Predicting Terry Stop Arrests in Seattle

Presented By: David Bartholomew

The Data Science Process: OSEMN

Obtain Data

The data for this project is provided directly from the Seattle police department. https://catalog.data.gov/dataset/terry-stops.

Scrub Data

Removing duplicate records, filling missing values, determining whether to keep or replace placeholder values, and converting to appropriate data types so that our model can interpret the data.

Explore Data

Visualizing features (such as Officer Race or Gender) that may have an impact on predicting arrests, creating new features as necessary, and determining which features to include to begin the modeling process.

Model Data

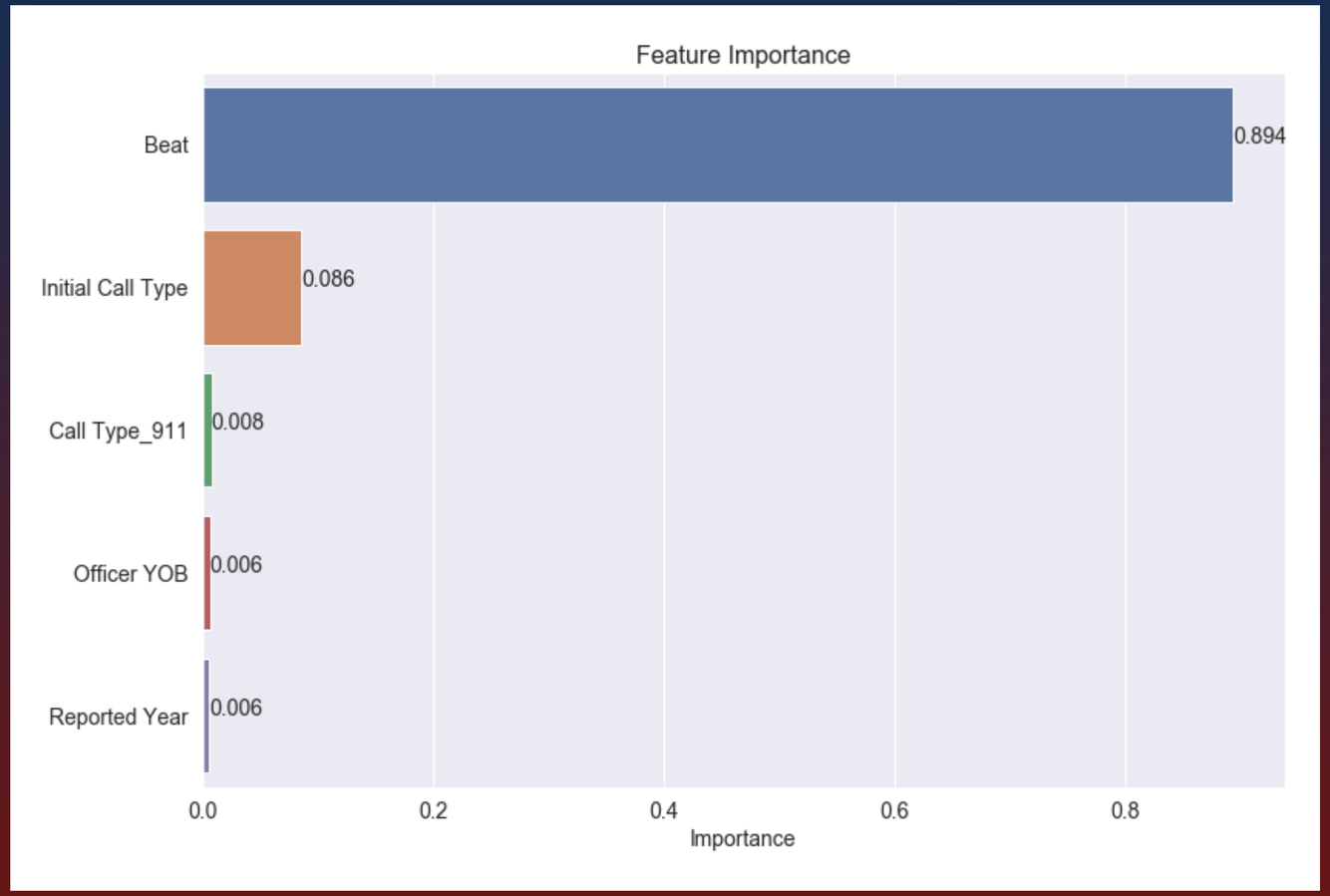
Utilizing several different algorithms to predict if a Terry Stop resulted in an arrest or not, then choosing the best algorithm based on model results

iNterpret Data

A final model was chosen based on a recall score of 95.5%, predicting the highest amount of arrests while minimizing false negatives and maintaining overall accuracy of predictions above 50%.

