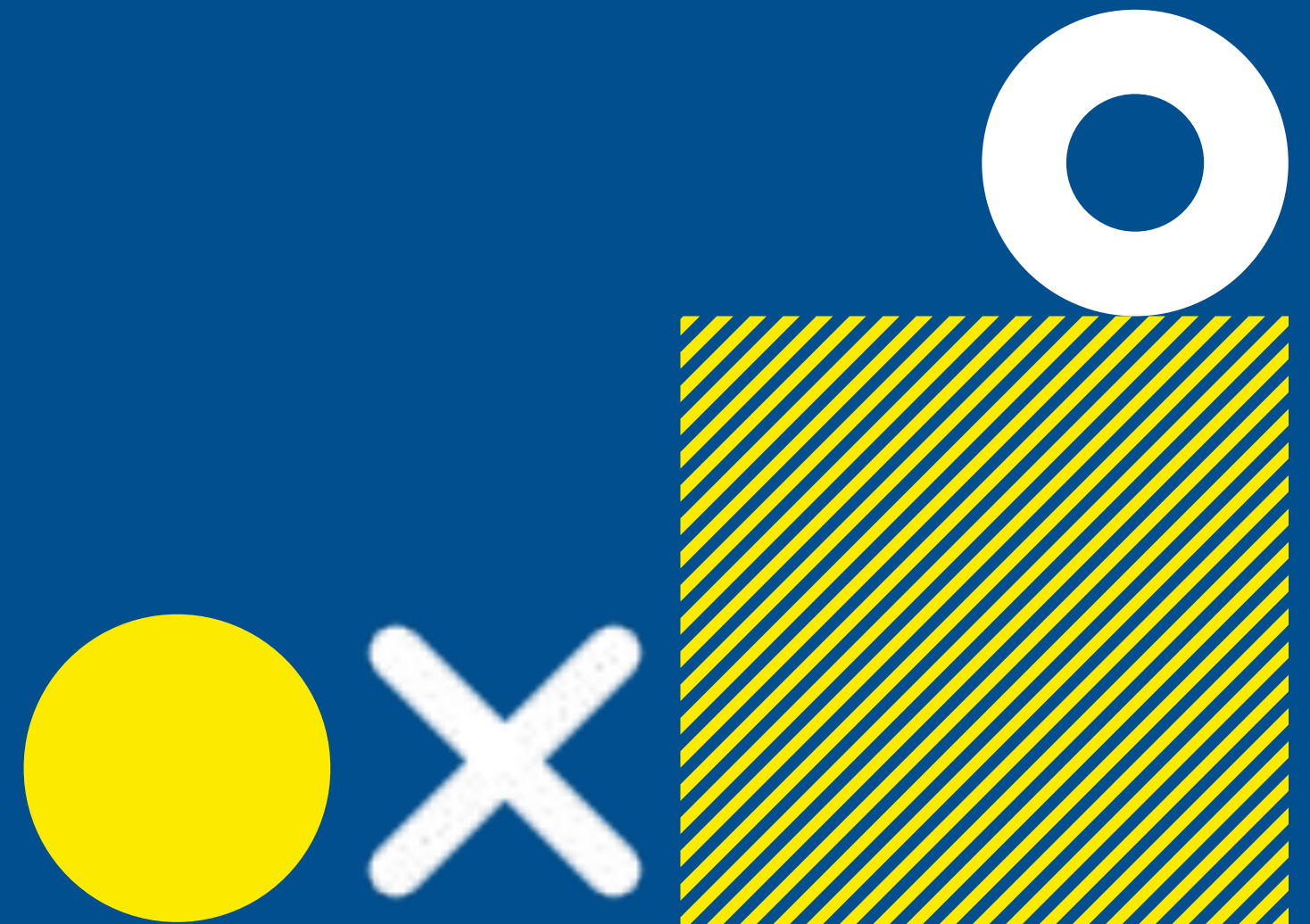# Predicting Recidivism

Using machine learning to build a fairer model for predicting reoffending risk

