

Defense Attorney Advisory Tool for Equity (DAATE)

Jackie Nichols, Hao Wu,
Robert Ling, Song Park



01.

Introduction

02.

Potential Impact

03.

Feasibility

04.

Approach

INTRODUCTION

- Recent research has shown that African (Black) Americans face harsher treatment than White Americans in the US criminal justice system; as much as **19.1% longer sentencing** than White male offenders
- Greater transparency into sentencing has the potential to provide additional checks-and-balances to the system and provide valuable information for the public and for individuals in the sector such as defense attorneys.
- There is public court/sentencing data across the US to support providing greater transparency into the criminal justice system.



Introduction

- No existing tools that use data science techniques like the ones we are considering which has prohibited defense attorneys from this approach in the past
- Ethical considerations like COMPAS – the data is just one aspect to the equity challenge the US criminal justice system faces
- DAATE – An advisory tool for defense attorneys to visit to gain information as they prepare to represent their client



Potential Impact

- Approximately 240k Defense Attorneys in the US and each lawyer has multiple cases per year
- With approximately **five** times more Black Americans being incarcerated than that of Whites, providing transparency into the criminal justice system towards promoting equity could have a substantial positive impact on Black Americans and in fact, all Americans
- Providing additional information from the US criminal justice system to a defense attorney can aid their case preparation to represent and to seek **fair and equal** treatment for their client.

DAATE MVP GOALS

Answer these three questions through an online advisory tool via visualizations

How do judges historically sentence individuals for a crime?

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)

Leverage DS techniques on real data and in new ways

Gain experience creating MVP's

Evaluate if this can be extended beyond capstone

DAATE 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.

DAATE EDA SNEAK PEEK

EDA

- 2004-2016 Sentencing Data from Florida
- Context Understanding
- Sanity Check
- ...

Demographic:

- Race
- Gender
- Date of Birth
- ...

Crime Description:

- Total Points
- Armed
- Location Proximity
- Crime Type
- Plea/Trial
- ...



"Outcome" Variable:

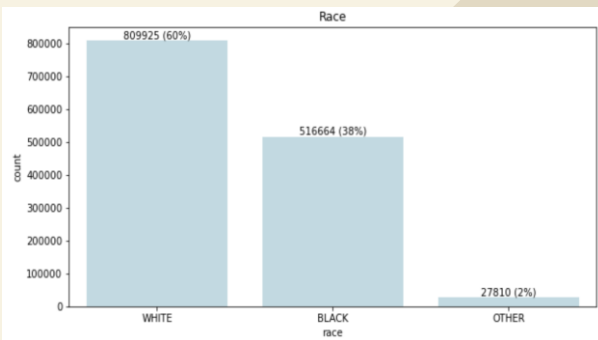
- Total Days Sentenced
- Severity of the Crime
- ☐ Degree in severity (life, 1st, 2nd...)
- ☐ Offensive Level (1-10)
- ...

DAATE EDA SNEAK PEEK

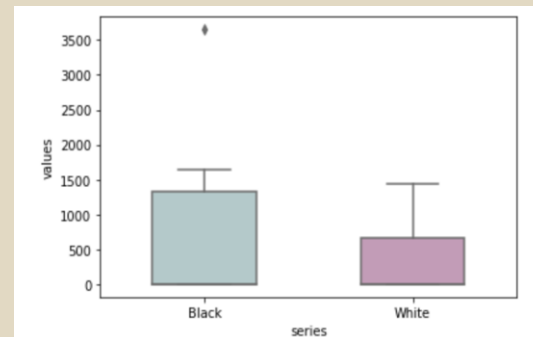
EDA

- Data Imbalance
- Missing Values and Outliers
- Interpretation and correlation
- ...

Total Crime distribution by Race



Total Sentencing time (days) by Race



Special Handling

- Extreme Values (99999)
- Systematic missing value
- Reality check

Correlation

- Linear vs Non-linear
- Regression based exploration
- Preventing "p" hacking
- Estimating model predictability



DAATE **FEASIBILITY**



DATA*

Working with two public datasets from Florida
Correction: DOC/ OBTS

Additional data from WA



MODELLING*

Determining best approach on
measuring bias and how to
measure magnitude/confidence.



ETHICAL*

Considering the ethical
implication of an advisory
tool and the language we
will need to include on
our website.



