



Welcome to DAATE

**Defense Attorney Advisory Tool for Equity
(DAATE)**

Jackie Nichols, Hao Wu,
Robert Ling, Song Park

01.

Introduction

03.

Data and EDA

05.

Challenges and Goals

02.

MVP

04.

Technical Approach

06.

Summary

DAATE Introduction

Our Mission

To empower legal professionals to realize fairness and equity by providing transparency into the criminal justice system using technology to promote fairness for all people.



Why DAATE?

Our goal is to provide greater transparency into sentencing in the United States (US) criminal justice system. It is intended as additional or complimentary information.



Incarcerations

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

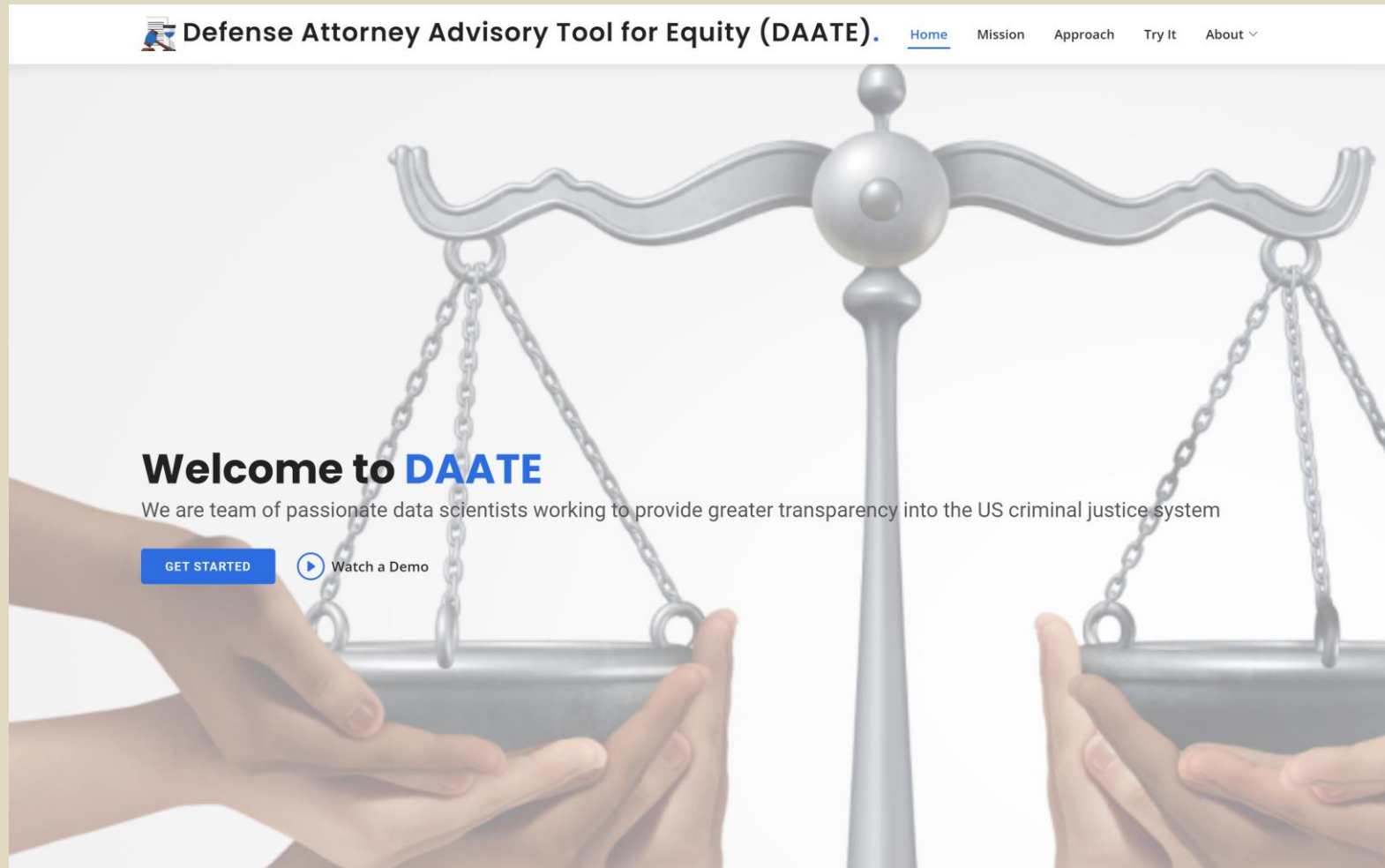


Sentencing

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

Impact: Approximately 240k Defense Attorneys in the US and each lawyer has multiple cases per year

DAATE MVP



How do judges historically sentence individuals for a crime?

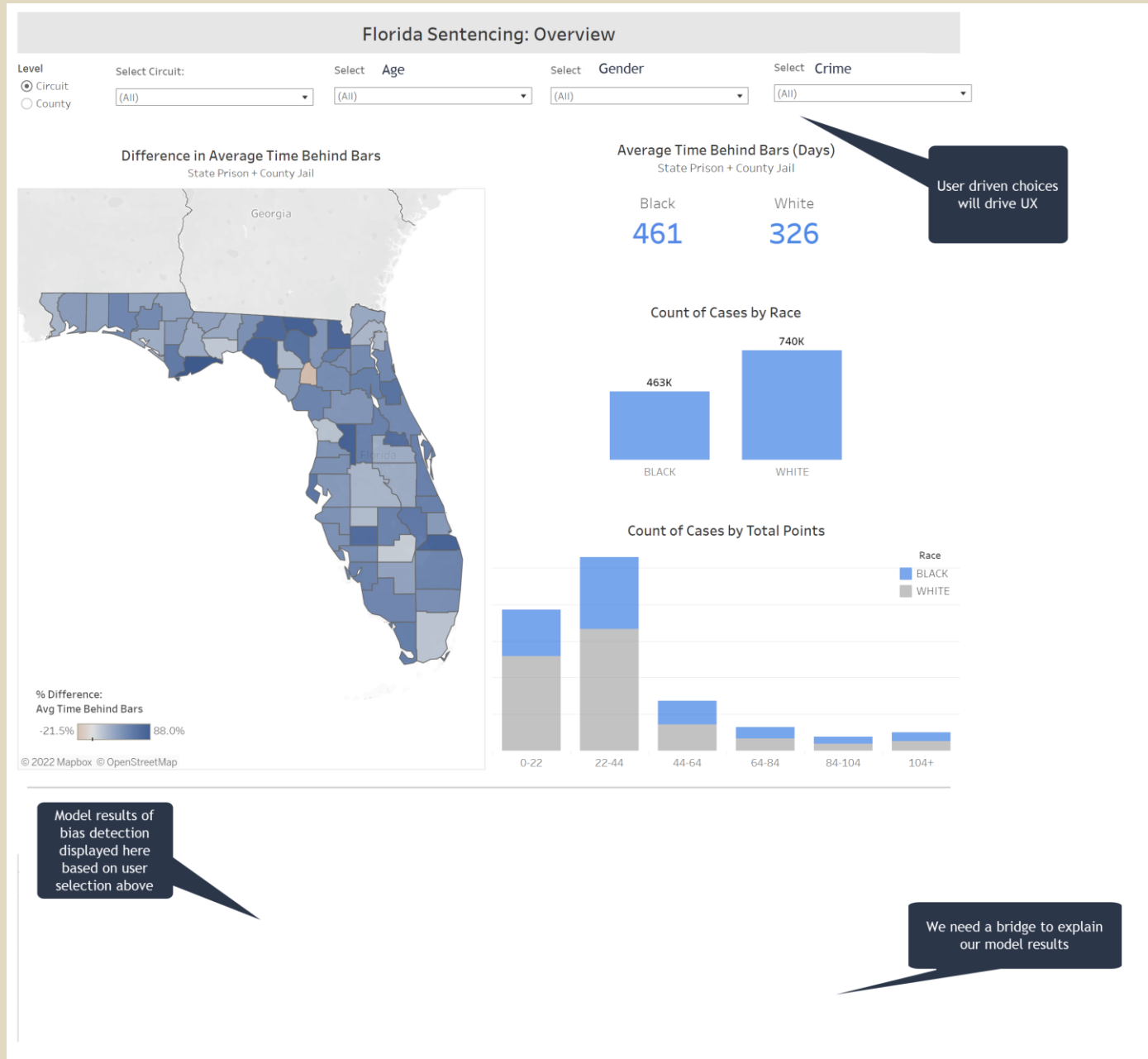
GOALS

What is the magnitude and confidence of the bias? (how significant is the bias compared to other counties)

Does the potential for bias appear in sentencing, if so what kind? (crime, race, gender, ruling towards certain sentence or plea)



DAATE MVP



How do judges historically sentence individuals for a crime?

