

A black and white photograph of three NYPD officers in riot gear. The officer on the left is seen from the back, wearing a helmet and a vest with 'POLICE' written on it. The officer in the middle is also seen from the back, wearing a helmet and a vest with 'POLICE' written on it. The officer on the right is facing forward, wearing a helmet, a balaclava, and a vest with 'POLICE' written on it. They are standing in a crowd of people, with a building visible in the background.

NYPD COMPLAINTS

**EXPLORING DECADES OF CIVILIAN COMPLAINTS
AGAINST THE AMERICA'S LARGEST POLICE FORCE**

PROJECT CONTEXT



The NYC Civilian Complaint Review Board (CCRB) is the oversight agency of NYPD.

It investigates complaints of misconduct against the NYPD.

PROJECT CONTEXT



The NYC Civilian Complaint Review Board (CCRB) is the oversight agency of the largest police force in the country.

The CCRB investigates complaints of misconduct against the NYPD.



CCRB data was inaccessible to the public until recently.

In the wake of national protests against police brutality, a state law protecting their secrecy was repealed in June.

PROJECT CONTEXT



The NYC Civilian Complaint Review Board (CCRB) is the oversight agency of the largest police force in the country.

The CCRB investigates complaints of misconduct against the NYPD.



CCRB data was inaccessible to the public until recently.

In the wake of national protests against police brutality, a state law protecting their secrecy was repealed in June.



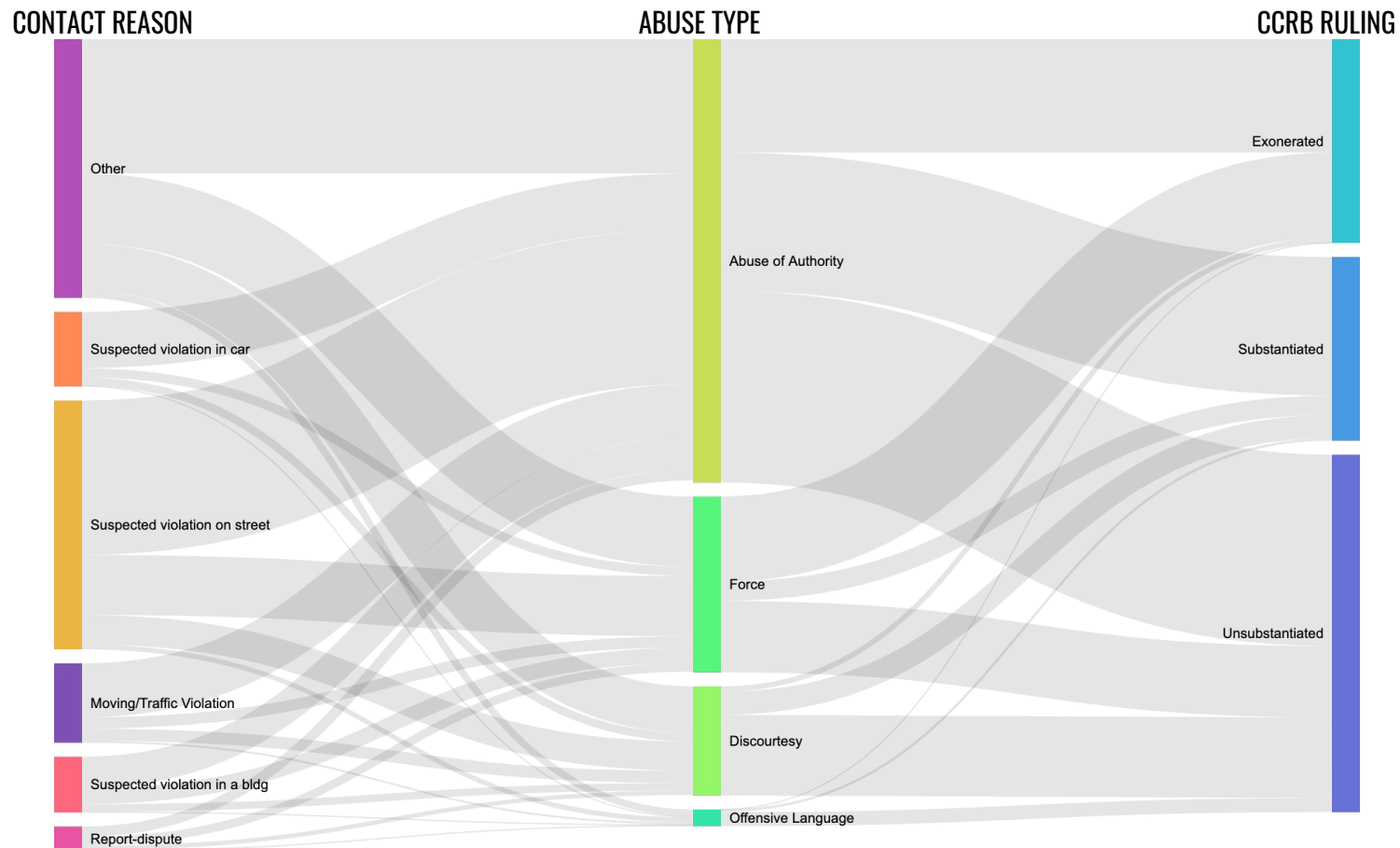
“The release of [this data] is an important step towards greater transparency and accountability, and is just the beginning of unraveling the monopoly the N.Y.P.D. holds on public information and officer discipline.”
- Christopher Dunn, Legal Director NYCLU

THE DATASET: QUICK STATS

- Covers complaints filed from 1985 through 2020
- Includes over 30,000 unique complaints, some from the same incident, which covers 4,000 active NYPD Officers
- Only includes complaints against current officers who have had at least one substantiated complaint against them

