# Predicting recidivism

## Background

Recidivism or the risk of someone reoffending is one of the most difficult decisions that judges and law officials in the US justice system need to consider. In the recent years, machine learning algorithms have become a common tool which helps judges and law officials assess the risk of recidivism for individuals arrested. In many cases, the risk score predicted from the algorithm is used in courtrooms to inform decision about the defendant sentence or their freedom. However, in 2014, the U.S. Attorney General Eric Holder warned that risk scores might be injecting bias into the courts.

ProPublica, a non-profit organization conducted an accuracy analysis on the algorithm by selecting one of the most popular commercial risk score tools made by Northpointe, Inc. called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions). They looked at the criminal records from the Broward County Clerk's office for the analysis.

The study focused on 3 main variables which were gender, age, and race, and defined recidivism as an arrest within two years of the previous arrest. By running a logistic regression, they found that being Black, of younger age, and male is more likely to have higher risk score than the opposite. They have further analysed the accuracy of the score and found that the algorithm is more likely to misclassify a black defendant as higher risk than a white defendant, while under-classified white reoffenders as low risk almost 70.5 percent times more often than black reoffenders.

## Hypothesis

- There is a strong bias on COMPAS scores because of gender, age and ethnicity. Propublica study was conducted around Race, we want to check if other factors like age and gender have bias as well.

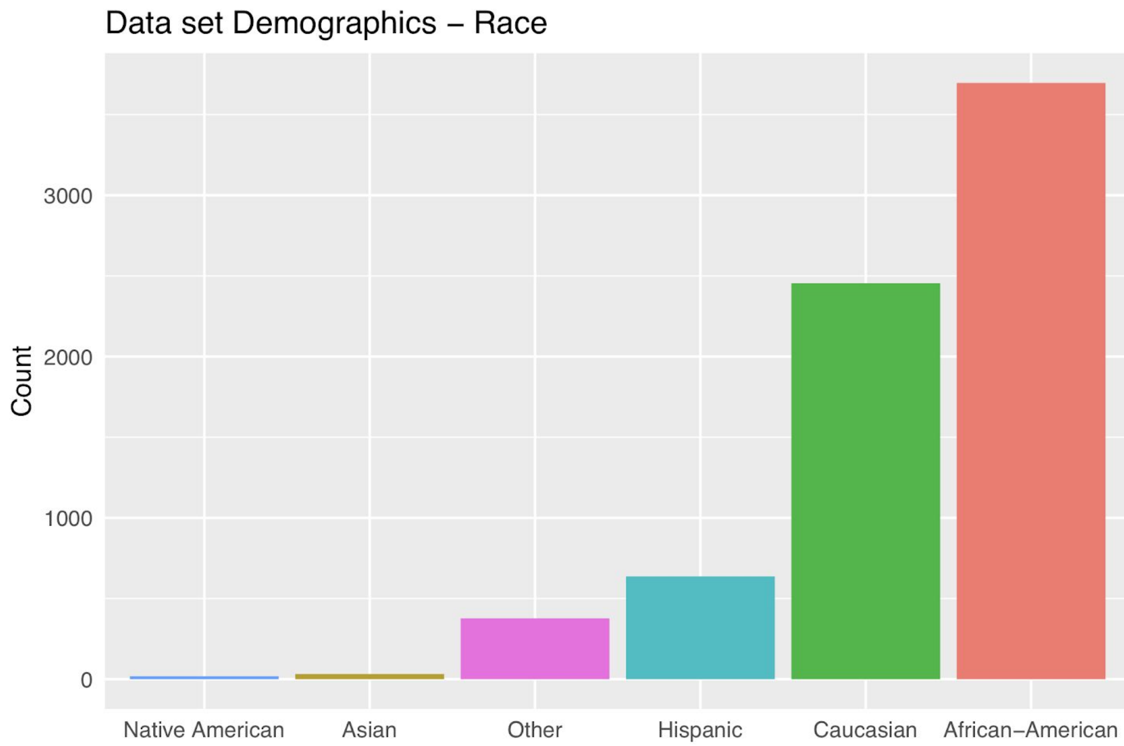
## Research Question

- What is the bias of different factors on recidivism
- Whether the past crime information is a better indicator and is enough to predict the risk of recidivism

