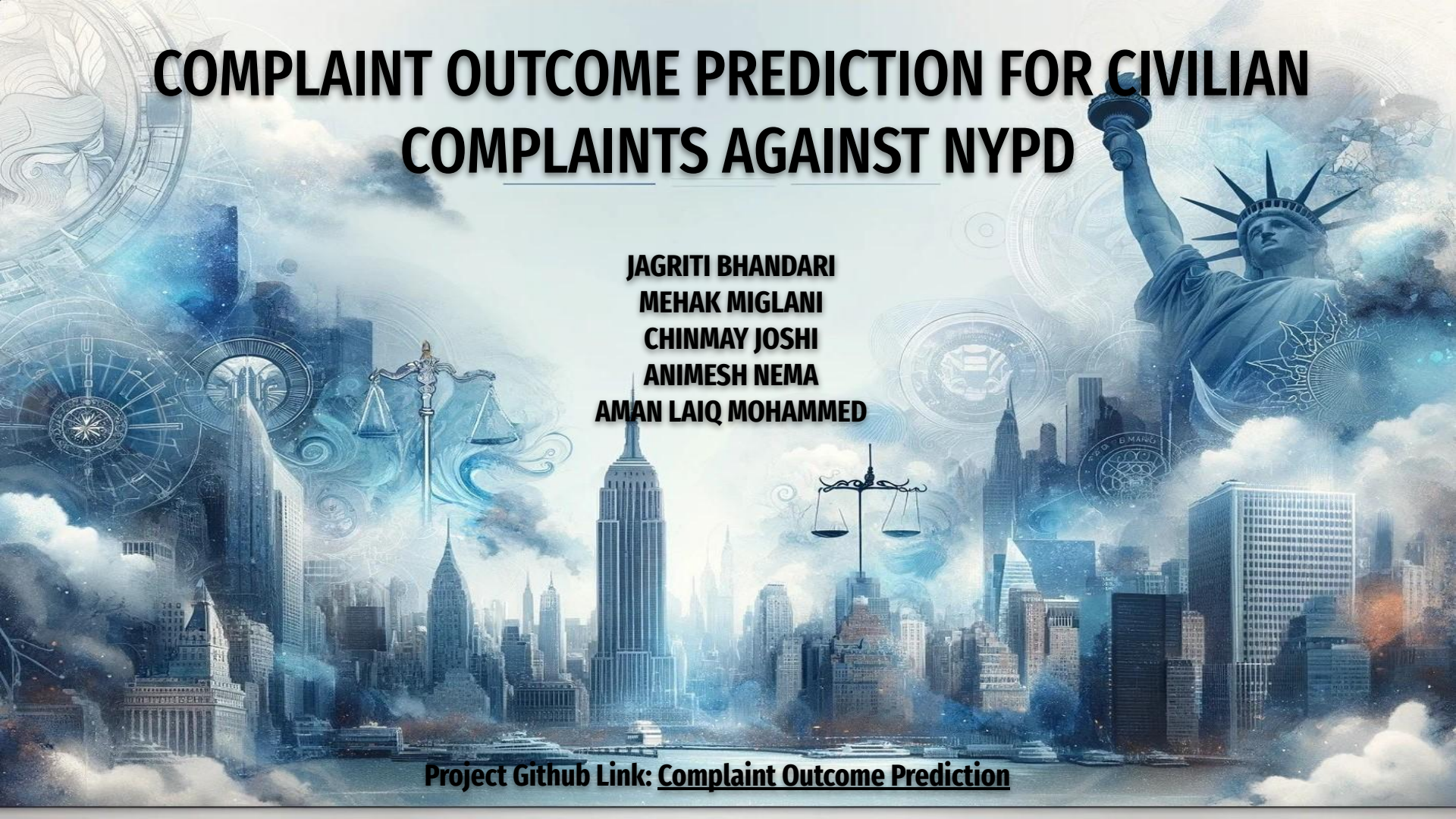
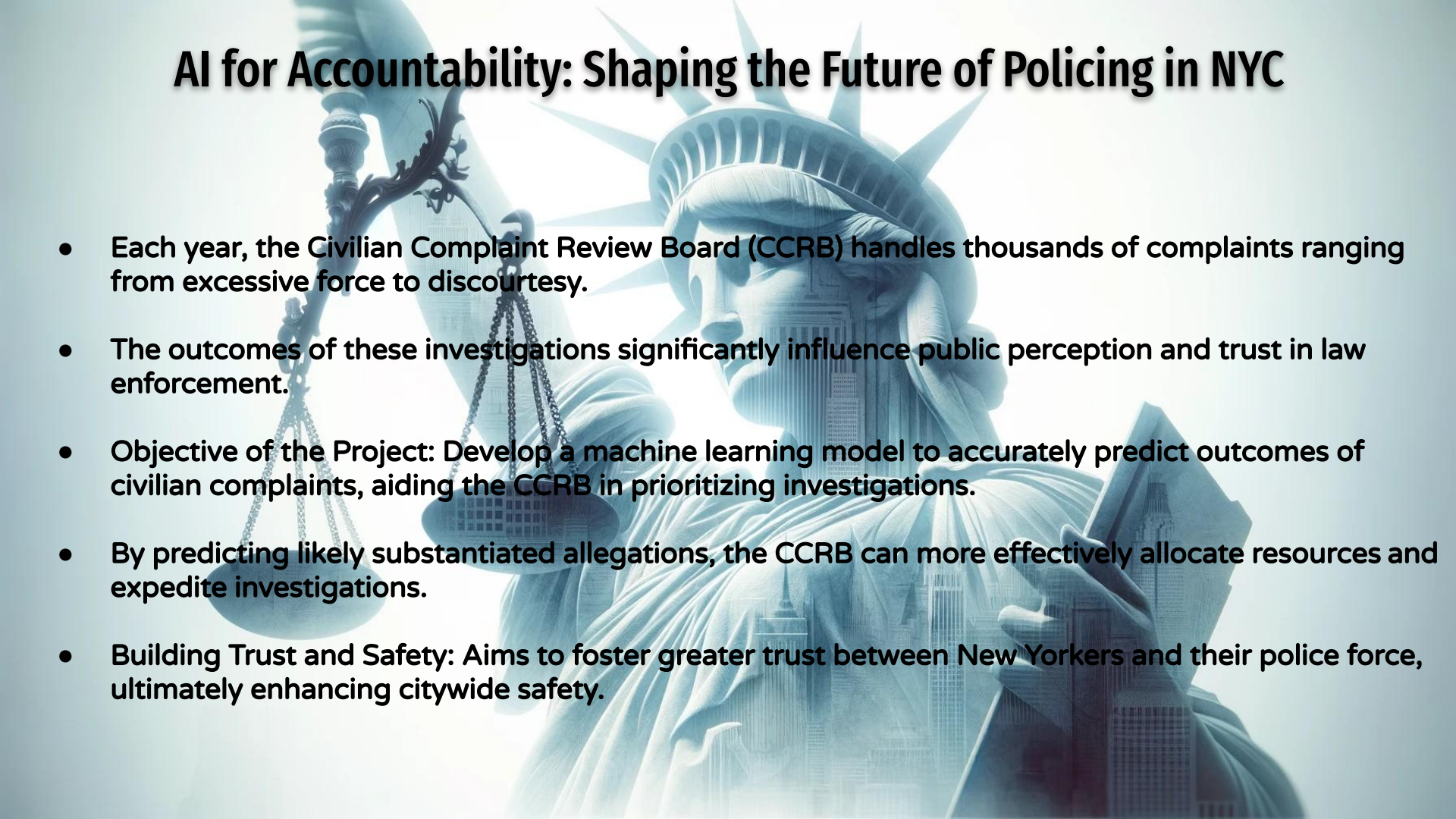
COMPLAINT OUTCOME PREDICTION FOR CIVILIAN COMPLAINTS AGAINST NYPD

