



Predictive Modeling of Intimate Partner Violence Using NYC Police Data

Ash Todd, Eaint May, Nora Phelan, Zack Taylor

Introduction

Background: What Is
Intimate Partner Violence?

- Abuse appears in a variety of forms including: physical, emotional, economic, psychological, sexual, and more.
- IPV is a subcategory of Domestic violence
- “[P]atterns of coercive behavior that influence another person within an intimate partner relationship” (OVW)

Motivation: Why did we
study IPV?

- Difficult to predict due to complex psychological and social factors.
- historical inequalities shaped by gender, class, and systemic neglect.
- Many cases go unreported Inadequate institutional responses Fear of retaliation or disbelief

Research Questions (Intro cont.)

Primary Question:

How can police data be used to develop predictive modeling for combating Intimate Partner Violence and what are the potential biases and ethical concerns that can arise in such use?

Sub Questions:

- How does bias shape the data and affect predictive accuracy?
- What gets lost or distorted when IPV is viewed through predictive lens?
- Who bears the risk when predictive tools are used in already-vulnerable communities?

Data

- Publicly available domestic violence reports from the NYC Police Department from 2020-2021

Key variables:

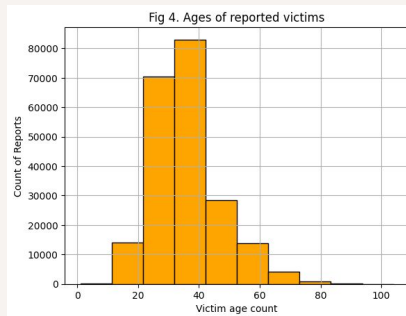
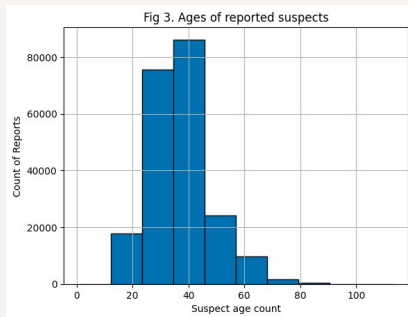
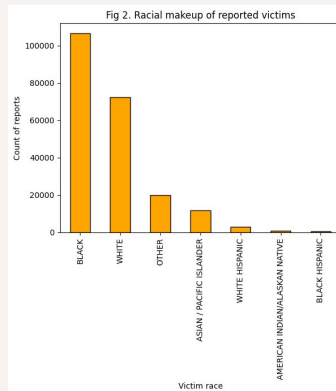
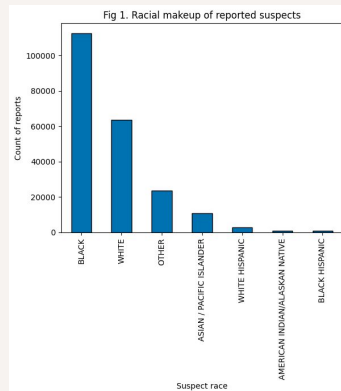
- Offense Type: Domestic Incident Report (DIR), felony rape, or felony assault
- Victim and Suspect Information: Age, sex, race
- Location: Precinct code, borough

Data cleaning:

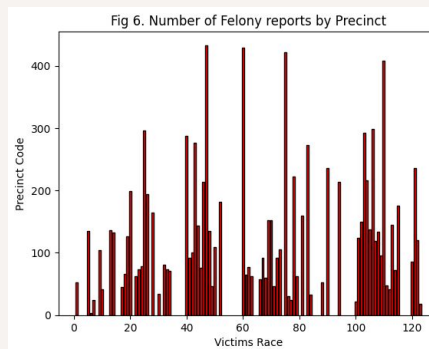
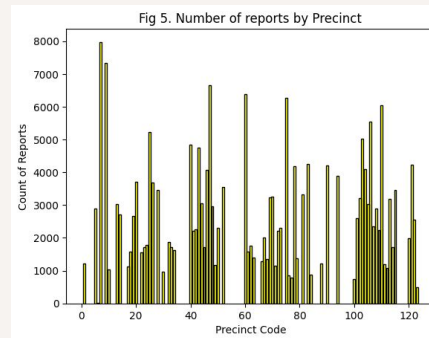
- Missing Data: Missing values in financial indicators and ages were replaced with column averages.
- Data Types: Age and sex columns were converted to integers; report date converted to datetime.
- Felony Classification: A new variable, *Felony Offense*, was created to classify offenses as felonies (rape, felony assault) or non-felonies (DIR).

Highlights from EDA *Data Exploration*

Suspect and Victim Analysis



Reports by Precinct



Methods

- Types of Predictive Policing:
 - Person-based - Analyzes datasets and generates a list of individuals likely to commit a crime
 - Place-based - Analyzes datasets to predict where crimes are likely to occur

Our Three Methods:

- Place-based -
 - Found the ten precincts with the highest number of DIRs in a given month
 - Found the ten precincts with the highest number of felonies in a given month
 - Compared the lists to identify precincts commonly appearing in both

Methods Cont.

- Racial Makeup –
 - Created chart for racial makeup of suspects in each precinct
 - Created similar chart showcasing racial makeup of victims in each precinct
- Socioeconomic Standing –
 - Calculated the percent of community districts in each borough that are economically disadvantaged
 - Calculated the percent of reports in each borough that occurred in these community districts

Results - Model

Placed-Based Model: Feb - April 2020

Precinct	Reports	High Felonies
46	293	True
75	252	True
40	246	True
43	237	True
47	235	True
73	225	False
44	203	True
48	186	False
52	185	True
67	183	True

Table 1. Precincts with Highest Reports in February 2020

Precinct	Reports	High Felonies
46	324	True
75	285	False
73	283	False
43	259	True
47	238	True
40	227	True
42	225	True
44	203	True
52	201	True
67	190	False

Table 2. Precincts with Highest Reports in March 2020

Precinct	Reports	High Felonies
75	261	False
46	231	True
73	215	True
43	213	True
40	206	True
42	205	False
44	201	True
48	185	False
47	182	False
52	158	False

Table 3. Precincts with Highest Reports in April 2020

Results - Model Cont.

