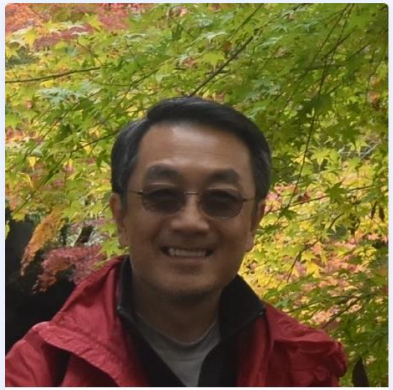
Defense Attorney Advisory Tool for Equity (DAATE)

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

AMERICAN EQUITY
& JUSTICE GROUP

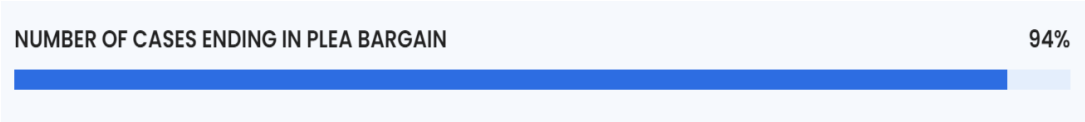
AEJG President and Board Member, Defense Attorney – Kim Gordon

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.



MVP IMPACT

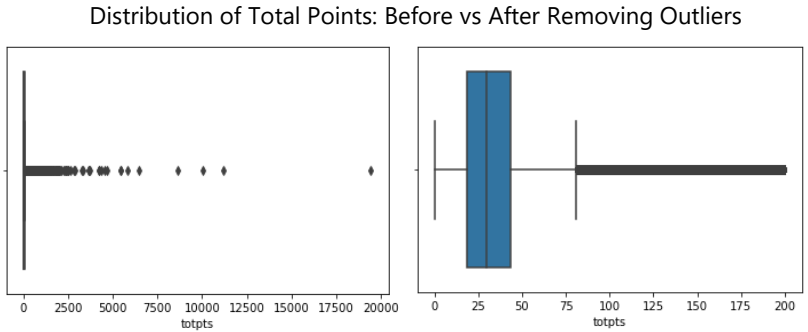
1. Use as input during **plea bargain** to **negotiate a fair sentence**

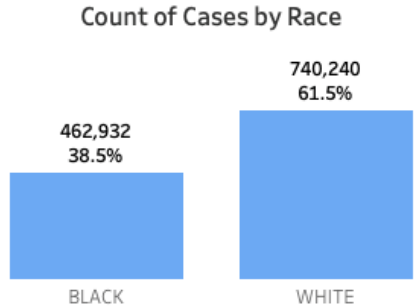
2. Additional input when **appealing a sentence or circuit**