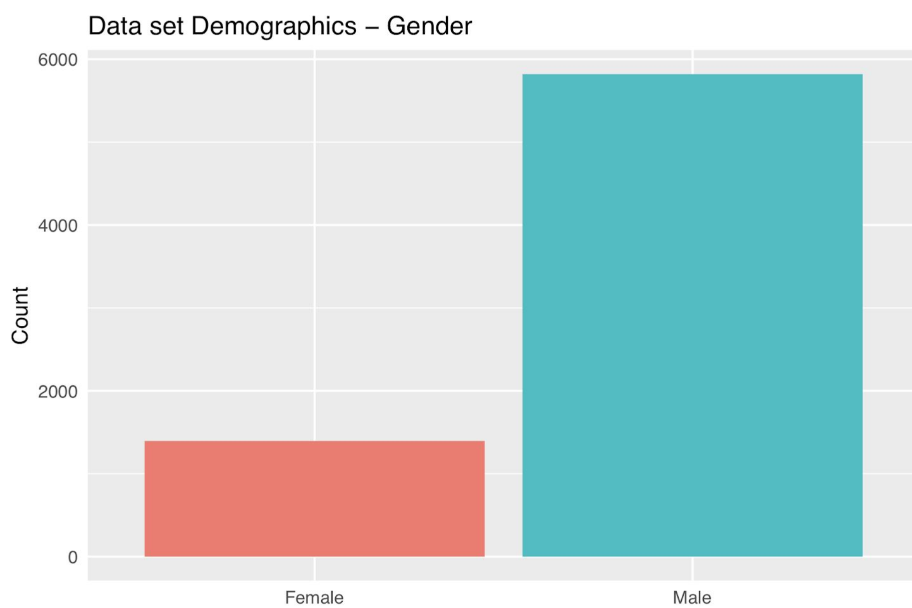
## Data

- Two-year COMPAS score
    - This includes 2 years of records with information on recidivism
    - Includes general information: name, age and age group, sex, date of birth
    - Includes other information: juvenile records, prior arrest count and others
    - The dataset contains 7,214 defendants records out of which 7,158 were unique person records
    - The COMPAS screening date was between 2013-2014

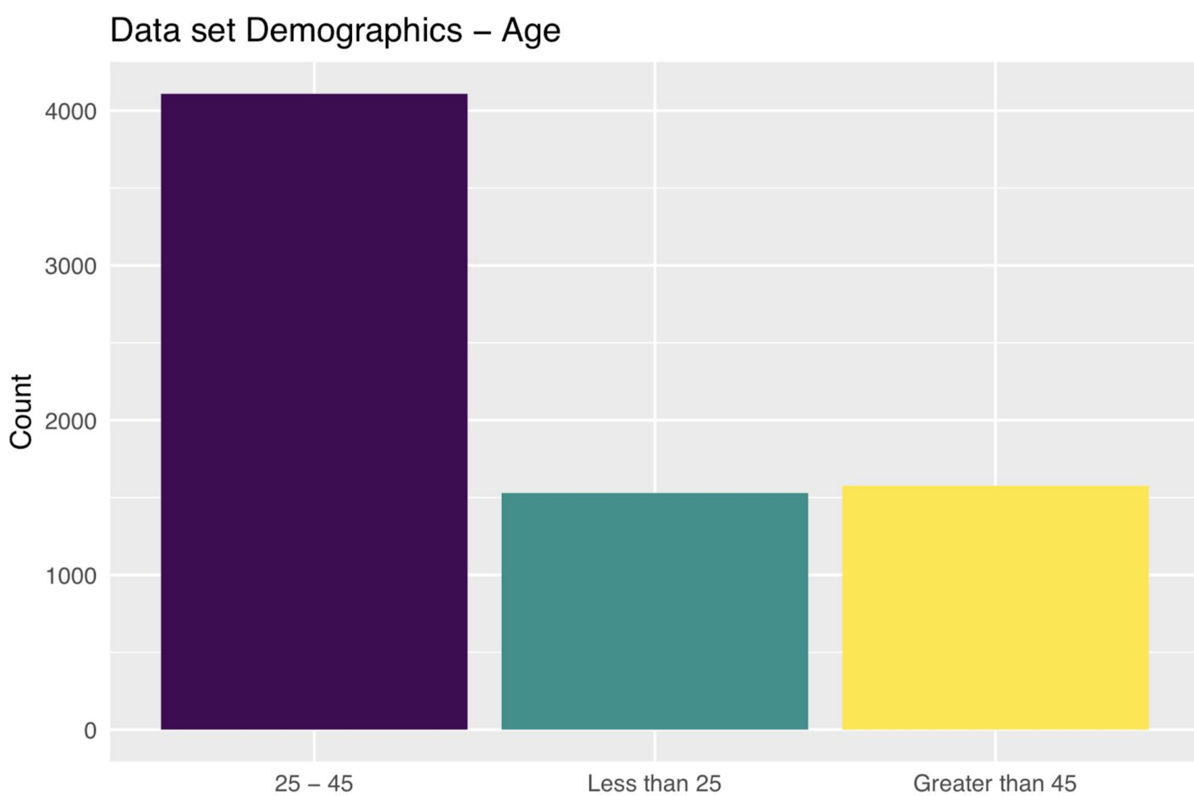
We conducted basic exploratory data analysis