SUBJECT MATTER EXPERTS

The DAATE team has access
to defense attorneys, judges
and investigative journalists
that focus on equity in the
criminal justice system

*** Team has intro level to deep knowledge in these areas**

DAATE MODELLING ALGORITHMS & Techniques

A

Sampling Data

B

Bucketing Outcome
Variables

C

Predictive Modeling
&
Propensity Score
Generation

D

- Causal Construct
- Propensity Score Matching
 - Regression Discontinuity Design

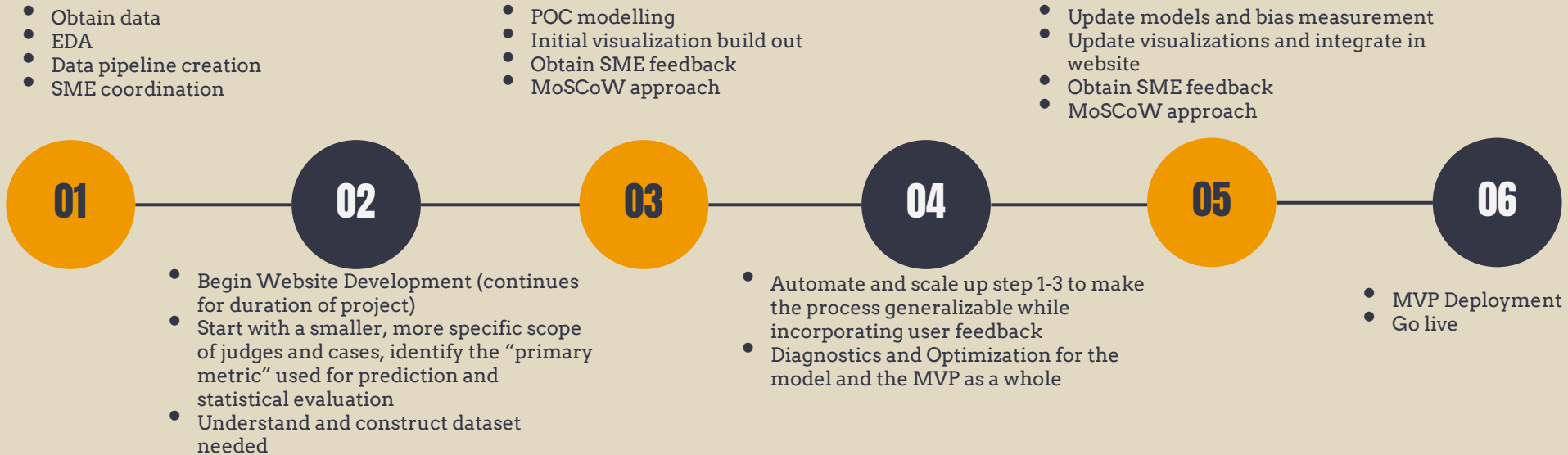
E

Treatment Effect
