PROJECT PROPOSAL
Analyzing Policing Patterns and Demographic Disparities in U.S. Traffic Stops

Name: Isabelle Krampert
Note: I will be working solo on this project!

INTRODUCTION

Main Question
Can the likelihood of a traffic stop be predicted based on race, age, or sex?

Problem
This project seeks to examine traffic stop patterns across driver race, age, and sex. I'll be using data from the Stanford Open Policing Project – it provides datasets for almost every state, and each set has fields such as "Stop date", "Stop time", "Driver race", "Driver sex", and "Driver age" (state datasets vary in the number of fields they have). The goal of my project is to build a predictive model trained on this data to determine whether a driver's demographic influences traffic stop patterns in any given state, and if so, whether this correlates with policing bias or demographic disparity.

Novelty and Importance
Although a major study has already been conducted on these datasets concerning racial disparity and policing patterns, I wanted to further investigate if factors like age and sex also influence the likelihood of a stop. This project focuses solely on the chance of a stop alone based on these factors, and does not consider if an arrest or citation is made afterward, unlike the main study. The importance of this project lies in the fact that it is focused on intersectional demographic analysis – for example, contrasting the percentage stops between younger white men vs. older black women – and if this data can improve predictive performance while highlighting police bias in any state. By doing this, the project will hopefully show existing disparities (if any), and whether they are prevalent to make the likelihood of a traffic stop predictable.

Related Works
This is the citation for the publication I was previously referring to:
E. Pierson, C. Simoiu, J. Overgoor, S. Corbett-Davies, D. Jenson, A. Shoemaker, V. Ramachandran, P. Barghouty, C. Phillips, R. Shroff, and S. Goel. "A large-scale analysis of racial disparities in police stops across the United States". Nature Human Behaviour, Vol. 4, 2020.
Link to PDF: https://5harad.com/papers/100M-stops.pdf

PROJECT PLAN

Data Source

I'll be using publicly available data from the [Stanford Open Policing Project] and the [United States Census Bureau]. The Stanford Open Policing Project includes over 200 million records from a multitude of local and state police departments across the United States, all concerning traffic stops. For the positive examples, I will only be incorporating these fields into my reframed data: Driver race, Driver sex, Driver age, Stop date. For negative examples, I'll be using data collected at county-level from the U.S. Census Bureau (many datasets in the Stanford Open Policing Project are at city-level). If I find that the Census Bureau datasets are a poor fit for negative examples however, my backup plan is to construct artificial population/demographic data based on the cities in the Stanford data. By including negative examples, my model will have something to compare positive traffic stops against, making outcomes binary.

Data Storage and Management

I'll be cleaning and processing the data with pandas. To ensure that demographic categories are standardized in my own data set, I will consolidate each demographic category in each state, as there may be differences between how each category is labeled. I will also group age ranges to further capture any trends, and incorporate U.S. Census data into my own data set to intermix both stopped and non-stopped individuals. Because I may have to combine multiple states' data, I'll likely have to use a small-scale SQLite database to store everything.

Models and Techniques

Logistic Regression:

I would like this to be my main predictive model, as it fits binary classification problems (stopped vs not stopped), provides probability estimates, and results in coefficients denoting how each demographic variable impacts the likelihood of a stop.

Naive Bayes:

I'd also like to train a Naive Bayes model, so I can compare its accuracy to the Logistic Regression model. I feel this model would be a good fit for this project in addition to a Logistic Regression model, as it handles categorical data effectively and is simple.

Implementation Process

1. Obtain Data → Download both datasets mentioned previously.
2. Clean data → Standardize demographic fields, remove rows missing values, incorporate people who were stopped (Stanford) and not stopped (U.S. Census)
3. Analysis → Observe basic distributions with my own dataset. I may use simple plots to visualize highest stop rates amongst counties, states, and demographics.

4. Train the Model → Fit Logistic Regression and Naive Bayes models, and observe model performance.
5. Evaluation of Models and Predictions → Compare performance of both models, noting which one better predicts the likelihood of an individual getting stopped. Additionally, assess whether models have large gaps in predicted stop likelihood, and ensure models are consistent and accurate with their predictions.
6. Written Report and Further Analysis

<u>Evaluation and Success Metrics</u>

- Accuracy: percentage of correct predictions
- Precision or Recall: how well model identifies that a person is likely to be stopped without any false positives (Must be balanced)
- ROC-AUC Score: to check how well our binary classification models work in terms of separating positive and negative cases. (Score well above 0.7)
- Fairness: Measure differences in
    - Demographic Equality (Do some groups have a higher rate of receiving a stop prediction than others?)
    - Errors (Does the model tend to make more errors for any given demographic?)
    - Group Percentages (What's the percentage of positive predictions for each demographic?)
    - Large disparities have to be observed and explained.

<u>Tools and Libraries</u>

- pandas, numpy
- scikit-learn
- fairlearn
- matplotlib or seaborn
- SQLite
- Python scripts