Welcome to DAATE

We are team of passionate data scientists working to provide greater transparency into the US criminal justice system

Defense Attorney Advisory Tool for Equity (DAATE)

Jackie Nichols, Hao Wu,
Robert Ling, Song Park

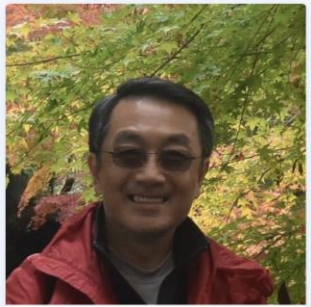
DAATE - Welcome

TEAM

Our Hardworking Team

The team that created DAATE

Jackie Nichols
Chief Architect



Robert Ling
Data Engineer

Song Park
Data Scientist

Hao Wu
Data Scientist

DAATE Mission Statement

Our Mission

To empower legal professionals to realize fairness and equity for their clients by providing transparency into sentencing in the US criminal justice system using data science techniques.

Why DAATE?

Our MVP goal is to provide greater transparency into sentencing of Black and White Americans in the United States (US) criminal justice system, through simple yet impactful analytics beginning first with the Florida Department of Corrections (DOC). In addition to analytics, the DAATE MVP provides statistical guidelines for evidence of sentencing bias through exploring causality, as well as providing easily interpretable, statistical model-based sentence time predictions.

DAATE is intended to be an advisory tool for defense attorneys to gain additional insight about potential inequity in sentencing, and to aid in the attorneys case preparation to represent and to seek fair and equal treatment for their client.

Incarcerations

Approximately **five times more Black Americans are being incarcerated** than that of White Americans

Sentencing

Recent research suggests that **Black Americans receive as much as 19.1% longer sentences** than White American offenders

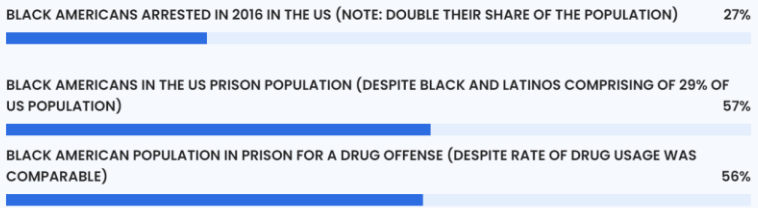
DAATE Potential Impact

1
Million Active Attorney's in the US in 2021

240
Thousand active Defense Attorneys in the US

9
Million Criminal Cases Filed in the US in 2020

41
US States Providing Data



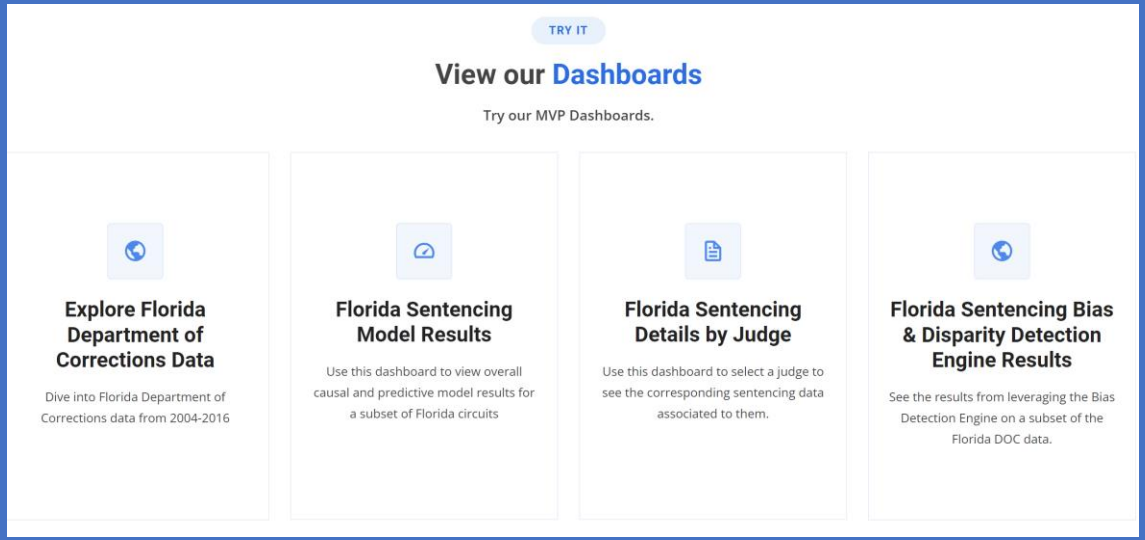
DAATE DEMO

MVP Goals

- Show simple yet impactful analytics for all the data
- Provide statistical guideline for evidence of disparity beyond analytics (causal model)
- Provide easily interpretable, statistical model-based sentence time predictions (predictive model)