Estimation

F

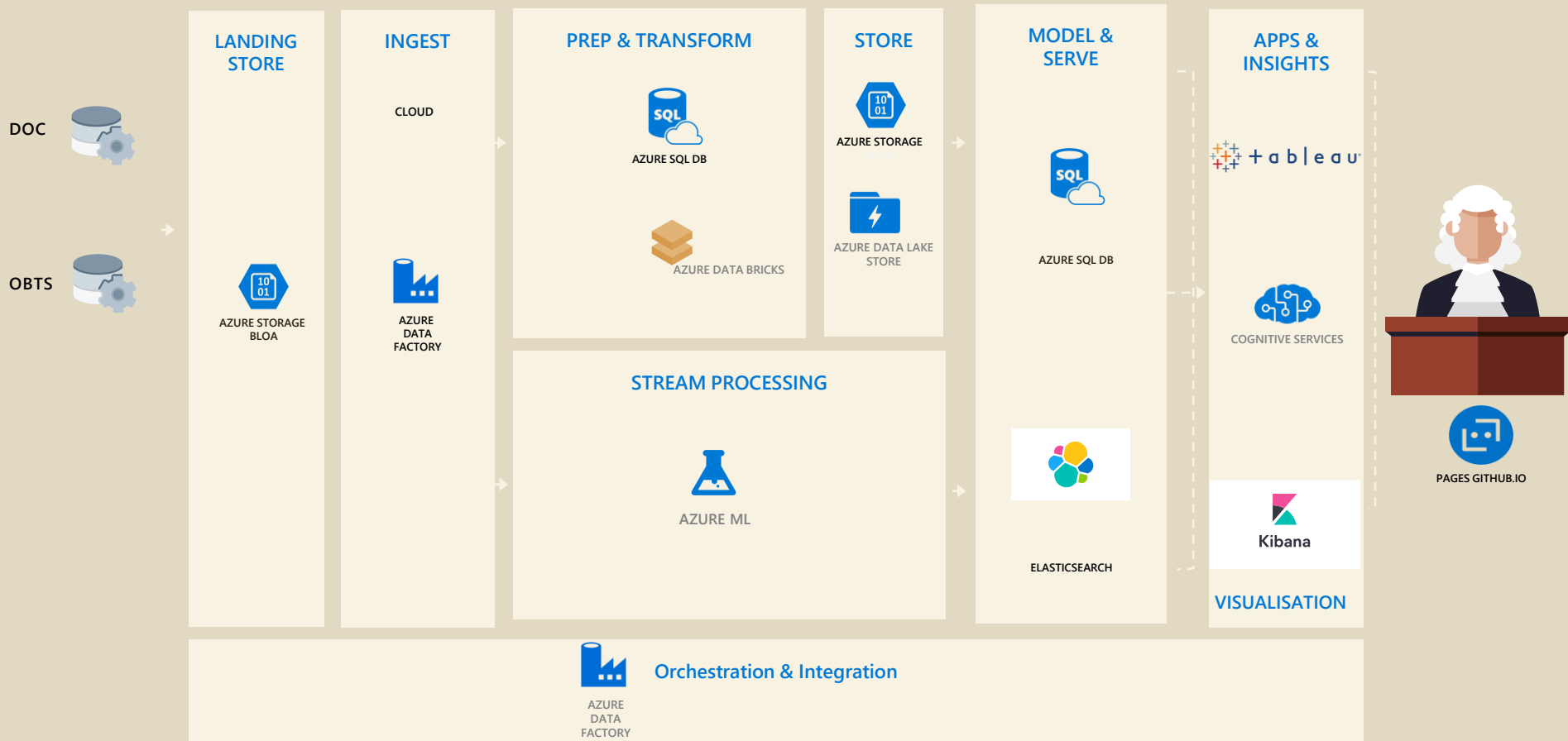
Automation
&
Upscaling

DAATE MVP APPROACH



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 PRODUCT ARCHITECTURE/PIPELINE V.01



WHAT SETS DAATE **APART?**

DAATE v0.1

Leveraging data science techniques to examine criminal justice data:

- Evaluate Different Models to score propensity of a judges sentencing history for bias
- Identify type of bias (race, gender, etc.)
- Measure magnitude and confidence of bias

DAATE v0.2

.....

