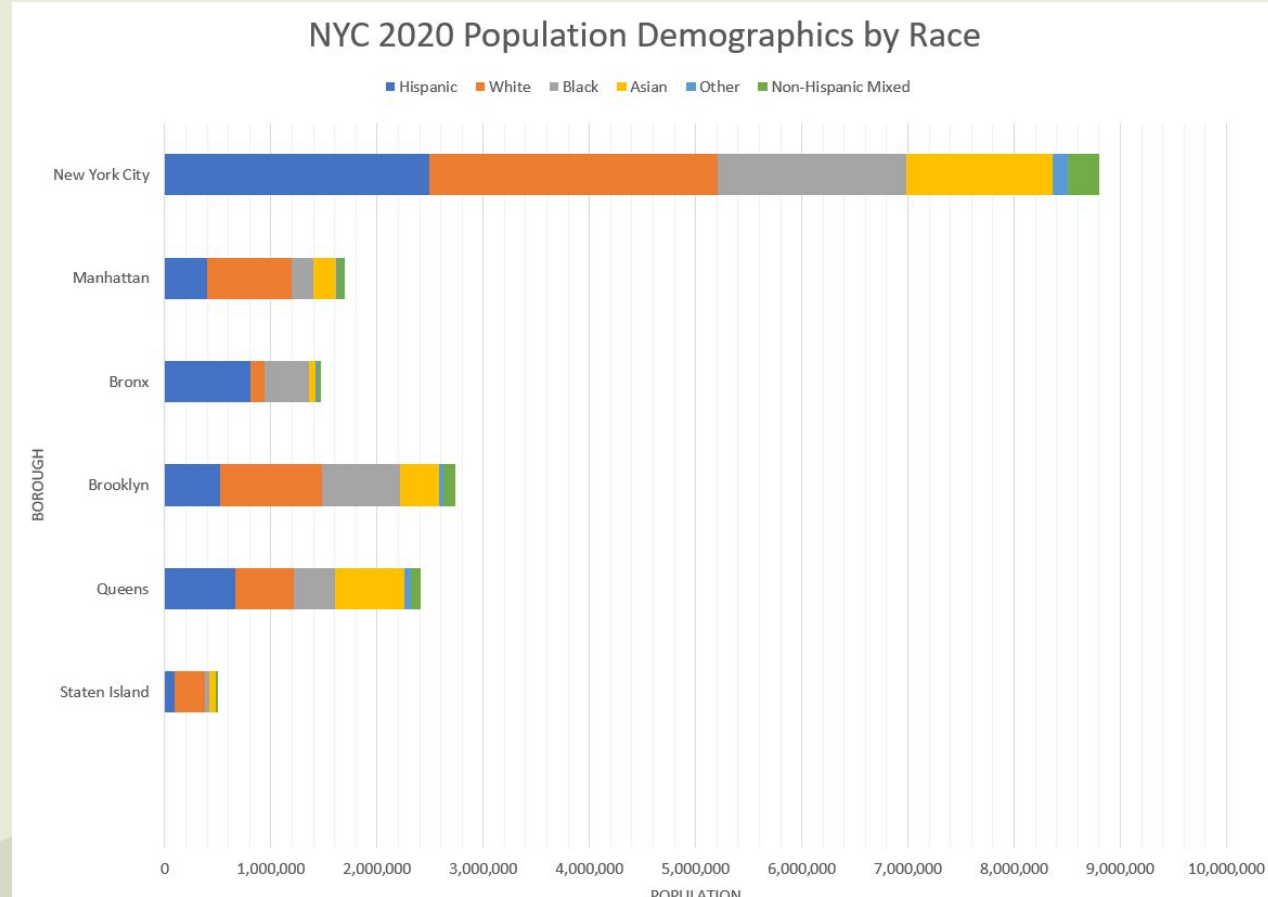
1. Hispanics
2. Whites
3. Blacks

Despite being the 3rd largest demographic in NYC, black people make up about half of total arrests