GOALS

What is the magnitude and confidence of the bias? (how significant is the bias compared to other counties)

Does the potential for bias appear in sentencing, if so what kind? (crime, race, gender, ruling towards certain sentence or plea)



DAATE MVP Data

- Source: Florida Department of Corrections (DOC)
- Publicly available data
- Sentencing data from 2004-2016
- 1.35m rows
- No data dictionary – made sense by series of SME interviews & search

The screenshot shows the Elastic Discover interface. At the top, the Elastic logo and a search bar are visible. Below the search bar, the 'Discover' tab is selected. The search query 'DUI* OR AMPHET* OR *' is entered in the search bar. The results show 1,354,399 hits. A callout bubble points to the search bar with the label 'Search'. Another callout bubble points to the hit count with the label 'ALL the Raw Rows'. A third callout bubble points to the raw data rows with the label 'ALL the Raw Rows'. A fourth callout bubble points to the 'Available fields' list with the label 'All the Raw Columns'. The 'Available fields' list includes fields like 'amrsn1', 'amrsn2', 'aocnts1', 'aoflst1', 'clfely', 'fstnm', and 'lstnm'. The 'clfely' field is expanded, showing a 'Top 5 values' section with a bar chart and a table of values.

Field	Value	Percentage
3RD DEGREE		47.6%
2ND DEGREE		38.2%
1ST DEGREE		9.2%
1ST/LIFE		4.4%
LIFE		0.6%

DAATE Florida Sentencing Scoresheet

1

- Date, County & Judge
- Offender Demographics
- Information about the offense (primary + additional)

2

- Additional factors for sentencing (e.g. prior criminal record)
- Final "sentencing points"

3

- Sentencing results
- State prison, county jail, probation, etc

Rule 3.992(a) Criminal Punishment Code Scoresheet
The Criminal Punishment Code Scoresheet Preparation Manual is available at: <http://www.dc.state.fl.us/pub/cpm/cpm/index.html>

1. DATE OF SENTENCE 10-26-16	2. PREPARER'S NAME SATO	3. COUNTY Duval	4. SENTENCING JUDGE Healey
5. NAME (LAST, FIRST, MI.) Forest, Ronnie A.	6. DOB 9-20-1969	7. RACE M <input checked="" type="checkbox"/> W <input type="checkbox"/> O <input type="checkbox"/> OTHER	8. PRIMARY OFFENSE DATE 9-8-16
9. GENDER M <input checked="" type="checkbox"/> F <input type="checkbox"/>		10. PRIMARY DOCKET # 16 C88085	
11. PRIMARY OFFENSE: If Qualifier, please check: A <input type="checkbox"/> S <input type="checkbox"/> C <input type="checkbox"/> R (A=Attempt, S=Solicitation, C=Conspiracy, R=Reclassification)		12. PLEA GUILTY <input checked="" type="checkbox"/> TRIAL <input type="checkbox"/>	

FELONY DEGREE F.S.# DESCRIPTION OFFENSE LEVEL POINTS
F3, 893.13, Possess Cocaine, 3, 16

Prior capital felony triples Primary Offense points ☐

II. ADDITIONAL OFFENSE(S): Supplemental page attached ☐

DOCKET#	FEL/MM DEGREE	F.S.#	DESCRIPTION	OFFENSE LEVEL	QUALIFY A S C R	COUNTS	POINTS	TOTAL

(Level - Points: 1=4, 2=10, 3=16, 4=22, 5=28, 6=36, 7=56, 8=74, 9=92, 10=116)
Prior capital felony triples Additional Offense points ☐

Supplemental page points II. 0

III. VICTIM INJURY:

	Number	Total		Number	Total
2nd Degree Murder	240 x		Slight	4 x	
Death	120 x		Sex Penetration	80 x	
Severe	40 x		Sex Contact	40 x	
Moderate	18 x				

IV. PRIOR RECORD: Supplemental page attached ☐

FEL/MM DEGREE	F.S.#	DESCRIPTION	NUMBER	POINTS	TOTAL
		S/M/D Cocaine 100'	1	14	14
		S/M/D Cocaine	2	36	72
		GTA	1	24	24
		Fleeing	1	16	16
		Vanols	15	0.2	36

V. Legal Status violation = 4 Points
☐ Escape ☐ Fleeing ☐ Failure to appear ☐ Supersedeas bond ☐ Incarceration ☐ Pretrial intervention or diversion program
☐ Court imposed or post prison release community supervision resulting in a conviction

VI. Community Sanction violation before the court for sentencing
☐ Probation ☐ Community Control ☐ Pretrial Intervention or diversion

☐ 6 points for any violation other than new felony conviction x _____ each successive violation OR
☐ New felony conviction = 12 points x _____ each successive violation if new offense results in conviction before or at same time as sentence for violation of probation OR
☐ 12 points x _____ each successive violation for a violent felony offender of special concern when the violation is not based solely on failure to pay costs, fines, or restitution OR
☐ New felony conviction = 24 points x _____ each successive violation for a violent felony offender of special concern if new offense results in a conviction before or at the same time for violation of probation

VII. Firearm/Semi-Automatic or Machine Gun = 18 or 25 Points

VIII. Prior Serious Felony - 30 Points

Subtotal Sentence Points 00

IX. Enhancements (only if the primary offense qualifies for enhancement)

Law Enf. Protect.	Drug Trafficker	Motor Vehicle Theft	Criminal Gang Offense	Domestic Violence in the Presence of Related Child (offenses committed on or after 3/12/07)	Adult-on-Minor Sex Offense (offenses committed on or after 10/1/14)
_____ x 1.5	_____ x 2.0	_____ x 2.5	_____ x 1.5	_____ x 1.5	_____ x 2.0

Enhanced Subtotal Sentence Points 0

TOTAL SENTENCE POINTS 44.2

If total sentence points are greater than 44: 44.2 minus 26 = 18.2 x .75 = 13.65

total sentence points 44.2

If total sentence points are 60 points or less and court makes findings pursuant to both Florida Statute 948.20 and 397.334(3), the court may place the defendant into a treatment-based drug court program.

The maximum sentence is up to the statutory maximum for the primary and any additional offenses as provided in s. 775.082, F.S., unless the lowest permissible sentence under the Code exceeds the statutory maximum. Such sentences may be imposed concurrently or consecutively. If total sentence points are greater than or equal to 363, a life sentence may be imposed.

maximum sentence in years 5

TOTAL SENTENCE IMPROVED

Years 0 Months 13 Days 0

☒ State Prison ☐ Life
☐ County Jail ☐ Time Served
☐ Community Control
☐ Probation ☐ Modified
Please check if sentenced as ☐ habitual offender, ☐ habitual violent offender, ☐ violent career criminal, ☐ prison releasee reoffender, or a ☐ mandatory minimum applies.
☐ Mitigated Departure ☐ Plea Bargain ☐ Prison Diversion Program
Other Reason _____

JUDGE'S SIGNATURE _____

Effective Date: For offenses committed under the Criminal Punishment Code effective for offenses committed on or after October 1, 1998, and subsequent revisions.

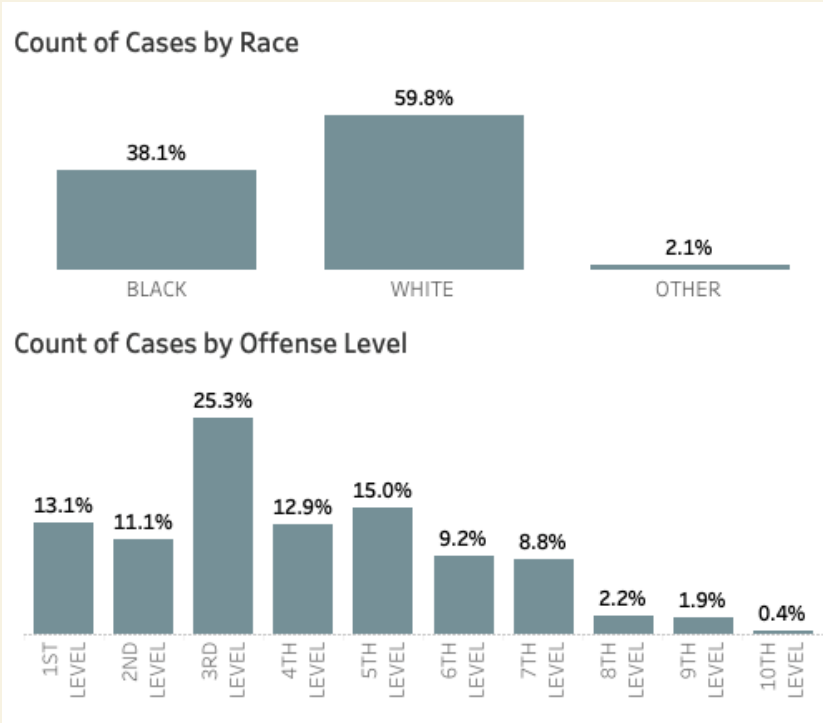
DAATE Florida Sentencing Scoresheet

👍 All of the information from the scoresheets are stored in a dataset and are public data