JAGRITI BHANDARI
MEHAK MIGLANI
CHINMAY JOSHI
ANIMESH NEMA
AMAN LAIQ MOHAMMED



Project Github Link: Complaint Outcome Prediction

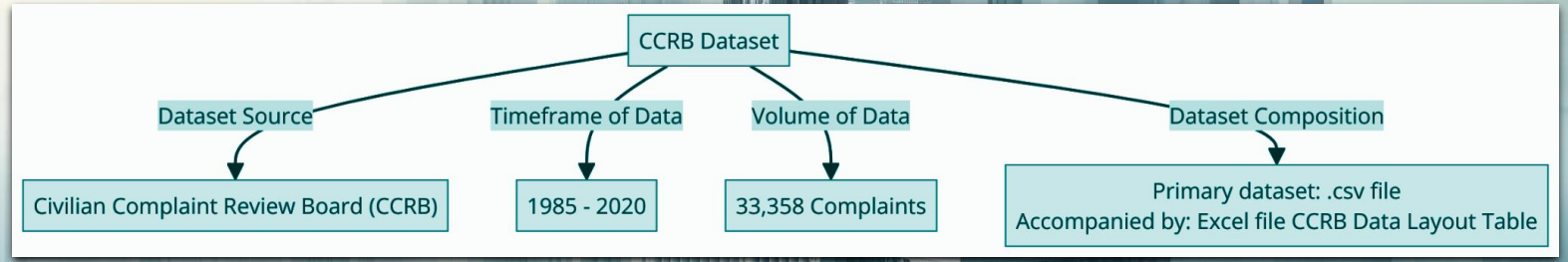
AI for Accountability: Shaping the Future of Policing in NYC

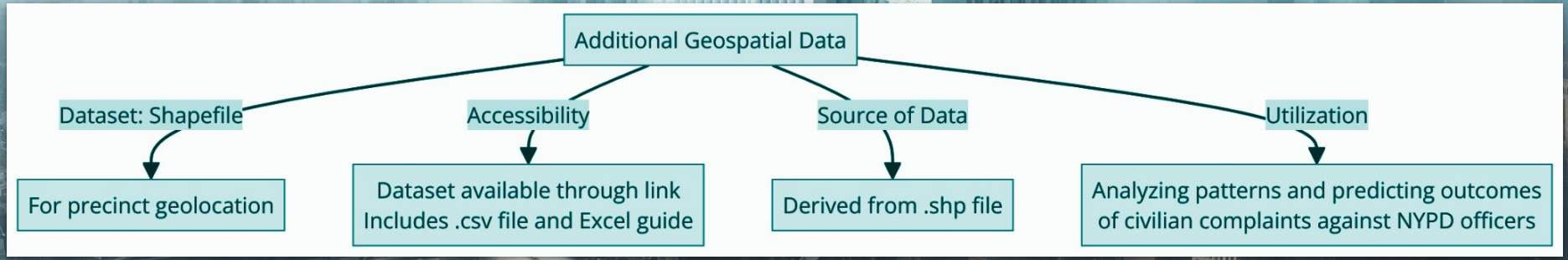


- Each year, the Civilian Complaint Review Board (CCRB) handles thousands of complaints ranging from excessive force to discourtesy.
- The outcomes of these investigations significantly influence public perception and trust in law enforcement.
- Objective of the Project: Develop a machine learning model to accurately predict outcomes of civilian complaints, aiding the CCRB in prioritizing investigations.
- By predicting likely substantiated allegations, the CCRB can more effectively allocate resources and expedite investigations.
- Building Trust and Safety: Aims to foster greater trust between New Yorkers and their police force, ultimately enhancing citywide safety.

Decades of Data: Unveiling the Patterns Behind NYPD Complaints

The data collected has been from two different sources - from CCRB which is the data about the complaints and the geospatial data which is the data about the police precincts' locations.



