



Defense Attorney Advisory Tool for Equity (DAATE)

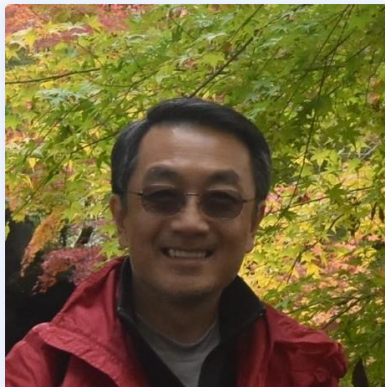
Jackie Nichols, Robert Ling, Song Park, Hao Wu

DAATE TEAM



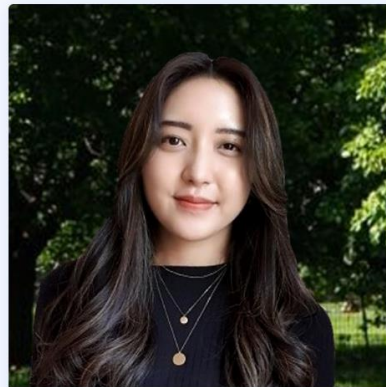
Jackie Nichols

Chief Architect



Robert Ling

Data Engineer



Song Park

Data Scientist



Hao Wu

Data Scientist

PROBLEM STATEMENT

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



Is there racial disparity in sentencing in the US criminal justice system?

OUR SUBJECT MATTER EXPERTS



Pulitzer Winning Journalist
Michael Braga



AEJG President and Board
Member, Defense Attorney –
Kim Gordon

MISSION STATEMENT

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.

IMPACT

Our DAATE MVP provides historical sentencing analytics, a view into sentencing disparity through **causality**, and provides sentence time **predictions** that a defense attorney can use to prepare for plea bargain or trial sentencing negotiations for their clients.



9

Million Criminal Cases Filed in the US
in 2020

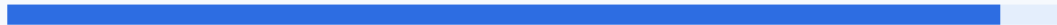


240

Thousand active Defense Attorneys in
the US

NUMBER OF CASES ENDING IN PLEA BARGAIN

94%



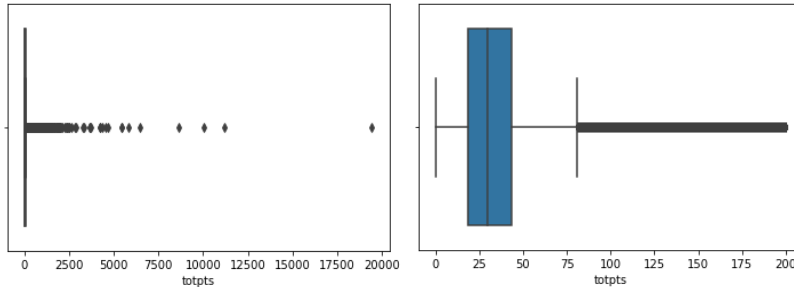
MVP IMPACT

- 1 Use as input during **plea bargain** to **negotiate a fair sentence**
- 2 Additional input when **appealing a sentence or circuit**
- 3 Being better informed about the situation your client is currently in through **full data transparency**

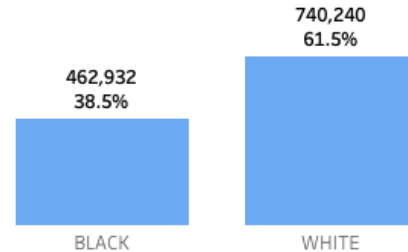
DATA OVERVIEW

- 2004-2016 public data from the Florida Department of Corrections (DOC)
- Removed 9.34% of rows after data cleansing (1.35m → 1.23m rows)
- Important factors for modeling:
 - Imbalanced data (race, gender, sentencing points)
 - Potential human error in manually entered data