Placed-Based Model: Feb - April 2021

Precinct	Reports	High Felonies
75	272	True
46	271	True
43	218	False
73	210	Flase
67	210	True
47	207	True
44	199	True
42	189	True
40	181	False
52	178	False

Table 4. Precincts with Highest Reports in Feb 2021

Precinct	Reports	High Felonies
75	299	False
46	294	True
73	283	True
43	240	True
47	234	True
44	229	True
42	228	True
40	218	True
67	189	False
52	187	False

Table 5. Precincts with Highest Reports in March 2021

Precinct	Reports	High Felonies
46	333	False
75	297	True
73	264	True
43	249	True
40	237	True
67	196	False
47	196	True
42	192	False
52	190	True
105	188	False

Table 6. Precincts with Highest Reports in April 2021

Results - Racial

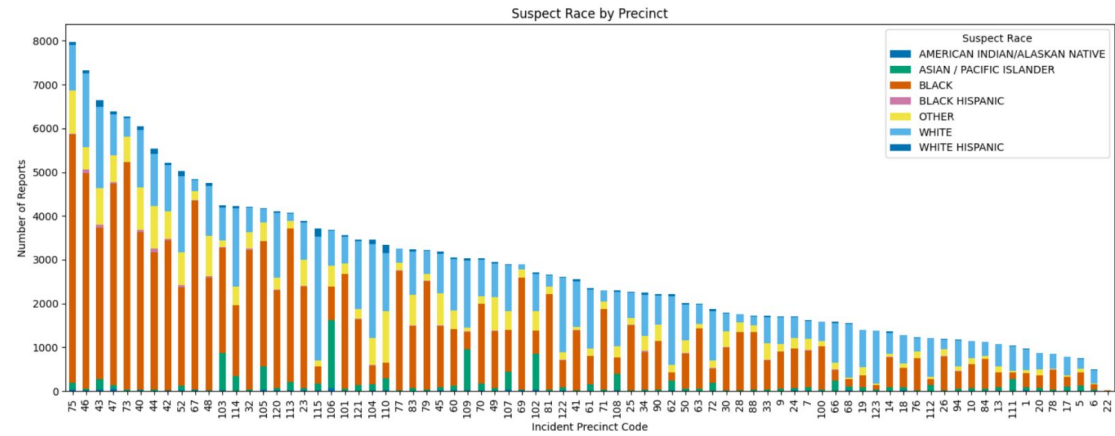


Fig. 1. Bar graph of racial makeup of suspects

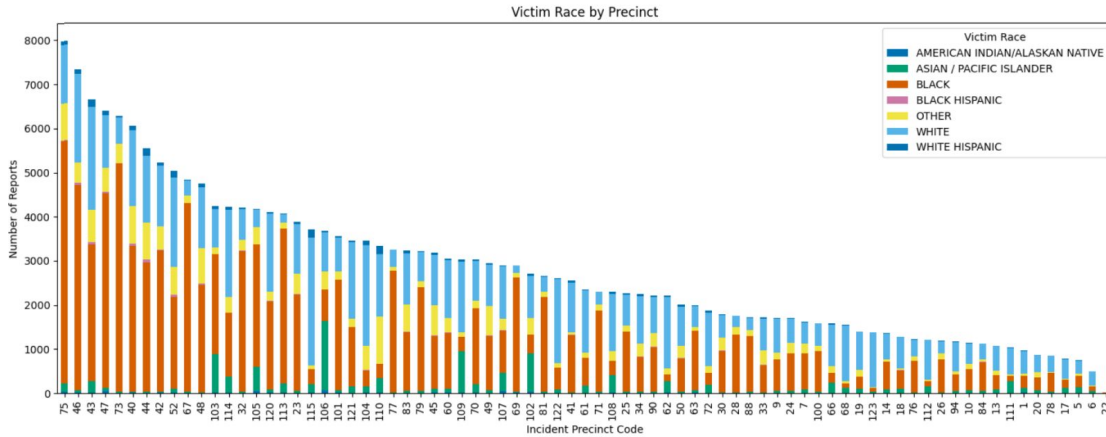


Fig. 2. Bar graph of racial makeup of victims

Results - Socioeconomic

	Poverty	Unemployment	Median Income
Bronx	74.7846	90.9770	74.7846
Manhattan	46.6064	47.0966	46.6064
Brooklyn	39.7933	9.9839	44.6663
Queens	0.0629	16.7266	0.0629
Staten Island	0	0	0

Table 7. Percent of Reports Occurring in Economically Disadvantaged Community Districts

	Poverty	Unemployment	Median Income
Bronx	53.3333	66.6667	53.3333
Brooklyn	25.0000	5.0000	30.0000
Manhattan	23.0769	30.7692	23.0769
Queens	0	5.5556	0
Staten Island	0	0	0

Table 8. Percentage of Community Districts at an Economic Disadvantage

Discussion

Key Takeaways

- Consistently high report numbers from specific precincts -> disproportionate targeting
 - Brooklyn and the Bronx
- High number of reports from economically disadvantaged districts may indicate systemic bias due to a higher police presence
- Systemic racial and economic inequalities -> unethical predictive policing model
 - Police data alone is not enough

Limitations

- Racial disparities in policing
- Absence of population data
 - Community districts, jurisdiction of precincts