Ursula Guo, Devi Mahajan, Nidhi Pareddy

CONTENT

Problem

—

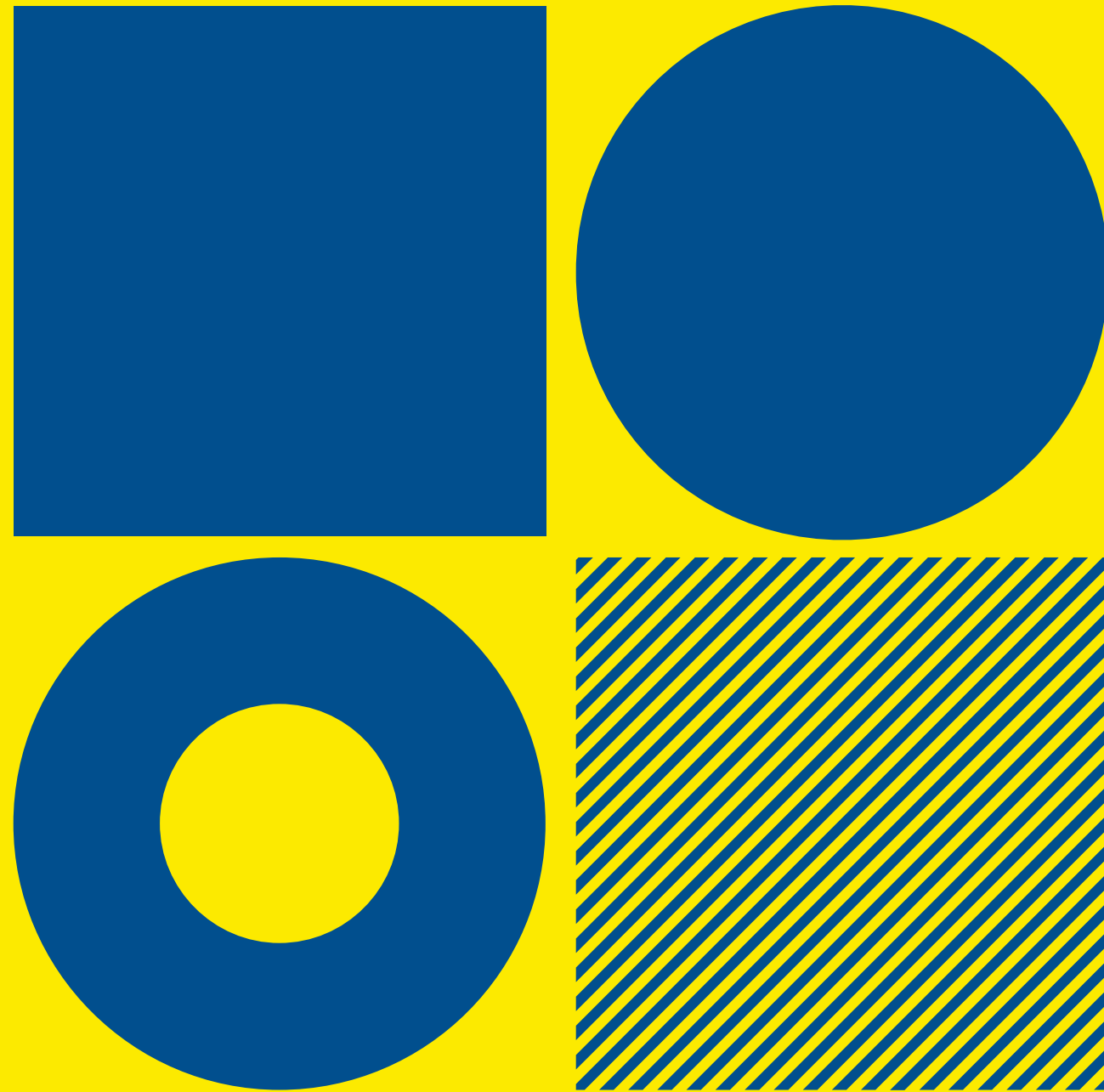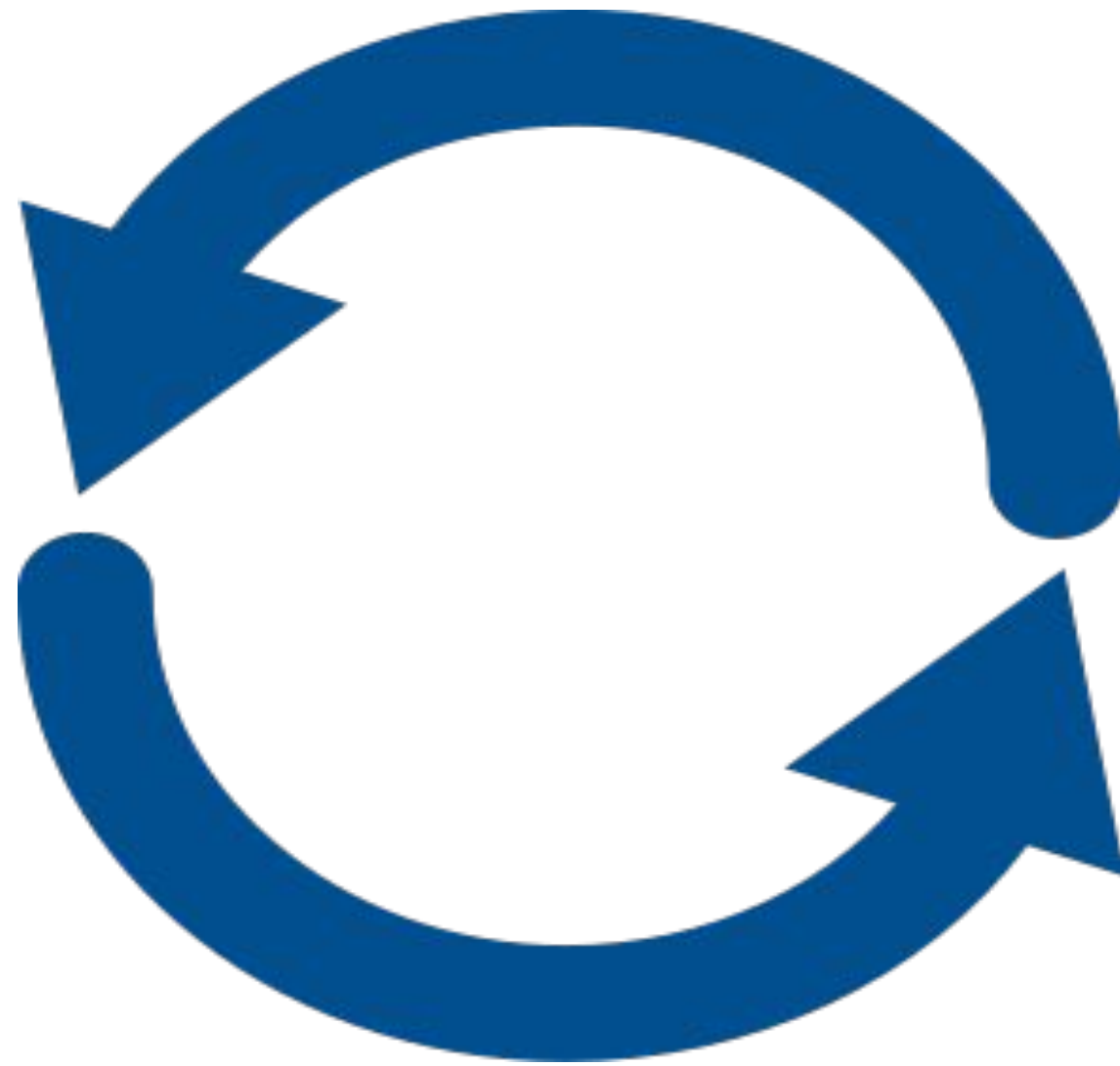What's wrong with current AI-driven