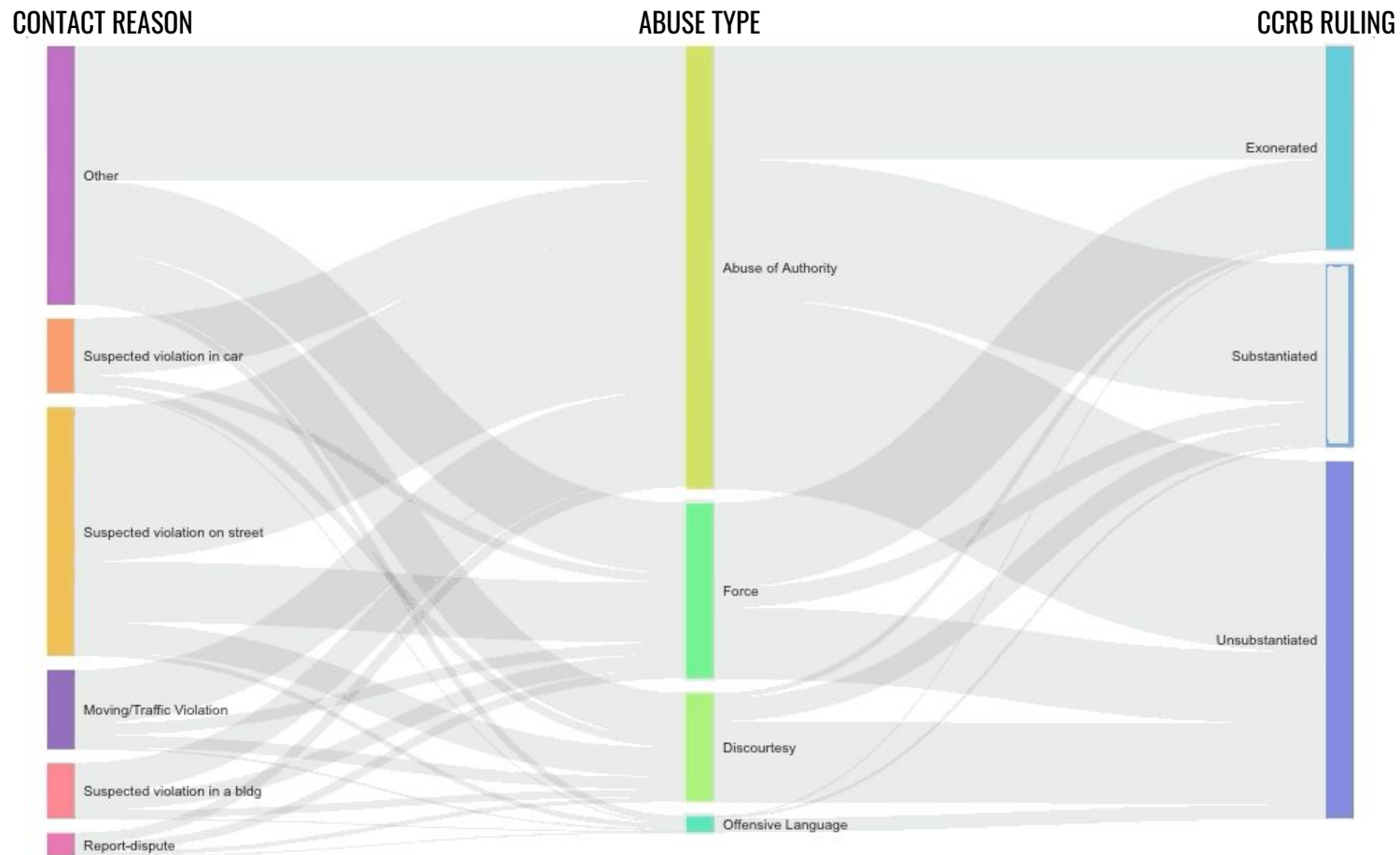
EXPLORING THE DATA

HOW COMPLAINTS FLOW THROUGH THE CCRB



See the interactive version [here](#). Credit to Jarrett Meyer for the template.