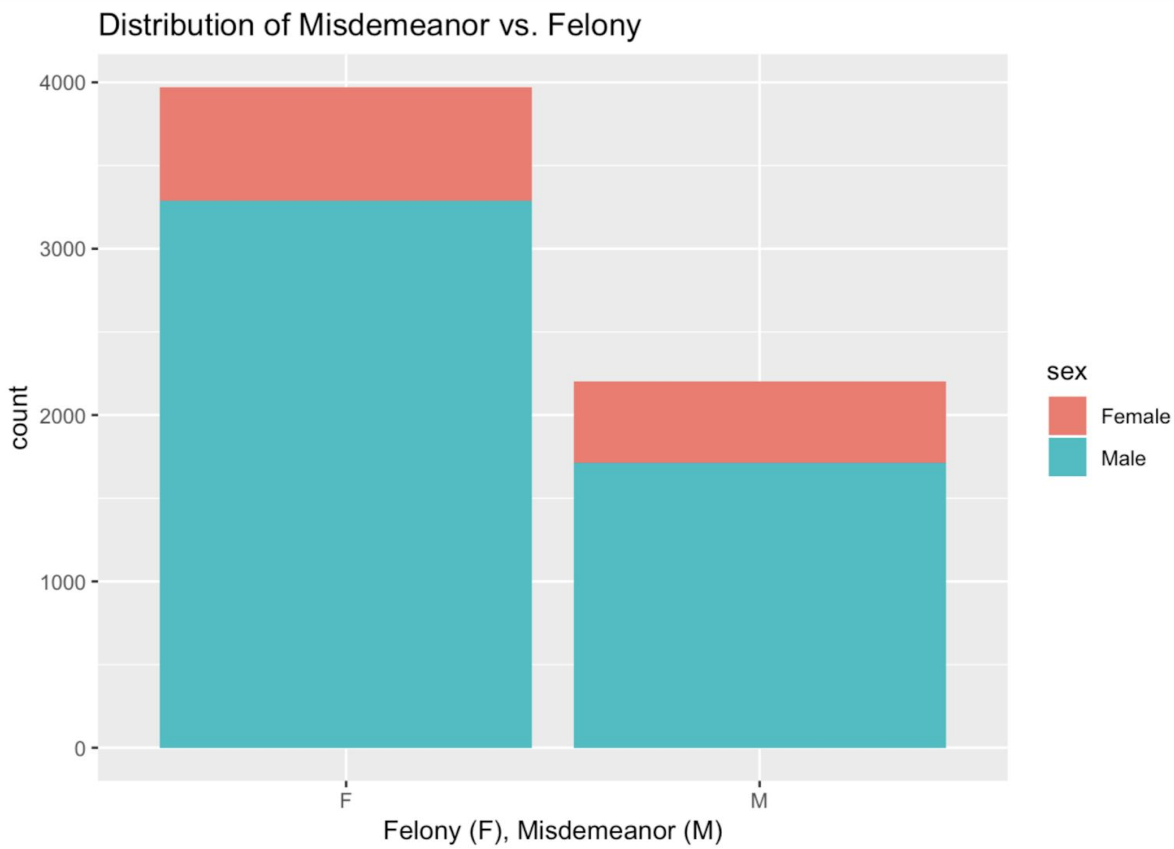

- The people present in the dataset were mostly from the African- American Community, followed by Caucasians.
- People from Native American community were very less.