recidivism predictors?

What is Recidivism?

BACKGROUND

AI-based **risk assessment algorithms** are used in the U.S. criminal justice system to inform bail, sentencing, and parole decisions. Studies such as ProPublica's analysis of COMPAS have exposed **fairness issues**, particularly in how these tools **disproportionately misclassify** defendants based on **race**.

PROBLEM STATEMENT

Current AI assessment tools such as COMPAS show significant **racial bias** and misclassify Black and White defendants, undermining the fairness and accuracy of the criminal justice system. Our goal is to build a **transparent, fair recidivism prediction model** that improves public safety while meeting ethical and regulatory standards.

OUR APPROACH

We aim to train three different machine learning algorithms to attempt to classify reoffenders without relying on race as a primary factor.

Why This Matters: Racial Disparity in US Jails

- **Black Americans have historically faced significantly higher incarceration rates than other races:** In 2021, Black individuals were imprisoned at a rate nearly five times that of white individuals

- **In New York City, racial disparities in jail admissions have worsened over time:** By 2021, Black individuals were incarcerated at a rate 11.6 times higher than white individuals, up from 4.8 times in 2016

- Disproportionate incarceration can lead to **disruption of family structures** and **educational/development challenges for children** of incarcerated parents as well as disinvestment in Black communities and neighbourhoods, resulting in long-lasting harm

Racial Disparities Persist in Many U.S. Jails

Despite narrowed gap in incarceration rates, Black people remain overrepresented in jail populations, admissions—and stay longer on average