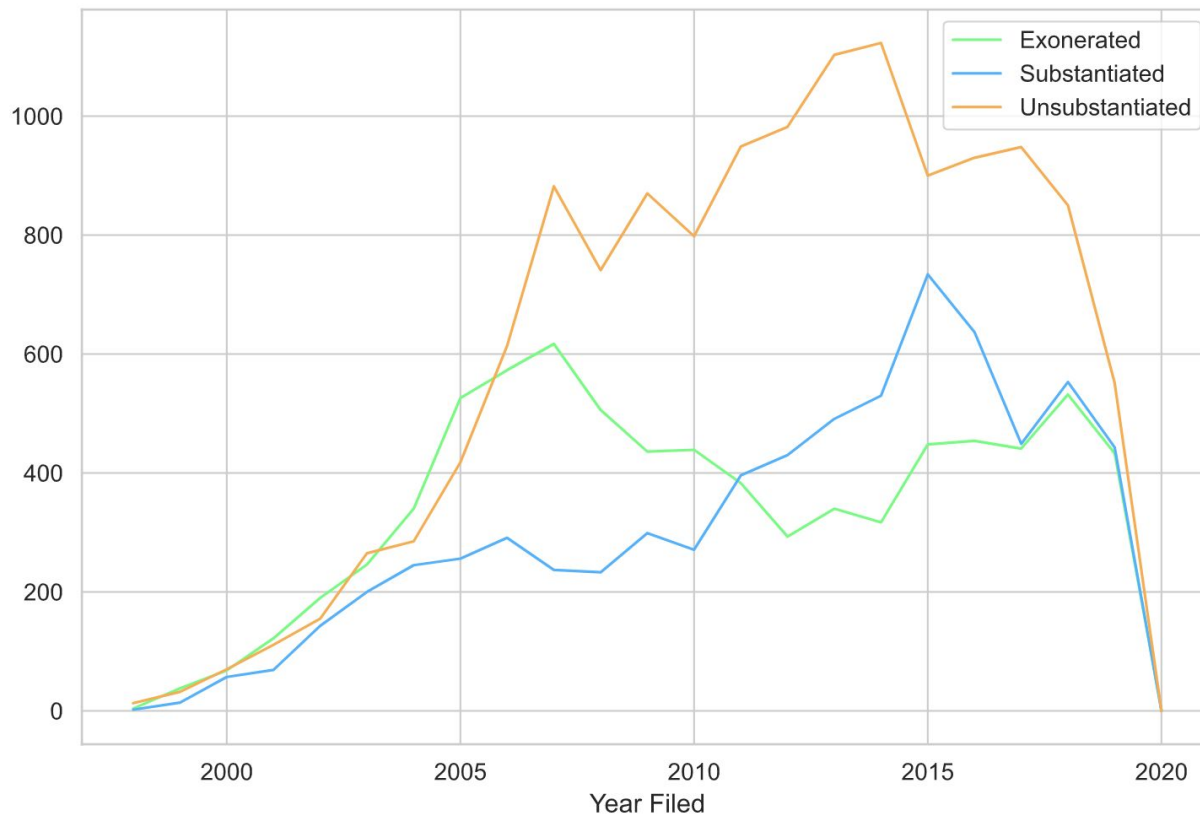
HOW COMPLAINTS FLOW THROUGH THE CCRB



See the interactive version [here](#). Credit to Jarrett Meyer for the template.

