Repo:

Particular Feedback:

- Cohesion
- Analysis Making Sense

Introduction

- Does the introduction provide appropriate background context for a general reader?

    - Yes. The introduction offers clear and accessible background on domestic violence, the complexities of its investigation, and how policing intersects with it. It successfully anticipates a general reader's needs by defining key terms like IPV and contextualizing the social and legal challenges surrounding domestic violence reports.

- Does the introduction provide a clear motivation for the project?

    - Yes, this project provides clear motivation as to why this specific topic was chosen.

        - As a result, responses to domestic violence are often lackluster and do not do enough to prevent the issue from recurring. In many cases, this failure to produce concrete assistance incentivises other victims not to report violence, and therefore makes fixing the broader issues even tougher.
        - This project aims to decipher the ways in which police data may be used to allocate resources to combat intimate partner violence.
    - The project is motivated by a desire to improve predictive policing by taking a place-based approach and to critique how bias may shape outcomes.

- Are there at least some relevant citations included?

    - Yes, within the "Keywords" and towards the end of the introduction. The introduction references both academic and institutional sources, helping ground the work in broader literature and social policy context.

- Are the research question(s) clearly stated?

    - Yes, the research questions are clearly states towards the end of the introduction

- ■ What does it mean to make violence "predictable" using data that itself is shaped by uneven policing and systemic bias?
- ■ And who bears the risk when predictive tools are used in already-vulnerable communities?
- Do the research question(s) follow from the motivation?

  - The research questions somewhat follow from the motivation. The motivation, as it seems, is that domestic violence responses aren't documented, reported, or given sufficient investigation, therefore making it harder to fix the issue overall. The research questions then go in a direction of how current police efforts to document DV then "predict" later offenses might prove to be difficult, with the possible bias existing in the system.

- Are the expectations or hypotheses for the research questions clearly stated?

  - No explicit hypotheses are provided. However, there is an implicit expectation that predictive models based on biased data may reinforce racial and economic disparities.

- Is the goal of the report clearly stated?

  - Yes. This project aims to decipher the ways in which police data may be used to allocate resources to combat intimate partner violence.

Data

- Is the source of the data stated with an appropriate citation?

  - Yes. The dataset comes from the NYPD and is publicly available. However, more background on how the NYPD collects and structures this data (e.g., limitations of DIRs or the reporting process) would strengthen transparency.

- Is it clear when and how the data was collected?

  - Yes. The dataset includes data from 2020–2021, and the explanation of Domestic Incident Reports (DIRs) is thorough and well-placed.

- Is data preparation described clearly (missing data, creation of new variables, etc)?

  - The description of their data cleaning and preparation was clear and easy to understand. The cleaning process is well documented: missing data was handled with imputation (e.g., mean values for age), NaNs were replaced with binary flags, and datatypes were appropriately converted. A new variable for felony

classification was created, showing thoughtful feature engineering.

- Does the data look clean and tidy?

    - Yes the data is easy to read, however, it was not on their preliminary pdf. They handled missing data well and transformed data types.

- Are the cases and relevant variables described?

    - Yes. The report provides a clear summary of each major variable, including derived variables like "Felony Offense," and ties them to research goals.

Methods

- Do the visualizations correspond to the stated research question?

    - Yes. They focus on spatial patterns, race distribution across precincts, and economic hardship vs. report rates.

- Are visualizations effective and do they follow clear visualization principles (including elements like titles, labels, appropriate for the type of data, etc)?

    - Partially. While the visuals are conceptually sound, the graphs in the PDF suffer from small fonts and lack clear axis labels or legends in some cases (especially Figures 5 and 6). In the notebook, visualizations would benefit from larger font sizes, clearer titles, and more accessible color schemes.

- Are the analytic methods clearly described?

    - Yes. They use a precinct-based comparison of felony rates, socioeconomic statistics, and racial demographics.

- Is the choice of analytic method/approach justified? Are the methods appropriate for the research questions?

    - Yes. They justify the use of place-based models by referencing existing predictive policing frameworks. Their decision to avoid person-based models helps reduce risk of individual-level bias and aligns with their ethical critique.

Results

- Are the chosen techniques for answering the research question appropriate for the research context and type of data?

    - Yes. Their analysis methods (monthly precinct comparisons, racial bar charts, economic overlay) are well-suited to the questions.

- Is the research question answered effectively?

  - Yes. Results show consistent overrepresentation of certain precincts (primarily in Brooklyn and the Bronx) across multiple months, along with strong correlations between report frequency and economic disadvantage.

Discussion

- Is the answer to the research question summarized and supported by evidence?

  - Yes. The discussion synthesizes the results effectively and backs up their points with figures and data tables.

- Are limitations of the analysis clearly outlined?

  - Yes. They note missing population data, reliance on government-collected reports, and the potential for systemic bias.

- Do the authors reflect on any implications the nature of the training data has for the generalizability of the findings?

  - Yes. They warn that any model trained on this data may reinforce existing biases.

- Do the authors reflect on the ethicality and/or societal implications of their work?

  - Yes. They incorporate intersectionality, historical policing trends, and systemic inequalities into their ethical reflection.

Analysis Code

- Does the code used for analysis look generally correct?

  - Yes. Standard pandas and matplotlib methods are used effectively. One point of confusion is the x-axis label in Figure 6, which may be mislabeled as the victim's race and then displays numeric values.

- Is the code readable, written in proper style, and commented appropriately?

  - Yes, and most, if not all code is well commented, suggest adding more explanations within complex blocks.

- Are there markdown descriptions of the analysis throughout?

  - Some. There are markdown headers, but these could be expanded with narrative explanations of results or decisions between blocks of code.

- Are their code and output (especially plots) accessible?

  - Yes, it all outputs and functions correctly, could benefit from larger titles, clearer legends, and more color distinction for better readability.

General

- Is the writing clear (including elements like spelling, grammar, etc)? Are you able to follow what is being done?

  - Yes. The writing is academically appropriate and clearly structured. The keywords section at the beginning is particularly helpful.

- Is the code clear? Are you able to follow precisely what is being done?

  - The code can be followed, but a few sections would benefit from better documentation and markdown context.

- Are you able to reproduce all aspects of the report, including output, visualizations, etc?

  - Yes. All cleaning, transformation, and analysis steps are included. Yes. With access to the dataset, the analysis should be fully reproducible.

- Is the report well-formatted and readable (including layout but also only reporting relevant output, with no extraneous code, visuals, etc)?

  - Yes, the report is formatted well with a logical structure.

- Have they appropriately outlined the next steps with gaps clearly defined?

  - Yes. The authors acknowledge population data as a key gap, and caution against extrapolating results from biased datasets.

- Is their repository well-organized or has it been difficult to find certain files?

  - Their repository is well organized with proper and helpful folder names. All necessary materials are included there is no difficulty finding anything.

- Any suggestions for them moving forward?

  - Improve visual clarity of graphs (font size, labeling, legends).

  - Add more markdown context in the notebook to explain code decisions.

  - Expand on how the data source (NYPD) may introduce structural bias in reporting and classification.

Final Considerations

- What is one question you have for the group after reading their analysis?

  - How might your conclusions shift if population and racial demographic data for each precinct were integrated to normalize report counts?

- What is one thing the group has done especially well?

  - I think this group did an amazing job setting up the paper with sufficient background information and citations. Their discussion of ethical implications and systemic bias in predictive policing is nuanced, well-researched, and clearly connected to their findings