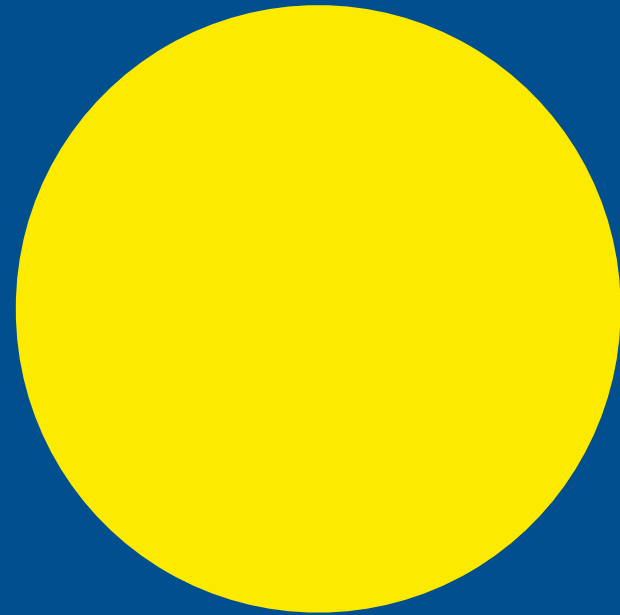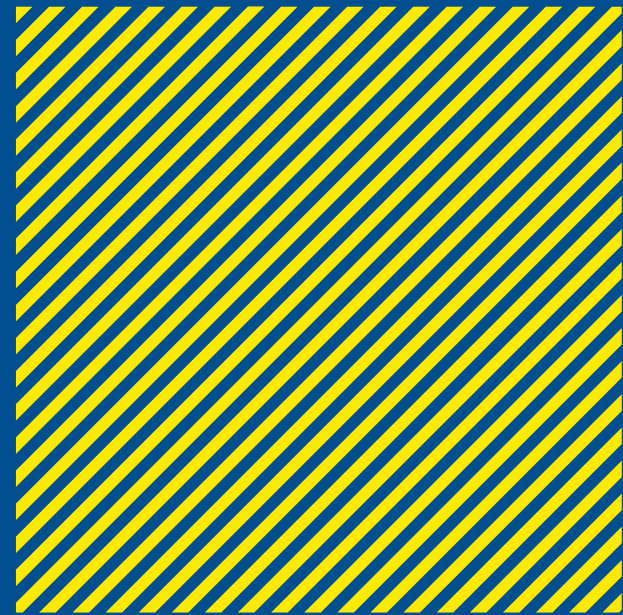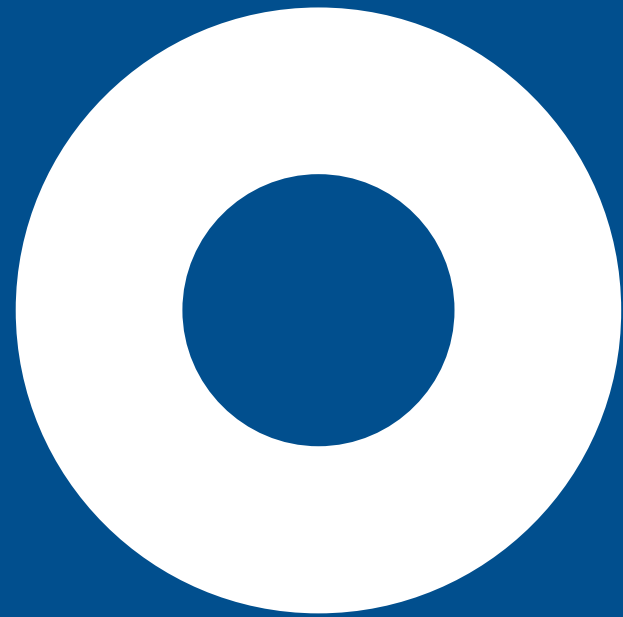| 2.2 million	32%	4x	73.9%
prisoners in the U.S., compared to 1.65 million in China, whose population is four times that of the U.S.	of Louisiana's population is Black, and yet 67.8% of prisoners are Black.	That means that in Louisiana, Black people are incarcerated at four times the rate of White people.	of people sentenced to life in prison in Louisiana are Black.