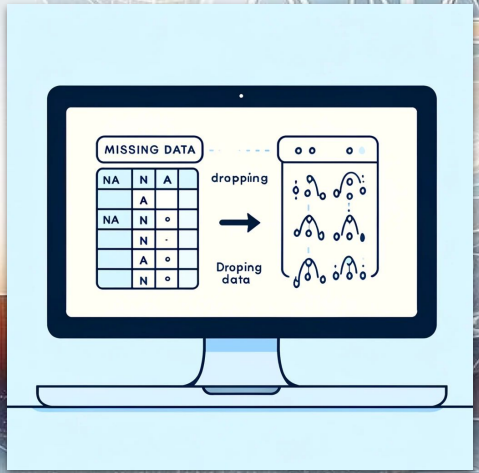


Data Cleaning

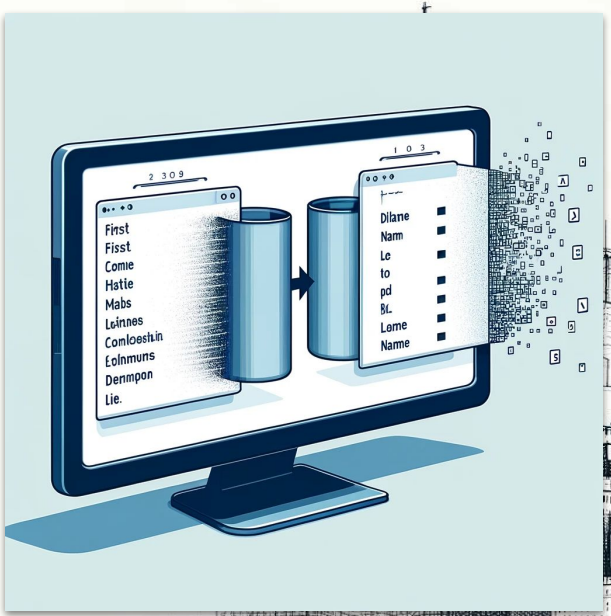
As a part of the data cleaning process, we have performed the following operations on our data:

1. **Dropping columns** - Dropping columns like ‘rank_now’, ‘rank_incident’, and ‘shield_no’.

2. Handling NAs - Certain rows containing null values were dropped and others were filled with operations like mean, mode, and percentile random values



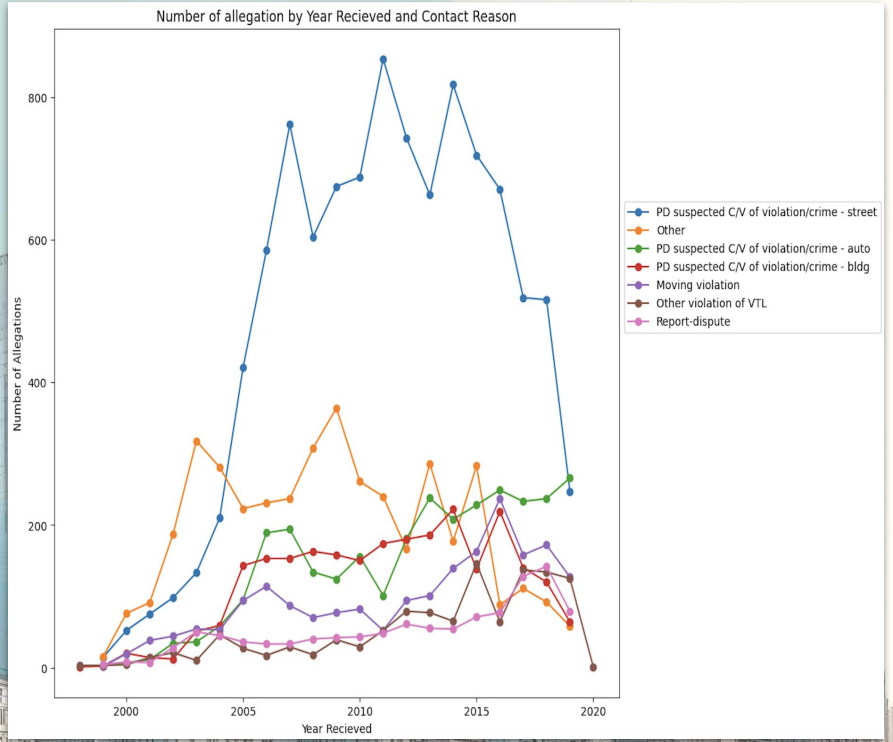
Data Cleaning (contd.)