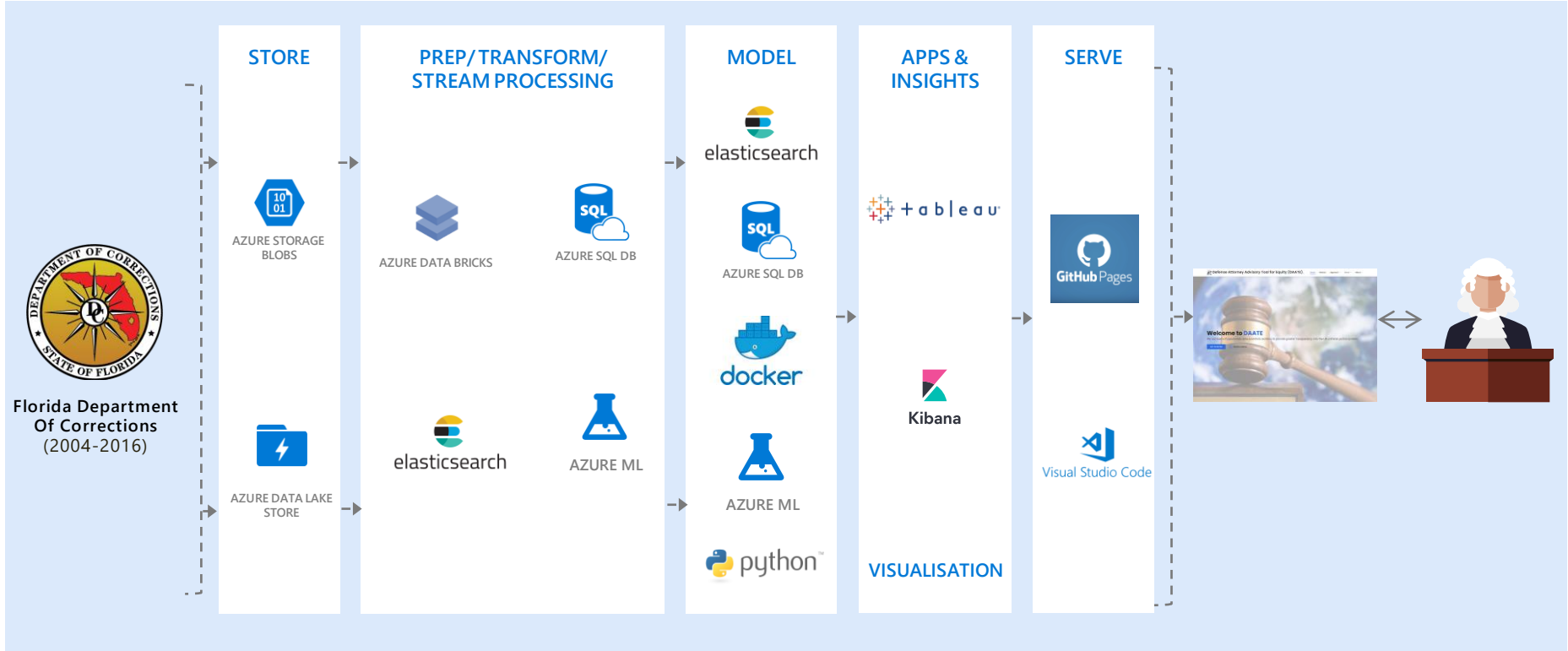
Distribution of Total Points: Before vs After Removing Outliers



