Examining Criminal Bias in Judicial Systems

MARYLAND
JUDICIARY

Written By
Yashika Bajaj
&
Riya Chanduka

Table of Contents

Overview

This report summarizes the analysis results of the risk assessment tool currently being used by St. Mary's county of Maryland. The data used in the analysis was sourced from the pre-trial risk-assessment tool provided by the county. The report presents how racial and demographic biases can affect the risk-assessment tool at any stage, i.e., development of the tool and its implementation; hence affecting the results of court system.

Introduction

The current judicial system in US requires a defendant to appear for a pre-trial hearing before a judge. Depending on various factors, such as previous criminal records, income, etc., the judge decides whether the person should be granted bail before their actual hearing. However, other factors such as age, gender, and race might introduce bias into the judge's decision. The intervention of technology can help reduce these biases and provide a fair decision for all, bringing America closer to its secular promise. Recently, many courts have started experimenting with machine learning algorithms, often called a "risk assessment tool" to make a probabilistic prediction about whether a defendant is likely to show up in the court for the actual trial if they were to be granted bail in the pre-trial hearing. These risk assessment tools were designed to be used by the courts in order to be fairer and to add scientific vigor to their decisions. However, these tools are not perfect and still have biases in them.

The risk-assessment tool is a questionnaire that generates a score which is the sum of the weights of options chosen by the defendant while filling out the personal information such as, age, gender, race, etc. As defined by the judicial courts at St. Mary's, the detention decision was based on the following points scored by the defendants:

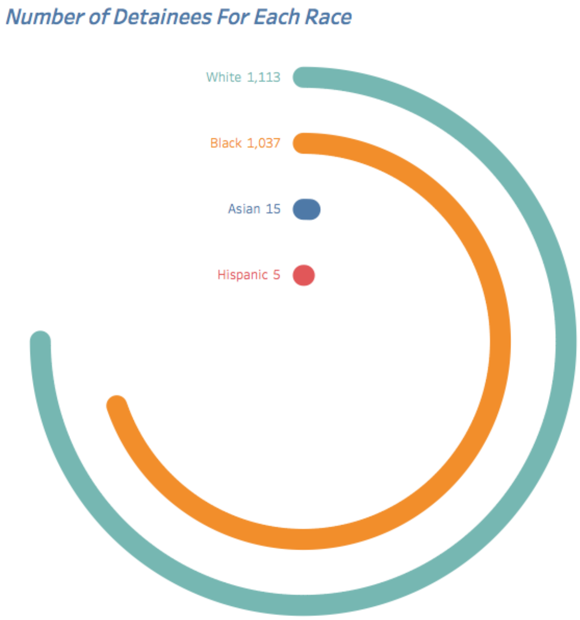
1. Detain - 14 or more point
2. Supervised Released - 6 to 13 points
    a. Level 1 - 6 to 7 points
    b. Level 2 - 8 to 9 points
    c. Level 3 - 10 to 11 points
    d. Level 4 - 12 to 13 points
3. Non-Supervised Release - 5 points or less

Dataset Overview

Defendant's Case Managers enter relevant information in an excel sheet. The source data provided to us are the accumulation of such records. There were four types of cases that the Case Managers handles:

1. Active Pre-Trial Cases
2. Inactive Pre-Trial Cases
3. Pre-Trial Releases
4. Pre-Trial Violations

We created a data corpus by merging the four case types. The final data corpus had 3049 records with 65 columns.



The above donut chart shows the number of detainees for each race for our dataset. The maximum number of defendants detained belong to the White and Black community.

Data Preparation

1. To remove any bias from our analysis we removed rows for which the risk assessment level had less than 5 number of records.
2. The original dataset had risk assessment level, points and extenuating circumstances all merged in one column. To simplify this, we split the column into three separate columns for the analysis
3. We have also simplified our analysis by restricting it only to the judges (top 4) who had the highest number of recorded cases. These judges are, Chesser, Abrams, Stamm and Densford.
4. The analysis showed that most of the cases are from district court. However, all of the district cases had been handled by Judge Chesser. The other three judges have handled mostly Circuit cases.
5. For some cases one defendant had multiple charges against him. To simply the analysis we have taken charge that was the highest level of charge among them.
6. The data corpus contains records from each month of 2016 to 2018. However, data from 2015 and 2019 only contains records spanning over few months. Hence, we have filtered out the years 2015 and 2019 for some part of the analysis.