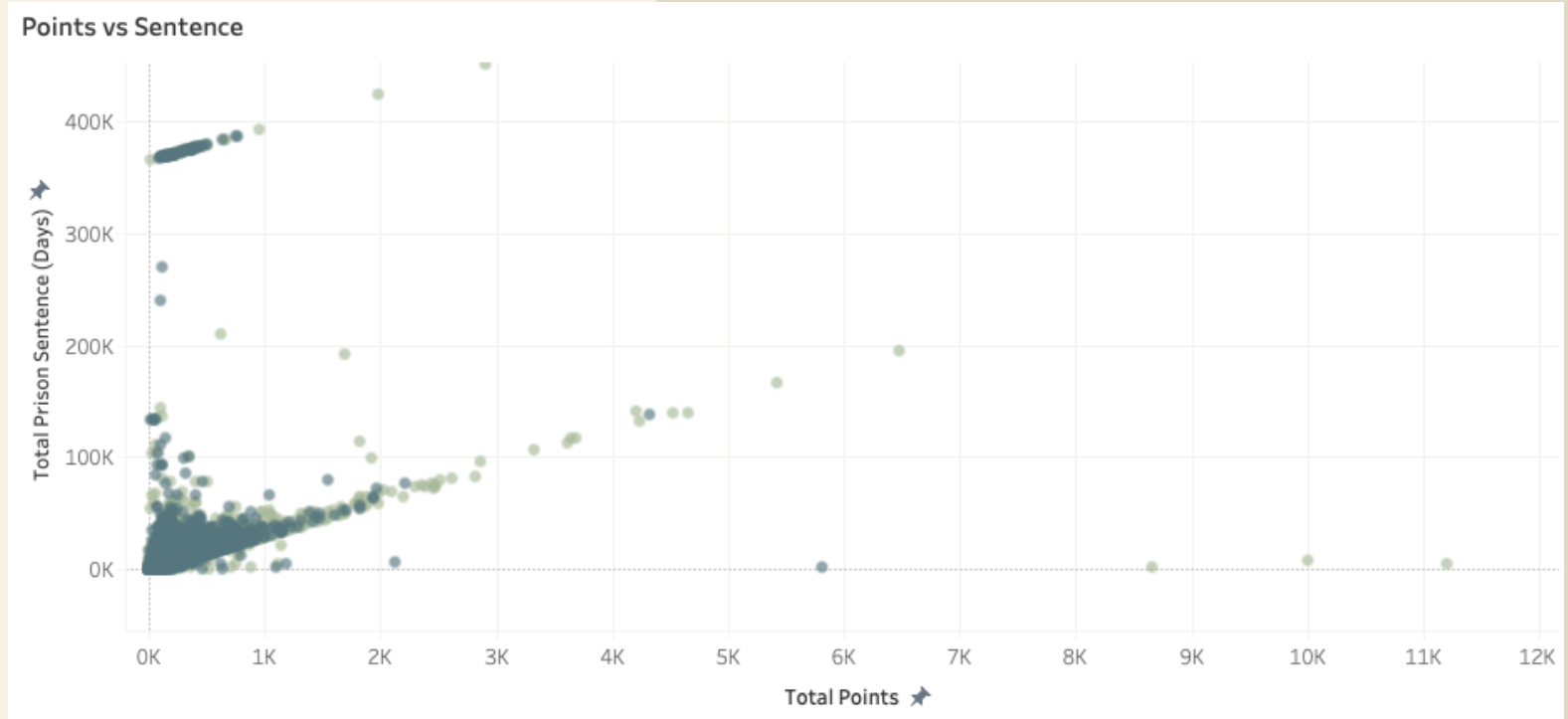
🗨️ Potential for human error in the dataset

sentdte	dornum	prepby	prepnm	county				judge		lstnm		fstnm		dob		obts		race		hispan	gender			
30OCT2009	Y17759	SAO	UNK	MIAMI-DADE				SOTO		MORGAN		TERRANCE		01MAR1980		NaN		BLACK		/	MALE			
05DEC2007	119602	SAO	UNK	DUVAL				MERRETT		WILLIAMS		MACK		21JUN1959		NaN		BLACK		/	MALE			
05OCT2005	X11898	SAO	HAT	OSCEOLA				LAUTEN		MANSO		EUGENE		07MAR1977		NaN		BLACK		/	MALE			
17FEB2011	667035	SAO	SCH	PALM BEACH				MILLER		FELICIANO		JOHNNY		18MAR1946		NaN		WHITE		/	MALE			
07JAN2009	T39753	SAO	CHA	HILLSBOROUGH				POMPONIO,D		KING		ROBERT		19MAY1983		NaN		WHITE		/	MALE			
07MAY2007	965880	SAO	STA	MADISON				JOHNSON,LEANDRA		DOVER		MICHAEL		01OCT1970		NaN		WHITE		/	MALE			
25JUL2014	V29355	SAO		vimodp	vislgh	vislgp	visexp	visxpp	visexc	visxcp	vitopt	prsupt	prtopt	pgsubt	p1subt	lsviol	rpviol	faviol	weapon	subtpt	enhanc	ensubt	totpts	
				0	0	0	0	0	0	0	0	0	57.0	0	0	0	0	0	0	113.0	NONE	0.0	113.0	
08MAY2007	R43285	SAO			0	0	0	0	0	0	0	0	42.0	0	0	0	0	0	0	98.0	NONE	0.0	98.0	
21SEP2006	M20788	SAO			0	0	0	0	0	0	0	0	28.0	0	0	0	0	0	0	64.0	NONE	0.0	64.0	
04FEB2000	085712	SAO			0	0	0	0	0	0	0	0	18.0	0	0	0	0	0	0	74.0	NONE	0.0	74.0	
					0	0	0	0	0	0	0	0	25.4	0	0	0	0	0	0	207.1	NONE	0.0	207.1	
					0	1	4	0	0	0	0	244	0	56.0	0	0	4	0	18	0	600.6	NONE	0.0	600.6
					0	0	0	0	0	0	0	0	77.0	0	0	0	0	0	0	186.2	LEP 2.5	465.5	465.5	
					0	0	0	0	0	0	0	0	23.2	0	0	4	0	0	0	65.4	NONE	0.0	65.4	
					0	0	0	0	0	0	0	0	27.0	0	0	0	0	0	0	64.8	NONE	0.0	64.8	

Basic Summary Data



Outliers in Total Points and Prison Sentence Length



- Data is not balanced: notably more data points for White defendants, more data for lower offense level crimes
- Scatterplot shows outliers for total points and sentence length

DAATE EDA: Removing Data

PART I: REMOVING INVALID/INCORRECT DATA

1 Life Sentences

- Created significant outliers in the sentencing outcome variable and showed signs of incorrect data
 - Only 0.26% of the raw data

2 0 Total Points

- Removed 3 cases with 0 total points, likely incorrect data

3 Race = 'Other'

- Unclear what 'Other' indicates
 - Only 2.1% of the raw data

4 44+ Points & No Prison/Jail Time

- Cases with 44+ points should usually receive a state prison sentence as a baseline
 - Manual checks indicated incorrect data

DAATE EDA: Removing Data

PART II: REMOVING OUTLIERS

4

Total Points Outliers

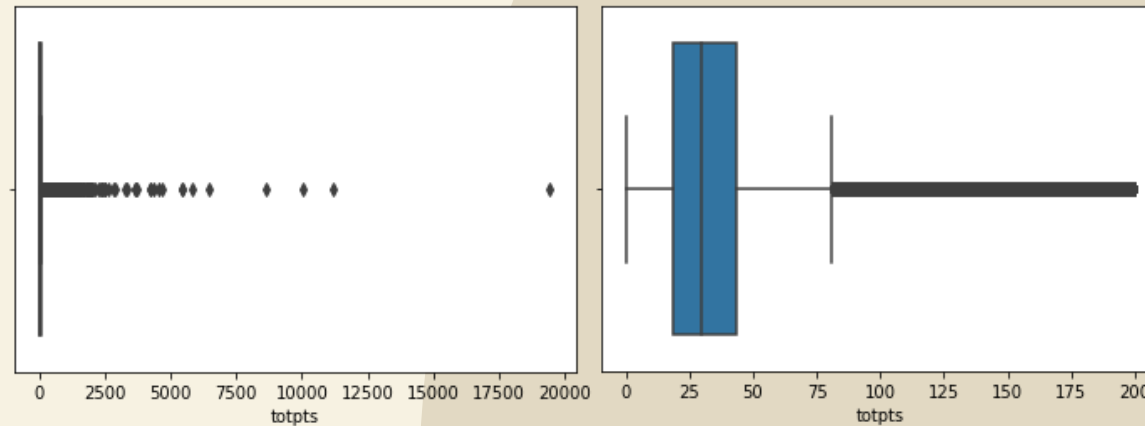
- Extreme values (thousands of points, max = 19,439) that are not feasible
- Use z-score method to identify and remove outliers (new max = 199.6)