3. Being better informed about the situation your client is currently in through **full data transparency**

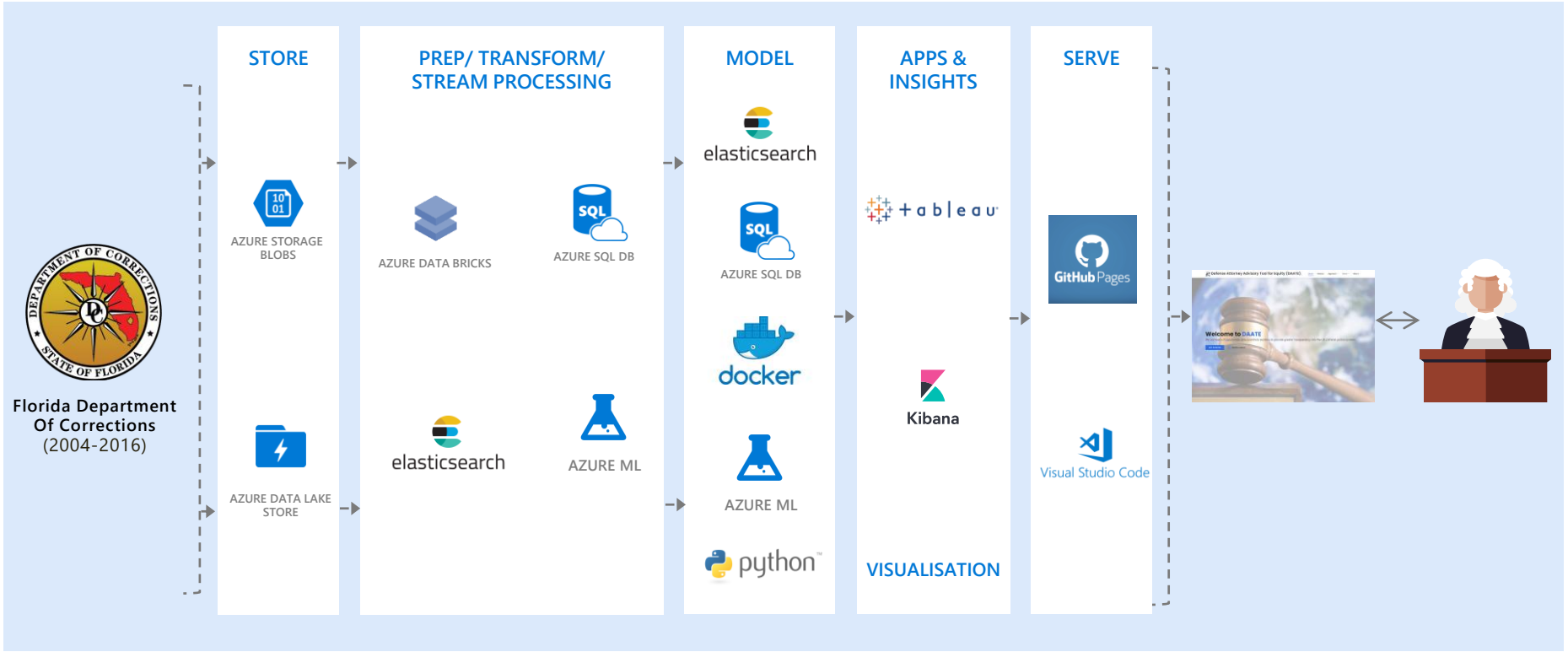
DATA OVERVIEW

- 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





MVP ARCHITECTURE



PREDICTIVE MODEL ARCHITECTURE

Florida Department Of Corrections (2004-2016)

Raw Data

Data Pre-Processing

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

Segmented by Crime Type, Circuit, Judge

huber, lightgbm, ridge regression models

Models chosen with slight loss accuracy for faster training times

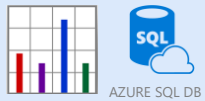
Feature Selection
+
Feature Engineering

Data transformation
+
Normalization and Regularization

Model Configuration
+
Hyper-Parameter Tuning

Predictive Results

Assuming no omitted variables



Model Output: Sentence in Days

PREDICTIVE MODEL RESULTS

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

What is the Predicted Sentence Time?

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

		Predicted Sentence (Black Defendants)	Predicted Sentence (White Defendants)	Predicted Difference
CIRCUIT 06 - CLEARWATER ALL JUDGES	DRIVER LICENSES	124	124	+0
	DRUGS	124	97	+27
	ROBBERY	172	157	+15
CIRCUIT 11 - MIAMI ALL JUDGES	DRIVER LICENSES	75	75	+0
	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