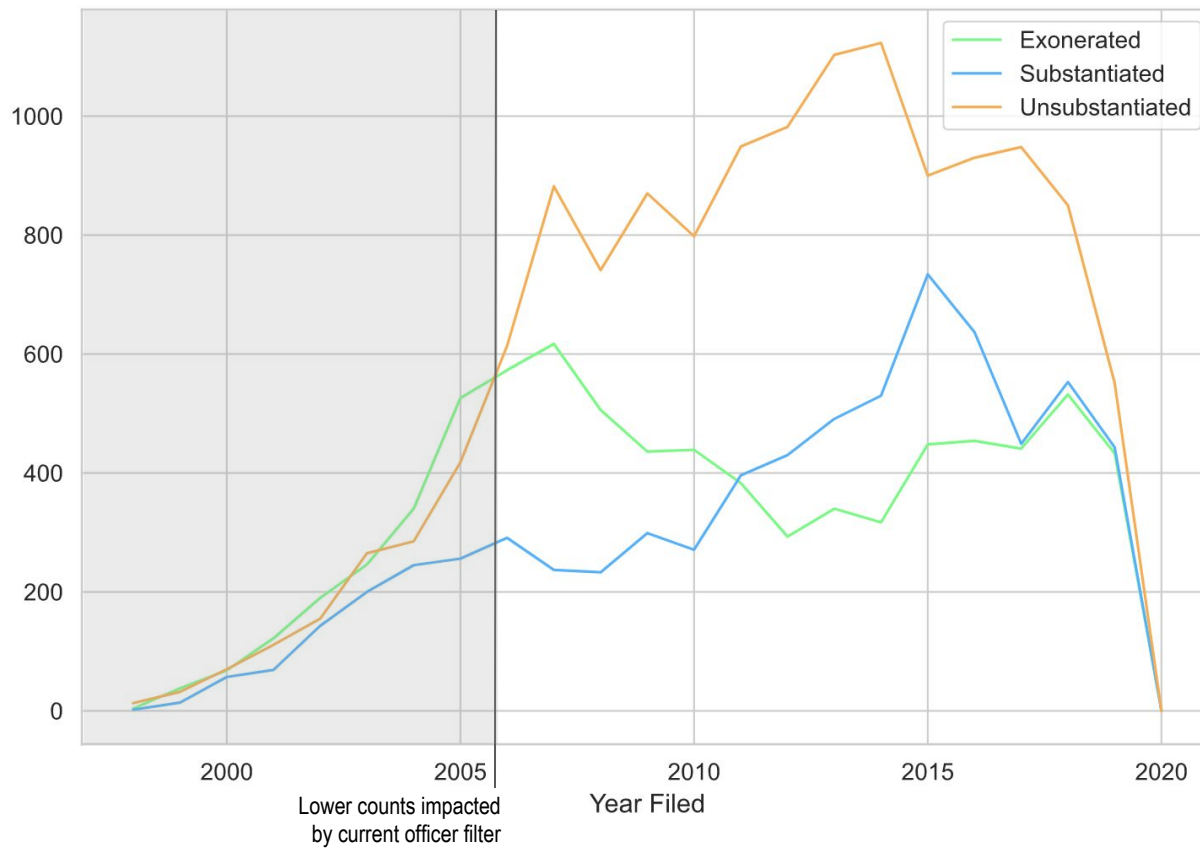
COMPLAINTS OVER TIME

COMPLAINTS BY YEAR AND CCRB RULING



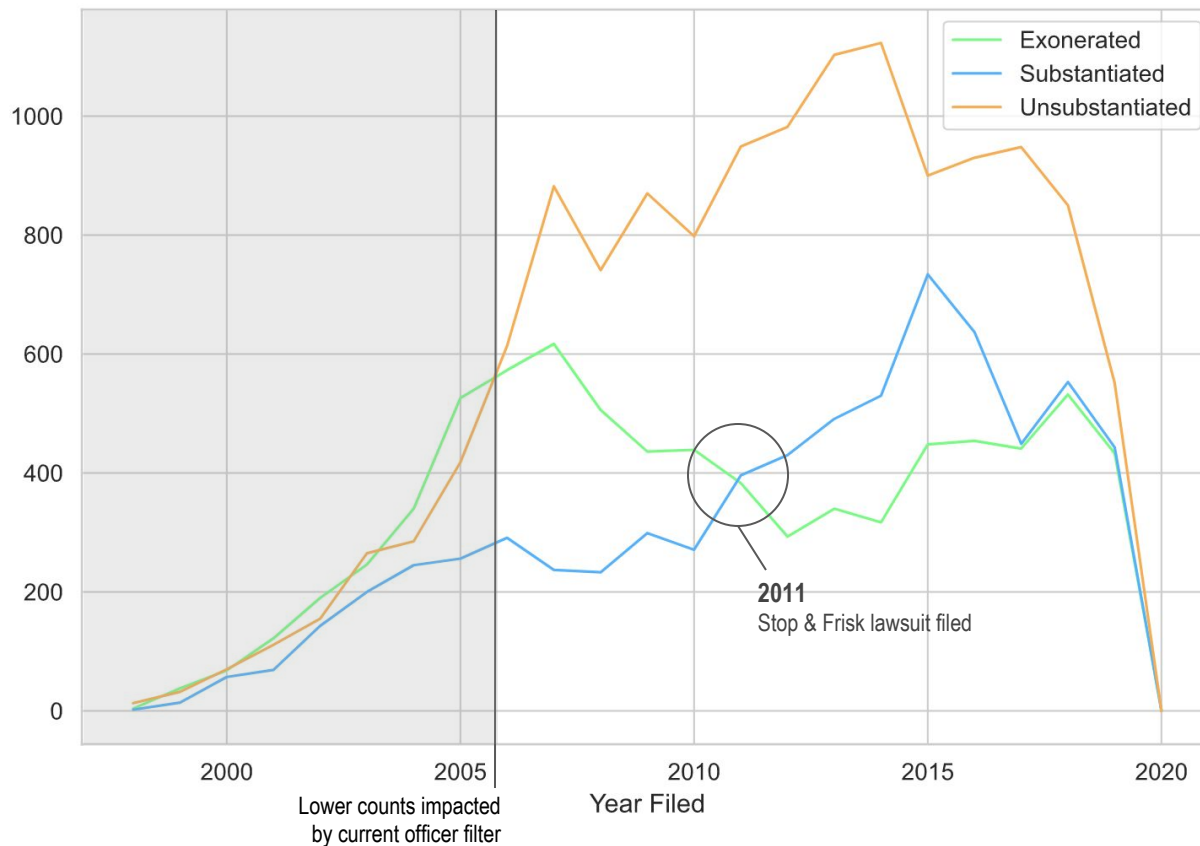
COMPLAINTS OVER TIME

COMPLAINTS BY YEAR AND CCRB RULING



COMPLAINTS OVER TIME

COMPLAINTS BY YEAR AND CCRB RULING



NUMBER OF COMPLAINTS PER OFFICER



2020

Note: shield numbers have been anonymized for the purpose of this presentation.

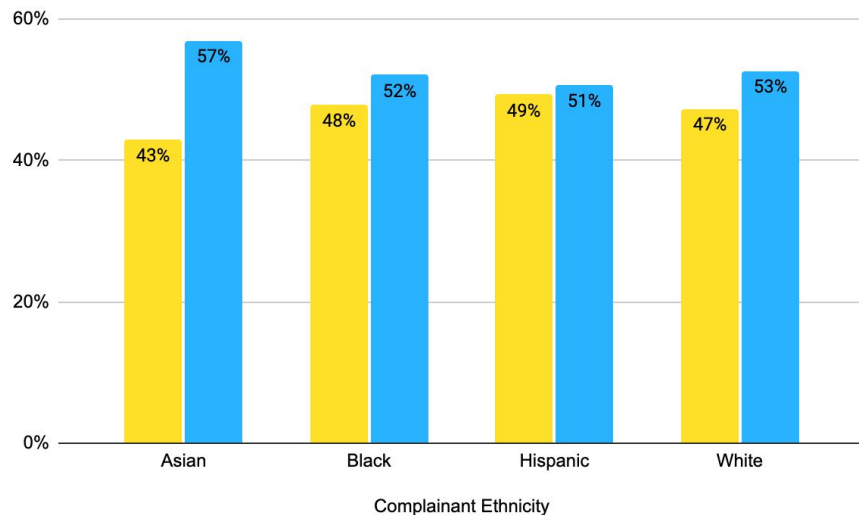
See the interactive version [here](#). Credit to Mike Bostock for the template.

DIFFERENCE IN BOARD RULING ACROSS RACE/ETHNICITY

DID THE CONDUCT OCCUR?

