NYPD Civilian Complaints

Moksha Poladi and Esha Desai

LITERATURE

Currently the topic of racial bias and discrimination is a prominent issue in the criminal justice system. There are many studies that show how black individuals are more likely to be stopped by NYPD than their white peers. For example, in an NYU research study, it was found that black drivers were searched about 1.5 to 2 times more often as white drivers even though they are less likely to carry illegal contraband compared to their white peers.
Although this is widely talked about topic, there isn't much research regarding how often complaints made by minority communities are substantiated.
We decided to tackle this question instead. Similar to other studies, we can see that there is a racial bias since the proportion of cases that are substantiated when the officer is white and when the complainants are non-white is disproportionately lower than that of the general population.

INTRODUCTION

This dataset contains complaints filed against NYPD officers till 2020. It gives us additional information about both the accused officer and the complainant such as their gender, ethnicity, and age. It also goes into more details such as the outcome of the complaint and whether action was taken against the other officer or if the complaint was dismissed. The details helped us conduct some interesting analyses which take into account the complainant and officer ethnicities and the outcome of the case.

After exploring the data, we decided to examine whether the ethnicity of the complainant and the police officer is associated with the outcome of the complaint filed. In our dataset, complaints can be of three types: exonerated, substantiated, and unsubstantiated. An exonerate complaint is when the complaint was dismissed, and the officer is not at fault. A substantiated complaint is when the officer is in violation and the complaint is valid. Finally, an unsubstantiated complain is when there is not enough information to make a decision regarding the complaint.

Link to the dataset:
https://urldefense.com/v3/__https://www.propublica.org/datastore/dataset/civilian-complaints- against-new-york-city-police-officers

Link to GitHub Repo:

https://github.com/M2606/DSC190-FinalProjectCode.git

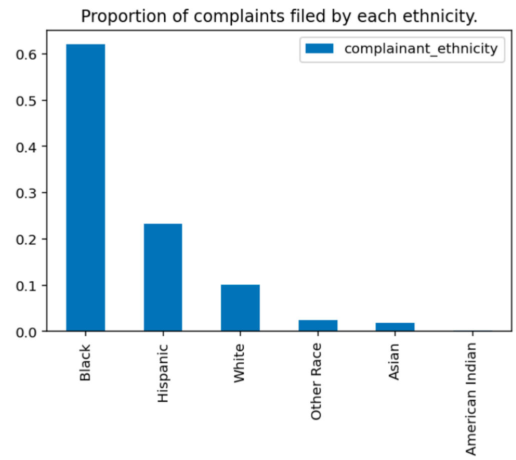
CLEANING AND EDA

Originally, the dataset had 33,358 rows however after dropping the null values, we ended up with a table that had 28,012 rows. This dataset did not contain too many null values but did have some messy data that needed to be converted to nans. We first examined the ages of complainants and we found that some were negative or very low values. We chose to convert all ages 7 and under to nan because based on the data generating process these values were unlikely. We continued to clean the data frame by looking where the ethnicity were Unknown and where individuals refused to state their race. We then changed those values to nans. We also checked where complainant gender was not described and changed those to nan in our data frame. We found the wait time for each complaint by converting the year and month received columns and the year and month closed columns to pandas datetime objects and subtracted them to get a time delta object. Within Substantiated there were multiple divisions describing the action taken against the officer. For our analysis we decided to just group all these divisions into one category that we called substantiated. In general, in cases where the values seemed inconsistent with the data generating process or there was missing information, we converted them to NaN values. For the most part, changing these values to NaN did not affect our results too much (specific details mentioned under each cleaning step below).

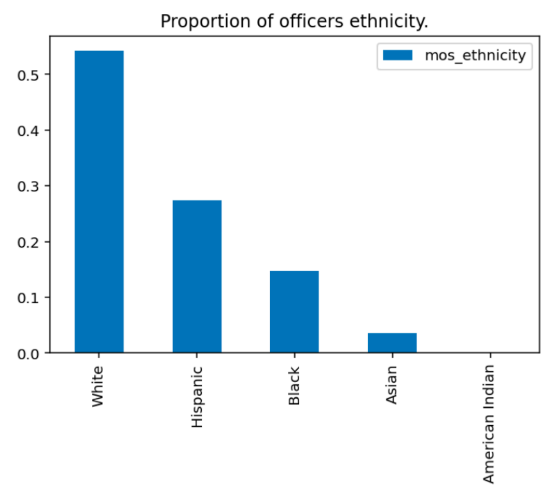
After this level of cleaning, we explored multiple questions about the data using visualizations. While our main focus was on ethnicities, we did multiple analyses which are shown in the code in Github.