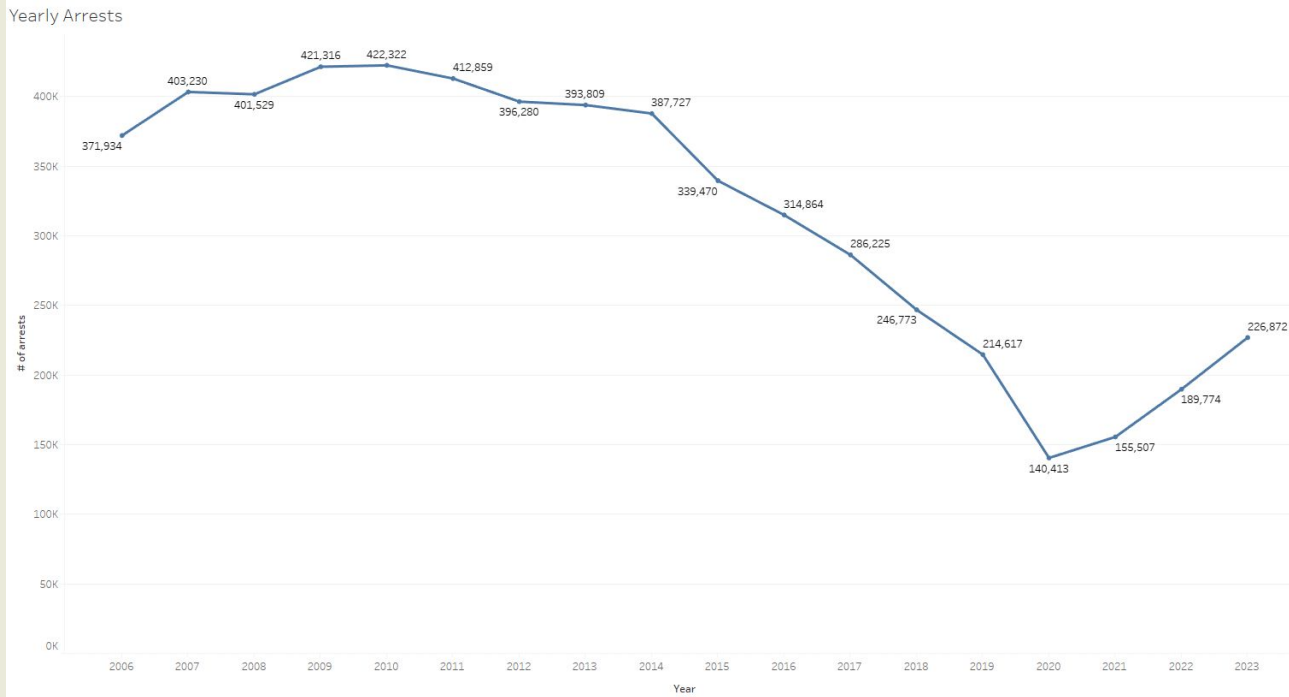


Boroughs ordered by # of arrests:

- Brooklyn – White dominant, then black
- Manhattan – White dominant, then hispanic
- Bronx – Hispanic dominant, then black
- Queens – Hispanic dominant, then white
- Staten Island – White dominant