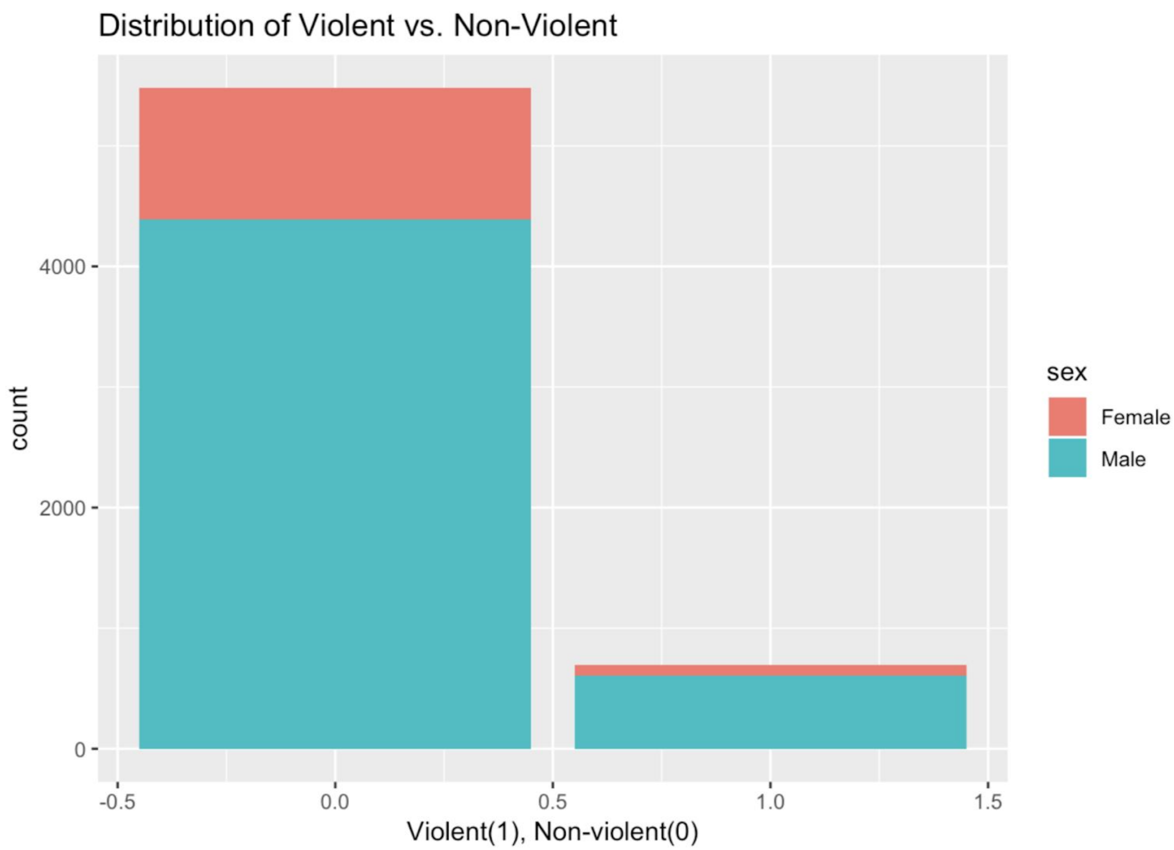
- With respect to the Male female divide, the male population in the dataset is larger than the female population.




- The people in the dataset were mostly in the 25-45 age category, followed by less than 25.




- Let’s look at the distribution of people for Misdemeanor and for felony crime.
- It looks like the felony crimes are more reported and present in the dataset.
- This distribution also shows the male female distribution for the dataset with respect to the type of crime.