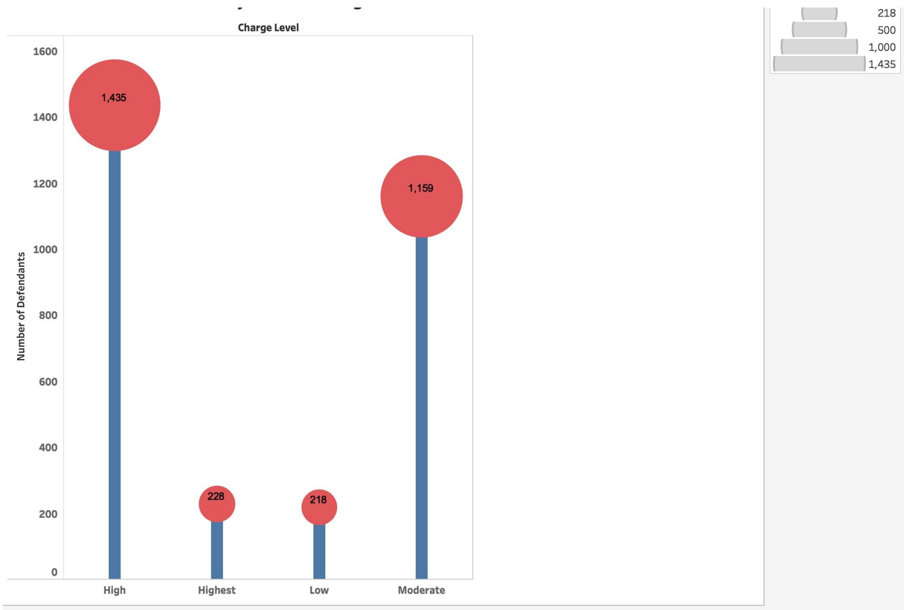
Assumptions

For the scope of this project we had to make various assumptions about the dataset. They are as follows: -

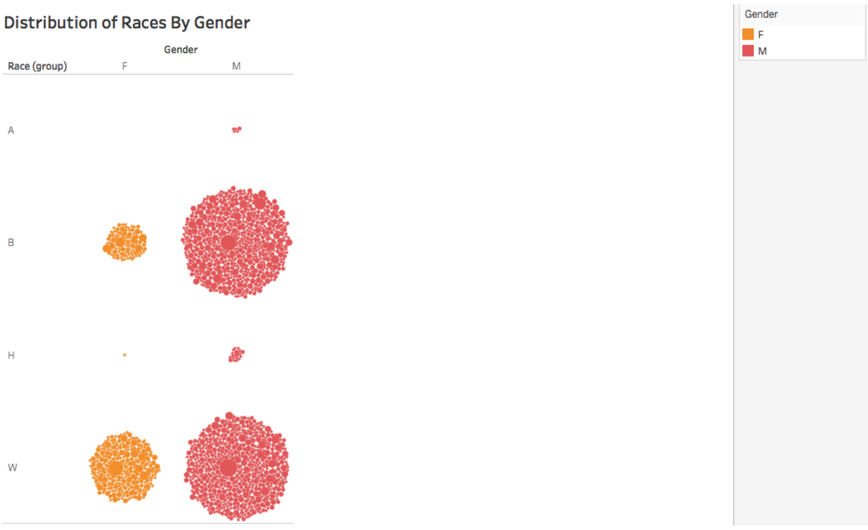
1. Created a new column ‘Charge Level’ by grouping the charges under different categories. During this process, we defined the level according to the main charge for instance if a charge was “vop-xyz” then the charge level was “Moderate” as “vop” is a “Moderate” level charge but there were some exceptions like when the “vop-assault” was a charge then we went with the level of “assault” and considered it to be “High”. We have assumed this for “High” and “Highest” level of charges.
2. For the analysis, we considered only 8 risk assessment levels to do the analysis with. For instance, we removed the “Sent by Judge” risk assessment level.
3. We have substituted missing values of the column “Points” by looking at the Risk assessment tool level given in the dataset.

Data Exploration

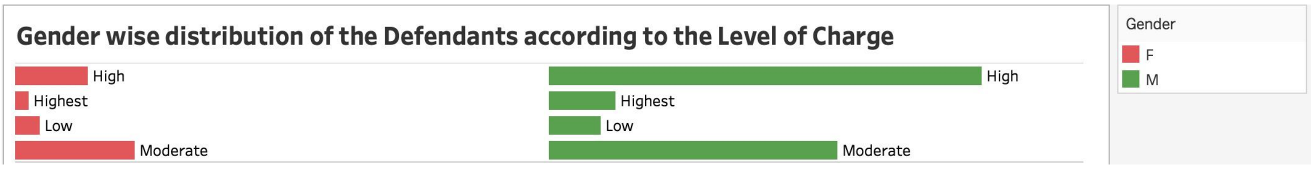
The data consisted of 371 Unique charges that the defendants were charged with. In order to do a more definite analysis, a new column defining the Level of these charges was added, i.e. “High”, “Highest”, “Low”, “Moderate”. These levels have been sourced from the St. Mary’s county report. The below graph shows the number of people with different charge levels. The data has mostly people charged with High levelled charges.



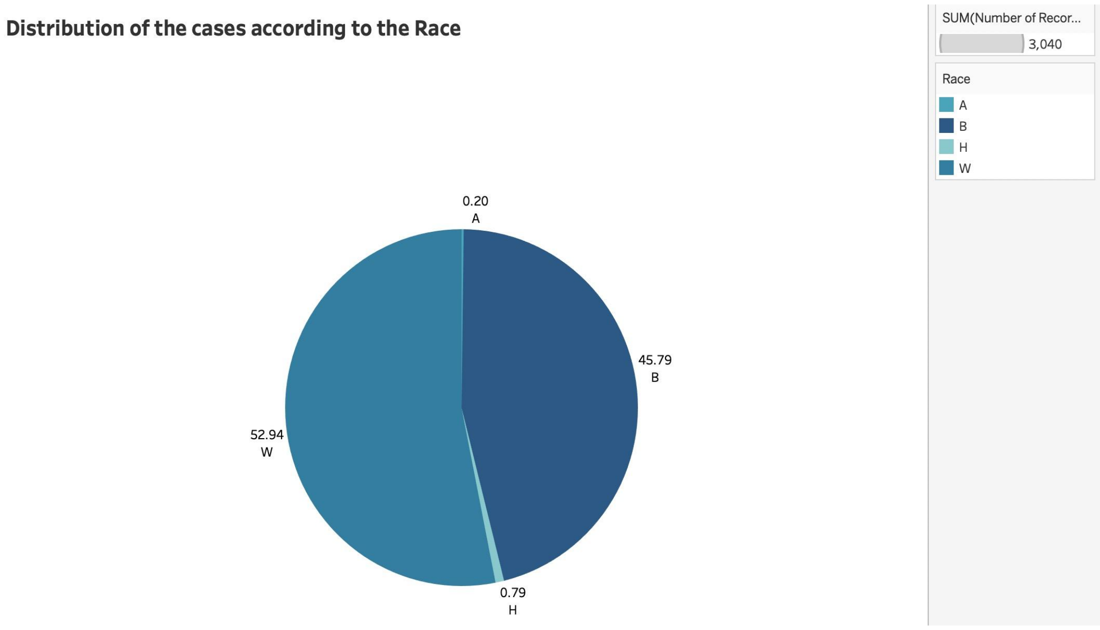
The below scatter plot shows the gender wise distribution of defendants according to different races. It is interesting to see that number of male defendants was greater than female defendants for all the races.