5

Sentence Time Outliers

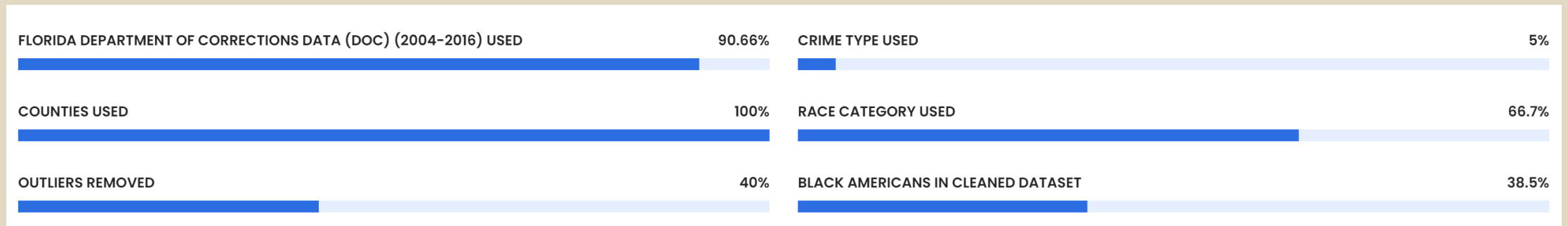
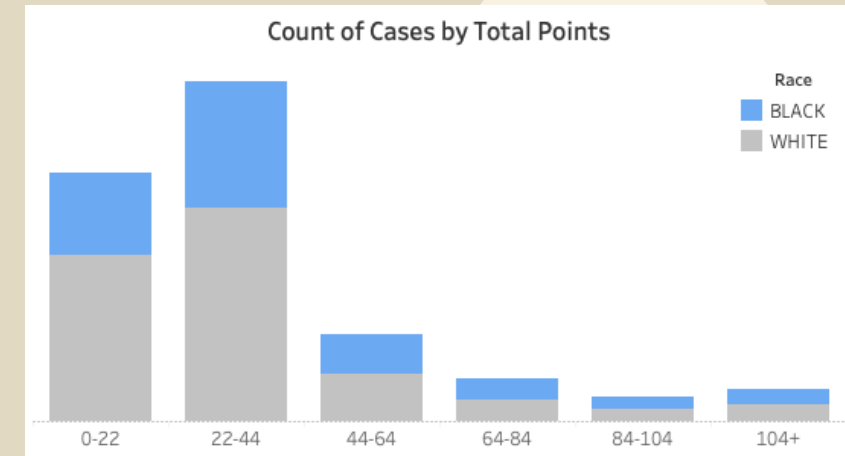
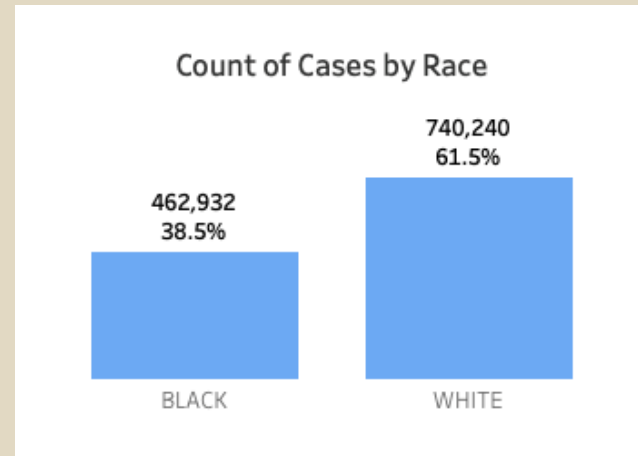
- Handle potentially incorrect data (e.g. very small points but extremely high sentence time)
- Bucket total points and removed sentence time outliers (via z-score method) within each bucket

Distribution of Total Points: Before vs After Removing Outliers



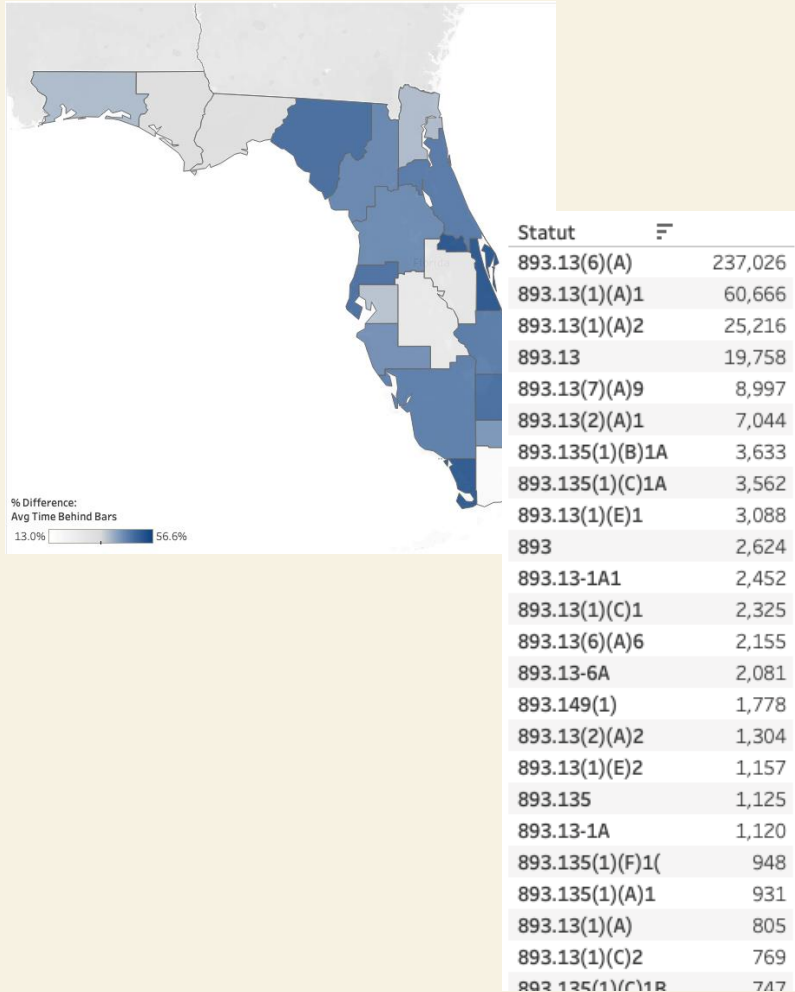
DAATE MVP Cleaned Data

- 1.35m rows originally → 1.23m rows after EDA (detecting unusual data) and identifying outliers
 - Removed 9.34% of data
- Still have imbalanced data



DAATE EDA: Challenges

The more specific (particular crime type, judge, etc.) the more directly applicable our product can be. However, in addition to potential negative implications on certain judges/counties, sample size and model accuracy are crucial.



Geographic Region

- 20 circuits, 67 counties
- Focus on select circuits/counties for MVP

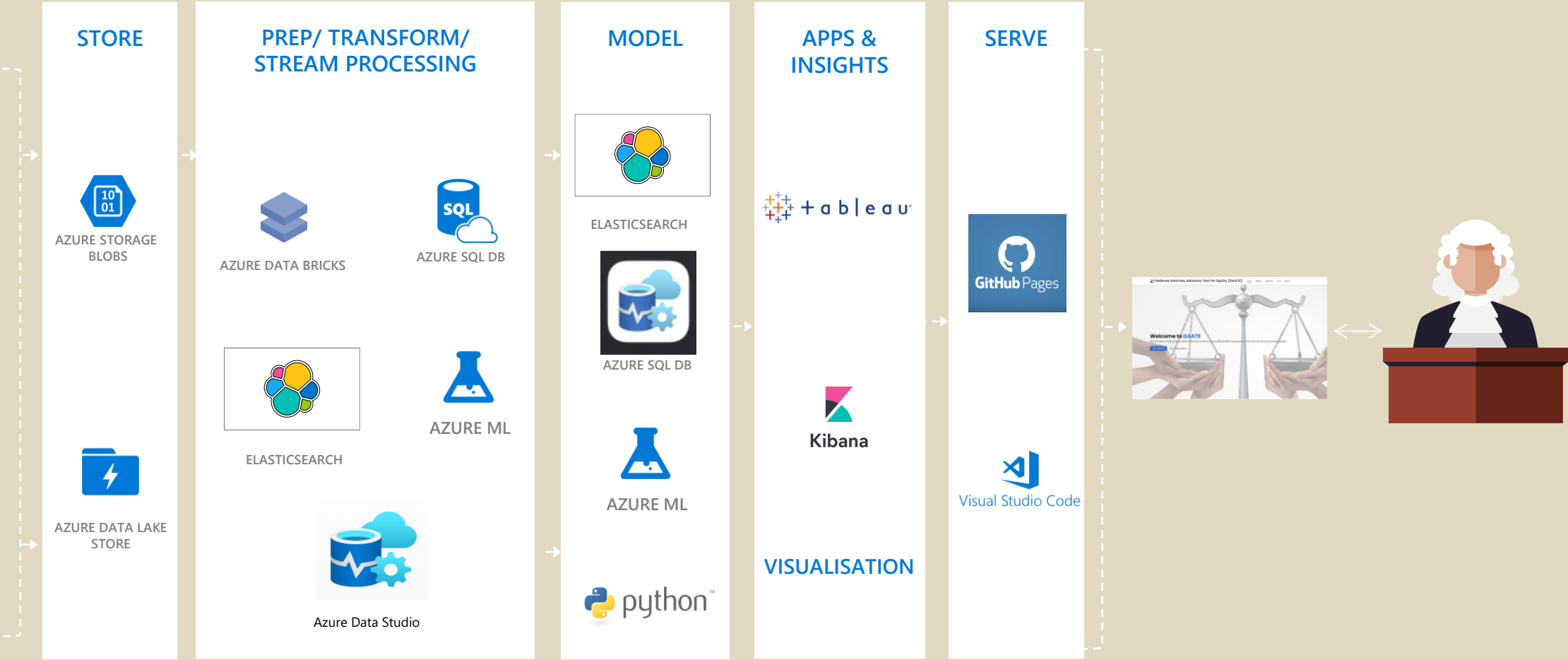
Crime Type (Statute)