Count of Cases by Race



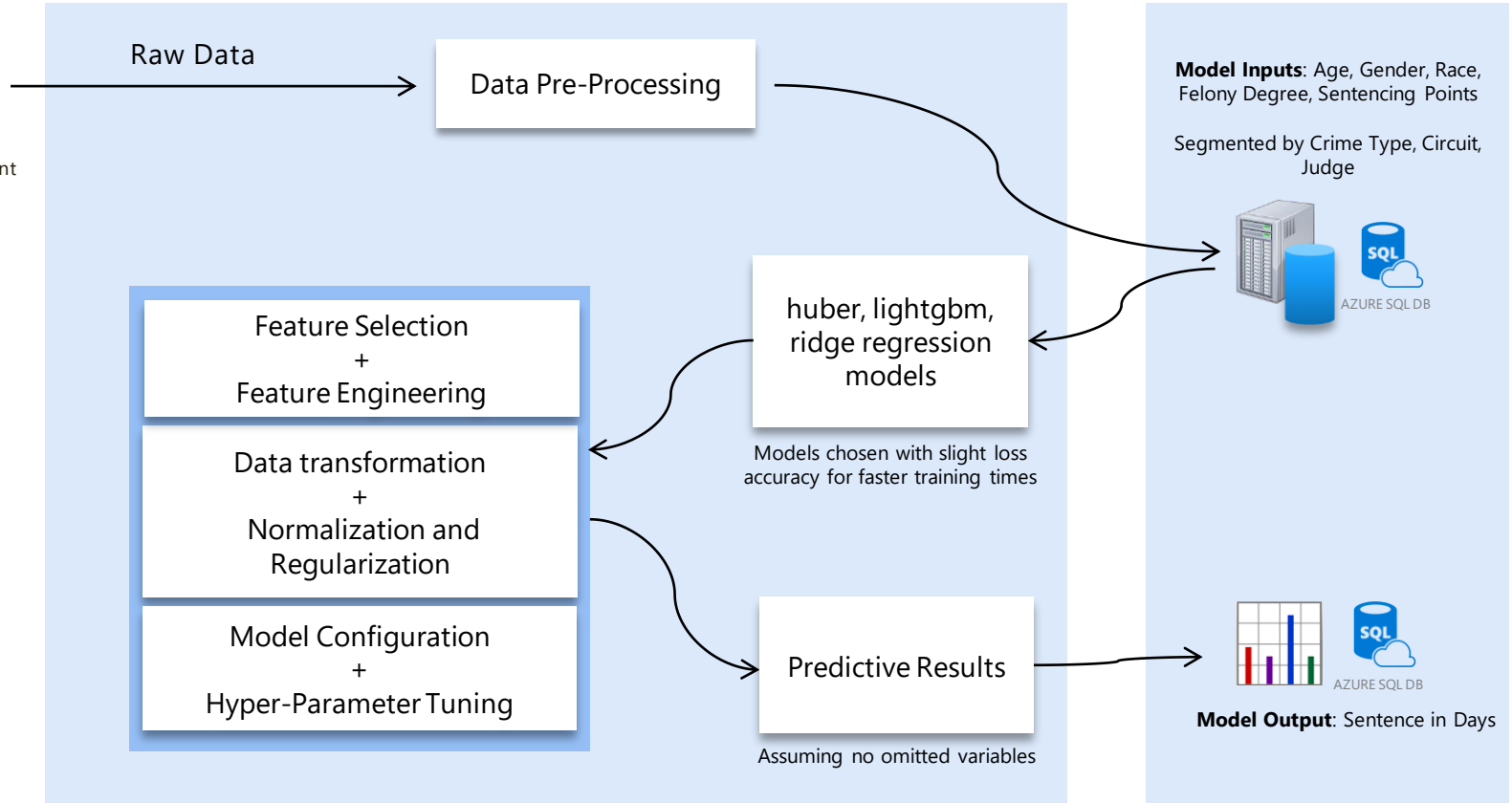
MVP ARCHITECTURE



PREDICTIVE MODEL ARCHITECTURE



Florida Department
Of Corrections
(2004-2016)



PREDICTIVE MODEL RESULTS

What is the Predicted Sentence Time?

Our prediction model is based on historical data for the selected filters.

For a given circuit, we can view sentencing predictions for all judges across various crime types

			Predicted Sentence (Black Defendants)	Predicted Sentence (White Defendants)	Predicted Difference		
CIRCUIT 06 - CLEARWATER	ALL JUDGES	DRIVER LICENSES	124	124	+0	Gender	MALE
		DRUGS	124	97	+27	Degree	3RD DEGREE
		ROBBERY	172	157	+15	Total Sentencing Poin...	22-44
CIRCUIT 11 - MIAMI	ALL JUDGES	DRIVER LICENSES	75	75	+0	Age	40
		DRUGS	10	25	-15		
		ROBBERY	196	198	-2		
CIRCUIT 17 - FT. LAUDERDALE	ALL JUDGES	DRIVER LICENSES	142	159	-17		
		DRUGS	132	106	+26		
		ROBBERY	157	115	+42		