Following graph represents the ratio of the defendants according to their gender depicting that mostly Males were charged with High Levelled charges while females were charged with Moderate Charges.

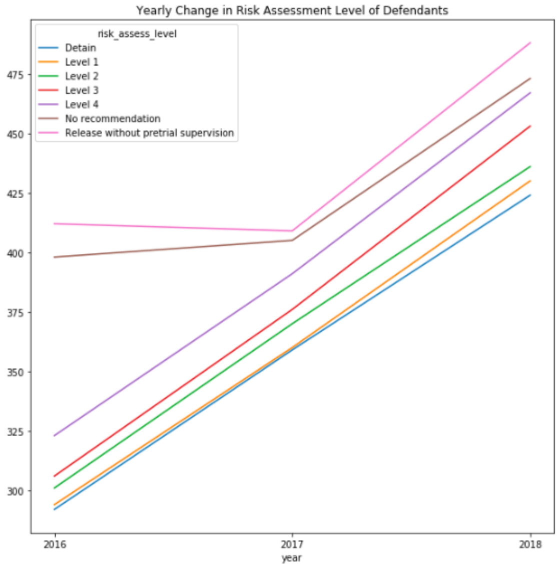


The next pie chart shows the distribution of the number of cases in the entire data per Race.

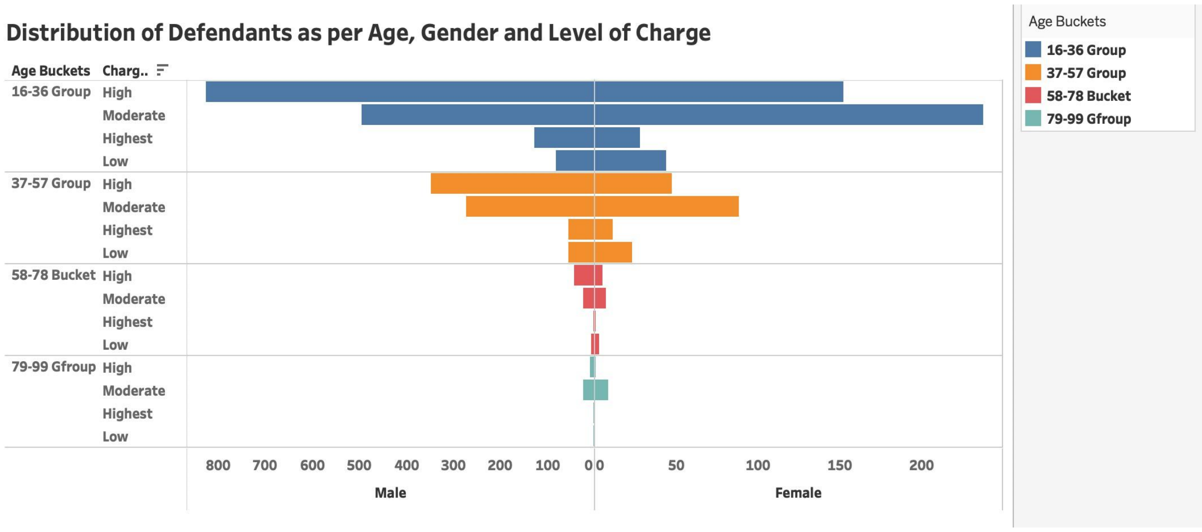


It was also seen that White population had more number of cases registered than the Black population for the year's 2016, 2017 and 2018. Even though we didn't have the complete data for the year 2019, the same trend was observed.

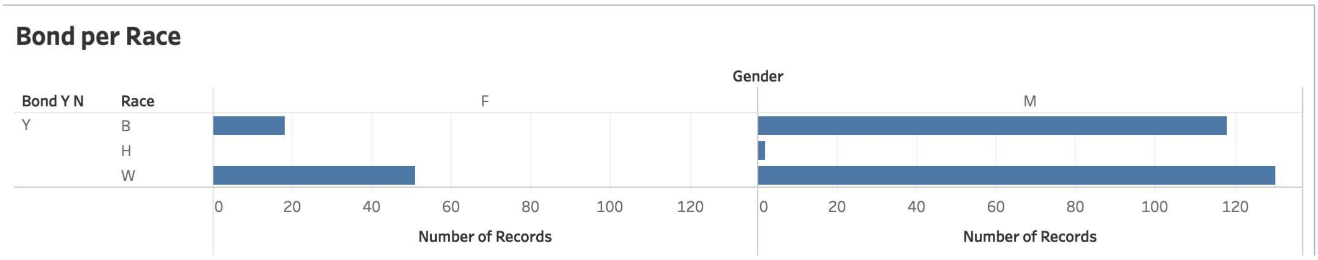
The below stacked line plot shows the yearly change in number of defendants for various levels of risk assessment. It can be seen that the number of detained defendants increase over the years while the cases with level “No Recommendation” and ‘Release without pre-trial supervision’ showed a little dip from 2016 to 2017 but had a sharp increase from 2017 to 2018.