Key Statistics:

- Percent of Black Americans in the general U.S. population: **13%** +
- Percent of people in prison or jail who are Black: **37%** +
- Prison incarceration rate for Native people vs nation as a whole: **763 vs. 350 per 100,000**
- Percent of people serving life, life without parole, or "virtual life" sentences who are Black: **48%** +
- Arrest rate for Black vs white Americans (2020): **4,223 vs. 2,092 per 100,000** +
- Yearly arrests of Black Americans (2020): **1.99 million** +
- Percent of people on probation or parole who are Black: **30%** +

Black Prisoners Face Higher Rate of Botched Executions, Study Finds

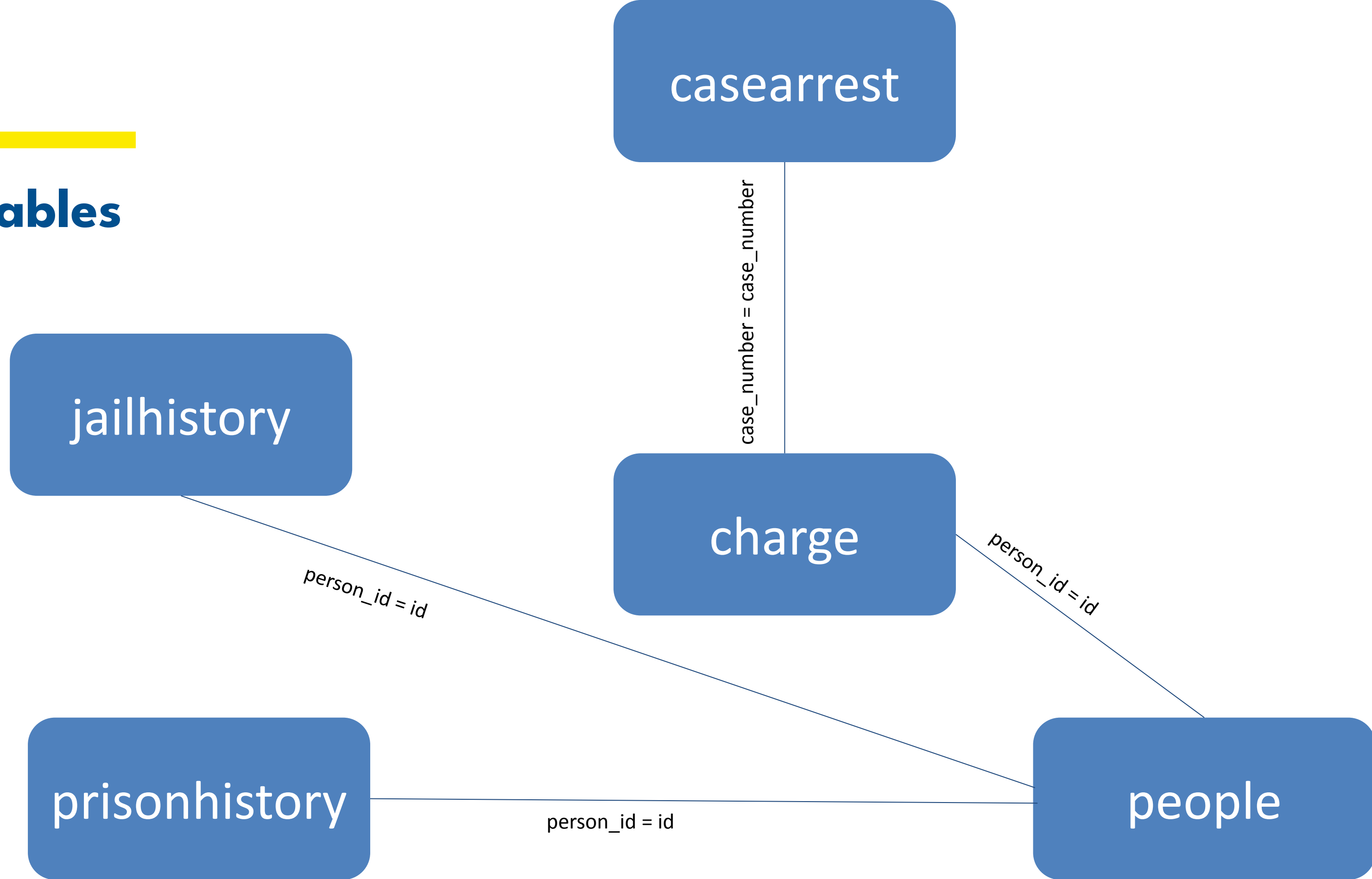**Racial disparities in incarceration**

Since 2020, the Bureau of Justice Statistics has reported on more racial and ethnic categories in jail populations, as well as prison populations by sex. These new statistics underscore the ongoing racial injustice of prisons, where the national incarceration rate of Black people is *six times the rate of white people* and more than twice the rate in every single state.