- Want to provide crime type (e.g. "robbery") as filter for use in our product
- Statute codes are very specific, so we need to find accurate groupings that also have meaningful sample size
- **Challenge #1:** statute field is manually entered, so there are differences in format
- **Challenge #2:** statute is based on the defendant's primary offense, and there could be many other reasons for individuals with the same primary offense to ultimately end up with a very different number of sentencing points

DAATE MVP Architecture/Pipeline v.01



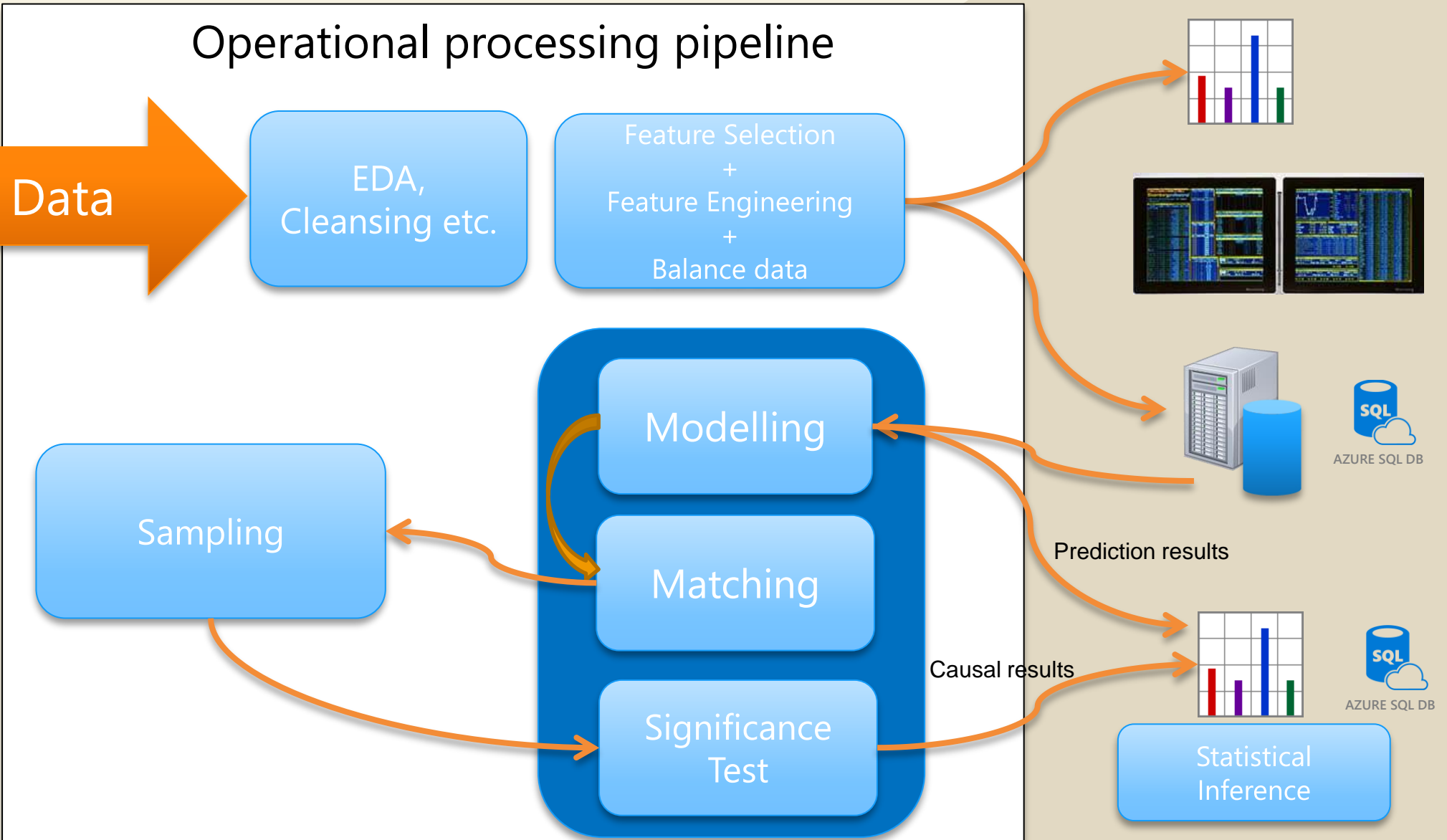
Florida Department
Of Corrections



DAATE MVP Model Architecture



Florida Department
Of Corrections



DAATE MVP Predictive Model

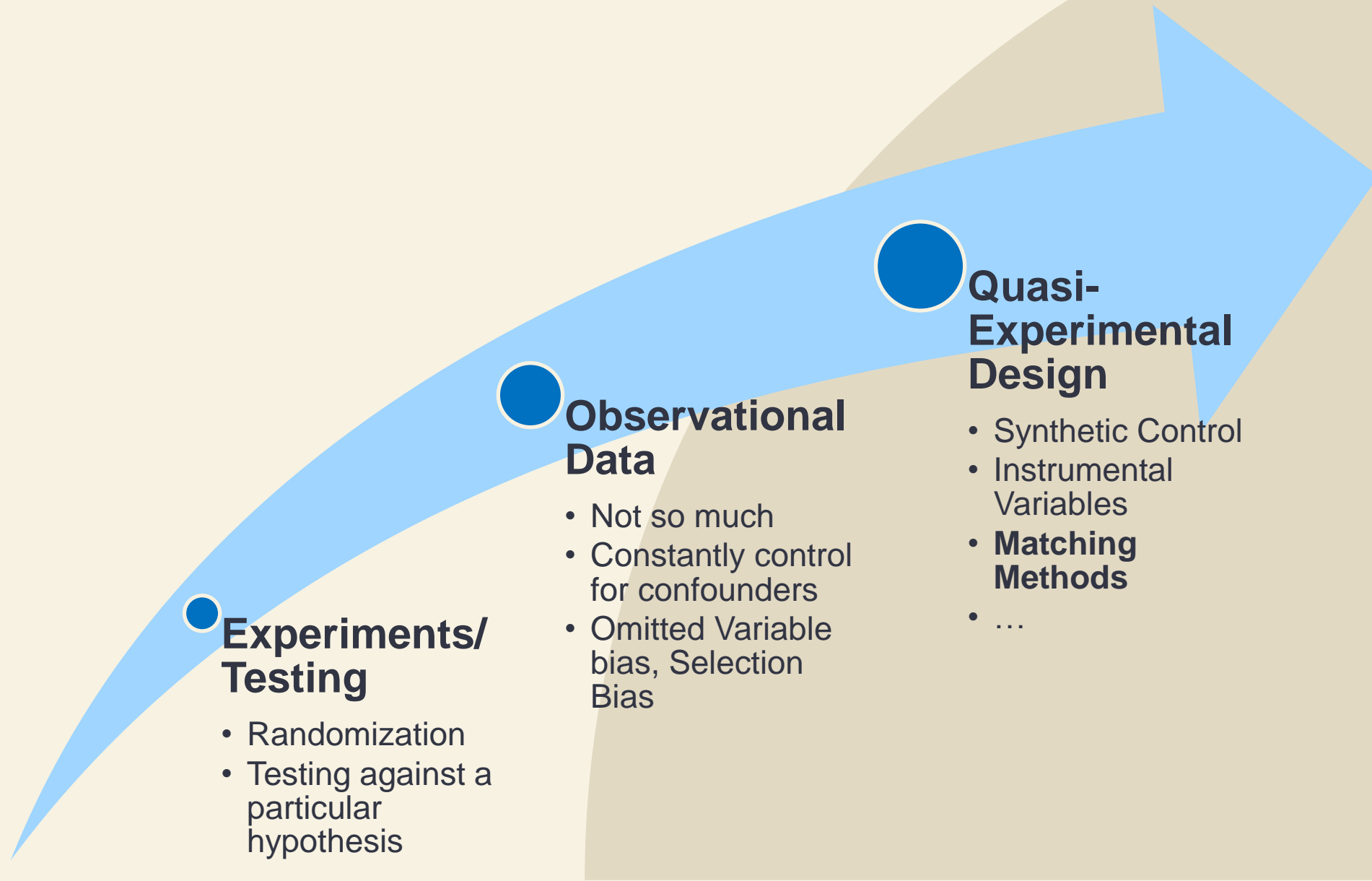
Overview

1. Segment data on Circuit & 3 top crime types
 2. AutoML with pycaret to evaluate 19 algorithms, 10 fold cv
 3. Pick a model based on metrics & training time
 4. Retrain on totality of test & validation, predict on hold-out set and report
- **Evaluation:** RMSE (same units as target, most intuitive to audience)
 - **Challenge:** Data set grew smaller for the 2nd & 3rd crime types. R2 decreases and anomalies
 - **Trade offs:** Chose models with slight loss of accuracy for faster training times
 - **Assumptions:** No OVB