Based on age of the defendants, the data was divided into age buckets. Most of the defendants belong to the age group 16-36. In this age bucket, majority of male population were charged with High levelled charges. The number of defendants decreased with increasing age.



To look at the bond amount more precisely, a new column was made with a Boolean value of Y(yes) or N(no) to see how many records per Race and per Gender were assigned with the bond amount. Interestingly the White Male population were the ones with highest number of bonds.



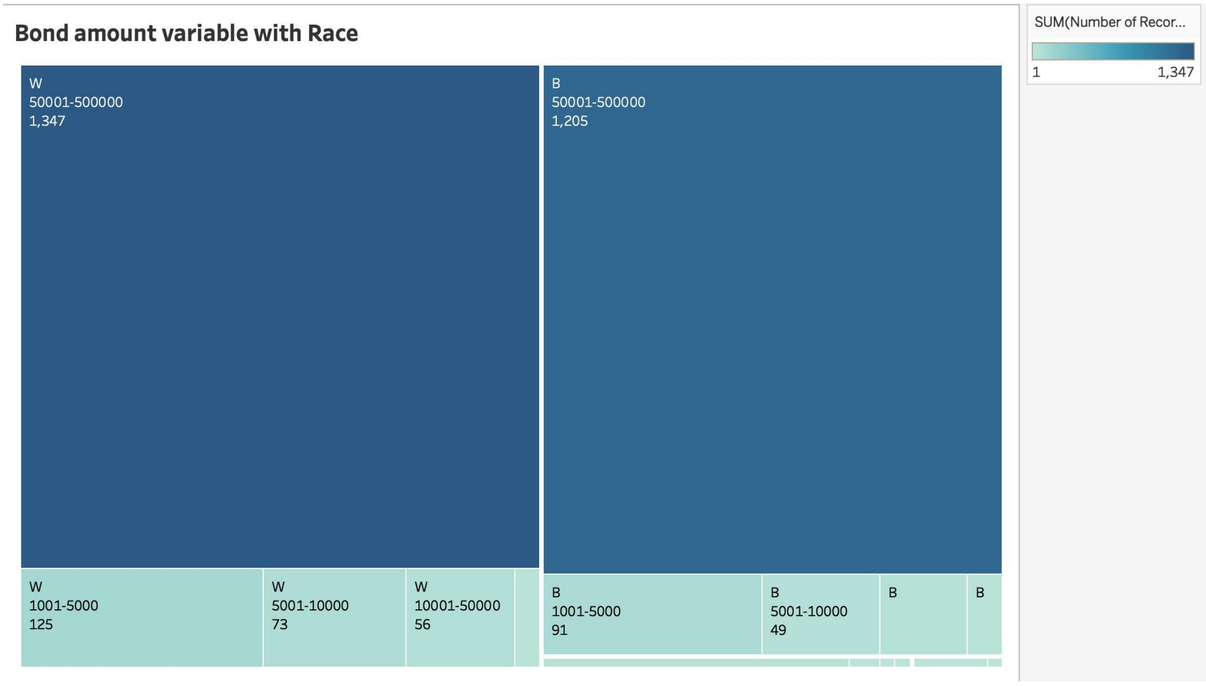
The below table gives information about the judges who ordered maximum number of Bond.

Judges who ordred Bond

Judge Name 1	Y B	Y H	Y W
Abrams	5	1	3
Chesser	75		125
Cooper	3		3
Densford	11		8
Devlin	2		
Duley	1		
Gasparovic	1		6
Harrington	1	1	
Robert Riddle	2		
Robert Riddle OR Robyn R..	2		6
Saunders	4		3
Slade	4		5
Stamm	11		9
Wilcox	1		2

SUM(Number of Recor... 1 – 125

In order to see the difference in the bond amount according to the Race, Bond buckets were created. This treemap shows that the large sum of bonds belongs to the White population.



Exploratory Analysis

The main goal of this project was to find any kind of bias present in the risk assessment tool based on the Defendant's information like race, gender, age, etc. The data had different kinds of offenses that defendants were accused of. St Mary’s county has defined a Severity of Offense Categories which consisted of Highest, High, Moderate and Low offense categories.

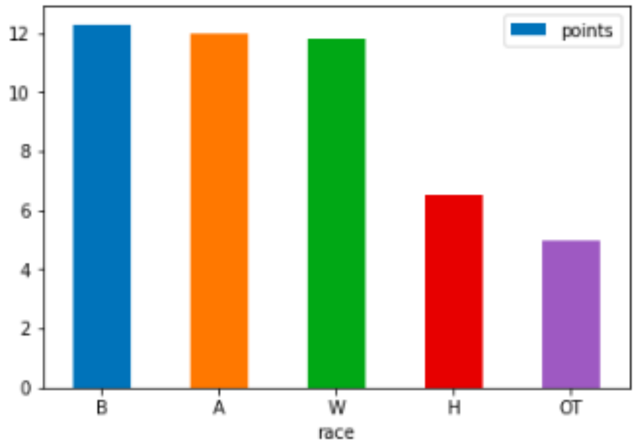
It was observed that for the same offense the risk assessment tool had different recommendations:

1) On the basis of Race- For almost 30% of cases the risk assessment tool had different recommendation for Black and White population
2) On the basis of Gender- For almost 82% of cases the risk assessment tool had different recommendation for male and female

The risk assessment tool calculated points based on the defendant's information and then calculated a score. Based on this score, the tool provided risk assessment recommendation.

