Midterm Report: Predicting Civilian-Officer Outcomes

Lucy Du (dd483), Elina Hvirtsman (eh582), Julia Ng (jen67)
November 8, 2020

1 Problem Statement

Our project's goal is to understand what factors most influence civilian and police interaction outcomes and determine if there is evidence to suggest racial bias in Minneapolis police precincts. Given demographic and situational factors, we hope to develop a model that can predict whether the civilian will be issued a citation in the event of being stopped by police, or which police force the civilian will experience in more severe situations.

2 Dataset

The two main datasets are sourced from Open Minneapolis. The goal of the Minneapolis Open Data Portal is to encourage access to data managed by the City of Minneapolis, making content available to be freely used, modified, and shared by anyone for any purpose. The datasets are refreshed on a daily basis by 9:30 AM. For the purpose of this project, we use the finalized dataset published on October 24, 2020.

The Police Stop dataset spans from July 6, 2017 to October 24, 2020, and the Police Use of Force dataset spans from January 1, 1970 to October 23, 2020. Both provide basic information about the civilian like race and gender, as well as information about each civilian-officer interaction like if there was a search of the person, the primary offense, and the specific problem, such as suspicious person, traffic law enforcement, attempt pick-up, etc. Police precinct, neighborhood, and latitudinal/longitudinal data is present too. Our outcome variable of interest is citation issued for Police Stop, which contains either yes or no. Our outcome variable of interest is force type for Police Use of Force, which contains values like bodily force, taser, chemical irritant, etc. There are 12 features and 116,165 examples in the first dataset, and 20 features and 29,787 examples in the second dataset.

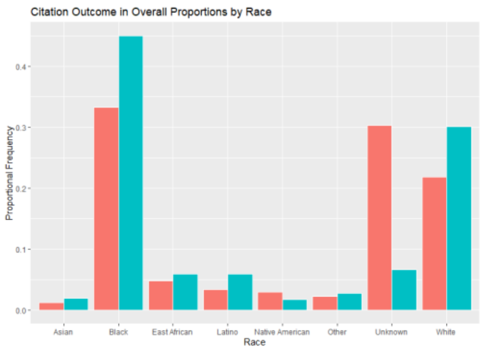
3 Data Preprocessing

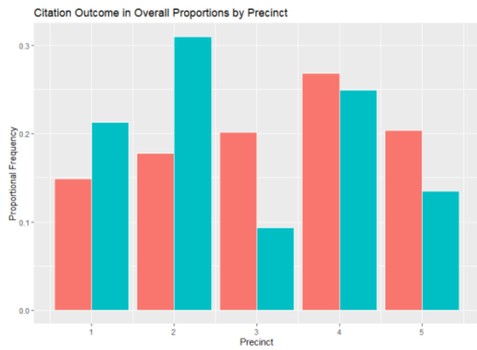
To make our datasets fit our problem statement, we dropped any observations with missing `citationIssued` or `ForceType` values. To make our dataset more approachable, we converted some nominal values into one-hot encoding. For example, the single `Race` column, which contains values like `Asian`, `Black`, `East African`, `Latino`, `Native American`, `Other`, `Unknown`, and `White`, was converted into eight columns with a 1 in the applicable column and a 0 in the others. For variables like `Neighborhood`, which have 86 possible entries, we keep the values for the top 5 most common and convert the remaining into a category `Other`. This column is then converted again through one-hot encoding. Finally, the aggregated `Date` column is separated into more usable `Year`, `Month`, `Day`, `Hour`, and `Minute` columns.

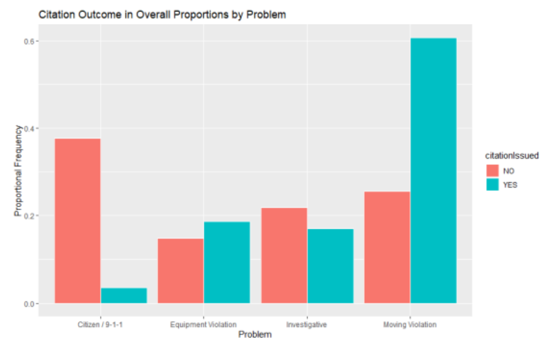
The datasets are comprehensive and complete with intuitive values. Histograms were created for the numeric variables like latitude and longitude to check for outliers. In the Police Stop data, there are around 700 observations with missing latitude and longitude, which are dropped given it is less than 1% of our entire dataset. Frequency tables were created for the nominal variables like race, problem, and neighborhood to check for outliers, missing, or erroneous data. Data validation was most likely employed in the data collection process as there are no known outliers in the data. There are now 115,388 examples in this dataset with around 80% no citation issued and around 20% citation issued.