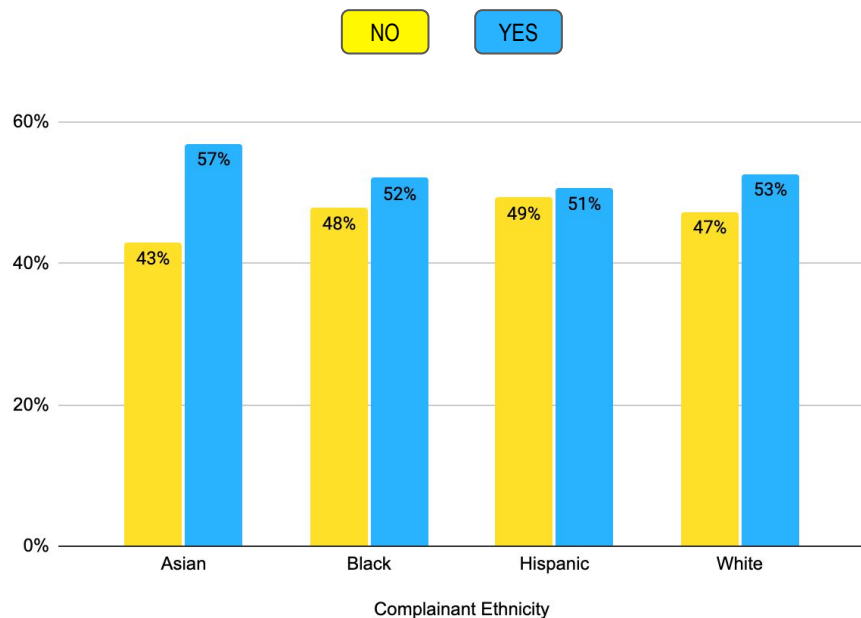
NO

YES

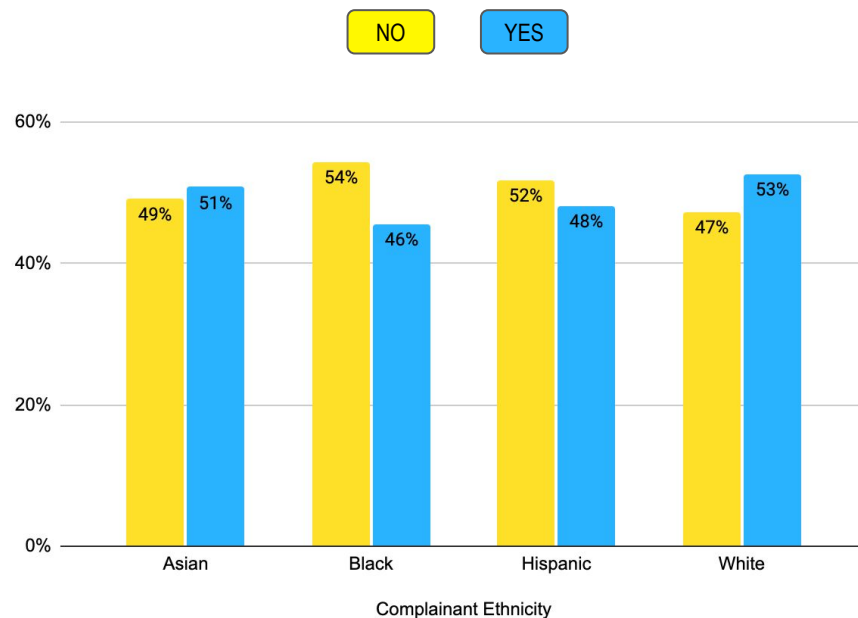


DIFFERENCE IN BOARD RULING ACROSS RACE/ETHNICITY

DID THE CONDUCT OCCUR?

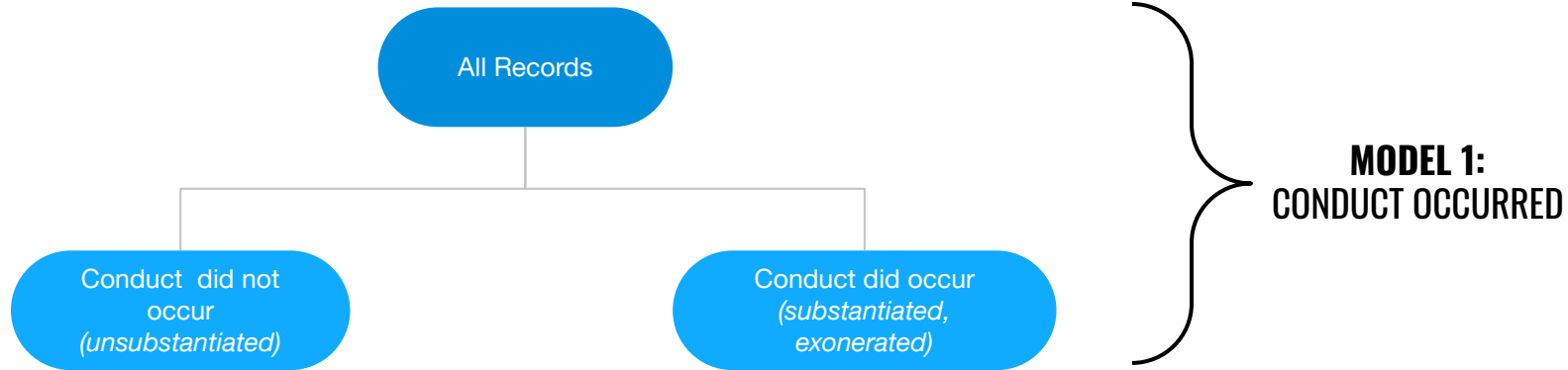


DID THE CONDUCT VIOLATE NYPD RULES?