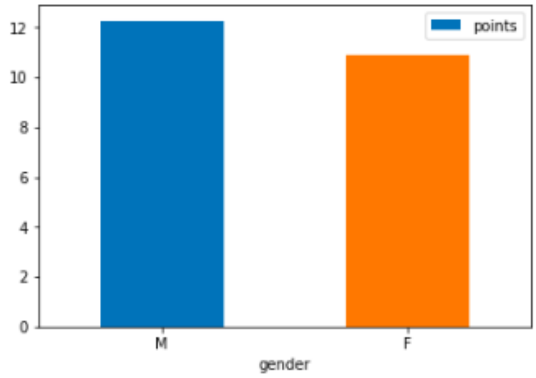
The following table lists out the average points calculated for each race. Black and Asian community had the highest average points.

	race	points
0	A	12.000000
1	B	12.295918
2	H	6.500000
3	OT	5.000000
4	W	11.798973



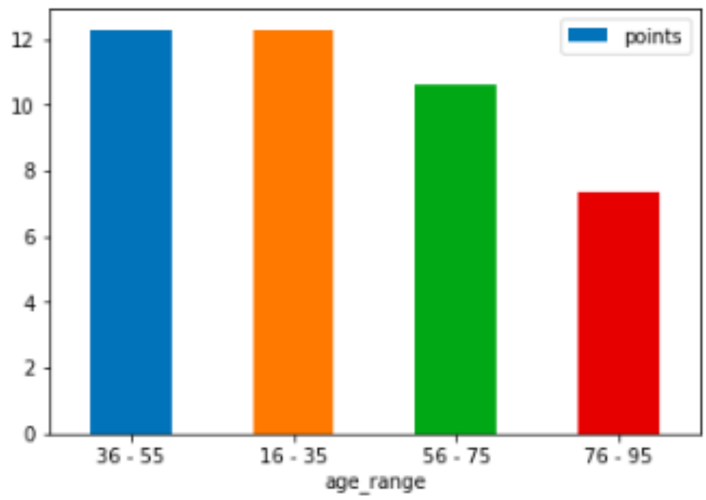
The average points for each gender was also calculated, the below table shows the distribution. It can be observed that male have higher average points than female.

	gender	points
0	F	10.915385
1	M	12.272790

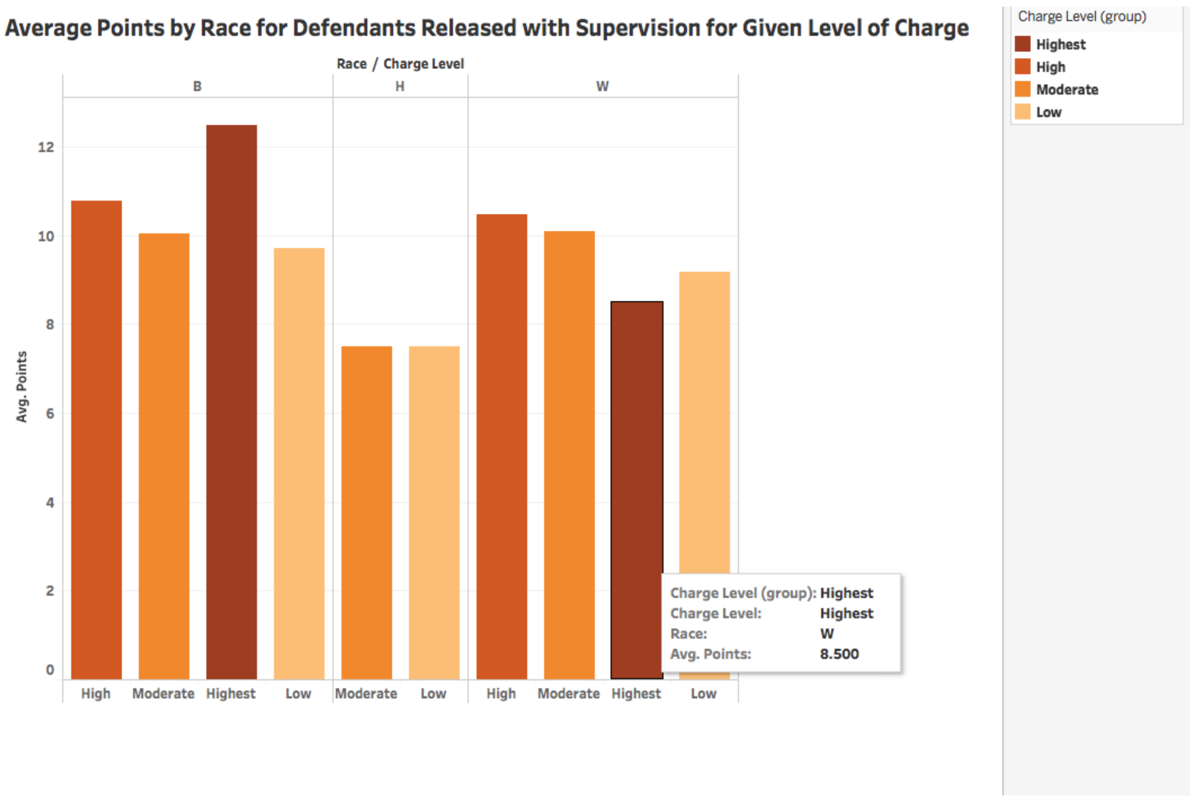


It was also observed that the defendants in the older age range had less average points than the defendants in the younger age range.