Applying User Feedback

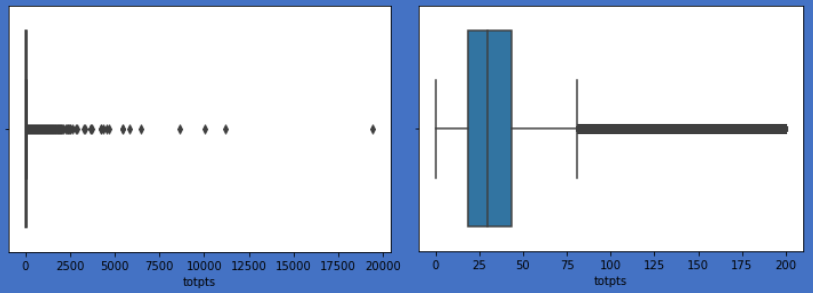
- Importance of including judge-level model results
- Including p-values and statistical jargon for transparency
- Allowing detailed visibility into each judge's sentence history

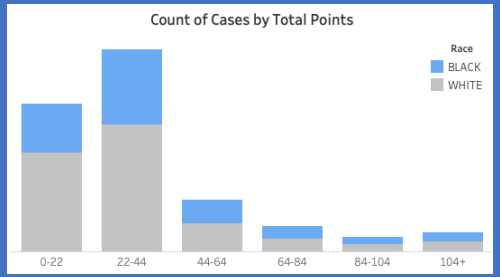


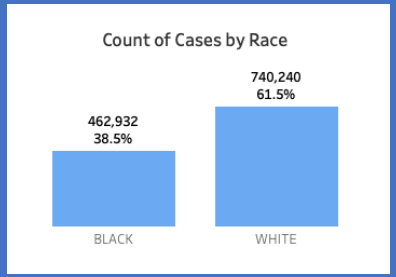
DAATE Data Overview

- Sentencing data from the Florida Department of Corrections
  - 2004-2016
  - Publicly available data
- 1.35m rows originally → 1.23m rows after EDA (detecting unusual data) and identifying outliers
  - Removed 9.34% of data
  - Outliers in total points and sentence time
- Important factors for modeling:
  - Imbalanced data (race, gender, sentencing points)
  - Potential human error in manually entered data

Distribution of Total Points: Before vs After Removing Outliers







DAATE MVP Architecture v.01

Florida Department Of Corrections

STORE

- AZURE STORAGE BLOBS
- AZURE DATA LAKE STORE

PREP/ TRANSFORM/ STREAM PROCESSING

- AZURE DATA BRICKS
- AZURE SQL DB
- elasticsearch
- AZURE ML

MODEL

- elasticsearch
- AZURE SQL DB
- docker
- AZURE ML
- python

APPS & INSIGHTS

- tableau
- Kibana

VISUALISATION

SERVE

- GitHub Pages
- Visual Studio Code

DAATE Predictive Model Architecture

Florida Department Of Corrections

Data

EDA, Cleansing etc.