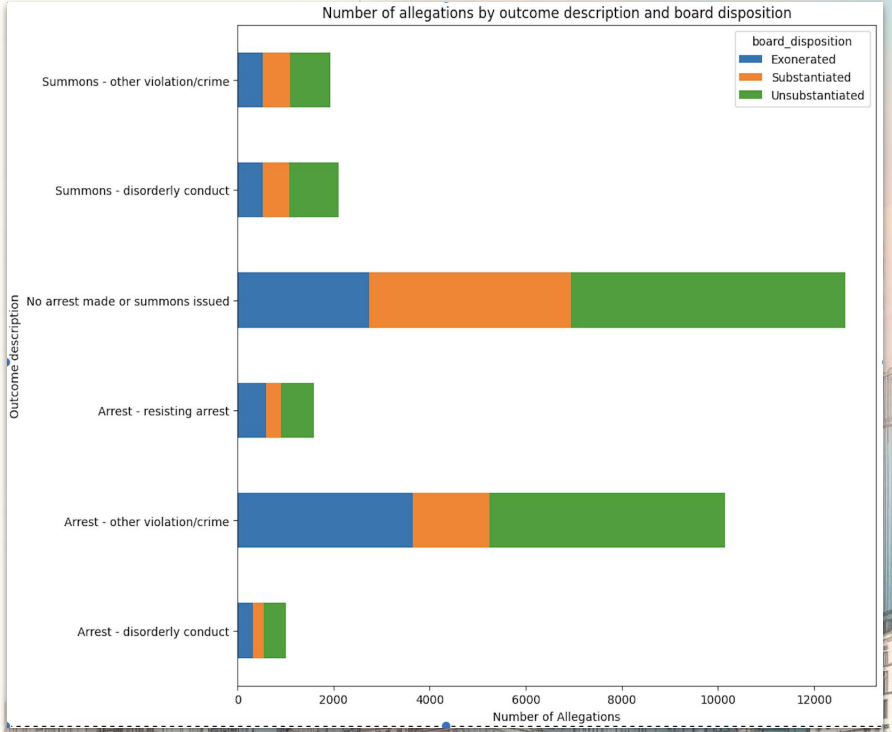
3. Combining columns and dropping the original ones - To simplify and make the data frame easier to read, columns like "first_name" and "last_name" were combined. "Name" is the new column's name.

The original columns were then dropped.

Data Deciphered: Visual Insights into NYPD Complaint Dynamics

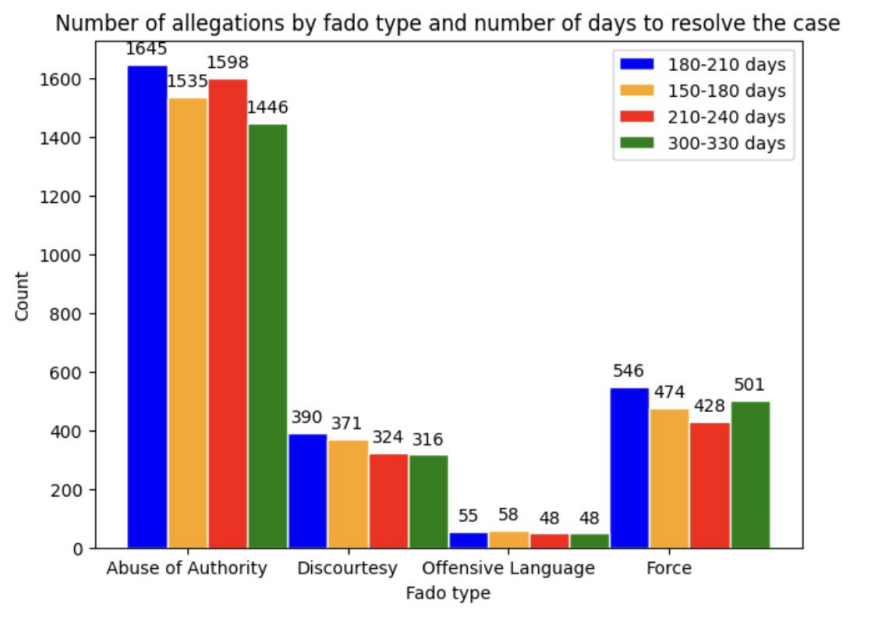


The surge in street-related allegations may reflect changes in policing strategies or community conditions between 2005-2015.