Our Data

Tables

casearrest

jailhistory

charge

prisonhistory

people

case_number = case_number

person_id = id

person_id = id

person_id = id

8

Data Overview

- SQLite database with 11,757 unique individuals obtained from the paper *How We Analyzed the COMPAS Recidivism Algorithm* by Larson at el. in 2016.
- 6 continuous features

| Variable Name	Description
juv_fel_count	Counts of juvenile felony convictions
juv_misd_count	Counts of juvenile misdemeanor convictions
juv_other_count	Counts of juvenile convictions other than felony or misdemeanor
prison_duration_days	Total duration of days spent in prison for one charge
age	Age of person as of today
priors_count	Total numbers of prior convictions

Data Overview

- SQLite database with 11,757 unique individuals obtained from the paper *How We Analyzed the COMPAS Recidivism Algorithm* by Larsen at el. in 2016.
- 6 continuous features
- 10 categorical features

Features	Description
arrest_charge_degree	Degree of the arrest charge
legal_status	Legal status of the person arrested
marital_status	Marital status of the person arrested
race	Race of the person arrested
sex	Sex of the person arrested
c_charge_degree	Degree of crime charge in sentencing
filling_type	Types of filing
filing_agency	Agency responsible for the filing
initial_charge	Initial charge type
initial_status	Initial status of the case