- The violent crimes are quite less in number compared to non violent crimes.
- The proportion of violent to violent crimes committed is less for female population than for the male population in the dataset

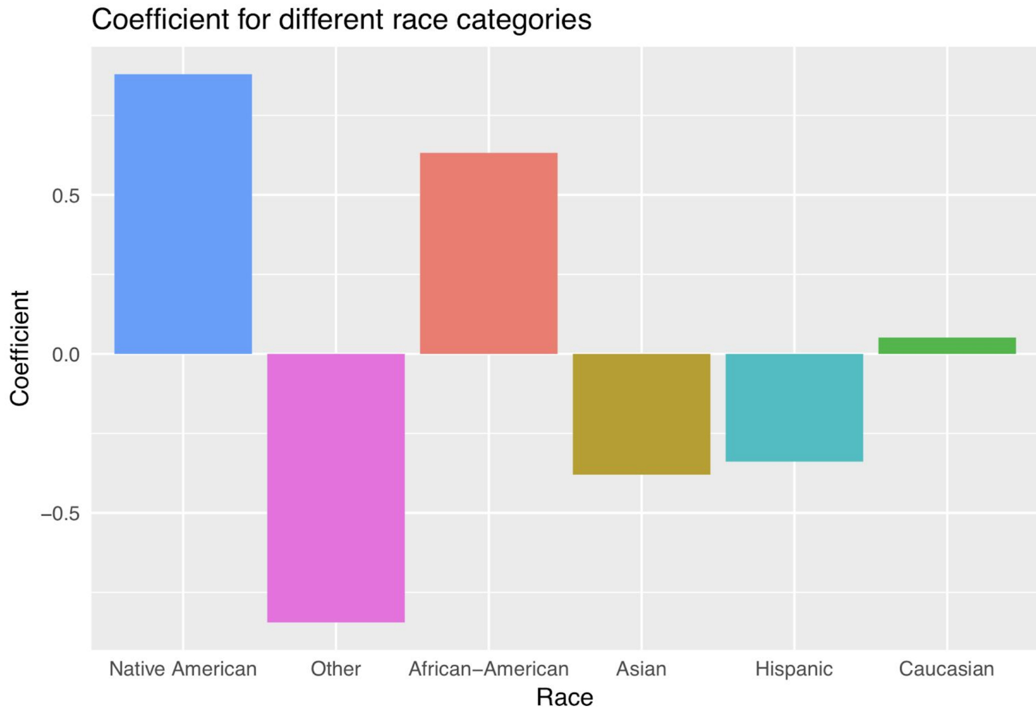
**Findings**

Our analysis reiterated ProPublica’s results

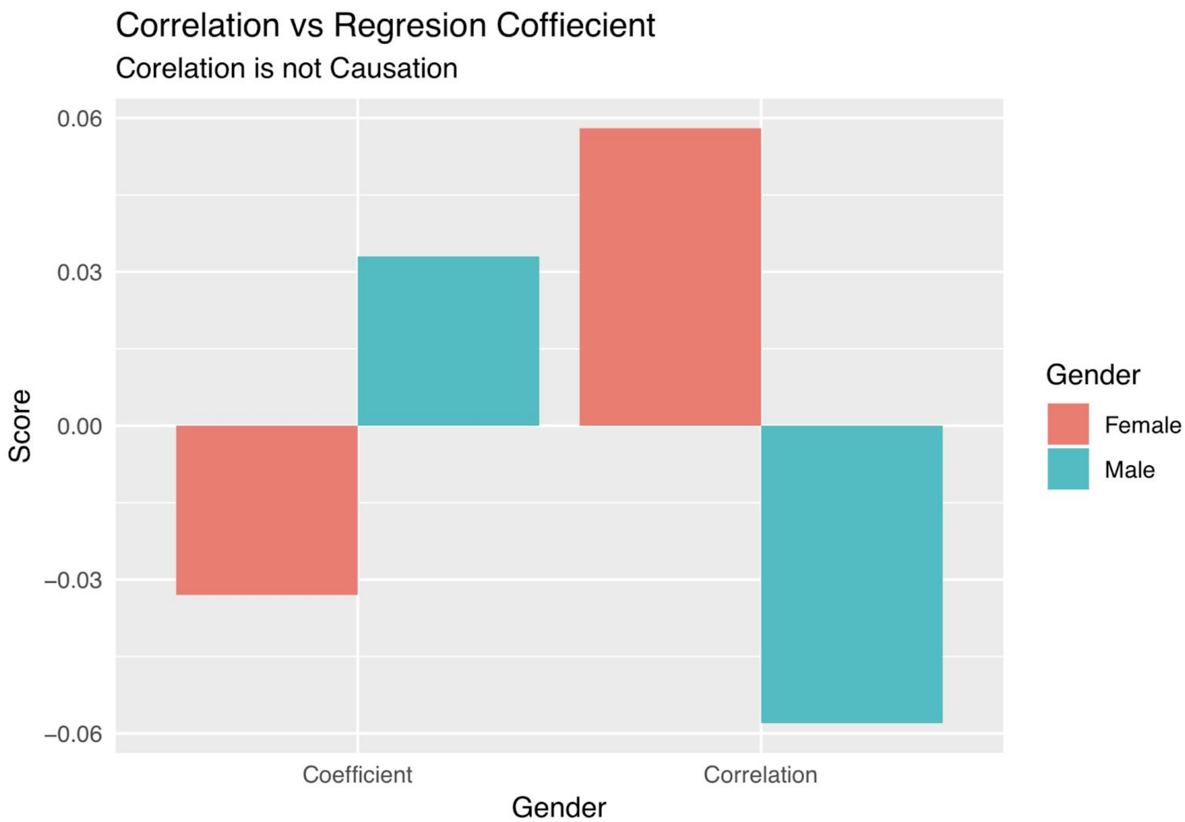
- Focused on gender, age, and race
- More likely to get higher risk score
    - Black
    - Younger
    - Male
- Accuracy test
    - more likely to misclassify a black defendant as higher risk than a white defendant
    - under-classified white reoffenders as low risk more often than black reoffenders