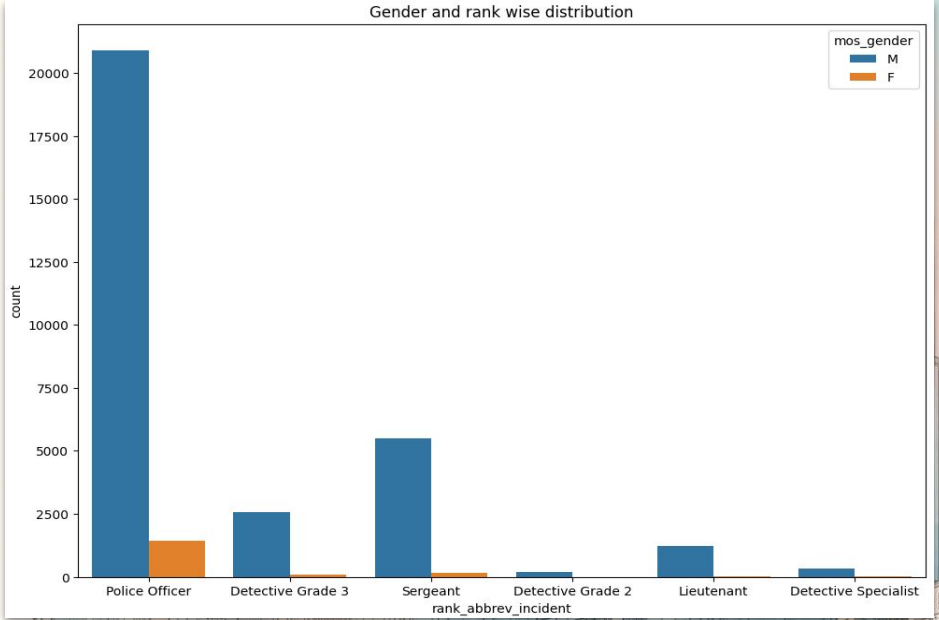


The high rate of allegations resulting in 'No arrest made or summons issued' and the high rate of unsubstantiated claims might suggest issues with the complaint vetting process or public perception of police conduct.