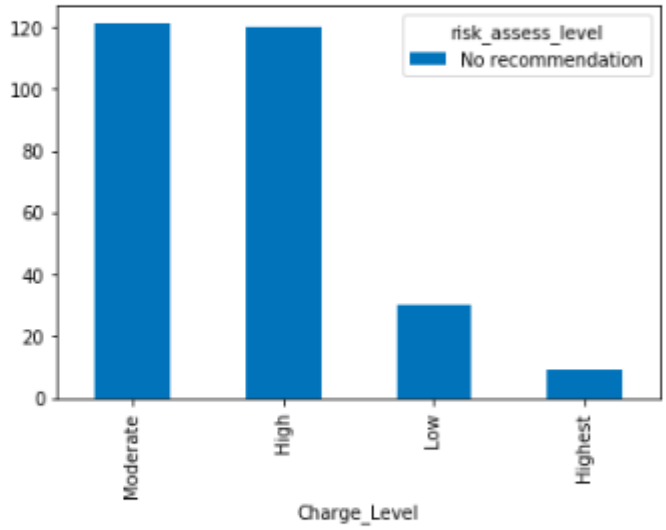
	age_range	points
0	16 - 35	12.249068
1	36 - 55	12.291580
2	56 - 75	10.635036
3	76 - 95	7.333333



For some cases the risk assessment tool recommended that the defendant can be released with supervision and will be supervised according to the assigned level. The following bar chart represents the average points calculated for each race that the tool recommended to be released with supervision.

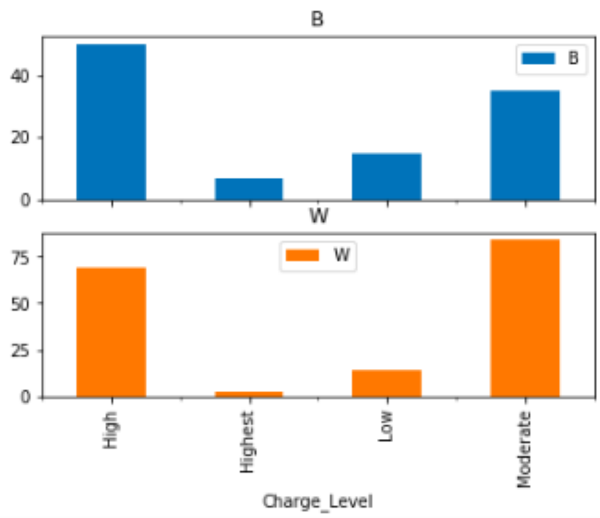


For some cases the risk assessment tool did not provide any recommendation. The following bar chart shows that the maximum cases that had no recommendation by the risk assessment tool was for Moderate and High level of charges.

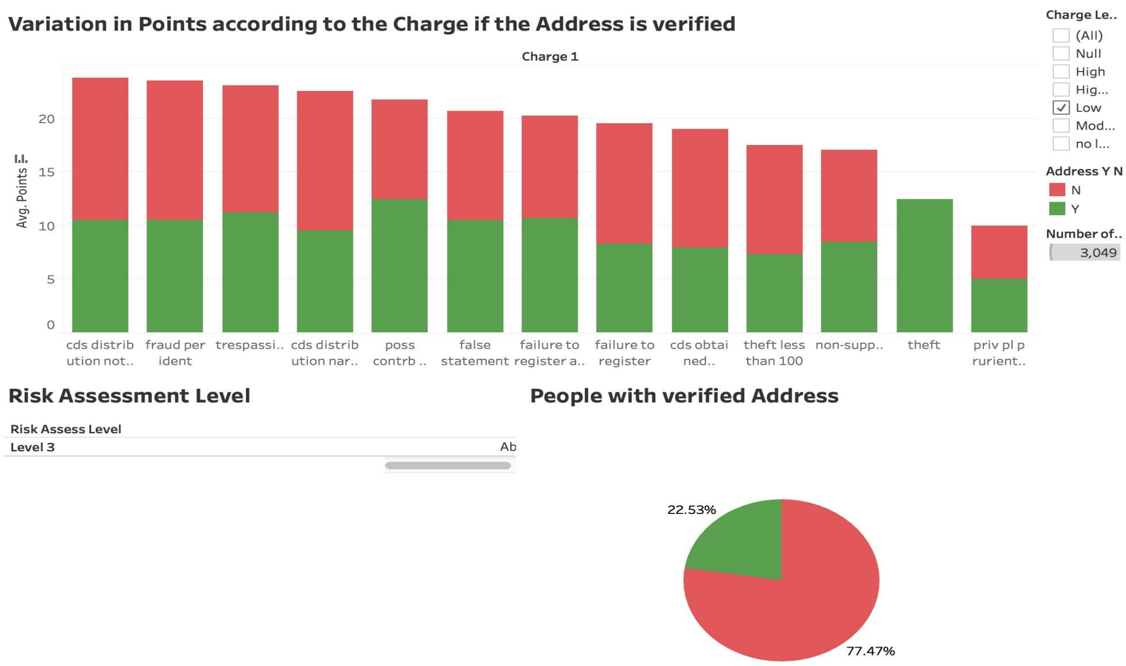


The risk assessment tool did not provide the maximum number of recommendations for Whites as compared to Blacks for Moderate and High Level of charges