To check what kind of bias Northpointe's algorithm had, a linear regression was conducted by taking the important features from the dataset. The target of this linear regression was set to be the decile_score given by the algorithm. So by doing this, we wanted to find out which feature was given most importance by the algorithm. The more the absolute value of the coefficient increased, the more the algorithm is biased towards in favor (if positive) and against (if negative).

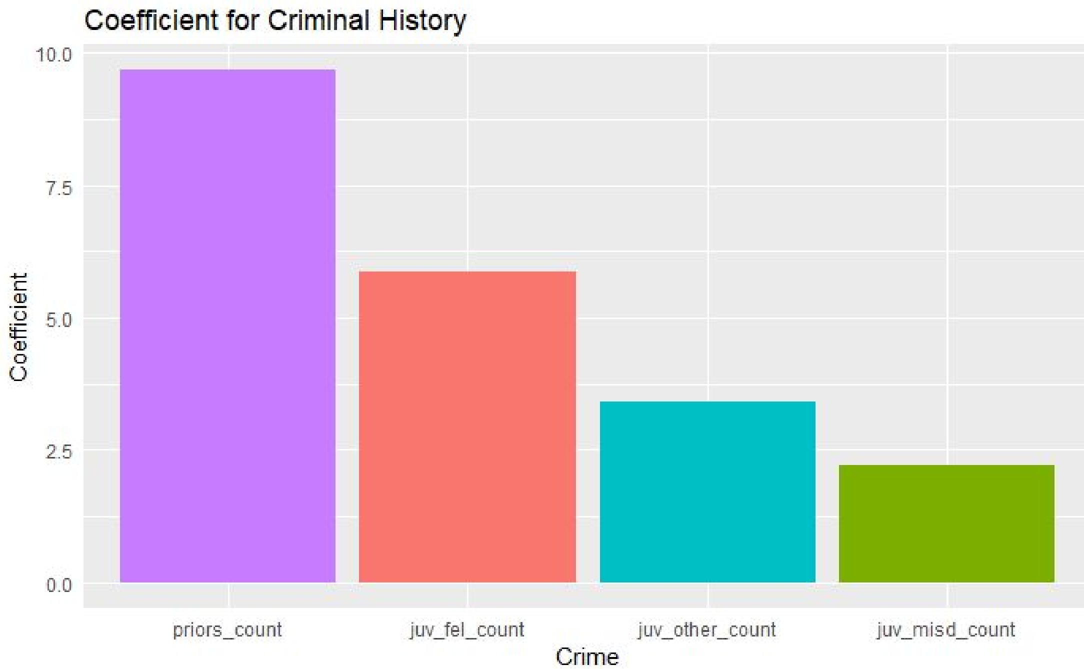
The data was normalised using a min max normalisation using python's scikit learn package. The linear regression model was also run using the same model.
Let us look at the findings.