Tune for Gender, Age, Felony Degree & Total Sentencing Points

0 = No evidence of disparity

+ = Longer sentence for Black defendants

- = longer sentence for White defendants

PREDICTIVE MODEL RESULTS

Prediction vs Historical Average by Sentencing Points

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

- Final RMSE ~300 days
- How is RMSE perceived by the user ?
 - Different point ranges affect sentencing differently
 - Omitted variable bias
 - Age variable not considered in historical analytics
- Results should be used as **additional input** for a defense attorney
- Future improvements: Step-wise models & oversample sparse point ranges

CAUSAL MODEL ARCHITECTURE



Florida Department
Of Corrections
(2004-2016)

Raw Data

Data Pre-Processing

Data Transformation
+
Normalization

Distance-Based Clustering
+
Cluster Method Optimization

Subgroup Matching
+
Stratified Sampling
+
Contrast Group Creation

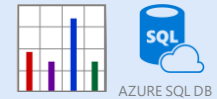
Feature Space
Identification

Assuming SUTVA and treatment
assignment is independent of
potential outcome

Significance
Testing

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

Model Inputs: Age, Gender, Total
Sentencing Points.
Segmented by Crime Type, Circuit,
Judge



Model Output: Sentence difference
in days, confidence interval and p-
value for the difference

CAUSAL MODEL APPROACH

1 Creating Contrasting Groups (Similar attributes, different only by Race)

Cluster Matching

Ensure data points within same cluster are similar in feature space except for race



Stratified Sampling

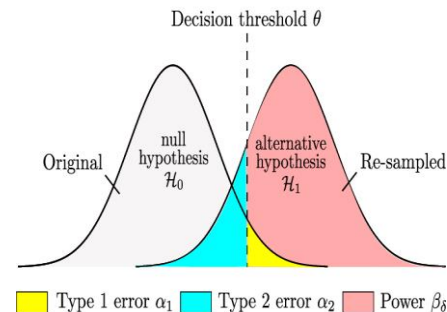
Ensure equivalent group size



Treatment
vs
Control



2 Quantify Disparity (Difference in Sentencing)



Statistical Testing

H_0 : No sentencing disparity between races
 H_1 : Evidence of sentencing disparity between races

CAUSAL MODEL EVALUATION

1 Determine optimal cluster number

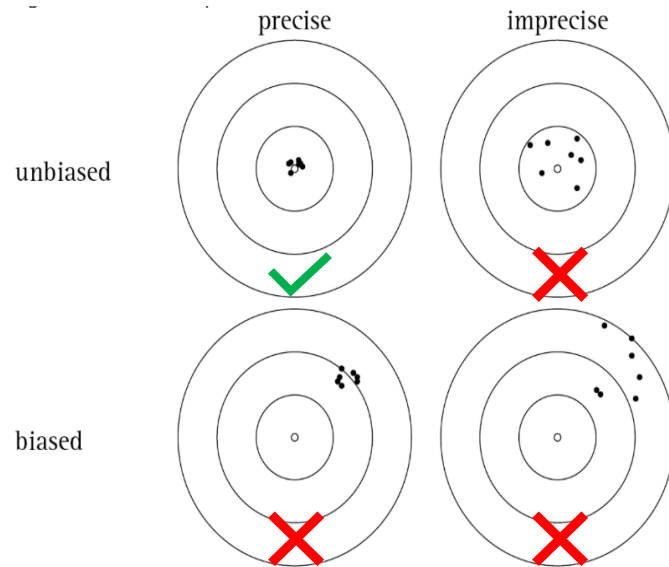
- identify reasonable data structure and category

2 Improve stableness of the estimator (p-value, CI)

- Choose stable clustering algorithm
- Averaging statistics

3 Unbiasedness & Explainability of the process

- Easy-to-understand inputs: Gender, Age, Total Points
- Distance based clustering
- Sentencing difference as the outcome (lower & upper bounds)