Feature Selection
+
Feature Engineering

Data transformation
+
Normalization and Regularization

Model Configuration
+
Hyper-Parameter Tuning

huber, lightgbm, ridge regression models

Models chosen with slight loss accuracy for faster training times

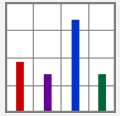
Predictive Results

Assuming no omitted variables

AZURE SQL DB

Model Inputs: Age, Gender, Race, Degree, Points

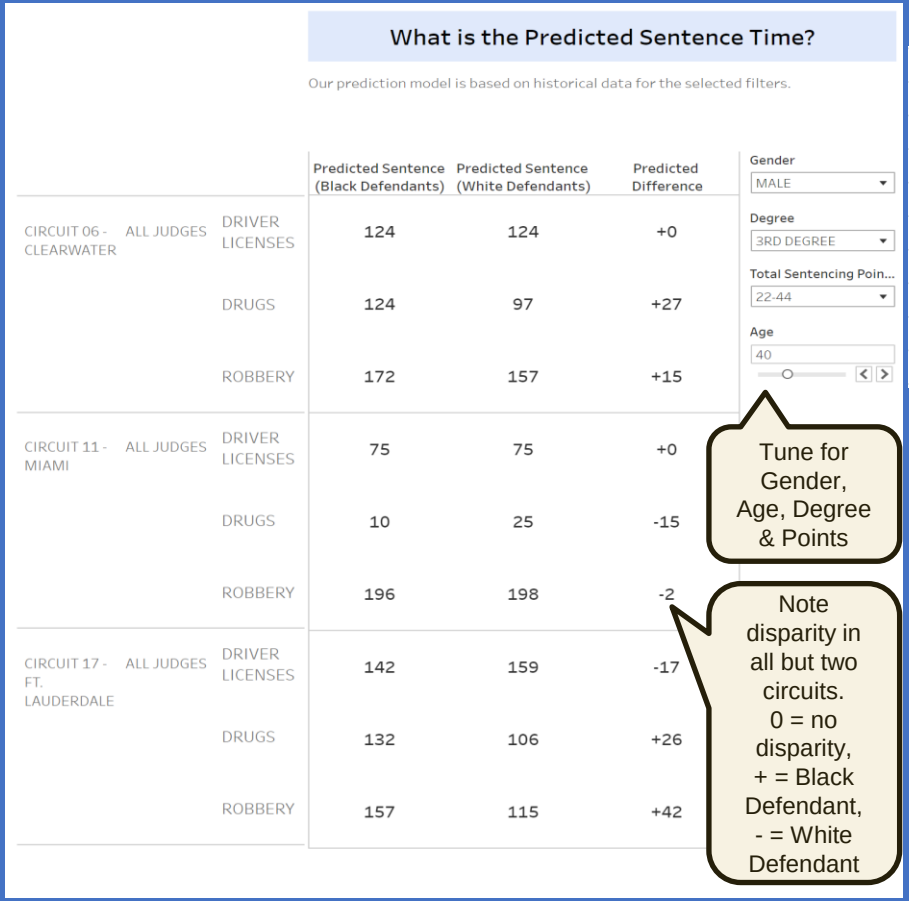
Segmented by Crime Type, geo-locations, Judge