Gender: MALE

Degree: 3RD DEGREE

Total Sentencing Poin...: 22-44

Age: 40

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

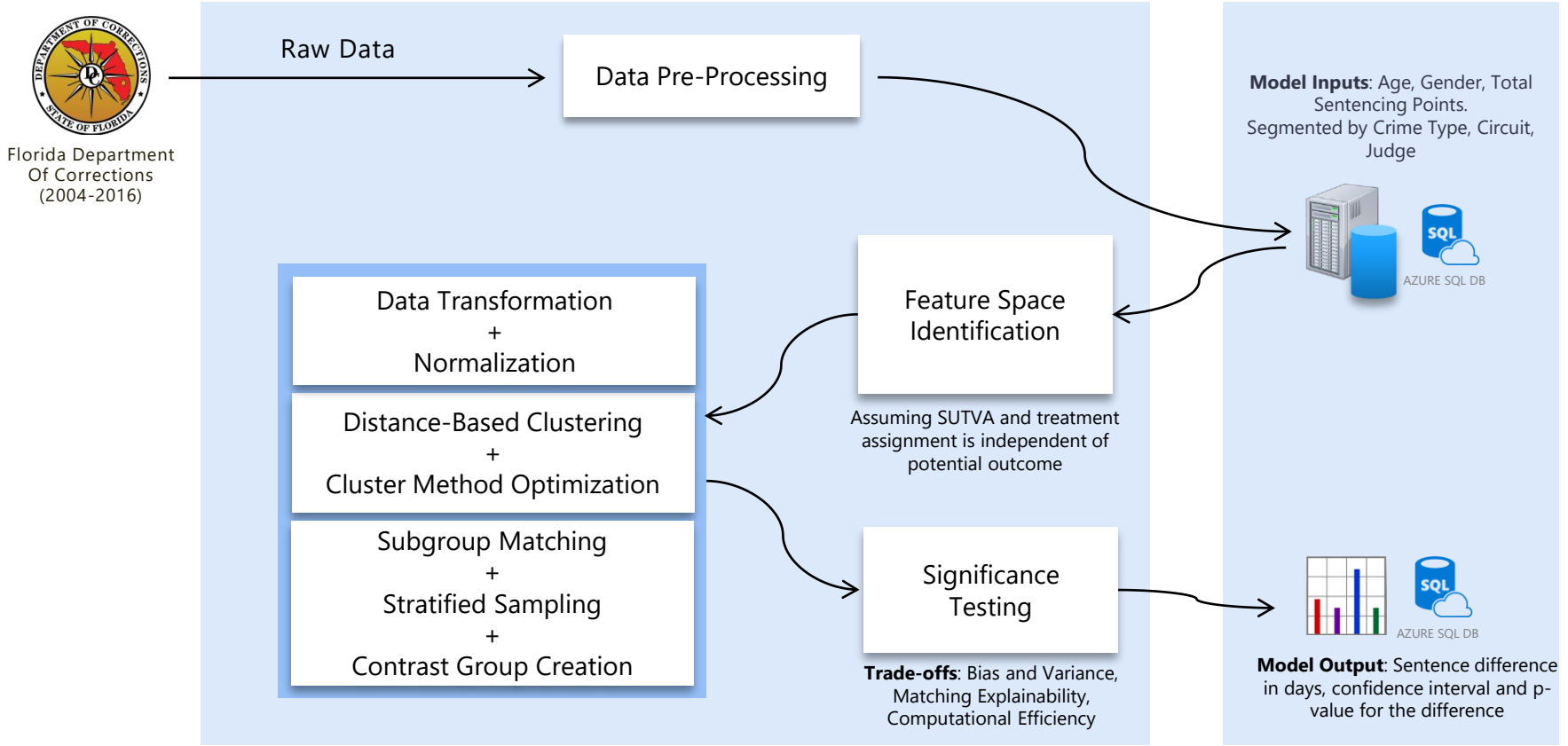
Black Defendant	22-44	44-54	104-114	114-124	194-204
20	-3	-26	-241	-433	-849
40	14	-9	-224	-417	-833
60	23	0	-215	-408	-824
80	30	6	-208	-401	-817
White Defendant	22-44	44-54	104-114	114-124	194-204
20	-22	-26	-51	43	-1506
40	-5	-9	-34	104	-1489
60	4	0	-25	113	-1480
80	11	7	-18	120	-1474

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



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

→

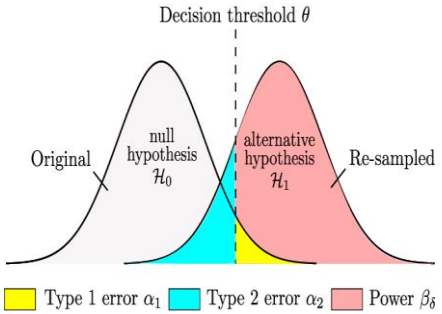
Stratified Sampling

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**

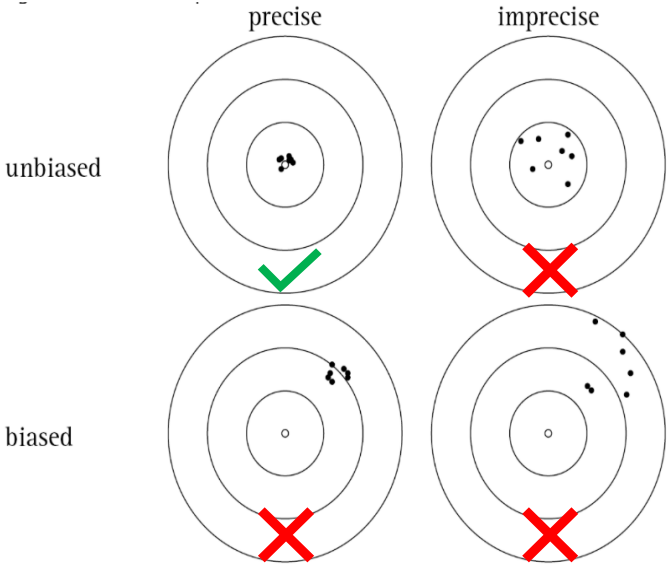
- identify reasonable data structure and category