CAUSAL MODEL RESULTS

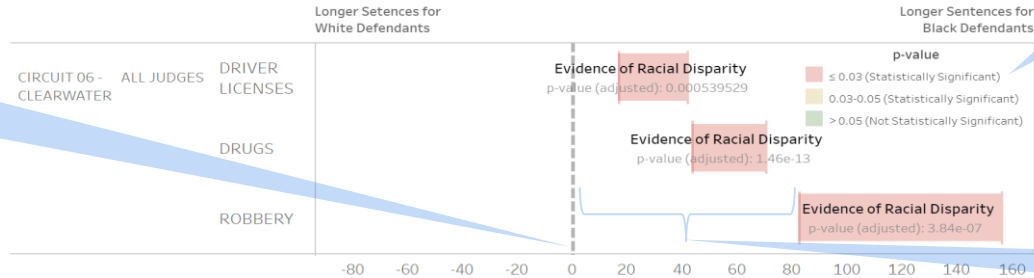
Select View:

☒ Circuit View

☐ Judge View

Is There Racial Disparity in Sentencing?

Our statistical model uses cluster based matching, taking into account various factors (such as age, gender, and total sentencing points) to accurately compare sentence times for Black vs. White defendants. A p-value of less than 0.05 indicates strong statistical evidence of disparity.



Color coding indicating statistical significance

Range of "colored box" is estimated interval for disparity

Evaluation

Visual interpretability

SME feedback

Error Adjustment

Multiple comparison adjustment

Multiple random sampling and average statistics

Future Improvement

Generalizability

Less Data Demanding Options

TAKEAWAY AND LEARNINGS

Novel Approach

Little to no examples of using DS to detect sentencing disparity

Data Complexity and Legal Terms

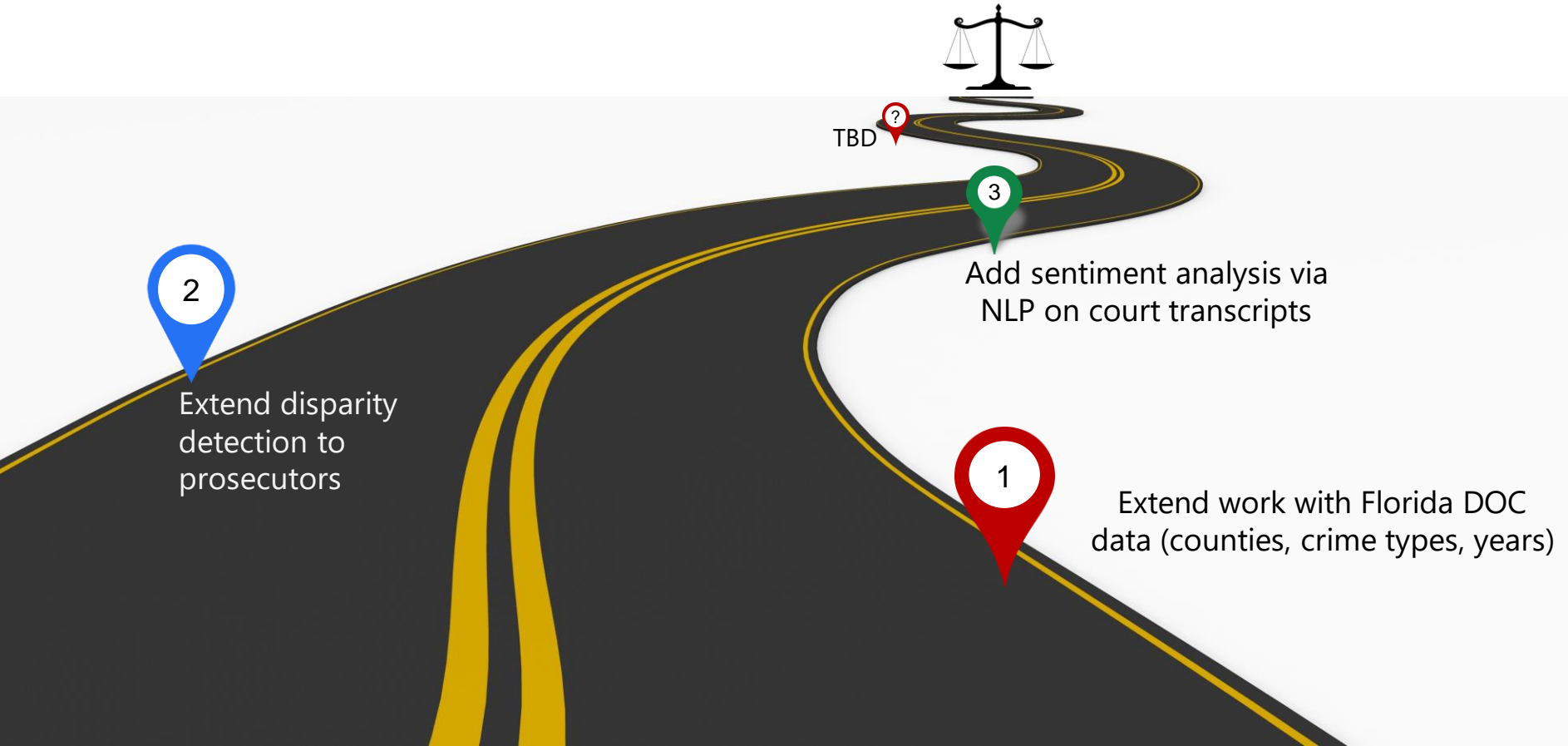
1.3M rows, 290 columns, 638 crime types and legal jargon