Data Deciphered: Visual Insights into NYPD Complaint Dynamics

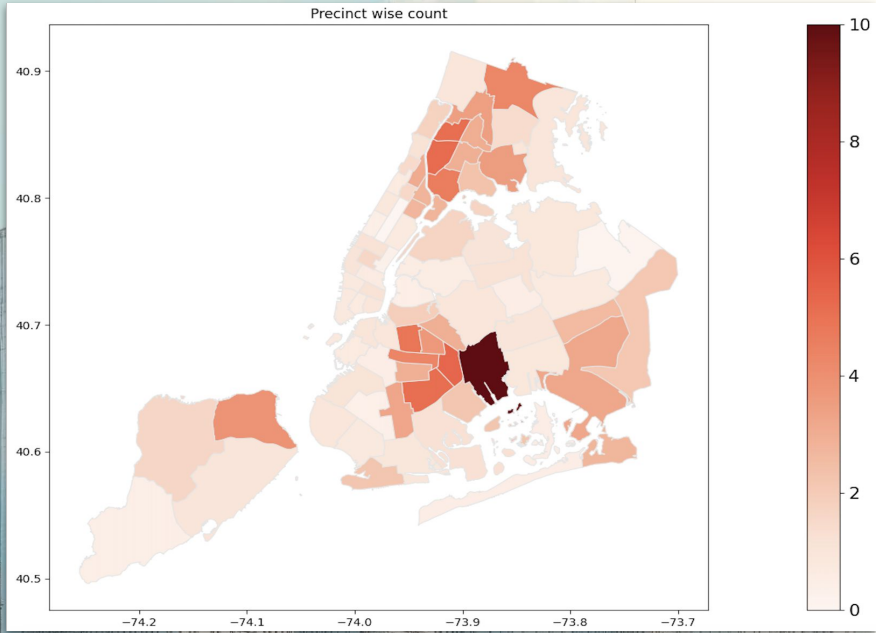


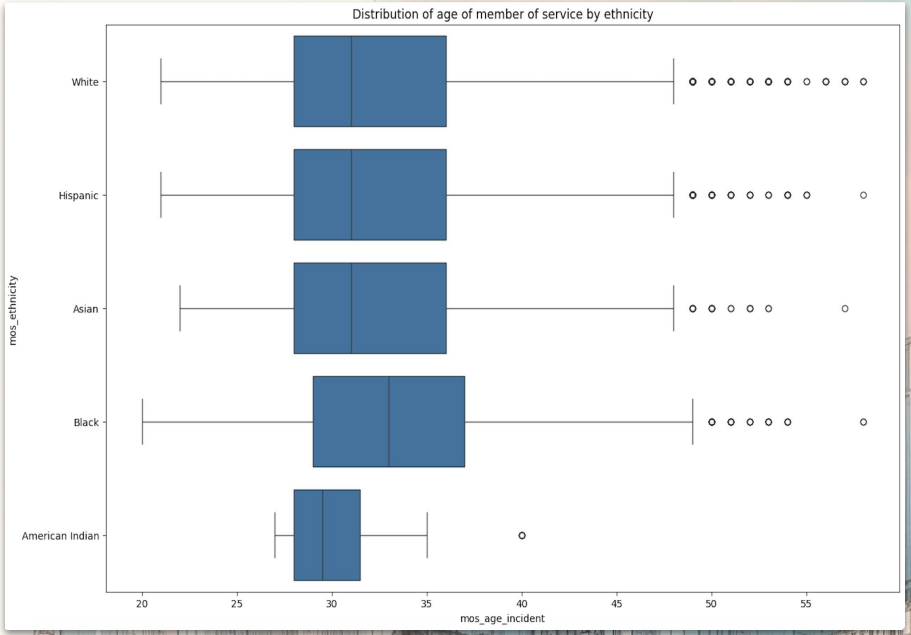
The longer resolution time for abuse of authority complaints might reflect their complexity and seriousness, pointing to potential inefficiencies in the review process that could benefit from streamlined procedures or additional resources.



The disproportionate number of complaints against male officers across all ranks suggests possible gender differences in policing style, behavior, or the nature of assignments, which could be vital for training and operational strategies.

Analyses on Age distribution and other factors





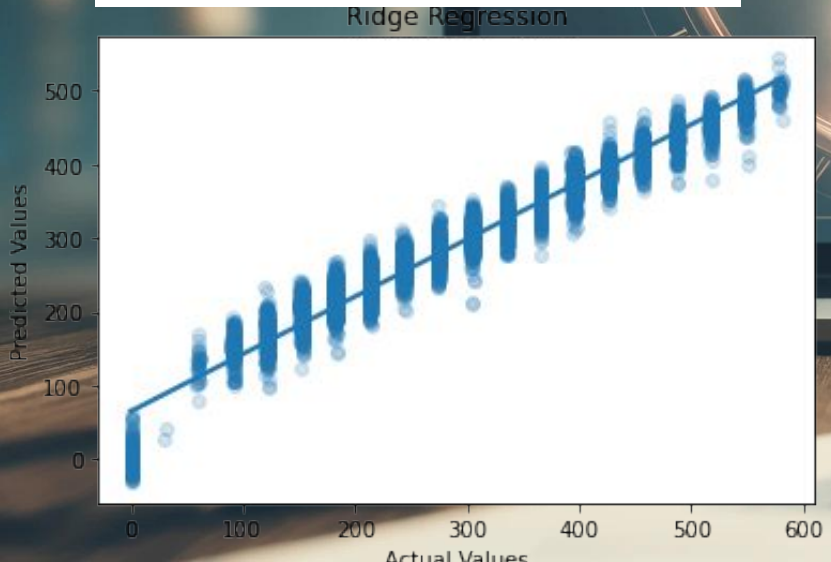
The concentration of complaints in East New York and broader Brooklyn could point to local factors such as socio-economic conditions, policing practices, or community-police relations that necessitate focused policy interventions.

The observed age range(27-37) and distribution among officers facing charges highlight the role of experience and age in incidents leading to allegations. The specific age demographics officers could reflect recruitment patterns or career progression within the force.