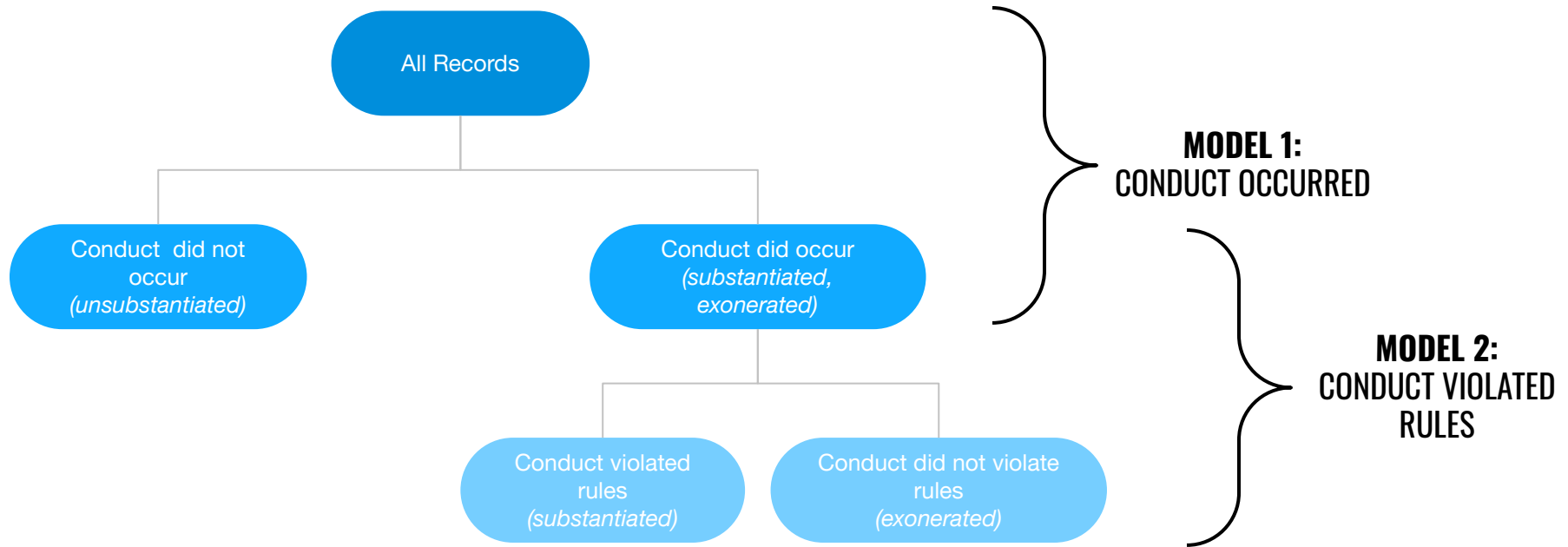


**HOW WELL CAN
CCRB RULINGS
BE PREDICTED?**

A TALE OF TWO MODELS



A TALE OF TWO MODELS

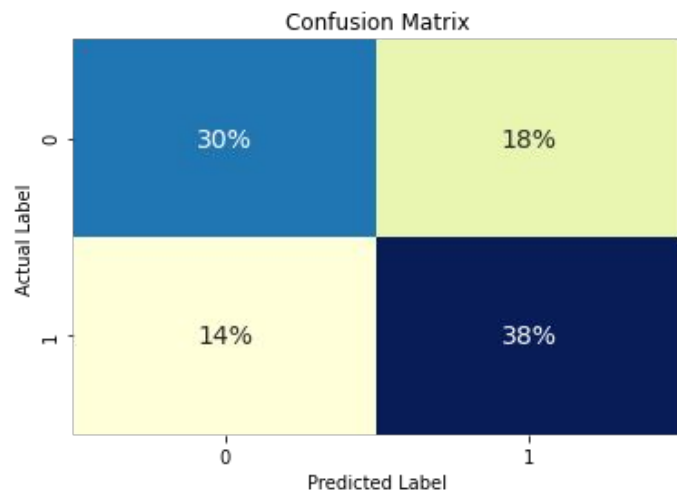


MODEL PERFORMANCE

MODEL 1: CONDUCT OCCURRED

Final model chosen:
Random forest with default configuration

68% Accuracy
on unseen
data

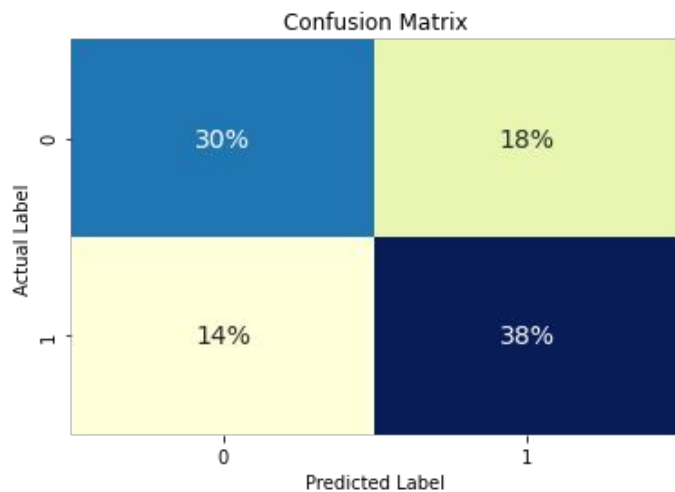


MODEL PERFORMANCE

MODEL 1: CONDUCT OCCURRED