AZURE SQL DB

Model Output: Sentence in Days

DAATE Predictive Model Results



Difference sentencing days Model vs Historical Averages

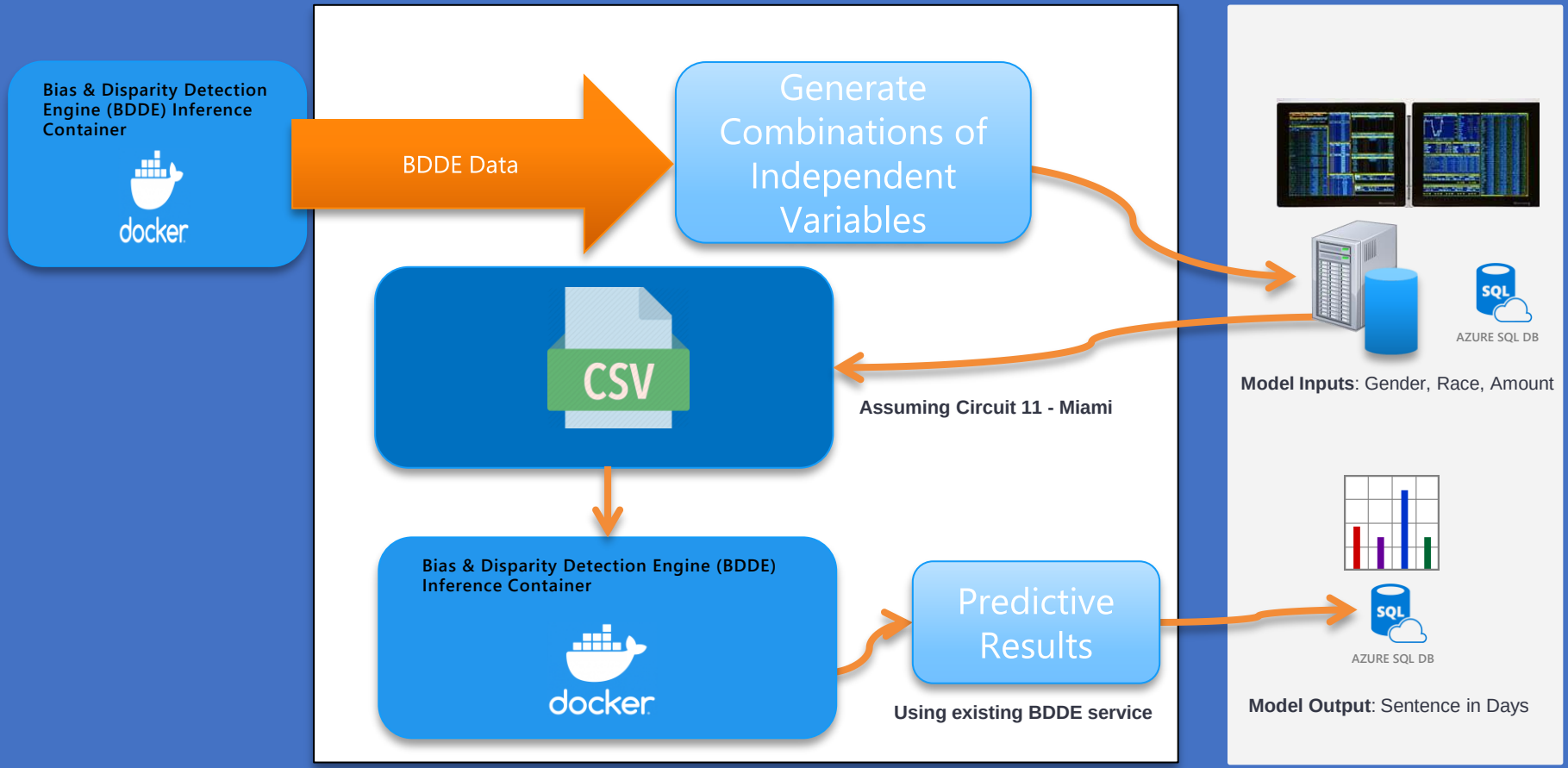
Black Defendant	22-44	44-54	104-114	114-124	194-204
20	-3	-26	-241	-433	-849
40	14	-9	-224	-417	-833
60	23	0	-215	-408	-824
80	30	6	-208	-401	-817
White Defendant	22-44	44-54	104-114	114-124	194-204
20	-22	-26	-51	43	-1506
40	-5	-9	-34	104	-1489
60	4	0	-25	113	-1480
80	11	7	-18	120	-1474

Overprediction

Underprediction

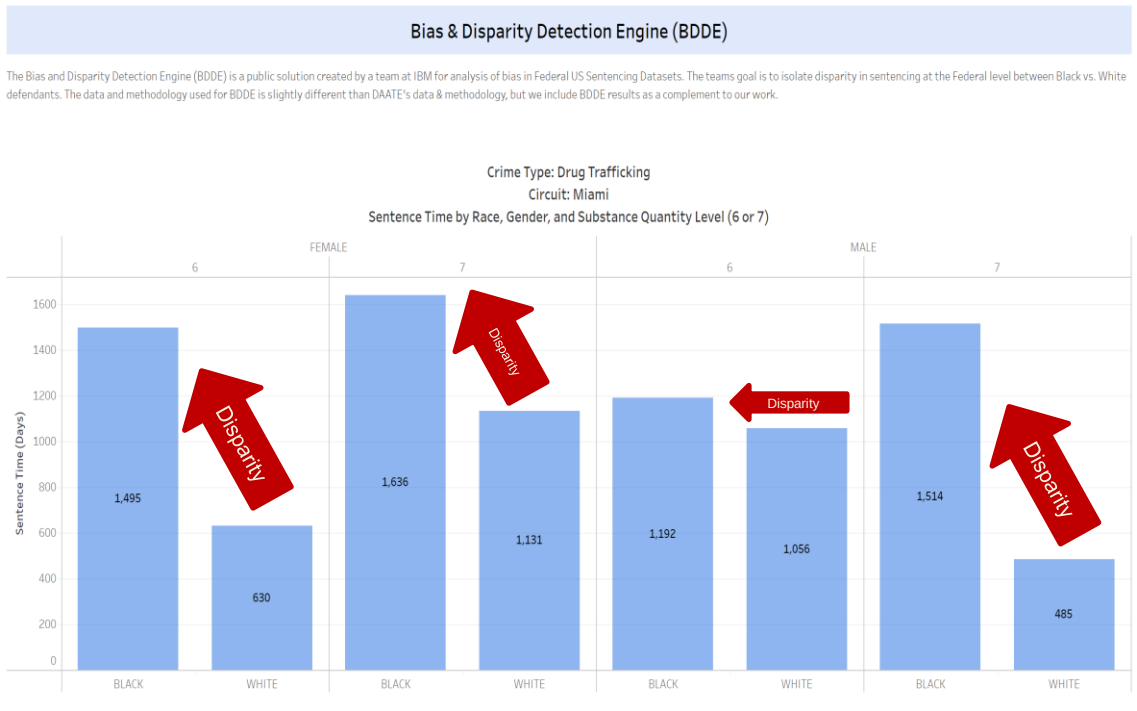
- Accuracy was a goal rather than interpretability: RMSE ~300.
- Errors: **Overprediction** at low end of point ranges with significant **underprediction** at high end of point ranges. Reasons: different effect of point ranges on sentencing, use age for analytics & OVB
- Pre-rendered predictions worked well with Tableau. Challenge was many misspellings of Judge Names (fixed by fuzzy matching & sanity checks)
- Implement the front-end with Streamlit & serve predictions via API