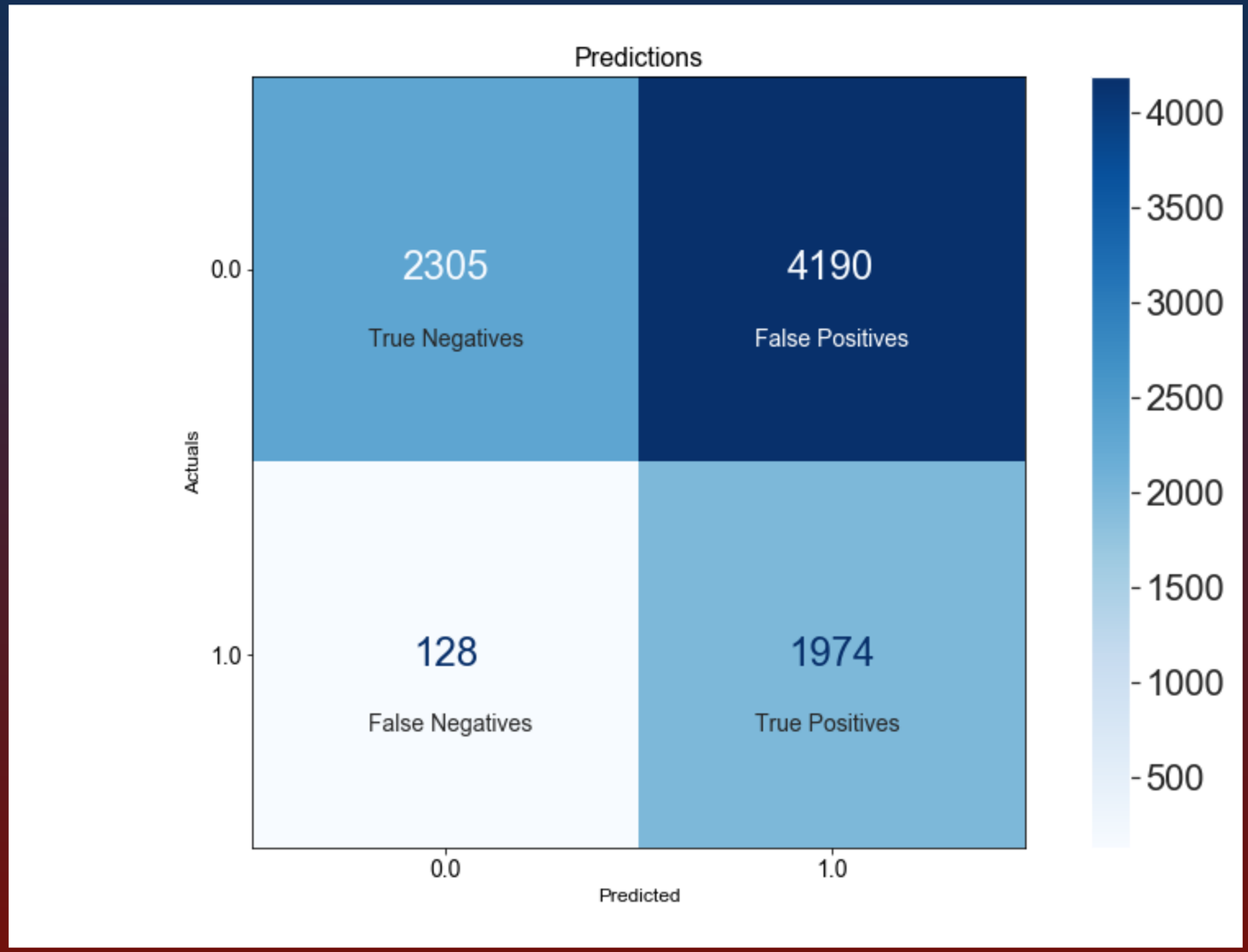
Final Features

- Beat: whether the address associated with the underlying Computer Aided Dispatch (CAD) event was reported or not
- Initial Call Type: whether an initial call type was reported or not
- Officer Year of Birth
- Call Type_911: call received by the communication center by dialing 911
- Reported Year