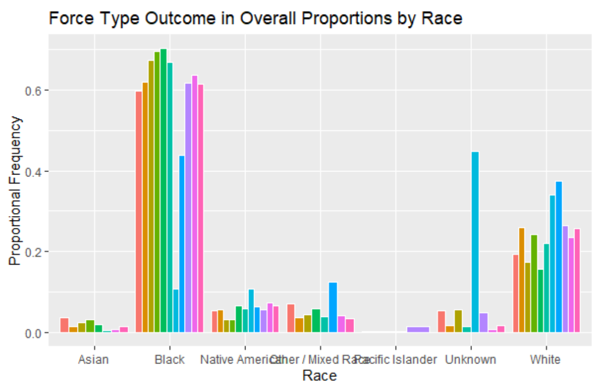
In the Police Use of Force data, there are around 100 observations with missing latitude and longitude, which are dropped given it is less than 1% of our entire dataset. Frequency tables were again created. There are around 300 observations with non-recorded race and 5 with non-recorded sex, which are dropped given it is around 1% of our entire dataset. Erroneous precinct observations are dropped too. There are now 29,251 examples in this dataset with `Bodily Force` as the most common `ForceType` at around 75% of all examples.

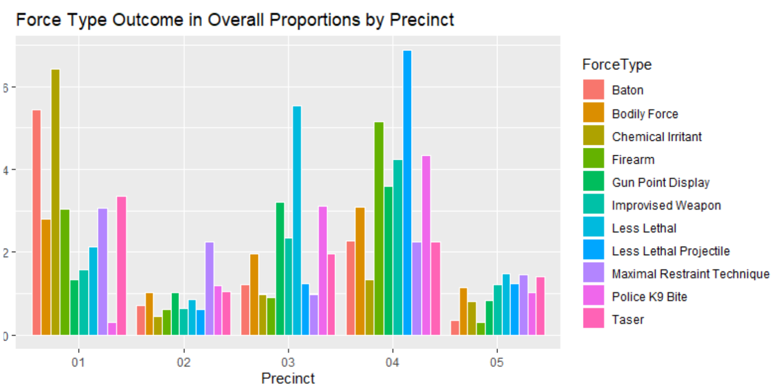
4 Data Visualizations











Before modeling, we conduct preliminary descriptive analysis to understand our dataset and the relationships between variables better.

For the Police Stop data, we plot frequency plots for citation outcome by race, precinct, and problem. The following plots show that Asian, Black, East African, Latino, Other, and White have relative greater citations issued than not within their own race category. Overall, Black, Unknown, and White have the absolute highest citations issued. Precincts 1 and Precinct 2 have relative greater citations issued than not within their own precinct category. Overall, Precinct 2 and Precinct 4 have the absolute highest citations issued. Finally, Moving Violation has the highest relative and absolute, and Citizen/911 has the lowest relative and absolute citations issued. This could be an important predictor in our later models.

For the Police Use of Force data, we plot frequency plots for force type outcome by race and precinct. The following plots show that Less Lethal and Less Lethal Projectile are most common Force Type outcomes for White, Unknown, Other/Mixed Race, and Native American. Comparatively, Gun Point Display and Firearm are most common for Black. Precincts also differ in Force Type outcomes, ranging from most common being Less Lethal or Less Lethal

Projectile for Precinct 3 and Precinct 4 to Chemical Irritant and Maximal Restraint Technique for Precinct 1 and Precinct 2.

5 Preliminary Model Fitting

For preliminary modeling, we fit three different models and implement cross validation where 80% of the data is used as training across each of the 5 iterations, then the model is applied on the remaining 20% to obtain a validation error. To prevent overfitting, two regularizers are fit, where the regularization parameters λ are found using cross validation.

Police Stop

5-fold Cross Validation Misclassification Error on Validation Fold

Model	**1**	**2**	**3**	**4**	**5**	**Average**
Logistic	.0207	.0226	.0216	.0204	.0222	.0215
Lasso	.0213	.0236	.0225	.0212	.0228	.0223
Ridge	.0223	.0246	.0237	.0221	.0238	.0233

Detailed analysis shows that false negatives are around 3.963 times more common than false positives in Ridge, 3.622 times in Lasso, and 3.275 times in Logistic.

Police Use of Force

For this particular dataset, we are trying to classify the ForceType, which is a nominal value with 11 different types of force police can use by using one-vs-all classification and multinominal logit. Through ProxGrad, we initially tried to encode each of these force types as scalars from 1 to 11, but quickly found that the errors we saw seemed high. For Mean Absolute Error, OvA Loss had training MAE of 17.586 and test MAE of 18.003. Multinomial Loss had training MAE of 27.077 and test MAE of 27.532. Since the Multinomial Loss had a larger error, we believe OvA Loss is better, but the error was still higher than expected. However, training and testing errors are not very different, so it does not seem like overfitting is occurring. As a result, we hope to find a better classification model, as it seems like the model may be underfitting.

6 Future Plans

Other models will be considered for our project, such as classification trees, and to prevent overfitting, we will consider pruning or ensemble methods like bagging and boosting.

Also, for the Police Stop data, there is an imbalance of our outcome variable being heavily 0 in Citation Issued. Future steps could be considering the implications of false positives vs. false negatives, and analyze different model evaluation metrics like precision or recall in order to determine the best model.

Since Logistic denotes which variables are statistically significant and Lasso performs feature selection, we can obtain and interpret important features from its model output, which will be included in the final report.