DAATE BDDE Model Architecture



DAATE BDDE Model Results

- IBM pre-trained model on Miami data. We included this as a POC of how other models fit into architecture.
- Results didn't match our predictive model: Reasons included possession quantity as additional variable, lack of age variable
- A small and tidy table of results were generated. Documentation wasn't clear for example 6,7 were the only amounts and units were unspecified
- Retrain the model on our datasets to ensure consistent results



DAATE Causal Model Architecture

Florida Department Of Corrections

Data

EDA, Cleansing etc.

Data transformation
+
Normalization

Distance-Based Clustering
+
Cluster Method Optimization

Subgroup Matching
+
Stratified Sampling
+
Contrast Group Creation

Feature Space Identification

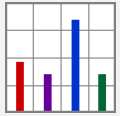
Assuming SUTVA and treatment assignment is independent of potential outcome

Significance Testing

Trade-offs: Bias and Variance, Matching Explainability, Computational Efficiency

AZURE SQL DB

Model Inputs: Age, Gender, Totpts.
Segmented by Crime Type, geo-locations, *Judge

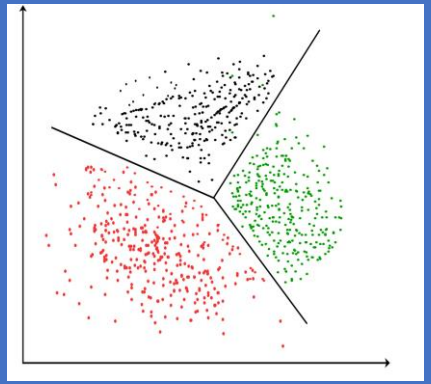


AZURE SQL DB

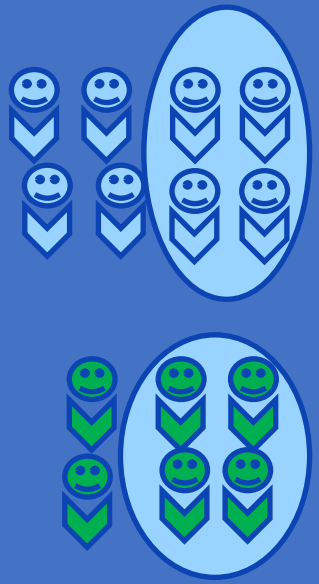
Model Output: Sentence Difference in Days, confidence interval and p-value for the Difference