American Equity and Justice Group

Leverages traditional data analytics to view criminal justice data



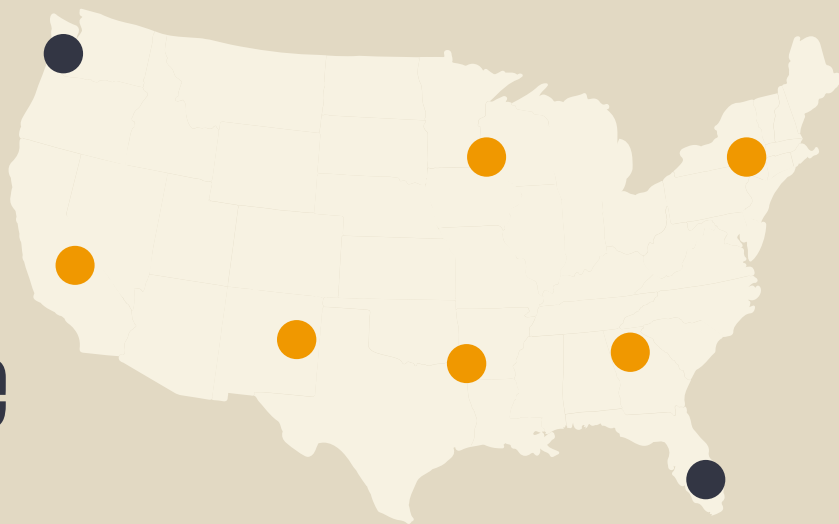
DAATE v.02 **EXPANSION**

+ **More US cities**

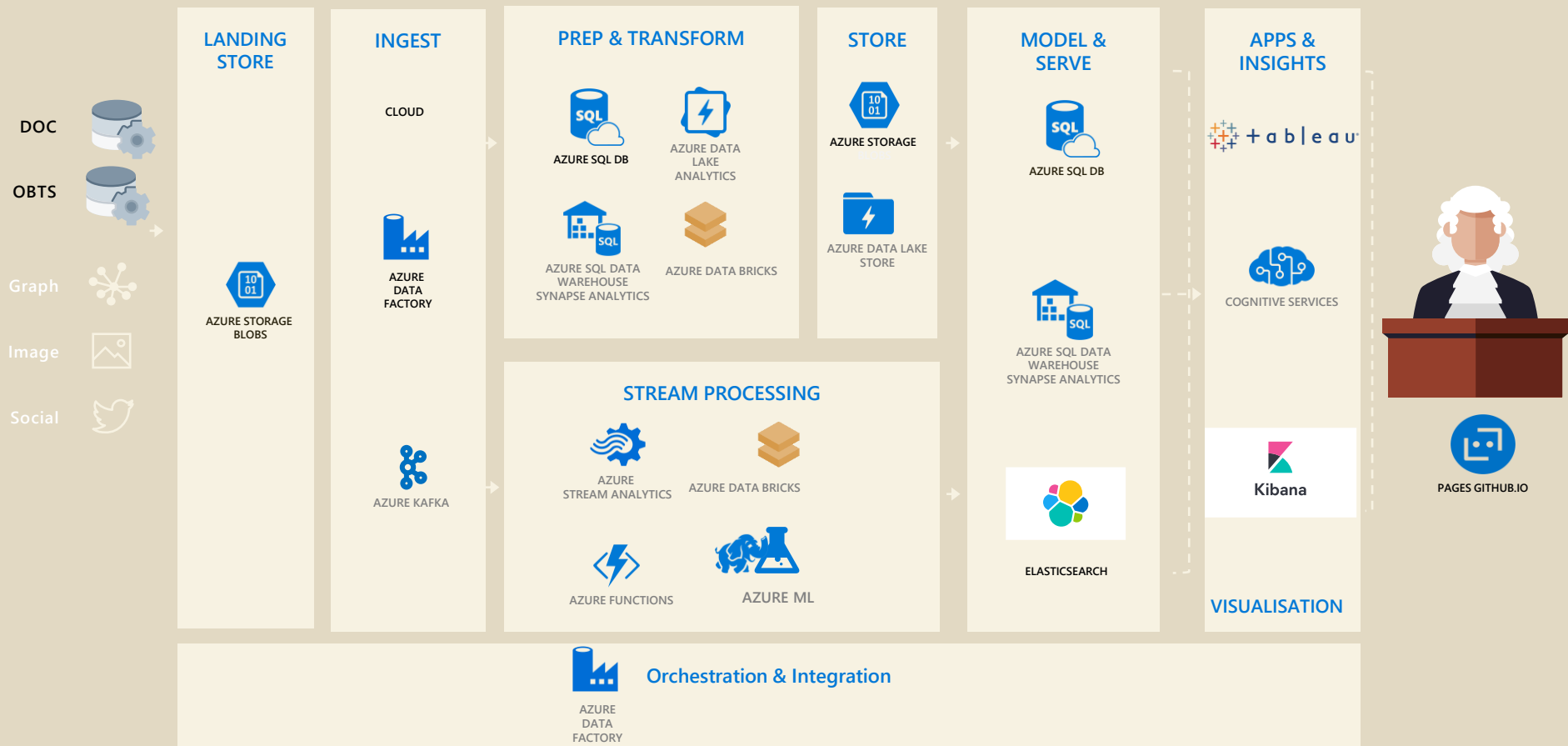
Continue to extend to US cities

+ **More Data Science**

- Bias in appearance
- NLP
- Real-time sentiment



DAATE PRODUCT ARCHITECTURE/PIPELINE V.02



DAATE CHALLENGES

- 1.3m rows-290 cols, 50GB-60m rows, Athena, Cost, Columnar
- Limits of AWS Glue & Data Studio – Load errors Strict Schema
- Document Store for EDA & Search. Elasticsearch.
- Could locally process the 1st dataset, 2nd needs Spark
- Understanding data and legal aspects (criminal codes etc.)
- Determine narrow enough scope to complete for MVP
- Translate Statistics into UX Lawyers can work with
- Algorithmic & Data Bias. Fairness Toolkit.