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

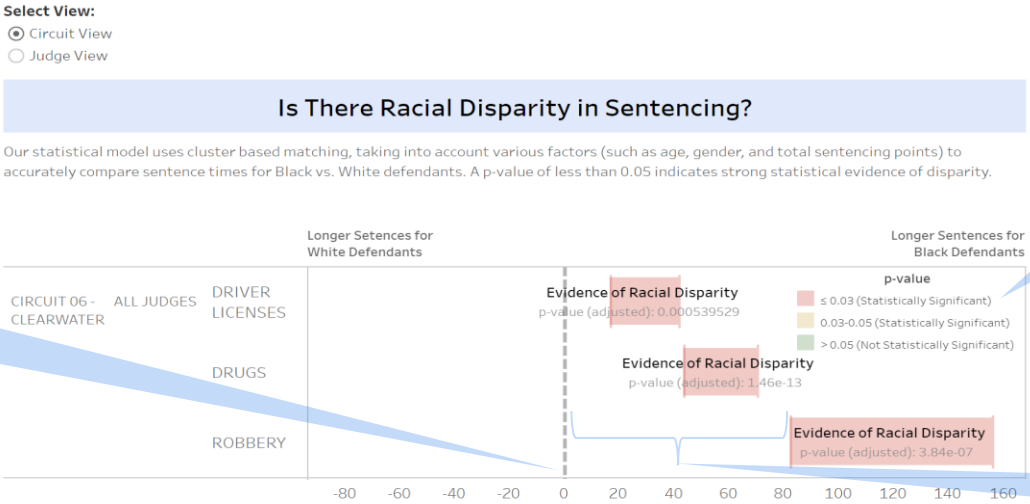
- Choose stable clustering algorithm
- Averaging statistics

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



Vertical "0" line is a reference for evidence of disparity

Color coding indicating statistical significance

Range of "colored box" is estimated interval for disparity

Evaluation	Error Adjustment	Future Improvement
Visual interpretability	Multiple comparison adjustment	Generalizability
SME feedback	Multiple random sampling and average statistics	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

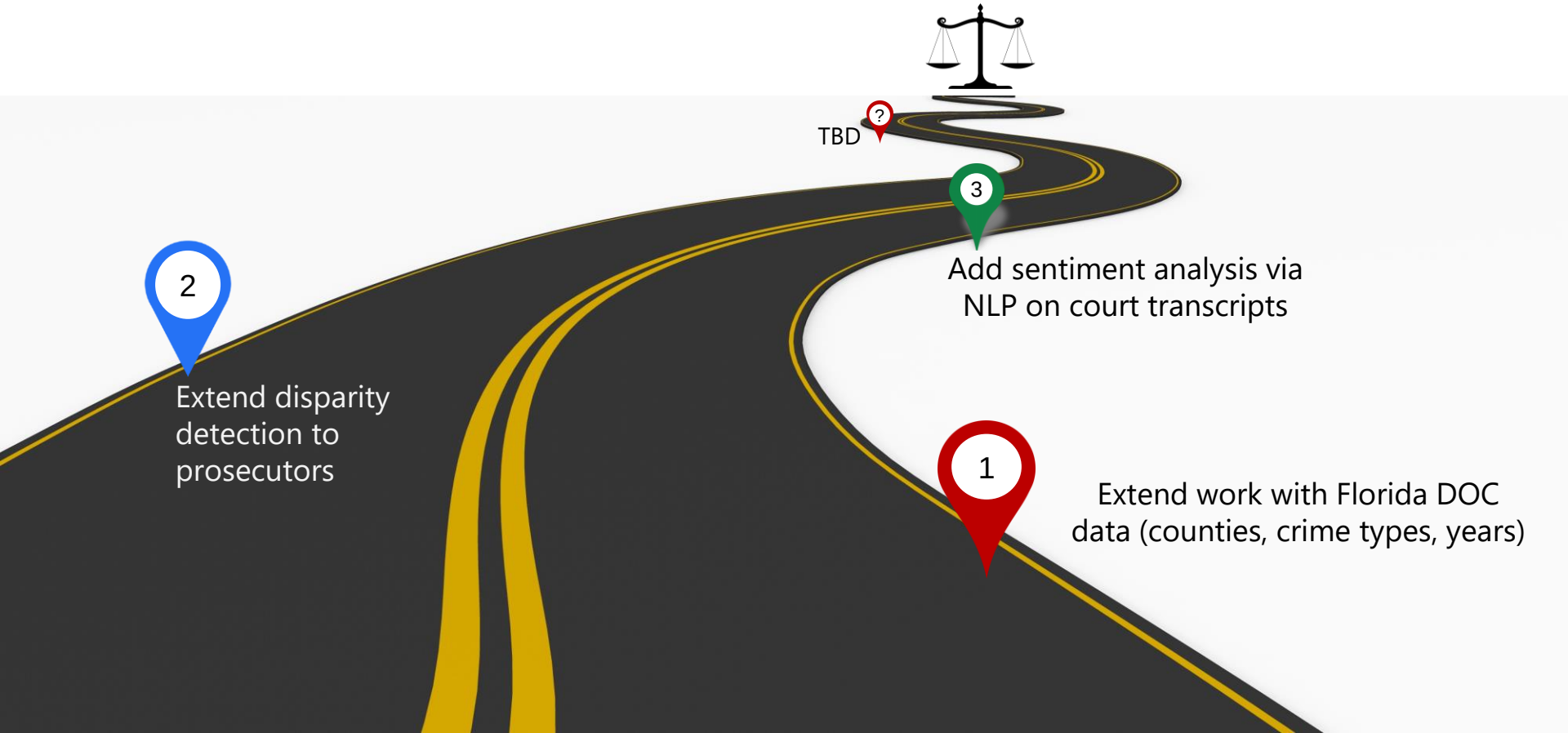
1.3M rows, 290 columns, 638 crime types and legal jargon