DAATE Causal Model Approach

Cluster Matching

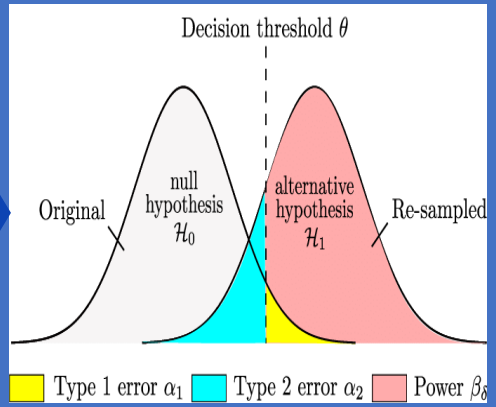


Stratified Sampling



Treatment vs Control

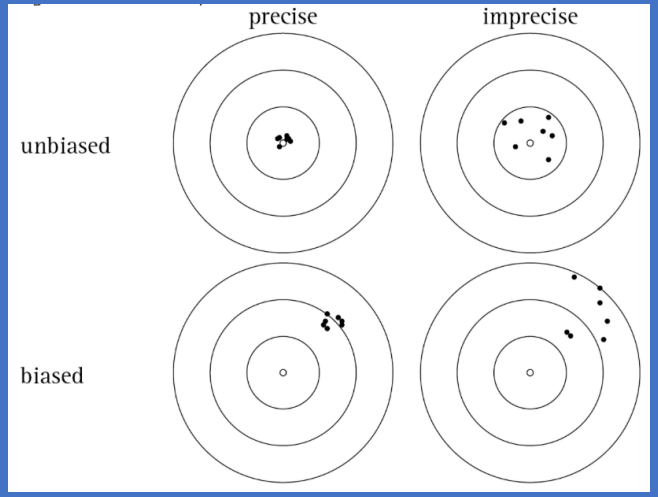
Statistical Testing



DAATE Causal Model Evaluation

Stableness of the Estimator

Unbiasedness of Estimator



Practical Interpretation
&
Explainability

Easy to understand input metrics
- Gender, Age Total Points

Distance based clustering

Sentencing Difference as Outcome
- With estimated lower bound and upper bound for interpretation

DAATE Causal Model Results

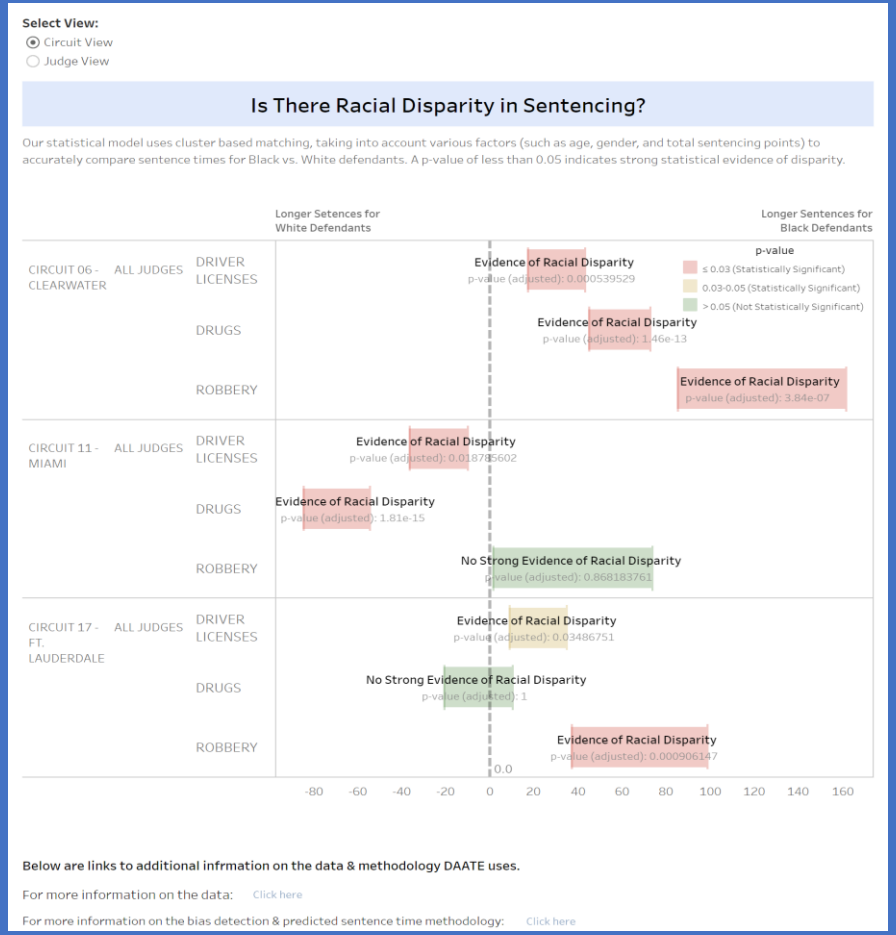
- **Vertical "0" line** - reference for evidence for disparity
- **Range of "colored box"** - estimated interval for disparity"
- **Color codes"** – Indicator for statistical significance