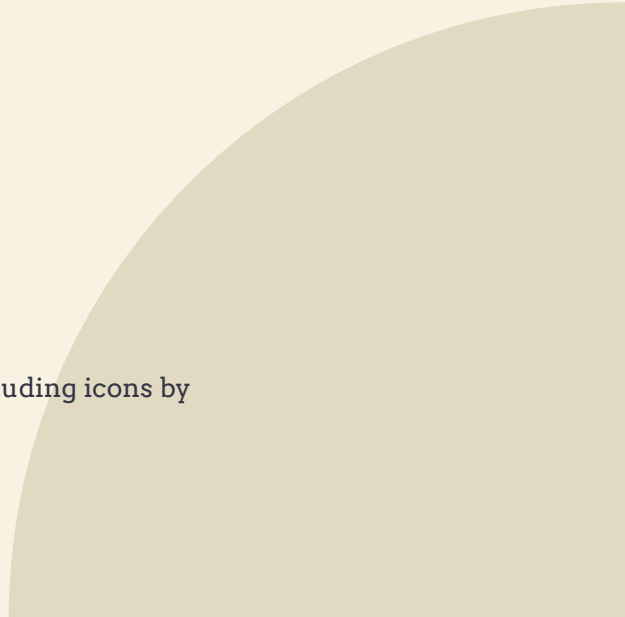


THANKS!

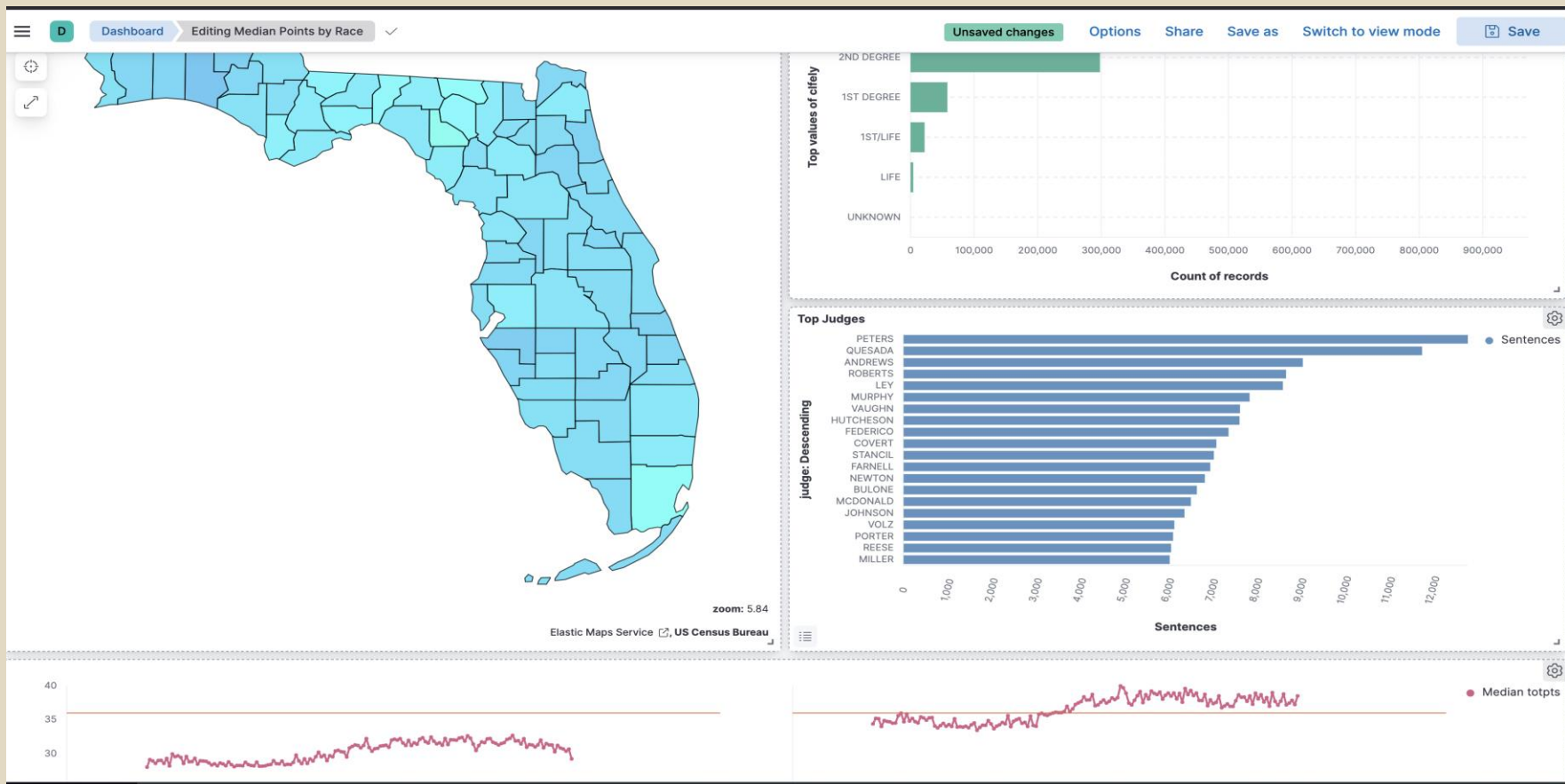
Do you have any questions?