Viable UX

Must be meaningful yet simple

Subject matter expert feedback is vital!

ROADMAP

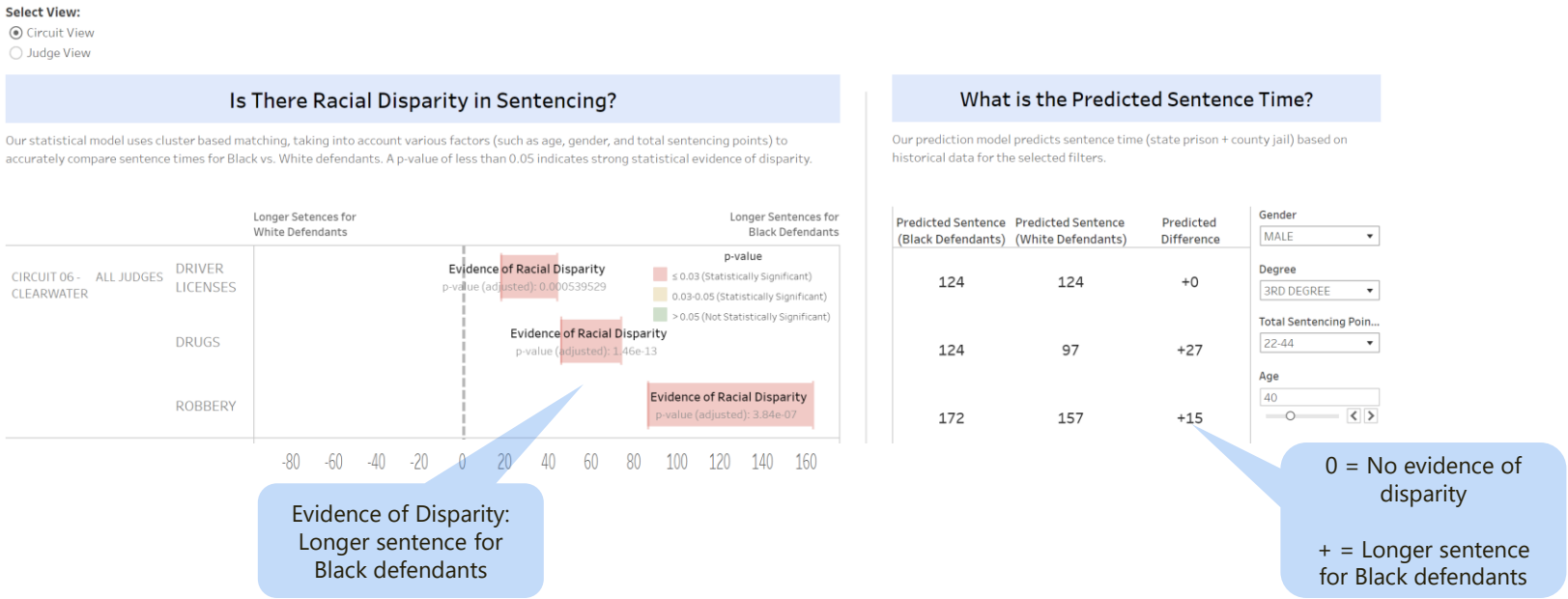

TBD
?
3
Add sentiment analysis via NLP on court transcripts
2
Extend disparity detection to prosecutors
1
Extend work with Florida DOC data (counties, crime types, years)

SME FEEDBACK

"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 of the injustices that have plagued the United States for hundreds of years." – Pulitzer Winning Investigative Journalist Michael Braga

SUMMARY

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



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