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
ridge	Ridge Regression	128.8533	96718.5872	309.0710	0.5754	3.0279	0.9443	0.0210
lightgbm	Light Gradient Boosting Machine	132.4459	96663.5388	309.0464	0.5748	3.0544	0.9034	0.7920
br	Bayesian Ridge	128.8935	97181.7357	309.8037	0.5733	3.0267	0.9431	0.0800
gbr	Gradient Boosting Regressor	133.2856	97134.4582	309.6884	0.5730	3.1364	0.9417	0.3370
lr	Linear Regression	129.2124	97443.1916	310.2528	0.5721	3.0284	0.9449	0.0270
lasso	Lasso Regression	131.5962	100928.3346	315.3701	0.5587	3.0610	0.9645	0.0530
par	Passive Aggressive Regressor	115.8579	103668.9573	319.8782	0.5457	2.3742	0.8438	0.1730
huber	Huber Regressor	118.3730	105475.1587	322.5171	0.5383	2.5091	0.8036	0.6390

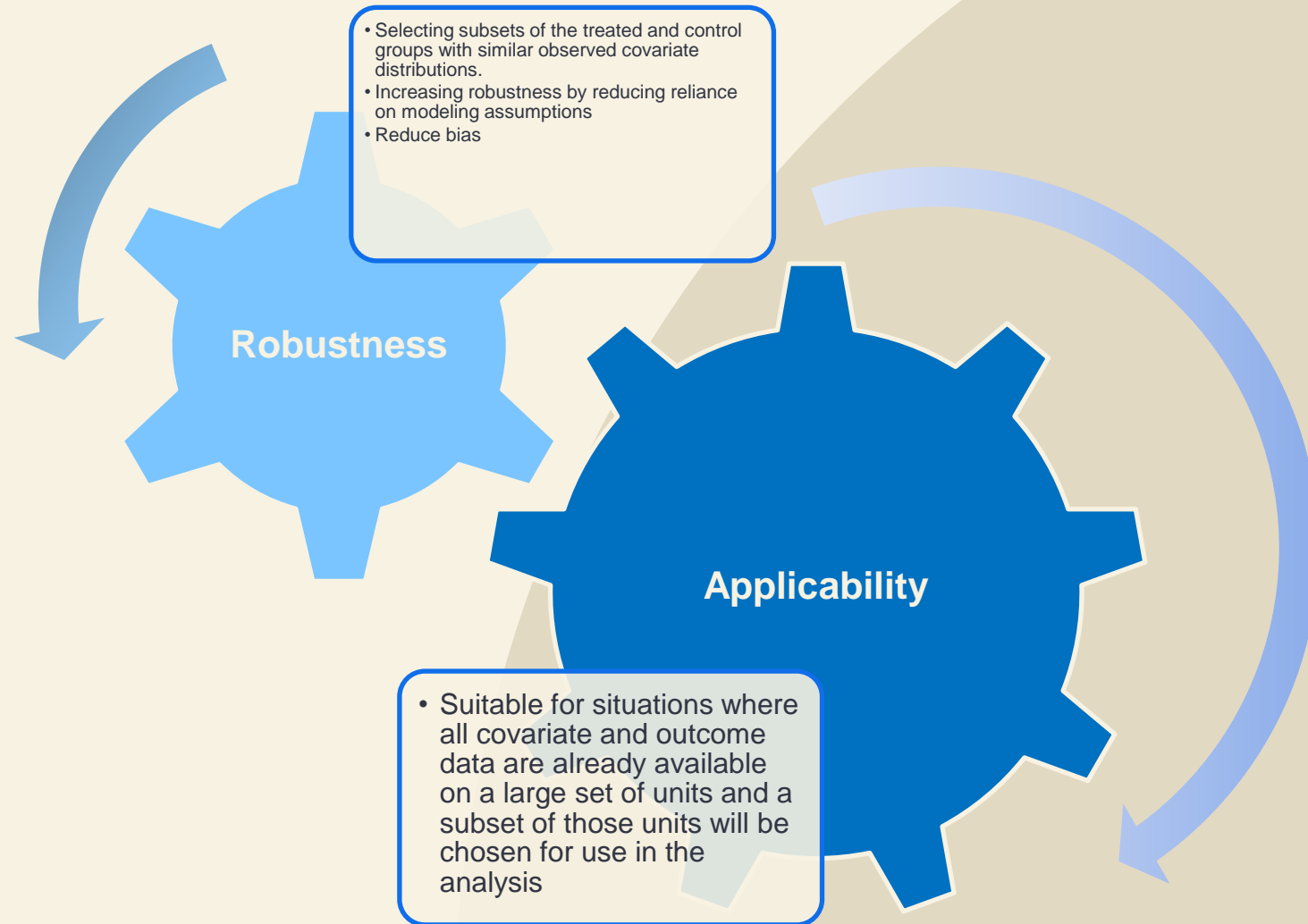
DAATE MVP Causal Model

MVP Hypothesis Testing vs. Observational Data



DAATE MVP Causal Model

Experimental Design - Matching Methods



DAATE MVP Causal Model

Quasi- Experimental Design - Matching Methods

Propensity Score Matching (predictive)

- Control for the likelihood of being in “group A” vs “group B” based on the given data attributes
- Requires matching through a binary variable

Clustering Based Matching (unsupervised)

- Generate data clusters based on feature proximity
- Distance based clusters requires explainability for categorical features
- **Less situational**

DAATE MVP Causal Model

Initial Results

Evidence of Racial Bias



P Values will be adjusted using multiple comparison adjustment, such as "Bonferroni correction", to minimize type-1 error

[37]:

	T statistic	df	pvalue 2 sided	Difference in mean	lb	ub	county	crime
0	4.173812	8602.129829	0.000030	48.236806	25.582286	70.891325	pinellas	drug
0	0.899278	3246.784702	0.368571	19.164822	-22.620217	60.949860	polk	drug
0	2.156391	463.695602	0.031567	79.059701	7.013535	151.105868	monroe	drug
0	-1.048804	3039.092106	0.294352	-29.345144	-84.206073	25.515784	escambia	drug
0	2.116456	2444.939459	0.034407	53.224673	3.911058	102.538288	volusia	drug
0	4.366550	3548.311605	0.000013	141.234824	77.818707	204.650942	pinellas	robbery
0	4.214603	1676.774760	0.000026	154.932914	82.830682	227.035146	polk	robbery
0	1.049870	120.736058	0.295875	65.437500	-57.962233	188.837233	monroe	robbery
0	3.619346	2787.351030	0.000301	133.875843	61.347330	206.404357	escambia	robbery
0	1.905694	977.621499	0.056983	92.475836	-2.751385	187.703058	volusia	robbery
0	4.766041	28192.710360	0.000002	49.327710	29.041549	69.613871	pinellas	burglary
0	1.121132	13158.606257	0.262252	18.366758	-13.744988	50.478504	polk	burglary
0	0.707595	1243.815442	0.479329	24.588141	-43.584731	92.761013	monroe	burglary
0	2.934266	14239.030094	0.003349	46.054442	15.289432	76.819452	escambia	burglary
0	2.119393	8161.319922	0.034087	43.827477	3.290854	84.364100	volusia	burglary