- For the Race category, The native american community had the highest positive coefficient. This meant that the algorithm was probably biased against the Native American community and being Native American increased the chances of getting a high score.  This is followed by the African American community
- If the person is from other communities, Asian or hispanic, Chances are that the algorithm may give a score less than usual.
- If in the future ,the an algorithm is developed without racial bias, then this graph will be flattened.




- Another important thing that we noticed was on the gender side. Though correlation between Gender and a high score was high against males, the regression model gave a different story altogether.
- The regression coefficient of Male was surprisingly negative and the it was positive for the female population.
- More investigation may be needed before we come to a conclusion on why this is happening

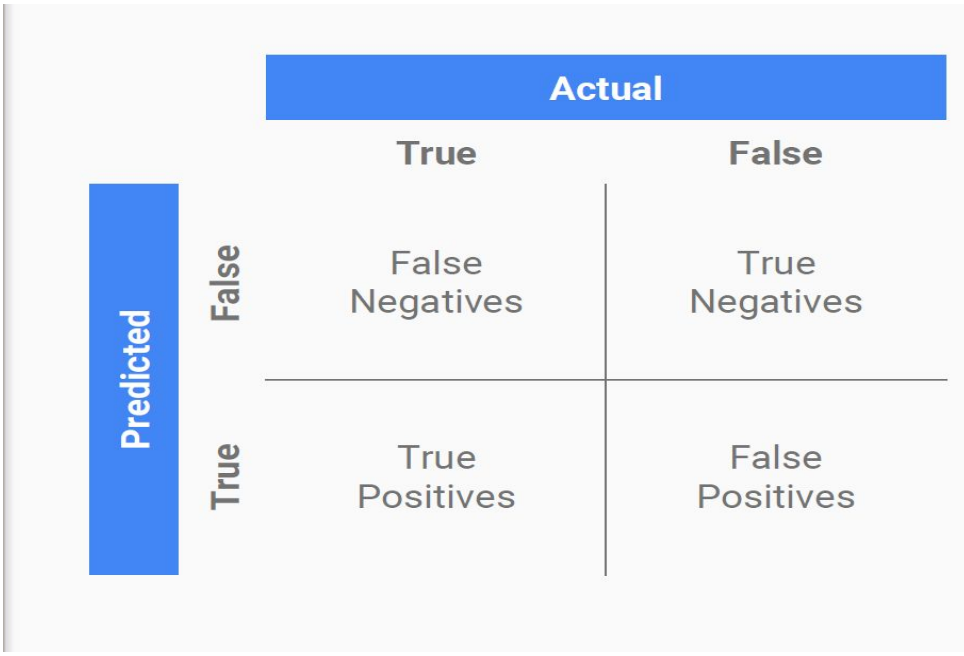



- According to the team, a better model will be look at the Criminal history, rather than the personal traits of the criminals.
- The prior count was a large factor in the scoring process given by the COMPAS algorithm
- We should also note that taking prior count and juvenile count will be biased against communities who are predominantly arrested in their juvenile years, which is sometimes biased due to the police system in America.

**Confusion Matrix**

We also looked at the confusion matrix which is a common visualization of model performance in machine learning. The name stems from the fact that it makes it easy to see if the system is confusing two classes.

We also looked to some commonly used ratios, defined as follows -
Accuracy rate = (TP + TN )/ (TN+FN+TP + FP) : Of all cases, how many do we accurately predict
False Discovery Rate = FP/(TP + FP) : Of people we predict to re-offend, how many do not
False Omission rate = FN/(TN + FN) : Of people we predict not to re-offend, how many do re-offend
False Positive rate = FP/(TN+FP) Of all the people who do not reoffend, who were predicted to do so




- Accuracy test was based on the COMPAS score. The Scenario that we took was to set the  medium to high risk as positive indication of recidivism or ”will reoffend” by the system and, and low risk as negative indication of recidivism “or will not reoffend” by the system.
  Now, let’s look at the individual cases. Note that the rows indicate Predicated and columns indicate Actual as given in the above figure. We will look at discrepancies within Race, gender and age

- Overall model

| Type | Positive | Negative |
|---|---|---|
| Positive | 2140 | 1177 |
| Negative | 1331 | 2566 |

  - Accuracy = 0.65
  - False Discovery Rate = 0.35
  - False Omission Rate = 0.34
  - False Positive Rate = 0.31

**Race**

Let us look at the bias of the model between the African-American community and the Caucasian Community. This was one of the main findings of propublica.