Evaluation of the Results

- Stats test result vs Observation
- SME feedback
- Usability of confidence interval provided
- Visualization Interpretability

Error Adjustment

- Multiple comparison adjustment
- Multiple random sampling and average statistics



DAATE MVP Takeaway and Learnings

Little to no examples of detecting disparity through the use DS

Novel Approach

Many external factors to consider

1.3M rows, 290 columns, 638 crime types

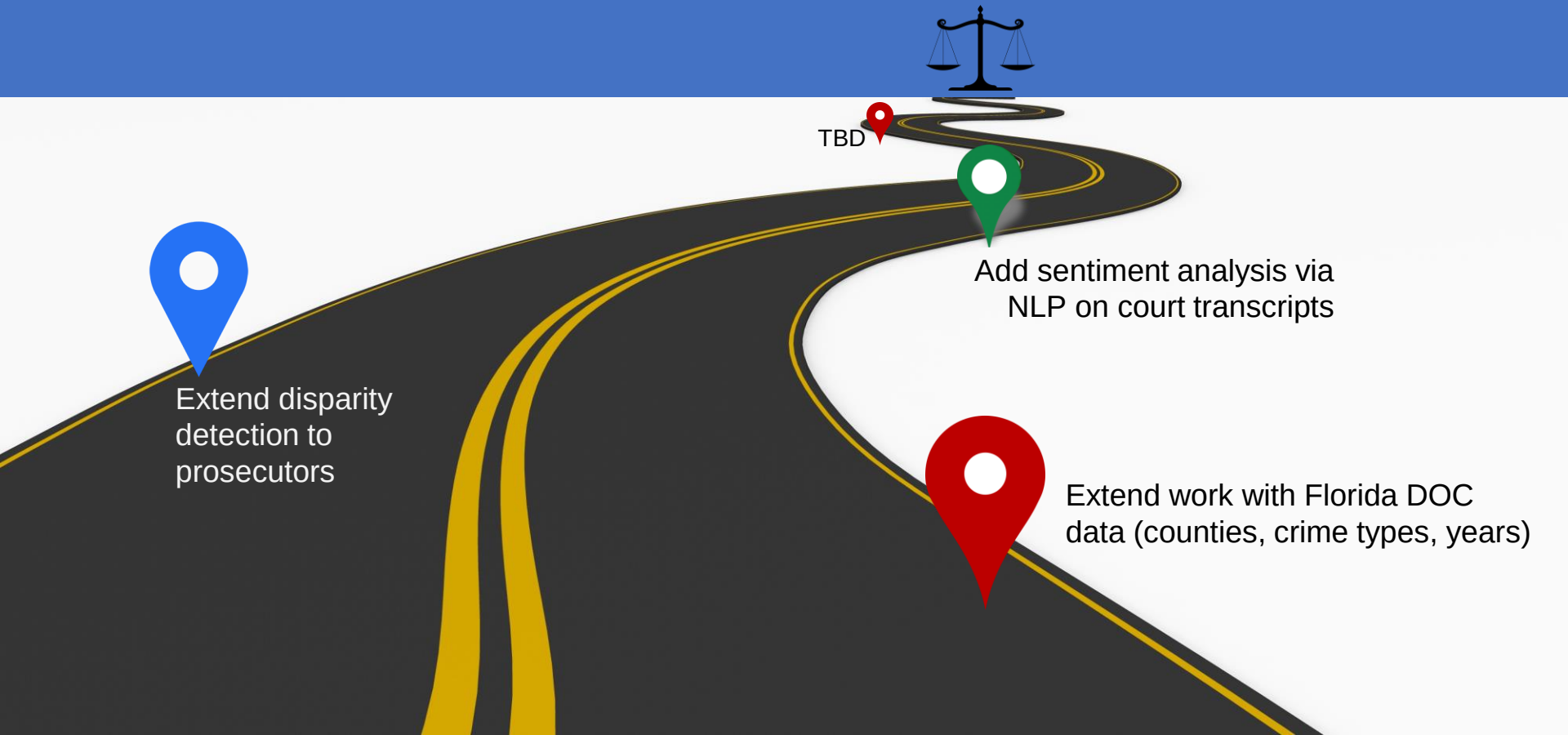
Data Complexity

No Data Dictionary

Legal Terms

Viable UX

DAATE Roadmap and Recommendations



DAATE Summary

Our Mission

To empower legal professionals to realize fairness and equity for their clients by providing transparency into sentencing in the US criminal justice system using data science techniques.