Viable UX

Must be meaningful yet simple

Subject matter expert feedback is vital!

ROADMAP



SME FEEDBACK

*“Our team of reporters and editors at the Sarasota Herald Tribune and at USA TODAY always believed that there was tremendous potential to use criminal justice data to help bring about equity in sentencing and to shine a light on judges, prosecutors and law enforcement officials who might not even understand the ways in which they are tilting justice in favor of one group or another. **We are grateful that the team that built DAATE recognized that potential and created the foundation for a tool that could be used by defense attorneys to right some the injustices that have plagued the United States for hundreds of years.**” – Michael Braga
(Pulitzer Winning Investigative Journalist)*

SME FEEDBACK

*"The work that the DAATE team has been able to accomplish in such a short time is remarkable. Building in statistical significance and p-values is such a great idea to help bring confidence to the disparity charts. **I'm hopeful that we can leverage this great work into the Equity Dashboard that the American Equity and Justice Group has been building, and to help us gain more momentum towards bringing more equity to the legal system nationwide.**"*

- Belinda Cheng (Board Member at American Equity Justice Group)

SUMMARY

Is there racial disparity in sentencing in the US criminal justice system?

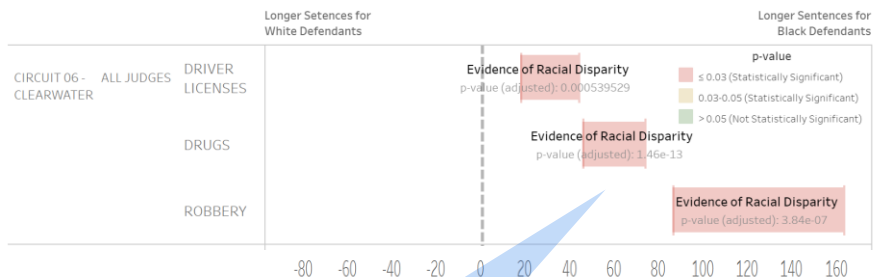
Select View:

☒ Circuit View

☐ Judge View

Is There Racial Disparity in Sentencing?

Our statistical model uses cluster based matching, taking into account various factors (such as age, gender, and total sentencing points) to accurately compare sentence times for Black vs. White defendants. A p-value of less than 0.05 indicates strong statistical evidence of disparity.



Evidence of Disparity:
Longer sentence for
Black defendants

What is the Predicted Sentence Time?

Our prediction model predicts sentence time (state prison + county jail) based on historical data for the selected filters.

Predicted Sentence (Black Defendants)	Predicted Sentence (White Defendants)	Predicted Difference	Gender MALE
124	124	+0	Degree 3RD DEGREE
124	97	+27	Total Sentencing Poin... 22-44
172	157	+15	Age 40

0 = No evidence of
disparity

+ = Longer sentence
for Black defendants

MISSION STATEMENT

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.

THANK YOU!

Do you have any questions?

mspuckit1@berkeley.edu

robling@berkeley.edu

hwu24@berkeley.edu

song.park@berkeley.edu