Final model chosen:
Random forest with default configuration

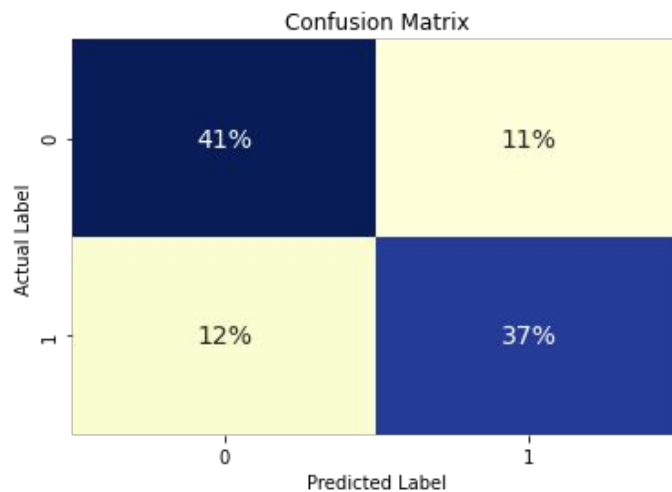
68% Accuracy
on unseen
data



MODEL 2: CONDUCT VIOLATED RULES

Final model chosen:
Random forest with default configuration

78% Accuracy
on unseen
data



**CCRB DECISIONS CANNOT
BE PREDICTED PERFECTLY**

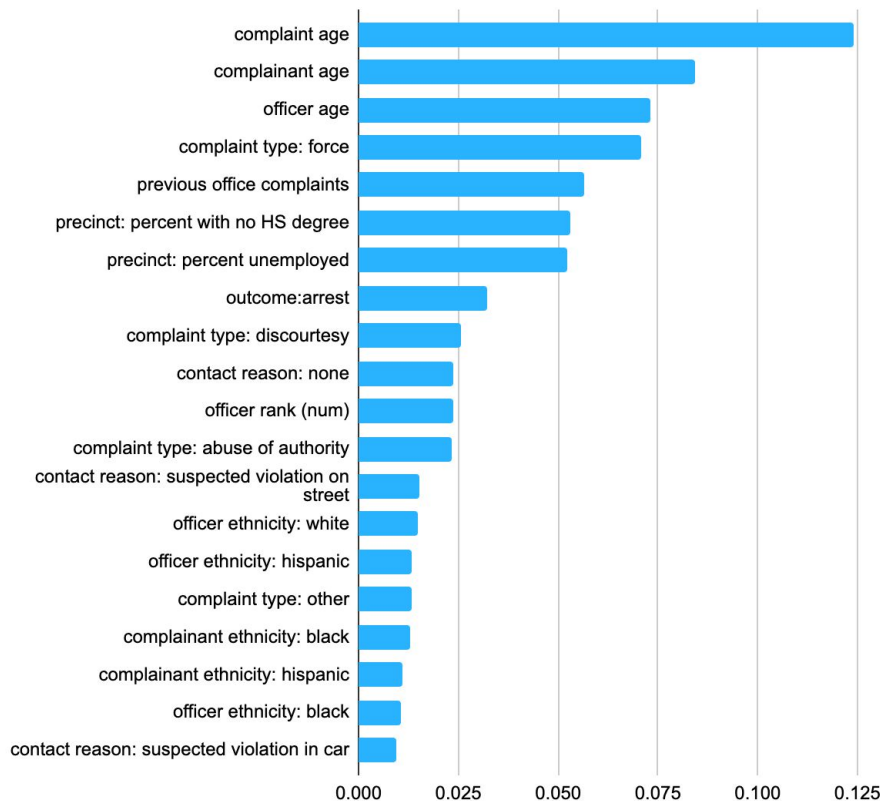
**CCRB DECISIONS CANNOT
BE PREDICTED PERFECTLY**

