Examining Criminal Bias in Judicial Systems

User Manual

Written By
Yashika Bajaj
&
Riya Chanduka

Tableau Manual

The data exploration and initial data analysis was done using Tableau. Different versions of data has been used for making visualization in Tableau

The analysis can be replicated for similar kind of data collected in other counties.The following explain in detail the worksheets, dashboards and storyboard that have been created in Tableau.

Worksheets:

Gender distribution by Race - This visualization shows the cluster distribution of Gender according to each race with SID's of the defendants forming the clusters.

Distribution of the cases according to the Race - This worksheet shows a pie chart for the number of cases for each Race in the data. The worksheet is again filtered for the nulls. The marks are changed to show the percentage of each category in the data.

Variation in Level of Charge by Year - This is an animated vertical bar chart which shows the change in the number of records charged for different level of charges from Year 2016 - 2019.

Distribution of Defendants by Level of Charge - This is a worksheet showing dual axis custom chart of the Number of people we have in our data for the 4 charge levels that we defined namely Highest, High, Moderate and Low as per the charge the defendants were charged with. This chart is called a lollipop chart and the size of the lollipop depicts the number of records. The markings and the colour combinations can be changed from the options under the "Mark" palette.

Gender wise distribution of the Defendants according to the Level of Charge - This horizontal bar graph shows the number of records as per Gender for different charge levels. The type of the graph can be changed from the "Show me" palette. The palette helps in auto identifying the graphs that can be made from the chosen Dimensions and Measures depending on the number of selected headers from each of the categories.

Variation in charge level by Age Group - This graph shows the count of the number of defendants within each Age Bucket for each charge level. For this sheet we made age buckets as a calculated field and used that.

Distribution of Defendants as per Age, Gender and Level of Charge - This sheet has a custom butterfly chart showing the distribution of the defendants as per age, gender and level of charge.

People with verified Job Information - As verified job information and Address were considered the mitigating factors for the risk assessment tool, we wanted to check that this was really the case. Hence, we checked the data for such instances. This sheet shows a packed bubble chart to identify the cases were the job information was verified.
People with verified Address - As explained above this worksheet showed a pie chart for the Address.
Distribution of Points per Race and Gender - This is a vertical bar graph showing the distribution of average points as per Race and Gender. The two dimensions are also in the filters as we have filtered out the null values from both. The filters can be edited for more close look.
Average Points by Race for Defendants Released with Supervision for Given Level of Charge - This graph shows the variation in the average points for the defendants who were released with supervision i.e. the risk assessment level gave "Release with PTS" as the judgement. The variation was shown depending on the Race and Charge Level.
Variation in Points if Address is verified - Now, coming to points we wanted to see the differences for the defendants who had verified address. Therefore, for high level analysis we made a horizontal stacked bar graph showing the difference in points for various charge level for both the cases.
Variations in the risk assessment levels by Race - This is one of the sheets used in one of the dashboards later explaining the variations in the judgement passed by the risk assessment tool for different Races.
Number of defendants for each Risk assessment level - We have shortlisted 8 risk assessment levels for our analysis. In this dual axis horizontal lollipop chart we are visualizing the number of records in the data for each of them.
Number of Detainees For Each Race - This is stacked donut chart that shows the number of detainees for each race. For this chart the data had to be manipulated therefore, we prefered to work on another copy of the data and called it Donut_Chart.
Number of cases as per the court - The data showed all the cases as a part of different type of courts. We wanted to identify the types thus made a pie chart and found that the data had most of the cases from District court.
Bond per Race - There was also some columns in the data depicting the bond amount as per the judge's decision. We wanted to see the difference in the number of defendants of each gender and race who were charged with bonds. The Yes and No bond buckets were created directly in the data and not in Tableau.
Judges who ordered Bond - The next worksheet showed a table which shows the names of the judges who ordered bonds as per Race. The table defines the number of instances for each of the categories.

Bond amount variable with Race - This sheet shows a treemap showing variation in the number of defendants with different bond amounts. We made the bond buckets as per the bond amount for each Race.

Judges decision when Pre Trial Recommends Detain - This is a pie chart showing the variation in the decision of the Judges when the risk assessment tool recommended “Detain”.

Judges and the Variations - This is an interactive dashboard consisting of 3 unexplained worksheets after the last worksheet. One is a donut chart essentially explaining the variations in the Risk assessment level with Race. We can choose Race and the chart arranges itself accordingly. This is a high level view of the other two worksheets. The bar graph shows the top 4 judges with maximum cases and the variation in the risk assessment levels for the cases they handled by Race. When you hover over one bar of this bar graph. The table in the other worksheet just below the above explained worksheet changes showing the Judge’s decision and charge level for the same. This interactively changes for each bar.

Variation with Age group and Gender - This is a dashboard with 3 worksheets. One already explained earlier. One of the worksheets shows the stacked bar graph for top 4 judges, the risk assessment tool’s decision and the information about the gender of the defendants for each of the case. Now if we hover over any part of the stack, the adjacent table shows the judge’s decision and the Level of charge.

Difference in Points with Address Verification - This dashboard again consists of 3 worksheets. A non interactive pie chart for the defendants who have verified address. A stacked bar chart of charges with the ability to be filtered according to the charge level depicting the average points if the address is verified and not verified. For the same charge if we hover over the Yes and No part we can see the risk assessment level given for each charge from the table in the sheet below.

Data Exploration - Finally, this is a story that we had created out of all the sheets and the dashboard to have a quick look of whatever we have done in Tableau for future use.

Python Manual

To find the biases present in the tool the analysis was conducted using python in a web-based notebook environment in Jupyter.

Installing the Jupyter Environment

For a local installation run the following command on the terminal:

$ pip install notebook

Running in a Local installation

Launching the jupyter environment

Launch with:

$ jupyter notebook

Then navigate to the required directory and run the notebooks containing the analysis

In total there are five notebooks that contain the analysis

1. **Compute age group-** Before the analysis the data was structured to the required format. This notebook contains grouping the age column into age buckets and then storing this new column in pandas frame which can be downloaded as a spreadsheet.
2. **Yearly Change in Risk Assessment Level of Defendants-** The date column was used to extract the year from it and then a new column year was created. This notebook contains the analysis and graphs showing the change in Risk Assessment Level of defendant through year 2016 to 2018
3. **Risk Assessment Analysis for all charges** - This notebook contains the analysis of risk assessment tool recommendation according to different racial and demographic variables for detailed offenses that defendants were accused of.
4. **Risk Assessment Analysis according to Charge Level**- The notebook contains the biases found in the risk assessment tool according to different racial and demographic variables according to different Charge Levels. Also contains the analyses according to average points calculated by the risk assessment tool for defendants for the same kind of offense.
5. **No Recommendation By Risk Assessment Tool Analysis** - The tool sometimes did not provide any recommendation for some cases. This notebook contains the analysis when the tool had no recommendation for different level of charges.