Since we were focusing on ethnicity, we decided to explore the following questions: What is the distribution of ethnicities? Which ethnicity has filed the greatest number of complaints? What is the distribution of ethnicities for officers? Which ethnicity do most officers belong to within this dataset? Is there any association between age and wait time? Does the severity of the accusation depend on the ethnicity of the complainant? Which gender has the highest average wait time? What are the average wait times by complainant ethnicity? Does the outcome of the case have anything to do with the complainant's ethnicity? For example, are complaints filed by people of color more likely to end up in the officer being exonerated? Are white-officer vs non-white complainant cases more likely to go against the complainant? Is the outcome of the complaint related to the rank of the officer? eg. are higher ranking officers more likely to be exonerated?

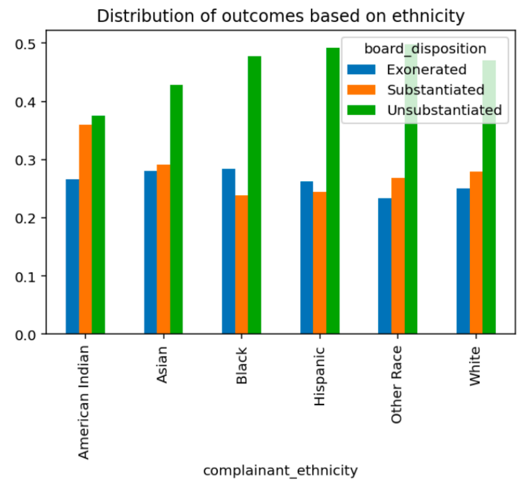
We first looked at the distribution of complainant ethnicities in the database. We found that most of the complaints were from African American individuals.



We then investigated the distribution of officer ethnicities in the database. Here we found that most officers were White.

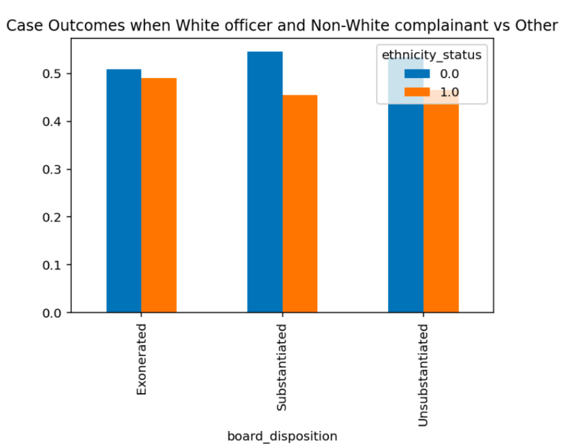


Based on these findings, we then decided to look at the distribution of complaint outcomes based on ethnicity.



We then asked the question: Are white-officer vs non-white complainant cases more likely to go against the complainant?

To do this, first we found where the complainant was not white and where the police officer was white. We put these bool values in the data frame with a column called ethnicity_status. We ensured that the places that originally had NaN values in complainant ethnicities continued to be NaN in the new column. We then plotted the case outcomes based on when the complainant is white, non-white, or other.



We found the difference between substantiated case proportions and exonerated case proportions based on the officer and complainant's ethnicity. When the complainant was not white and where the police officer was white vs all other outcomes the difference in proportions was greater for substantiated cases. This might indicate that complaints are more likely to go against the complainant when the officer is white, and the complainant is non-white. Based on these results, we decided to perform a permutation test to determine if the different in proportions were statistically significant.

PERMUTATION TEST

In order to check if white-officer vs non-white complainant cases are more likely to go against the complainant, we decided to perform a permutation test. We created a status column with value 1 if the complaint was against a white officer with non-white complainant and 0 otherwise. We then permuted this column repeatedly to see if the observed differences in distribution of outcomes for the two cases was simply due to randomness or if the outcome of the complaint was in some way dependent on the complainant's ethnicity.

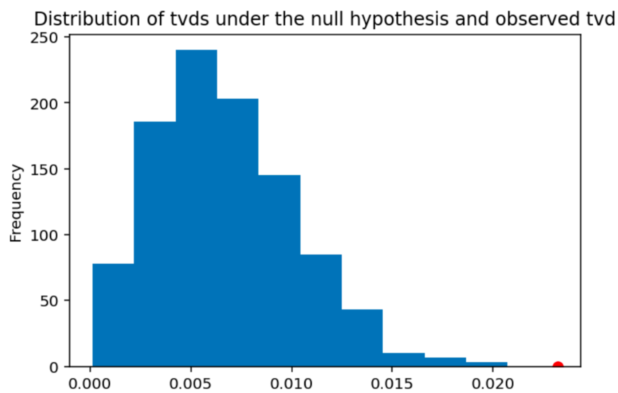
Null Hypothesis: White officer with non-white complainant vs those that don't meet that requirement come from the same underlying distribution.