**... AND THAT'S A GOOD
THING**

FEATURE IMPORTANCE: MODEL 2

RANDOM FOREST RELATIVE FEATURE IMPORTANCE

top 20 most important features

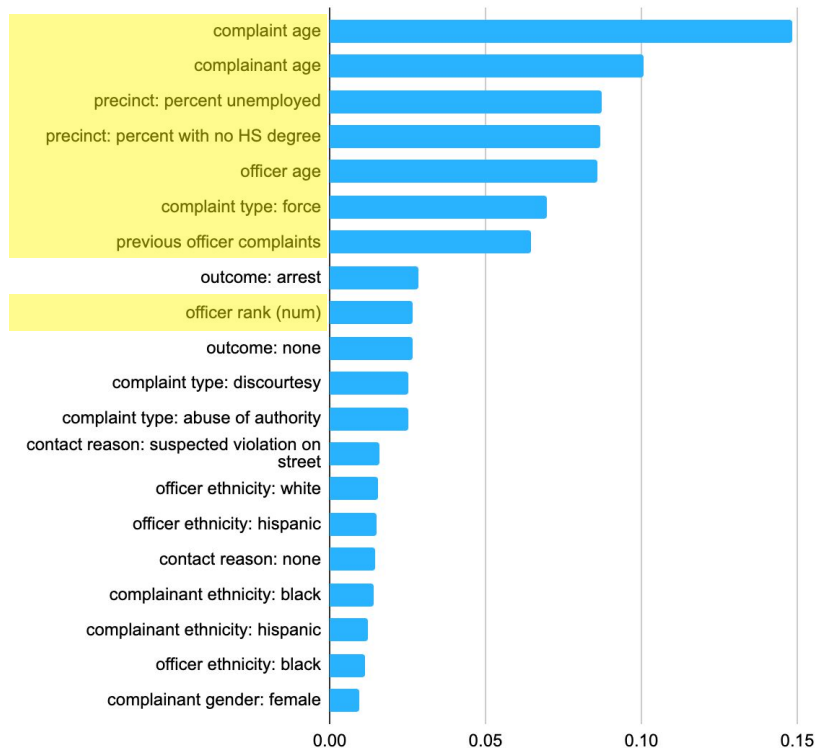


FEATURE IMPORTANCE: MODEL 2

RANDOM FOREST RELATIVE FEATURE IMPORTANCE

top 20 most important features

Numerical features
rose to the top

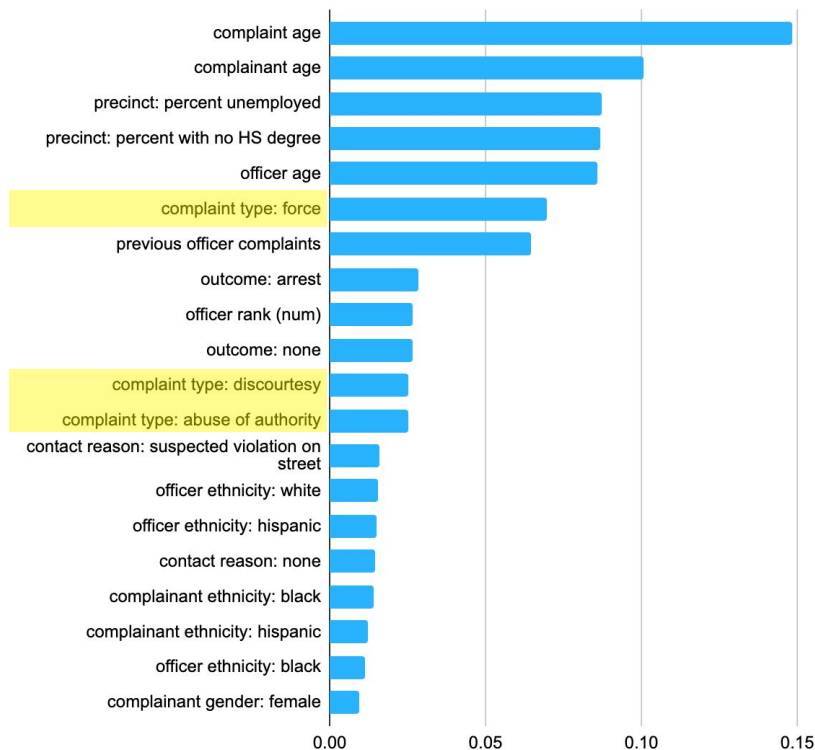


FEATURE IMPORTANCE: MODEL 2

RANDOM FOREST RELATIVE FEATURE IMPORTANCE

top 20 most important features

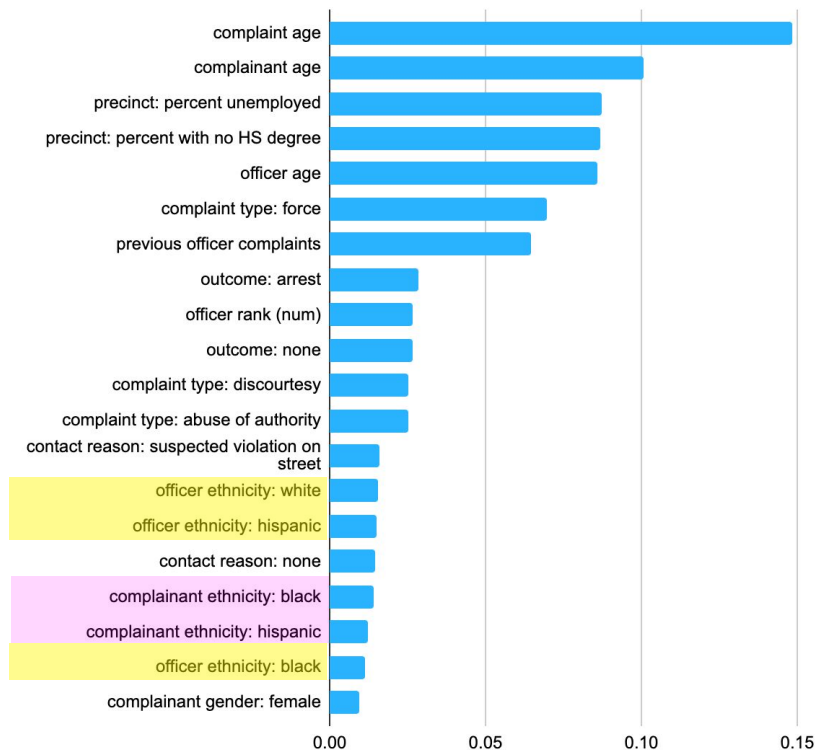
Complaint type
also important



FEATURE IMPORTANCE: MODEL 2

RANDOM FOREST RELATIVE FEATURE IMPORTANCE

top 20 most important features



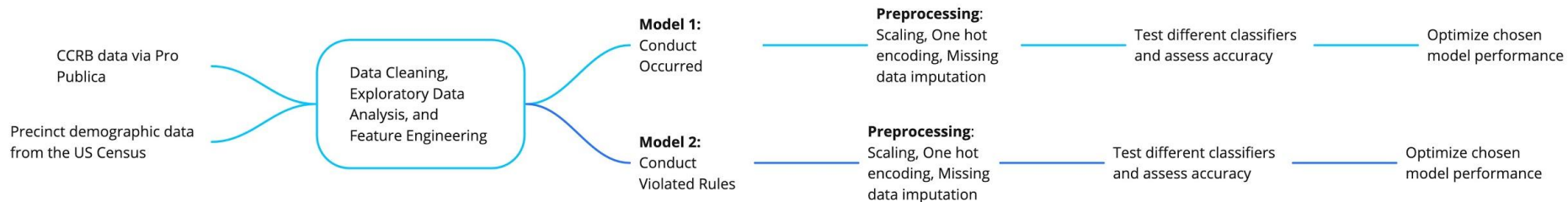
Officer and complainant ethnicity in the top features, but relatively unimportant

FUTURE WORK

- Improve model interpretability for categorical variables
- Add additional precinct-level data to improve model performance
- More nuanced integration of time series data
- Explore integrating the larger NYCLU CCRB database

APPENDIX

MODELING PROCESS



MODEL 1 CLASSIFICATION REPORT

	precision	recall	f1-score	support
0	0.68	0.63	0.66	2718
1	0.68	0.73	0.71	2946
accuracy			0.68	5664
macro avg	0.68	0.68	0.68	5664
weighted avg	0.68	0.68	0.68	5664

MODEL 2 CLASSIFICATION REPORT

	precision	recall	f1-score	support
0	0.78	0.79	0.79	1526
1	0.77	0.76	0.77	1420
accuracy			0.78	2946
macro avg	0.78	0.78	0.78	2946
weighted avg	0.78	0.78	0.78	2946

MODEL 2 IMPORTANT FEATURES HEATMAP