- Confusion Matrix for African-American community

| Type | Positive | Negative |
|---|---|---|
| Positive | 1445 | 729 |
| Negative | 591 | 931 |

  - Accuracy = 0.64
  - False Discovery Rate = 0.34
  - False Omission Rate = 0.39
  - False Positive Rate = 0.44

- Confusion Matrix for Caucasian Community

| Type | Positive | Negative |
|---|---|---|
| Positive | 523 | 331 |
| Negative | 502 | 1098 |

  - Accuracy = 0.66
  - False Discovery Rate = 0.39
  - False Omission Rate = 0.31
  - False Positive Rate = 0.23

Though the Accuracy was fairly similar, The False Positive rate for the African-American community was at least twice that of for the caucasian community, which shows the bias of the algorithm (One of the main findings of Propublica)

**Age**

Now lets us look if there was any age related bias of the COMPAS algorithm.

- Confusion Matrix for age less than 25

| Type | Positive | Negative |
|---|---|---|
| Positive | 669 | 330 |
| Negative | 244 | 286 |

    - Accuracy = 0.62
    - False Discovery Rate = 0.33
    - False Omission Rate = 0.46
    - False Positive Rate = 0.54

- Confusion Matrix for age 25-45

| Type | Positive | Negative |
|---|---|---|
| Positive | 1244 | 680 |
| Negative | 776 | 1409 |

    - Accuracy = 0.65
    - False Discovery Rate = 0.35
    - False Omission Rate = 0.36
    - False Positive Rate = 0.33

- Confusion Matrix for age greater than 45

| Type | Positive | Negative |
|---|---|---|
| Positive | 227 | 167 |
| Negative | 311 | 871 |

    - Accuracy = 0.70
    - False Discovery Rate = 0.42
    - False Omission Rate = 0.26
    - False Positive Rate = 0.16

Another major finding was that the False positive rate for people below the age of 25 is 54%. This goes to imply that America is losing more than half of the youth population arrested and sentenced without there being a risk of them actually committing a crime (according to the algorithm). The false positive cases for the youth is more than 3 times than that of False positive for people with age >45 which clearly indicates the bias and ineffectiveness of the algorithm

**Gender**

- Confusion Matrix for male

| Type | Positive | Negative |
|---|---|---|
| Positive | 1821 | 905 |
| Negative | 1120 | 1973 |

  - Accuracy = 0.65
  - False Discovery Rate = 0.33
  - False Omission Rate = 0.36
  - False Positive Rate = 0.31

- Confusion Matrix for female

| Type | Positive | Negative |
|---|---|---|
| Positive | 319 | 272 |
| Negative | 211 | 593 |

  - Accuracy = 0.65
  - False Discovery Rate = 0.46
  - False Omission Rate = 0.26
  - False Positive Rate = 0.31

- The Bias for the gender does not seem that prominent. Though there is difference in the FDR and FOR rates, Accuracy and False positive rates seem to be balance.

**Limitations**

1. Data induced bias.
2. The given data set made it difficult to test our idea of building a scoring system based on factors such as previous crimes and severity of those crimes. While some of these variables were in the data set, more complete data from the Uniform Crime Reports would expand our research ability

**Recommendation**

1. Data analytics based on personal demographics should not be used for cases of recidivism. As seen in the cases for race and especially for the age case, this tends to create a lot of bias. We believe that a person should not remain incarcerated just because he/she is of a particular race or of a particular age group.
2. The risk score should be based on the past actions. But sometimes, the data that tis received for past actions is also biased. Hence, unless a stable and better alternative is provided, this system should be done away with.
3. Crime history (crimes conducted before) should be a stronger indicator. Though this might induce more bias which we would need to check for.
4. Aim for a higher accuracy of the model. A 65% accuracy is a very poor rate to be making decisions on someone's life.
5. More transparency on the method used for calculating ratios - especially people making the decisions (judges or policy makers) should understand the biases in the data and limitations in methodology

**Team Contribution**

The team collaboratively worked on the approach. Ameya and Alex worked on the data analysis. Kulkanya and Nitasha worked on the presentation and policy paper