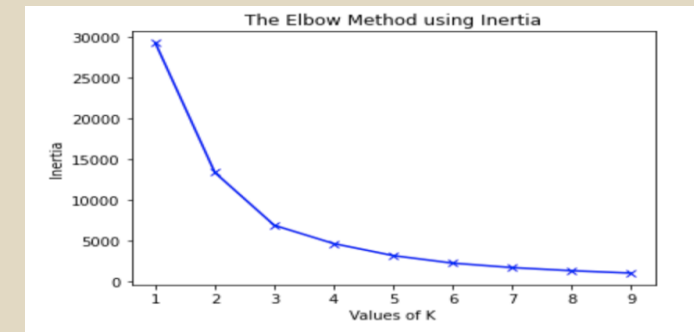
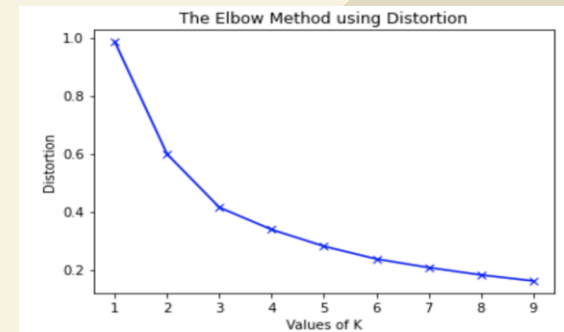
DAATE MVP Causal Model

Evaluation

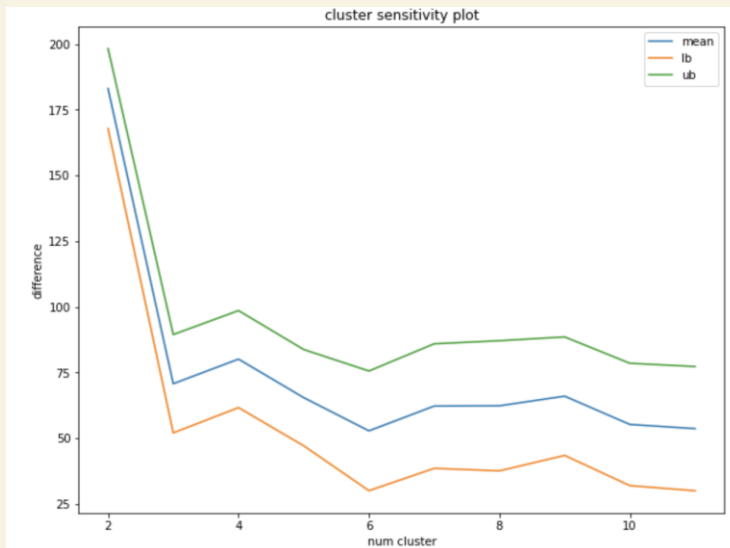
- SME feedback and domain knowledge checks



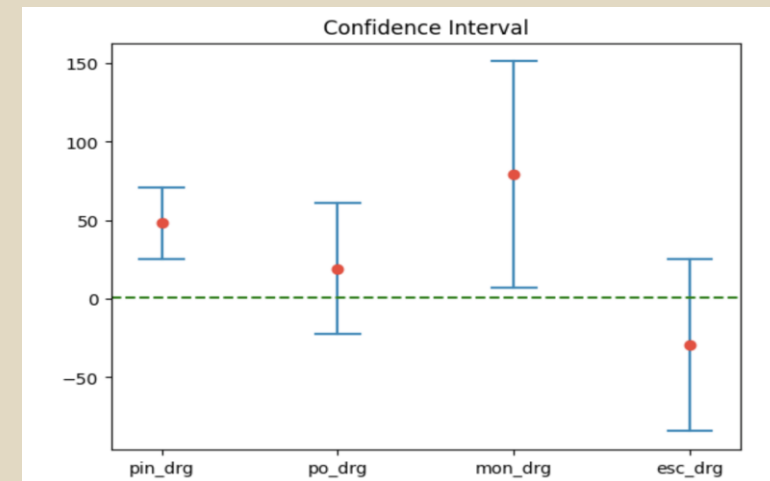
- Data driven method: 1) Elbow method 2) Empirical method 3) intra cluster variance



Sensitivity Analysis



- Confidence Interval and significance tests evaluation



DAATE MVP Causal Model

Trade Offs and Assumptions

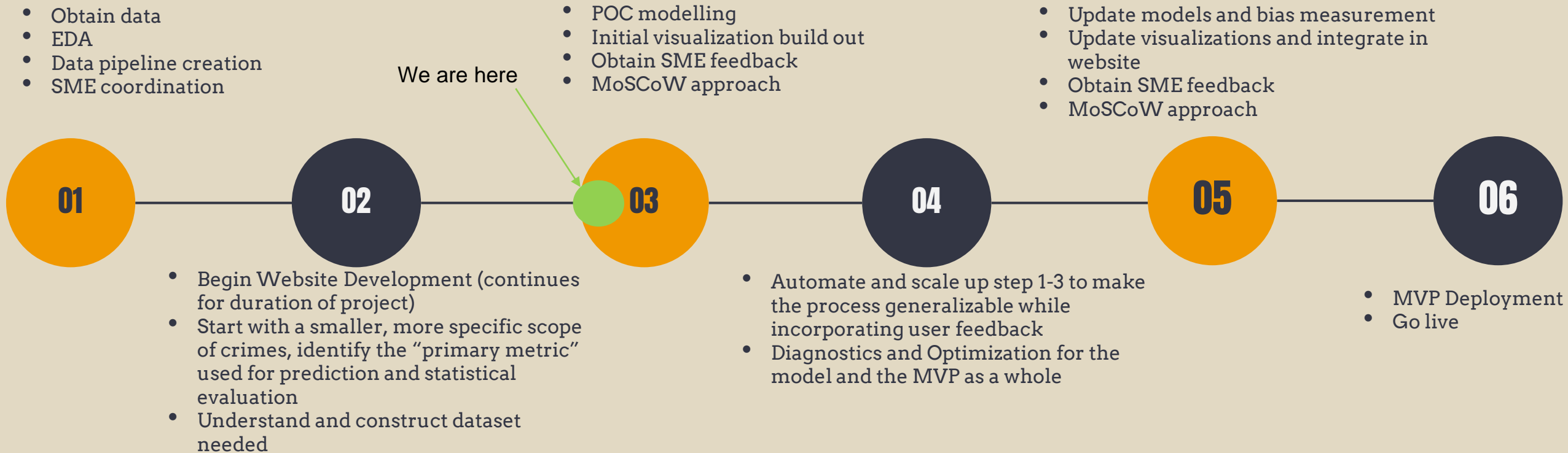
Trade Off

- Bias and Variance Trade Off
- Matching Explainability
- Computation Efficiency

Model Assumptions

- Stable unit treatment value assumption (SUTVA): treatment of one unit does not affect the potential outcome of other units
- Ignorability: treatment assignment is independent of the potential outcome

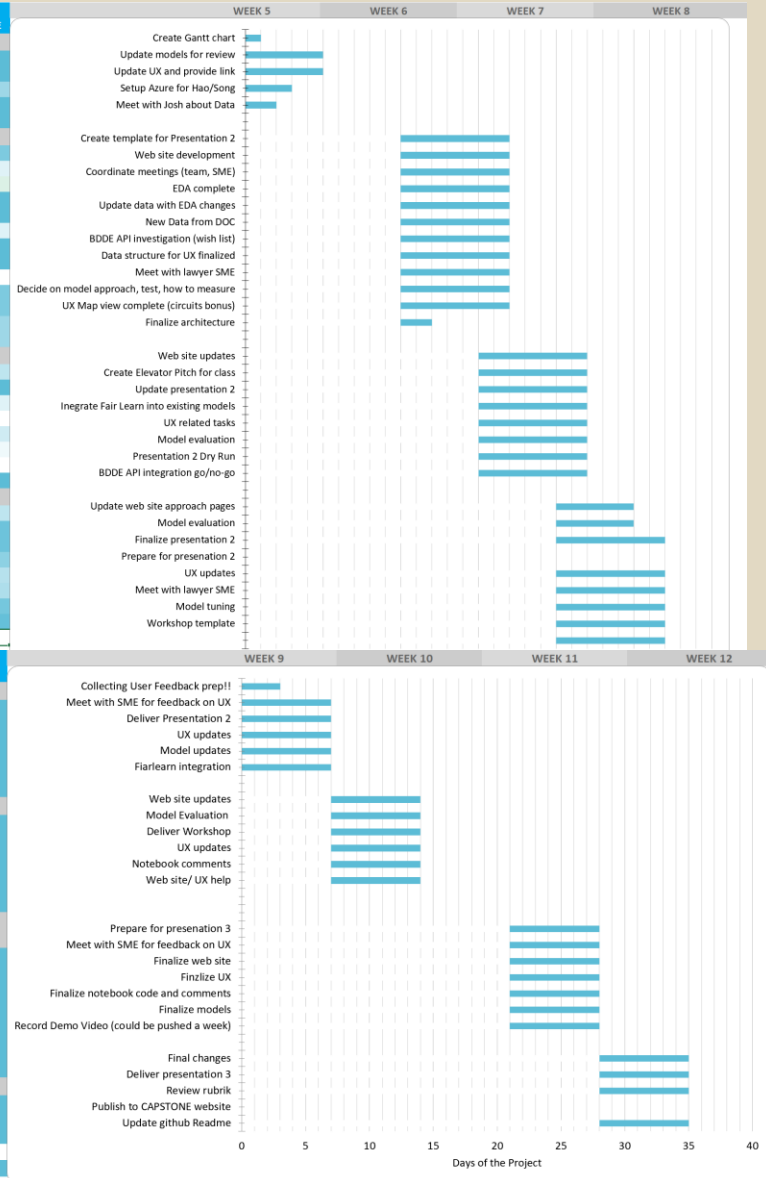
DAATE MVP Timeline



Note: For our MVP we will be focusing on White Americans and Black Americans defendants only for race and we may limit the initial list of crimes.