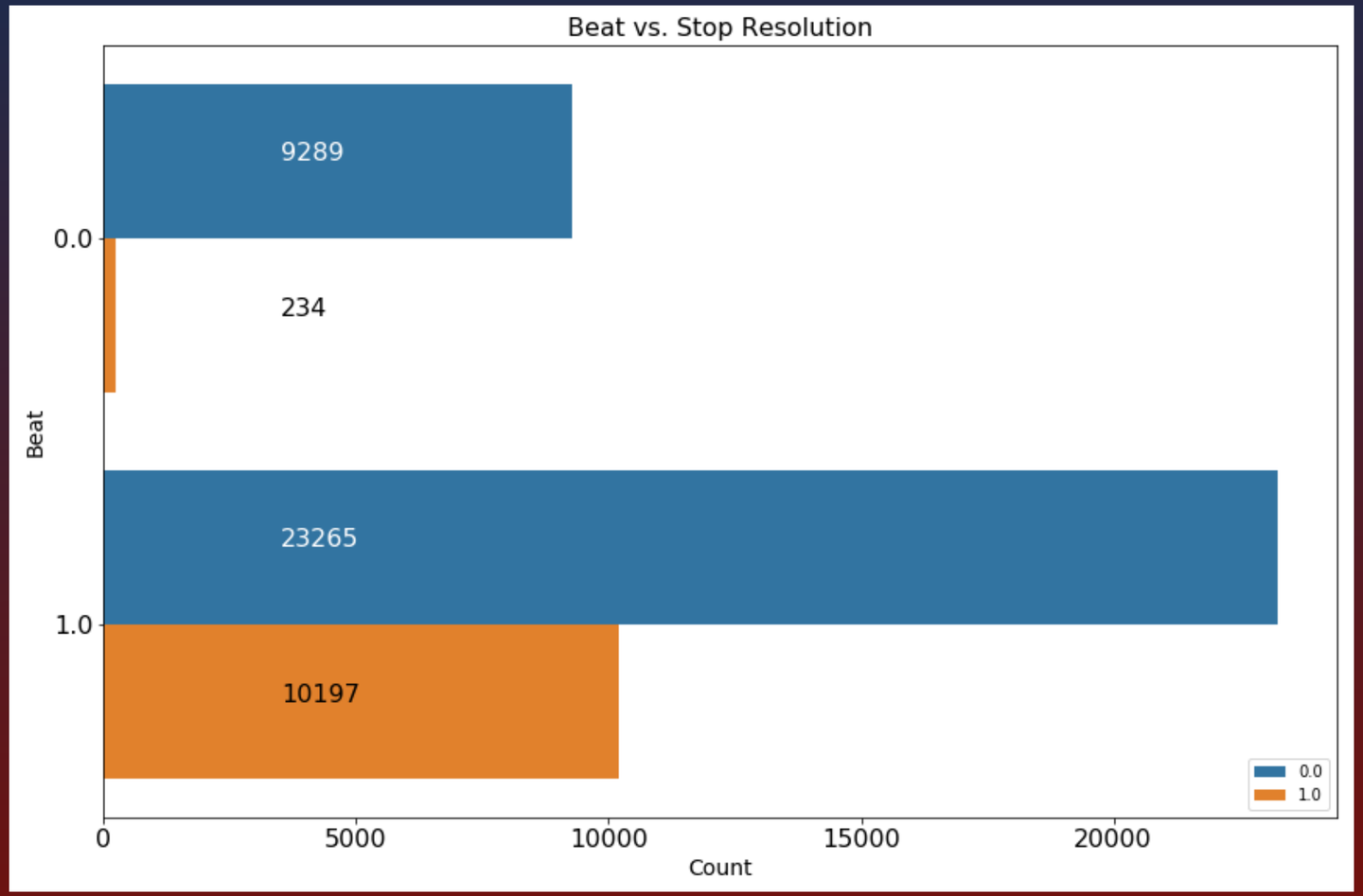
Model Results

- The final model resulted in accounting for 95.5% of total arrests.
- There are a lot of false positives, but this model was built to minimize false negatives and reduce unnecessary Terry Stops. In this case, 2,305 Terry Stops or just under 27% of overall stops may not be necessary to conduct.