Ridge Regression Model: "Streamlining Complaint Resolution: Insights from Ridge Regression"

Feature Selection: Analyzed temporal and case-specific data such as complaint dates and officer details.
Training: Utilized an 80/20 training/test data split, with the alpha parameter set to 0.5 to manage model complexity and prevent overfitting.

Performance: Achieved a 92% accuracy rate, with mean square error of 1426.51 for training and 1511.41 for testing data, indicating strong predictive reliability.



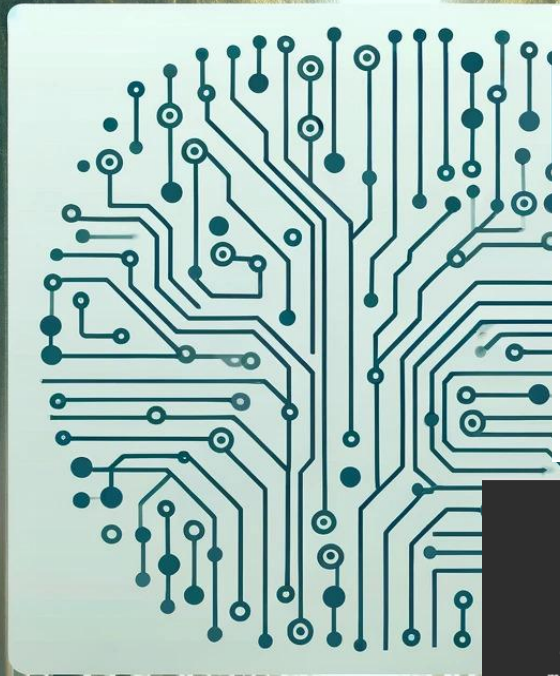
Enhancing Model Accuracy and Fairness in Complaint Resolution:

- Utilize model insights to identify and address biases, enhancing transparency and fairness in the complaint handling process.
- Provide complainants with accurate resolution timelines to set realistic expectations.
- Establish a feedback system to continuously improve the model's performance and utility.

Random Forest Classifier Model: "Decoding Officer Profiles: Efficacy of Random Forest Classification"

Feature Selection: Focused on officer demographics and complaint characteristics to predict the gender of the officer.

Training: Implemented oversampling to address class imbalance, particularly for improving the model's prediction of female officer involvement.



Data-Driven Fairness in Policing:

- Utilize analytics to identify and rectify gender disparities in law enforcement.
- Inform equitable recruitment strategies and foster inclusive policies.
- Enhance training programs to prevent offenses linked to gender bias.
- Promote transparency and accountability through predictive gender analysis.
- Expand analysis to include a wider range of demographic variables for systemic reform.
- Implement a feedback system for continuous improvement of the tool.

Performance: Reached an accuracy of 97%, with high precision in gender classification. Recognized the need for more balanced data to improve the recall for female officers.

	precision	recall	f1-score	support
F	0.94	0.38	0.54	341
M	0.97	1.00	0.98	6084
accuracy			0.97	6425
macro avg	0.95	0.69	0.76	6425
weighted avg	0.97	0.97	0.96	6425

```
0.9659143968871595
[[ 130  211]
 [   8 6076]]
```

Conclusions and Key Findings

Model Effectiveness and Practical Applications:

- **Ridge Regression Model:** Accurately estimated the resolution times of civilian complaints, demonstrating the capability to streamline case handling by prioritizing those likely to require extended investigation.
- **Random Forest Classifier:** Efficiently predicted the gender of officers involved in complaints, with a high accuracy rate. However, revealed challenges in identifying female officers, indicating a need for more gender-balanced data in training sets.

Data Insights and Impact on Policies:

- Both models uncovered significant patterns in complaint resolutions and officer behaviors, providing data-driven insights that could inform policy revisions and training programs focused on reducing complaint incidences and improving response times.
- Identified specific factors, such as officer rank and complaint type, that influence the duration and outcomes of complaint investigations. These insights can be used to target areas where policy and training adjustments are most needed.

Recommendations for Further Research and Development:

- Expand data collection to include more comprehensive demographic details of both officers and complainants to refine predictions and identify subtle biases.
- Continue refining models by incorporating feedback loops from model predictions to real-world outcomes to enhance accuracy and reliability.

Any Questions?


THANK YOU