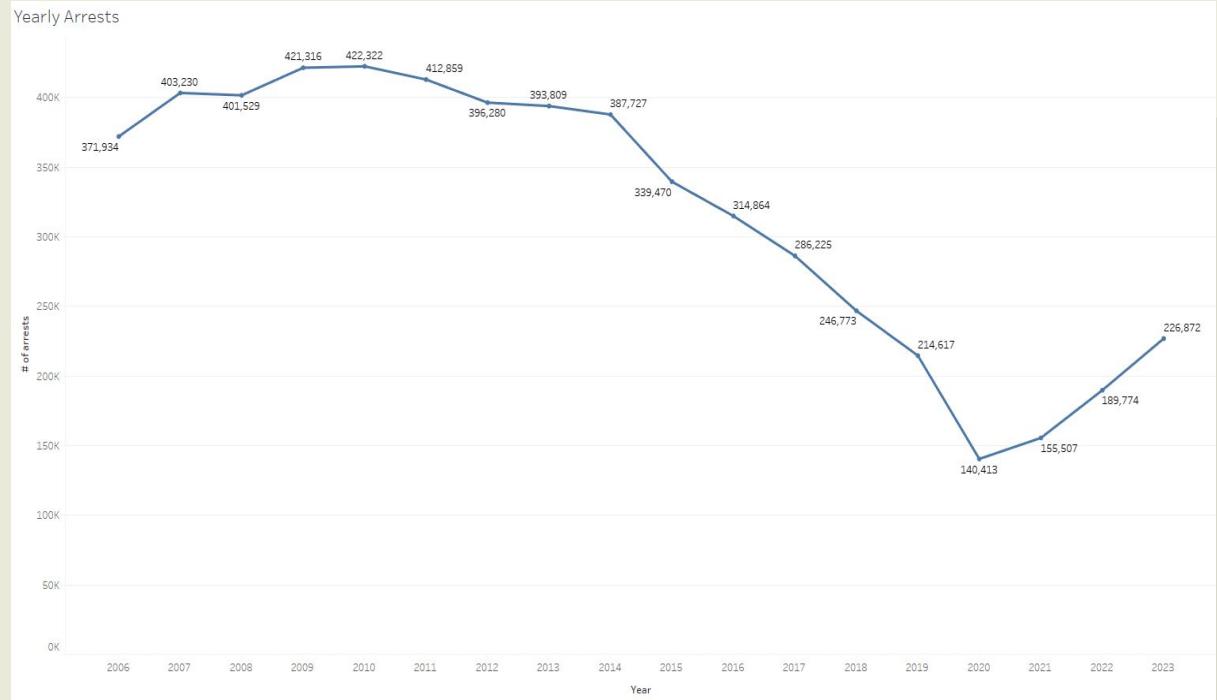


# A look over the years 2006 - 2023





- Arrest rates began trending downwards rapidly in the mid 2010s
- Until 2020, when arrest rates began to rise
  - in these years assaults were the most common offense type

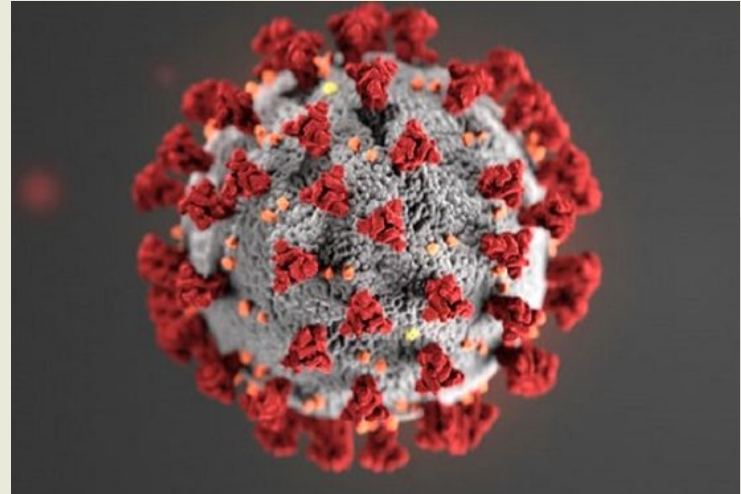


---

# A Post-Covid World

COVID-19 pandemic may have been a factor in crime, particularly violent crime

The pandemic lockdowns also caused severe economic distress, which is an intersecting variable with crime



- Another example of economic distress affecting crime may be shown during the peak years of 2009, 2010, and 2011
- Post 2008 financial crisis



---

# 2021 Census Median Household Income

By intersecting population density, demographic data shown previously with economic data, it creates a more contextualized picture of what conditions lead to crime.

Bronx having such a high arrest rate comparative to its population density makes more sense when seeing the median income is far below other boroughs. Inversely, Staten Island's low arrest rate is correlated with its high income levels.

Staten Island	\$86,504
Manhattan	\$84,435
Queens	\$73,262
Brooklyn	\$67,567
Bronx	\$43,011


---

# Conclusions

The black demographic accounted for about half of all arrests made despite being the 3rd largest population size. This may be indicative of racial bias in policing.

The rising arrest rates post 2008 financial crisis and post COVID pandemic is indicative of the link between economic conditions and crime rates. Under periods of severe economic stress, crime seems to skyrocket.

Further research into economic development programs in lower income/impoverished areas should be done.



---

# Thanks!

This project originally began as part of my first group project at COOP. It started with a smaller dataset, sampling ~1000 rows from 2021 NYPD arrest data. I found the contents of the project to be interesting, so I decided to revisit it here by extracting raw historic datasets to flesh out more of my own analysis.

**CREDITS:** This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