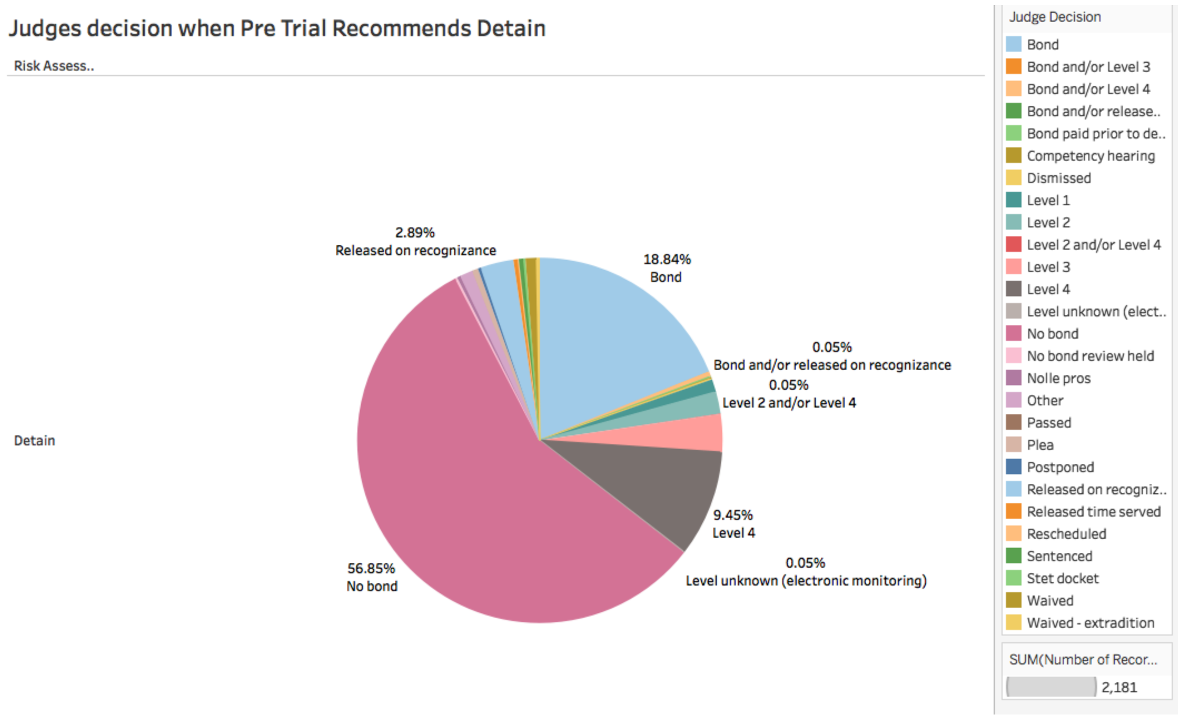
race	B	W
Charge_Level		
High	50.0	69.0
Highest	7.0	2.0
Low	15.0	14.0
Moderate	35.0	84.0



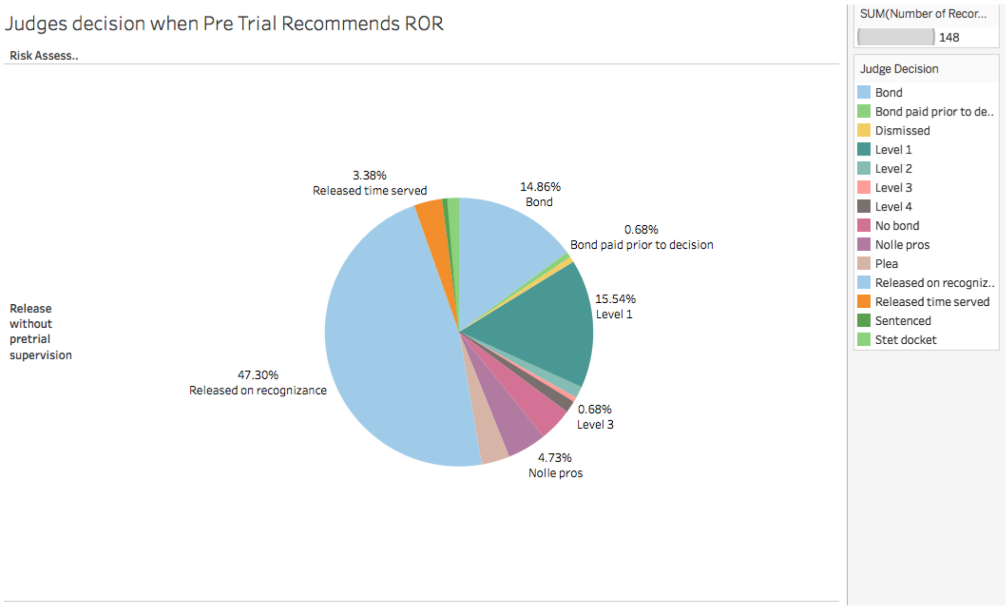
The following dashboard represents the variation in points calculated by the risk assessment tool for defendants when they provided a verified address or not. For most of the cases the risk assessment tool gave a more severe recommendation when the defendant did not provide a verified address. For instance, the following dashboard shows that for low level of charges the risk assessment tool gave recommendation Level 3 for defendants with a verified address and gave Detain as recommendation when the defendant did not have a verified address