EDA
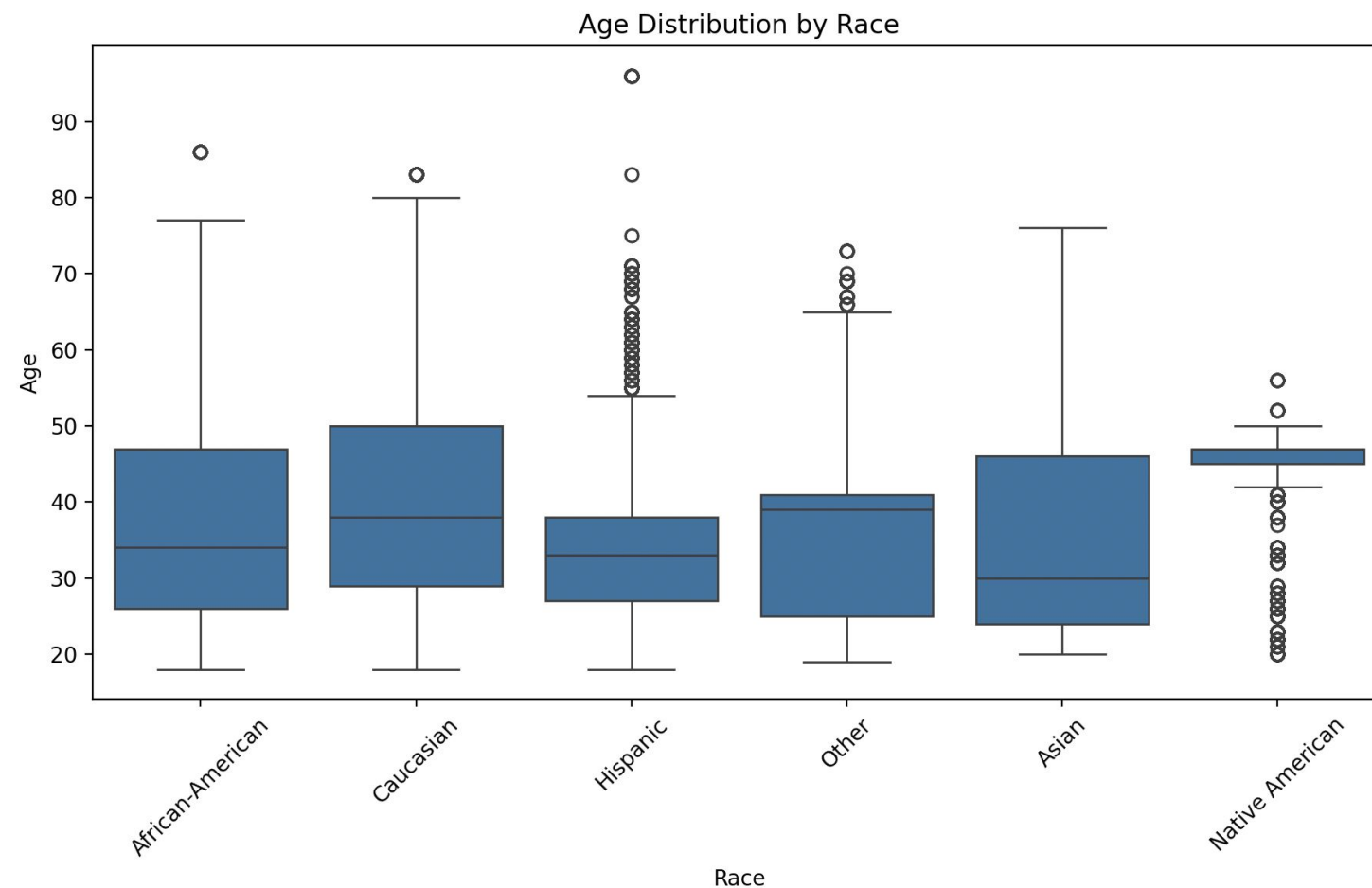

Distribution of Race

Distribution of Age
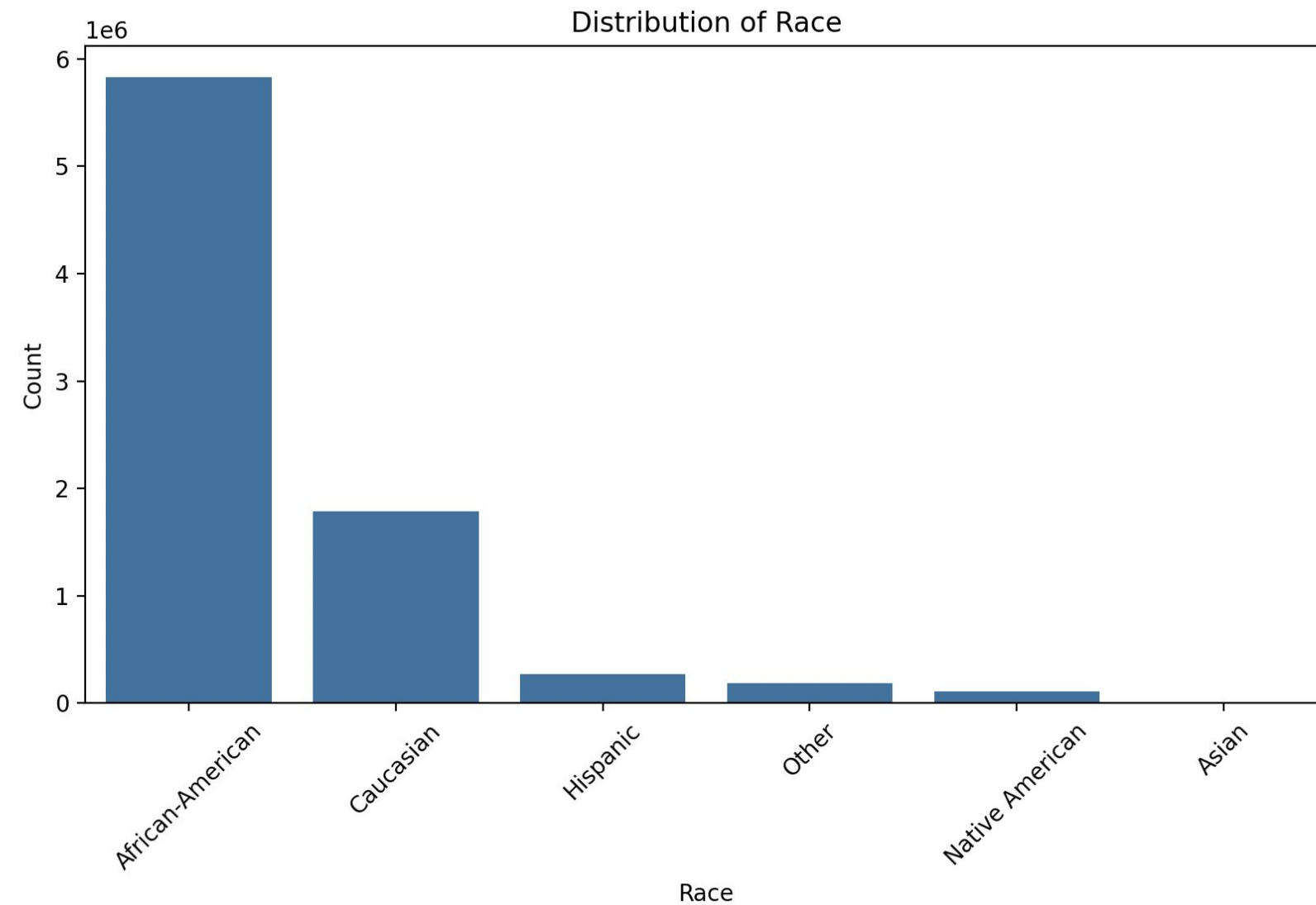
The original dataset is disproportionally