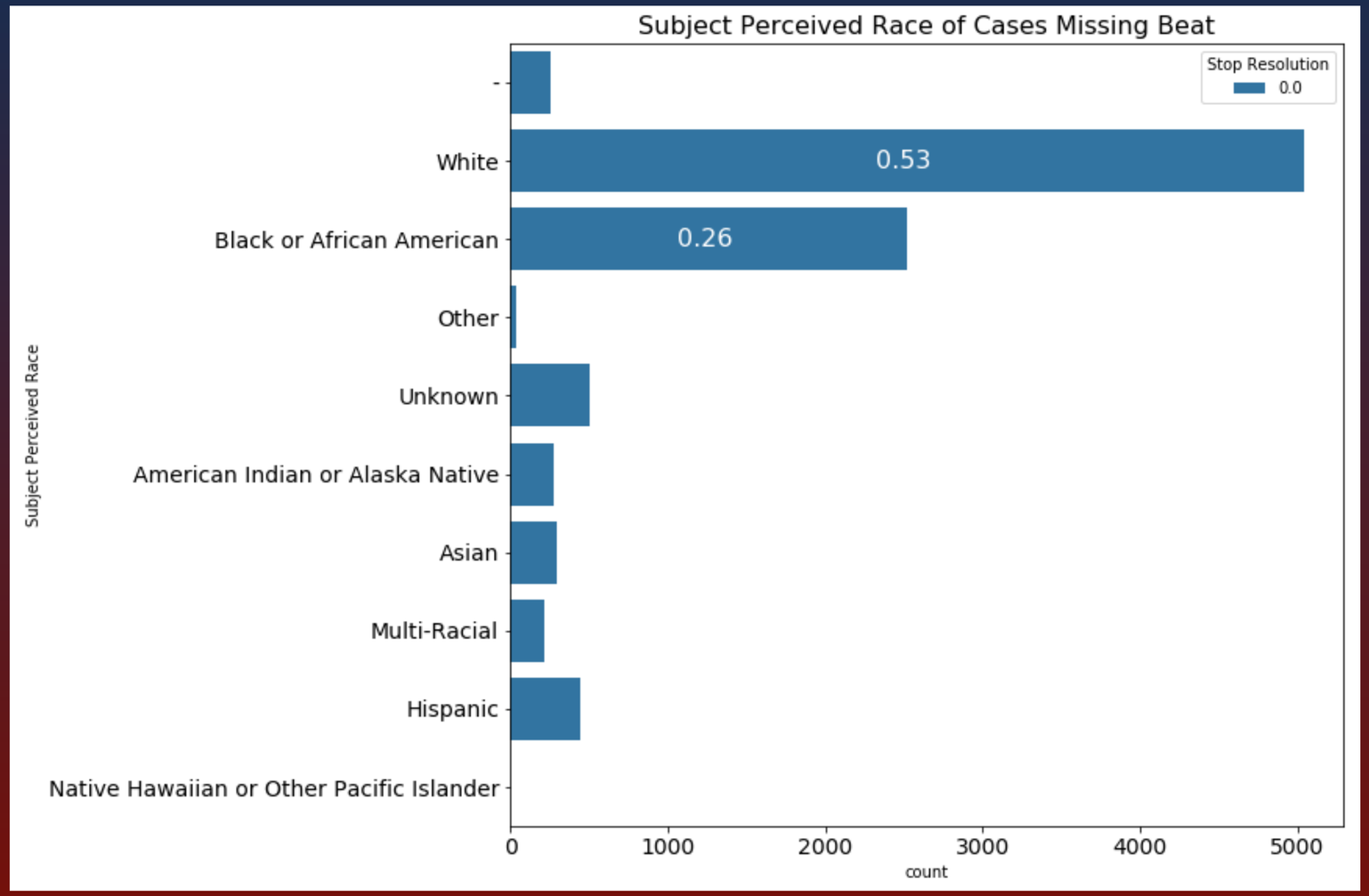
How reliable is the reporting process for Terry Stops?

- Beat is the address associated with the underlying Computer Aided Dispatch (CAD) event
- Out of 9,523 cases in which beat was not reported, there were only 234 arrests. With 97.5% of these cases resulting in no arrest, there should be additional information on why these Terry Stops were conducted.