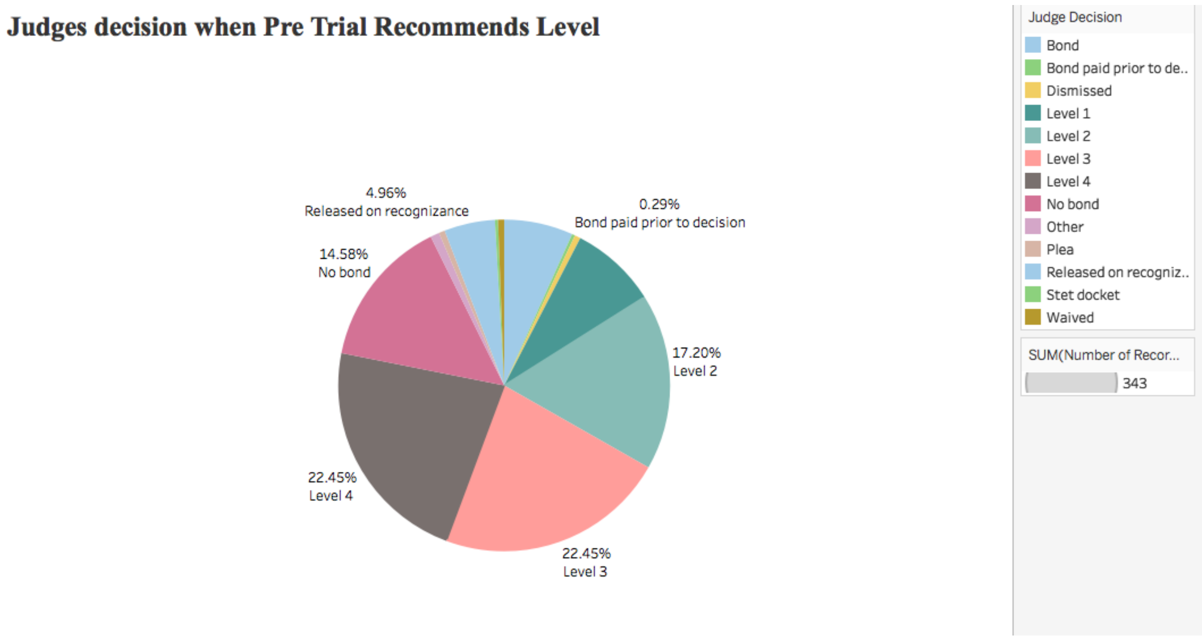
Judges selectively applied the tools' recommendations and passed their own judgements. The below pie chart indicates that when the risk assessment tool recommended Detain the Judges agreed for 57% of cases by deciding that there will be no bond and disagreed with the tool for 19% of cases by giving a lower level of recommendation.



The below pie chart shows that when the risk-assessment tool recommended Released on Recognizance of the Defendant, the Judges agreed for 48% of cases and disagreed for about 14% of cases



The following pie chart shows that when the risk-assessment tool recommended a Level, the Judges agreed for 72% of cases i.e. assigned a level and disagreed with it by giving a stricter recommendation for 15% of cases and disagreed with it by giving a more lenient recommendation for 5% of cases and assigned a bond for 7% of cases.



Key Insights

Through the analysis the following were they key insights that we found

1. For almost 30% of cases the risk assessment tool had different recommendation for Black community and the White community for the same offense that they were charged with.
2. For almost 82% of cases the risk assessment tool had different recommendation for Males and Females for the same offense that they were charged with.
3. For the same offense when Black defendants were recommended to be detained the risk assessment had No Recommendation for maximum number of cases for White defendants.
4. For the same offense when White defendants were recommended to be detained the risk assessment recommended that Black defendants should be Released under supervision with Level Three for maximum number of cases.
5. The risk assessment gave less severe recommendation to Females than Males for the same level of charge. For most of the cases when Males were Detained Females were either Released without any pre-trial supervision or given No Recommendation
6. The risk assessment tool scored on average more points to the age group between 16-35 than to defendants who were in the older age group range
7. When the risk-assessment tool recommended Detain, the Judges agreed with it for 57% of cases and disagreed for about 19% of cases
8. When the risk-assessment tool recommended Released on Recognizance of the Defendant, the Judges agreed for 48% of cases and disagreed for about 14% of cases
9. When the risk-assessment tool recommended Release with supervision and assigned a level for it, the Judges agreed for 72% of cases and disagreed giving a stricter sentence for 12% of cases

Conclusion

Through our analysis, we can say that the tool is sometimes biased towards one race, gender and age group. We think that the biases are mainly because of the data that has been used to train the model used for the prediction however, it would be wrong to say this as a fact because we have no knowledge of the input data or the scoring algorithm. Also, the biases may be present because some factors have been taken as mitigating factors by the tool. This is in a way introducing some inbuilt bias to the model which is in turn affecting the features being used in training the model. While using probabilistic and machine learning algorithms has the potential to make risk assessment decisions fairer it is not a definite way to remove all human biases because in the end it is the human that is designing this tool and implementing it according to his understanding. More detailed analysis is needed to make sure that these tools are fair to people of any cast and creed.