Alternative Hypothesis: White officer with non-white complainant vs those that don't meet that requirement come from two different distributions.

The test statistic we choose is the total variation distance (TVD) between the distributions when the officer is white, and the complainant is non-white vs. those that don't meet that requirement. We choose this statistic because it works best when we have two categorical distributions.

The significance level we chose is .05.

We got a p-value of 0.0 which indicates that with very high probability, the distribution of outcomes for white officers with non-white complainants vs those that don't meet that requirement is significantly different. Although this is helpful in showing an association between ethnicity and outcome, it does not say if white officers are more likely to be exonerated or substantiated when accused by non-white complainants because it is not a directional test.



We got a p-value of 0.0 which indicates that with very high probability, the distribution of outcomes for white officers with non-white complainants vs those that don't meet that requirement is significantly different.

While the permutation test was effective in establishing an association between complainant and mos_ethnicity and the outcome of the case, it does not answer the more specific question of whether white officers are more likely to be exonerated or substantiated when accused by non-white complainants because it is not a directional test.

PREDICTION MODEL

From our EDA and the results of our permutation test, we decided to create a model that uses the demographic information to predict if the officer was exonerated. While in our EDA we focused on ethnicity, we decided to add other features such as age, gender, and wait time to make our model more robust.

We are only predicting whether they are exonerated or not because helped us focus on one specific outcome. We decided to choose accuracy for our evaluation metric. We believed that in this case, since we are only predicting whether an officer would be exonerated from the complaint, we would treat incorrect and correct predictions equally, which is why we chose accuracy. That being said, because our data is slightly imbalanced, we did find the sensitivity and specificity as well in order to get a better overview of the model performance.

We created a baseline model using Decision Tree Classifier without doing any feature engineering and hyperparameter tuning. Our baseline model included a total of 9 features, 5 of which were nominal and the other four were quantitative. We One Hot encoded the categorical features. We chose these features because they were the most relevant to predicting the board disposition. We performed k-fold validation with 5 folds and observed an accuracy of: 0.7113934812235689. Our final test accuracy was: 0.7101380295097572. We also found, sensitivity: 0.4355188884064264, and specificity: 0.8138010162268481. We think the model worked fine as a baseline, but the relatively low accuracy was not ideal. The difference in sensitivity and specificity was also a little concerning, but for the reasons mentioned above, we focused on accuracy. We thought the overall accuracy could be improved by engineering some new features and possibly selecting a different model.

For our final model, we added a new column called wait time using the year_closed and year_recieved columns because we felt that if there was a higher wait time, the case was more likely to be unsubstantiated and not exonerated. We One Hot encoded the categorical features and we standardized our quantitative columns using StandardScalar(). We did this because the age, weight time, and the years were all on significantly different scalesThe model type we chose was a RandomForestClassifier(). Our model selection process involved performing cross validation for four classifiers: Decision Tree, RandomForest, KNeighbors, and Logistic Regression. We found that Logistic Regression and RandomForest had the best mean accuracy across 5 folds.

We chose the RandomForestClassifier because it would not overfit the training data and we were able to find the optimum parameters. We were also able to optimize its parameters using GridSearchCV(). After running GridSearchCV() we got max_depth': 18, 'min_samples_leaf': 2,' min_samples_split': 2, 'n_estimators': 200 as the best parameters and this helped improve the accuracy of the model. We then found the accuracy of the model using cross_val_score with 3 folds and got an accuracy of: 0.75 Our final test accuracy was: 0.74. While the accuracy improved, the sensitivity (0.28) decreased, and specificity (0.92) increased. This was a consequence of our focus on improving accuracy rather than the other metrics.

RESULTS

Our model does outperform the baseline. We wanted to optimize the parameters for our chosen RandomForestModel. We used GridSearchCV to do so. From our EDA analysis, we were able to find features that seem to be associated with the outcome of the complaint. We also checked for feature importance using the inbuilt Random Forest Method to valid the effectiveness of our features. We then created our final model using these optimized parameters causing the model to be more accurate than the baseline.