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**



APPENDIX



Sample Data

sentdt	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	GAR	PASCO	SIRACUSA, P.	DOUGHERTY	RYAN	21AUG1989	NaN	WHITE	/	MALE
08MAY2007	R43285	SAO	ROS	PINELLAS	LUCE	GAMBLE	DOMINIQUE	01APR1988	NaN	BLACK	/	MALE
21SEP2006	M20788	SAO	UNK	MIAMI-DADE	FREEMAN	PEREZ	LAZARO	23DEC1980	NaN	WHITE	/	MALE
04FEB2009	085712	SAO	UNK	MIAMI-DADE	BLITCHKO	COOPER	JEFFERY	20AUG1982	NaN	WHITE	/	MALE

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	0	57.0	0	0	0	0	0	113.0	NONE	0.0	113.0
0	0	0	0	0	0	0	0	0	0	42.0	0	0	0	0	0	98.0	NONE	0.0	98.0
0	0	0	0	0	0	0	0	0	0	28.0	0	0	0	0	0	64.0	NONE	0.0	64.0
0	0	0	0	0	0	0	0	0	0	18.0	0	0	0	0	0	74.0	NONE	0.0	74.0
0	0	0	0	0	0	0	0	0	0	25.4	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	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	0	23.2	0	0	4	0	0	0	65.4	NONE	0.0	65.4
0	0	0	0	0	0	0	0	0	27.0	0	0	0	0	0	0	64.8	NONE	0.0	64.8

Complex Schema's

AWS

Services

Search for services, features, blogs, docs, and more

[Option+S]

🔍

🔔

?

Oregon ▾

robling-ucb ▾

AWS Glue

Data catalog

Databases

Tables

Connections

Crawlers

Classifiers

Schema registries

Schemas

Settings

ETL

AWS Glue Studio New

Blueprints

Workflows

Jobs

ML Transforms

Triggers

Dev endpoints

Notebooks

Security

Security configurations

Name

Description

Database

Classification

Location

Connection

Deprecated

Last updated

Input format

Output format

Serde serialization lib

Serde parameters

Table properties

Schema

sentencing

w210

csv

s3://mids-w210-capstone-equity/dump/sentencing/

No

Sun Jan 23 20:00:07 GMT-800 2022

org.apache.hadoop.mapred.TextInputFormat

org.apache.hadoop.hive ql.io.HiveIgnoreKeyTextOutputFormat

org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe

field.delim ,

skip.header.line.count 1sizeKey 180066352objectCount 3

UPDATED_BY_CRAWLER sentencingCrawlerSchemaSerializerVersion 1.0recordCount 60106

averageRecordSize 1601CrawlerSchemaDeserializerVersion 1.0compressionType gzip

columnsOrdered trueareColumnsQuoted falsedelimiter ,typeOfData file

Showing: 1 - 100 of 290 < >

	Column name	Data type	Partition key	Comment
1	sentdte	string		
2	dornum	string		
3	prepby	string		

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

REFERENCES

- Pura, J (2015) Propensity Score Methods for Causal Inference Retrieved from https://biostat.duke.edu/sites/biostat.duke.edu/files/Causal_Inference_and_Propensity_Scores_0.pdf on January 20, 2022
- United States Sentencing Commission (2017) Demographic Differences in Sentencing Retrieved from <https://www.ussc.gov/research/research-reports/demographic-differences-sentencing> on January 6, 2022
- Defense Attorney (2021), Defense Attorney Demographics And Statistics In The US Retrieved from <https://www.zippia.com/defense-attorney-jobs/demographics/> on Feb 1, 2022

**A PICTURE IS
WORTH A
THOUSAND WORDS**