DAATE MVP Tasks

TASK NAME	START DATE	END DATE	START ON DAY*	DURATION* (WORK DAYS)	TEAM MEMBER	PERCENT COMPLETE
Week 6 (Feb 9)						
Create Gantt chart	2/7	2/8	0	1	Jackie	100%
Update models for review	2/7	2/11	0	5	Hao	100%
Update UX and provide link	2/7	2/11	0	5	Song	60%
Setup Azure for Hao/Song	2/7	2/11	0	3	Robert	100%
Meet with Josh about Data	2/7	2/9	0	2	Robert	100%
Week 7 (Feb 16)						
Create template for Presentation 2	2/14	2/20	10	7	Jackie	80%
Web site development	2/14	2/20	10	7	Jackie	20%
Coordinate meetings (team, SME)	2/14	2/20	10	7	Jackie	
EDA complete	2/14	2/20	10	7	Song	100%
Update data with EDA changes	2/14	2/20	10	7	Robert	100%
New Data from DOC	2/14	2/20	10	7	Robert/Jackie	20%
BDDE API investigation (wish list)	2/14	2/20	10	7	Robert	100%
Data structure for UX finalized	2/14	2/20	10	7	Hao	100%
Meet with lawyer SME	2/14	2/20	10	7	Team	0%
Decide on model approach, test, how to measure	2/14	2/20	10	7	Hao/Song	80%
UX Map view complete (circuits bonus)	2/14	2/20	10	7	Song	60%
Finalize architecture	2/14	2/20	10	2	Robert/Jackie	60%
Week 8 (Feb 23)						
Web site updates	2/21	2/27	15	7	Jackie	40%
Create Elevator Pitch for class	2/21	2/27	15	7	Team	100%
Update presentation 2	2/21	2/27	15	7	Team	20%
Inegrate Fair Learn into existing models	2/21	2/27	15	7	Robert	0%
UX related tasks	2/21	2/27	15	7	Song	30%
Model evaluation	2/21	2/27	15	7	Hao	10%
Presentation 2 Dry Run	2/21	2/27	15	7	Team	0%
BDDE API integration go/no-go	2/21	2/27	15	7	Team	100%
Week 9 (March 2)						
Update web site approach pages	2/28	3/6	20	5	Jackie	40%
Model evaluation	2/28	3/6	20	5	Team	90%
Finalize presentation 2	2/28	3/6	20	7	Team	90%
Prepare for presenation 2	2/28	3/6			Team	70%
UX updates	2/28	3/6	20	7	Song	45%
Meet with lawyer SME	2/28	3/6	20	7	Team	50%
Model tuning	2/28	3/6	20	7	Hao/Team	85%
Workshop template	2/28	3/6	20	7	Jackie	95%
	2/28	3/6	20	7	Robert	0%
TASK NAME	START DATE	END DATE	START ON DAY*	DURATION* (WORK DAYS)	TEAM MEMBER	PERCENT COMPLETE
Week 10 (March 9)						
Collecting User Feedback prep	3/7	3/13	0	3	Jackie	0%
Meet with SME for feedback on UX	3/7	3/13	0	7	Jackie	0%
Deliver Presentation 2	3/7	3/13	0	7	Hao	0%
UX updates	3/7	3/13	0	7	Song	0%
Model updates	3/7	3/13	0	7	Robert	0%
Fiarlearn integration	3/7	3/13	0	7	Hao	0%
Week 11 (March 16)						
Web site updates	3/14	3/20	7	7	Jackie	0%
Model Evaluation	3/14	3/20	7	7	Hao	0%
Deliver Workshop	3/14	3/20	7	7	Team	0%
UX updates	3/14	3/20	7	7	Hao	0%
Notebook comments	3/14	3/20	7	7	Song	0%
Web site/ UX help	3/14	3/20	7	7	Robert	0%
SPRING BREAK MARCH 21ST						
Week 12 (March 30)						
Prepare for presentation 3	3/28	4/3	21	7	Jackie	0%
Meet with SME for feedback on UX	3/28	4/3	21	7	Jackie	0%
Finalize web site	3/28	4/3	21	7	Team	0%
Finlize UX	3/28	4/3	21	7	Song	0%
Finalize notebook code and comments	3/28	4/3	21	7	Hao	0%
Finalize models	3/28	4/3	21	7	Robert	0%
Record Demo Video (could be pushed a week)	3/28	4/3	21	7	Team	0%
Week 13 (April 6)						
Final changes	4/4	4/10	28	7	Team	0%
Deliver presentation 3	4/4	4/10	28	7	Team	0%
Review rubrik	4/4	4/10	28	7	Team	0%
Publish to CAPSTONE website	4/4	4/10			Team	
Update github Readme	4/4	4/10	28	7	Team	0%



DAATE MVP Challenges and Goals

Challenges:

- Understanding data and legal aspects (criminal codes etc.)
- Determine narrow enough scope to complete for MVP (selecting specific crime types and counties)
- Create a visualization that is useable and meaningful for defense attorneys

Goals:

- Algorithmic & Data Bias. Fairness Toolkit
- Create a visualization that is useable and meaningful for defense attorneys

DAATE MVP Summary

Our Mission

To empower legal professionals to realize fairness and equity by providing transparency into the criminal justice system using technology to promote fairness for all people.

- Our goal is to create a MVP tool that satisfies our mission by
 - Understanding the data and the end users needs as well as the ethical implications
 - Providing transparency into the US criminal justice system through the use of Quasi- Experimental Design, and predictive data science techniques
 - Creating visually intuitive and explainable dashboards to our end users
- Feedback so far has been positive, so we are encouraged!

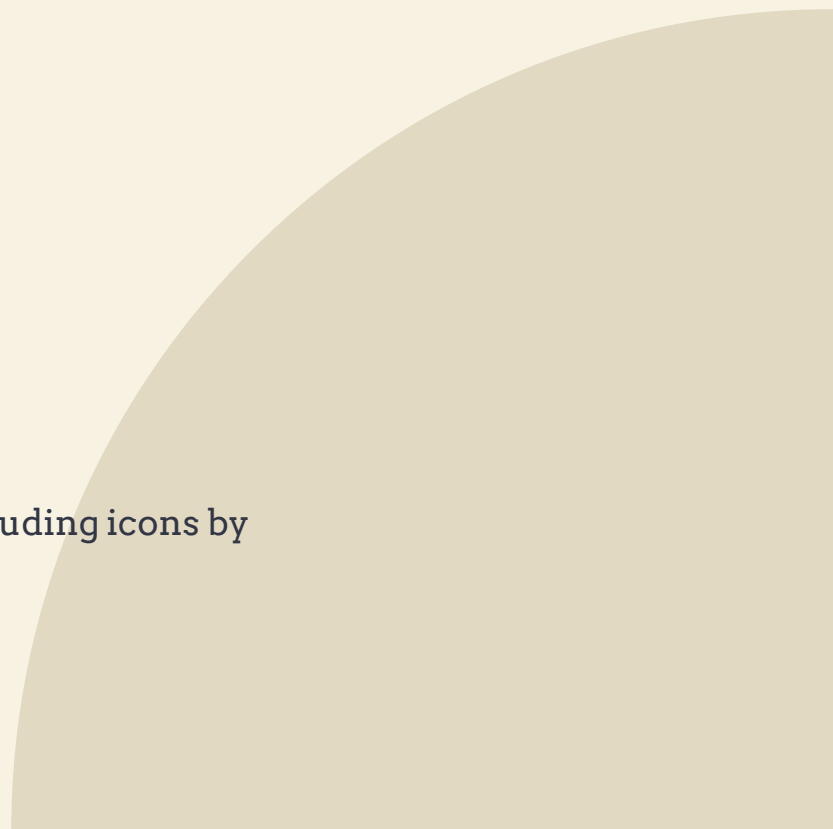


THANK YOU!

Do you have any questions?

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

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**



REFERENCES

- Alves, M (2021) 11 - Propensity Score Retrieved from <https://matheusfacure.github.io/python-causality-handbook/11-Propensity-Score.html> on January 22, 2022
- American Equity & Justice Group (2021) Retrieved from <https://americanequity.org/index.html> on January 5, 2022
- Salman, et. al (2016) Bias on the Bench Retrieved from <http://projects.heraldtribune.com/bias/> on January 2, 2022
- Number of lawyers in the United States from 2007 to 2021 (2021) Retrieved from <https://www.statista.com/statistics/740222/number-of-lawyers-us/> on January 20, 2022