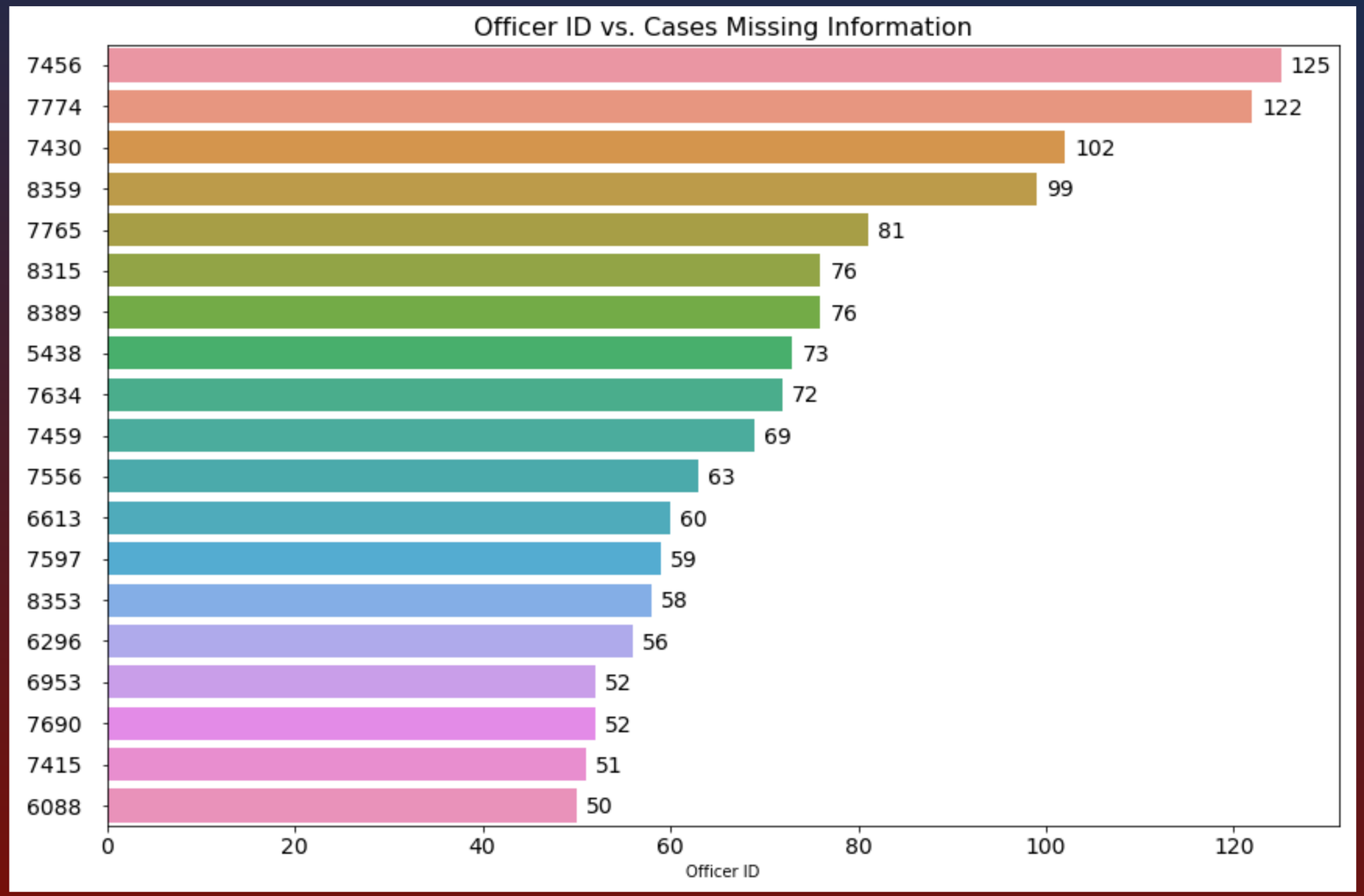
What is the breakdown of subject race for these unreported stops?

- 53% of these cases with unreported information are White subjects, 15% lower than the government census of the white population in Seattle at 68%.
- 26% of these cases with unreported information are Black subjects, 19% higher than the government census of the black population in Seattle at 7%.



Are there officers with repeat cases with unreported information?

- There are multiple (19) officers that have over 50 cases where information was not reported as a result of a Terry Stop.



Conclusion/Further Analysis

- Our model accounted for 95.5% of arrests, minimizing false negatives and showing that almost 27% of Terry Stops in Seattle may not be necessary.
- 97.5% of cases that did not have address information reported resulted in no arrest. Additional information is necessary to understand why these Terry Stops were conducted.
- 26% of cases with unreported information involve Black subjects. The census of the Black population in Seattle is 7%. White subjects are 53% of these cases, compared to being 68% of the population. At first glance, one could question if there is racial profiling based on these numbers.
- There are 19 officers in this data that have over 50 cases with unreported address information.
- Conclusions of racial disparity would be theoretical with this data alone, so rather than providing a theoretical conclusion, I recommend that the Seattle Police Department reduce questions of racial disparity by improving reporting in these cases, whether it be requiring officers to report this information themselves or automating processes further with Computer Aided Dispatch.

Thank you.

Full details of this project are available at this link:
https://github.com/dbarth411/dsc-mod-3-project-v2-1-online-ds-sp-000