Why DAATE?

Our MVP goal is to provide greater transparency into sentencing of Black and White Americans in the United States (US) criminal justice system, through simple yet impactful analytics beginning first with the Florida Department of Corrections (DOC). In addition to analytics, the DAATE MVP provides statistical guidelines for evidence of sentencing bias through exploring causality, as well as providing easily interpretable, statistical model-based sentence time predictions.

DAATE is intended to be an advisory tool for defense attorneys to gain additional insight about potential inequity in sentencing, and to aid in the attorneys case preparation to represent and to seek fair and equal treatment for their client.

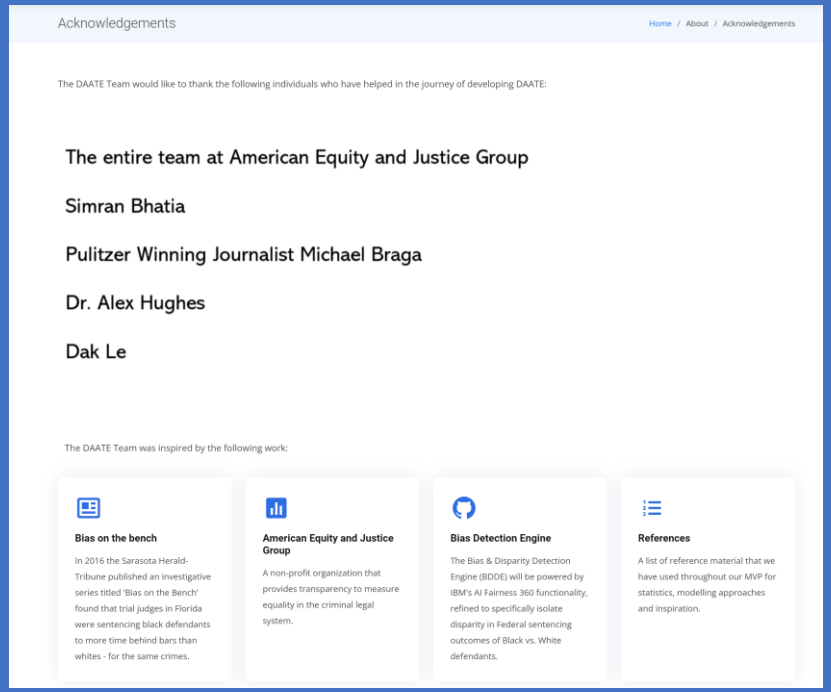
Incarcerations

Approximately **five times more Black Americans are being incarcerated** than that of White Americans

Sentencing

Recent research suggests that **Black Americans receive as much as 19.1% longer sentences** than White American offenders

DAATE Acknowledgements



Acknowledgements

Home / About / Acknowledgements

The DAATE Team would like to thank the following individuals who have helped in the journey of developing DAATE:

The entire team at American Equity and Justice Group

Simran Bhatia

Pulitzer Winning Journalist Michael Braga

Dr. Alex Hughes

Dak Le

The DAATE Team was inspired by the following work:

Bias on the bench

In 2016 the Sarasota Herald-Tribune published an investigative series titled 'Bias on the Bench' found that trial judges in Florida were sentencing black defendants to more time behind bars than whites - for the same crimes.

American Equity and Justice Group

A non-profit organization that provides transparency to measure equality in the criminal legal system.

Bias Detection Engine

The Bias & Disparity Detection Engine (BDDE) will be powered by IBM's AI Fairness 360 functionality, refined to specifically isolate disparity in Federal sentencing outcomes of Black vs. White defendants.

References

A list of reference material that we have used throughout our MVP for statistics, modelling approaches and inspiration.

THANK YOU!

Do you have any questions?

mspuckit1@berkeley.edu
robling@berkeley.edu
hwu24@berkeley.edu
song.park@berkeley.edu