skewed in race, which might raise concern for how our models uses race categories as a high indicator of recidivism
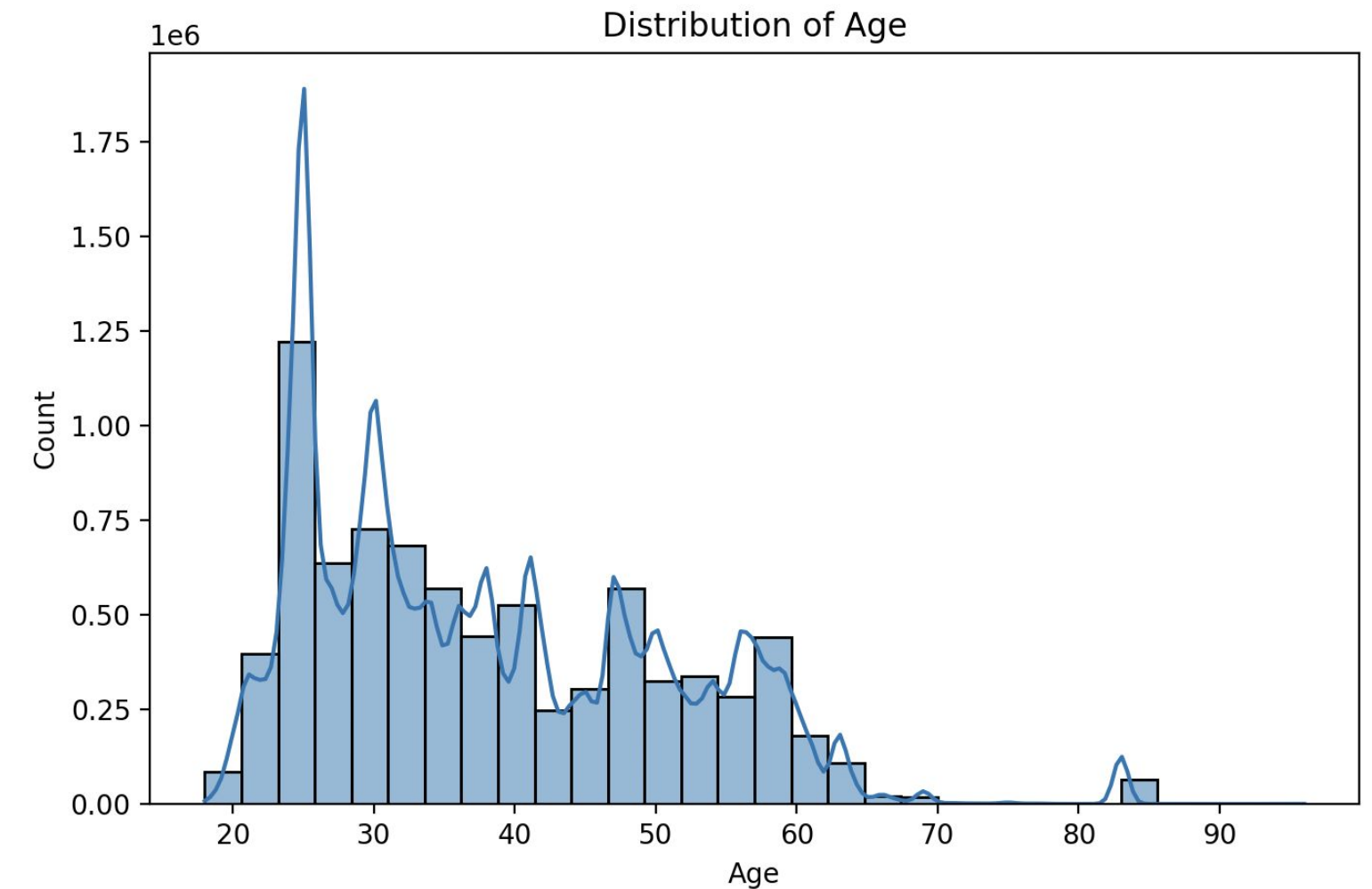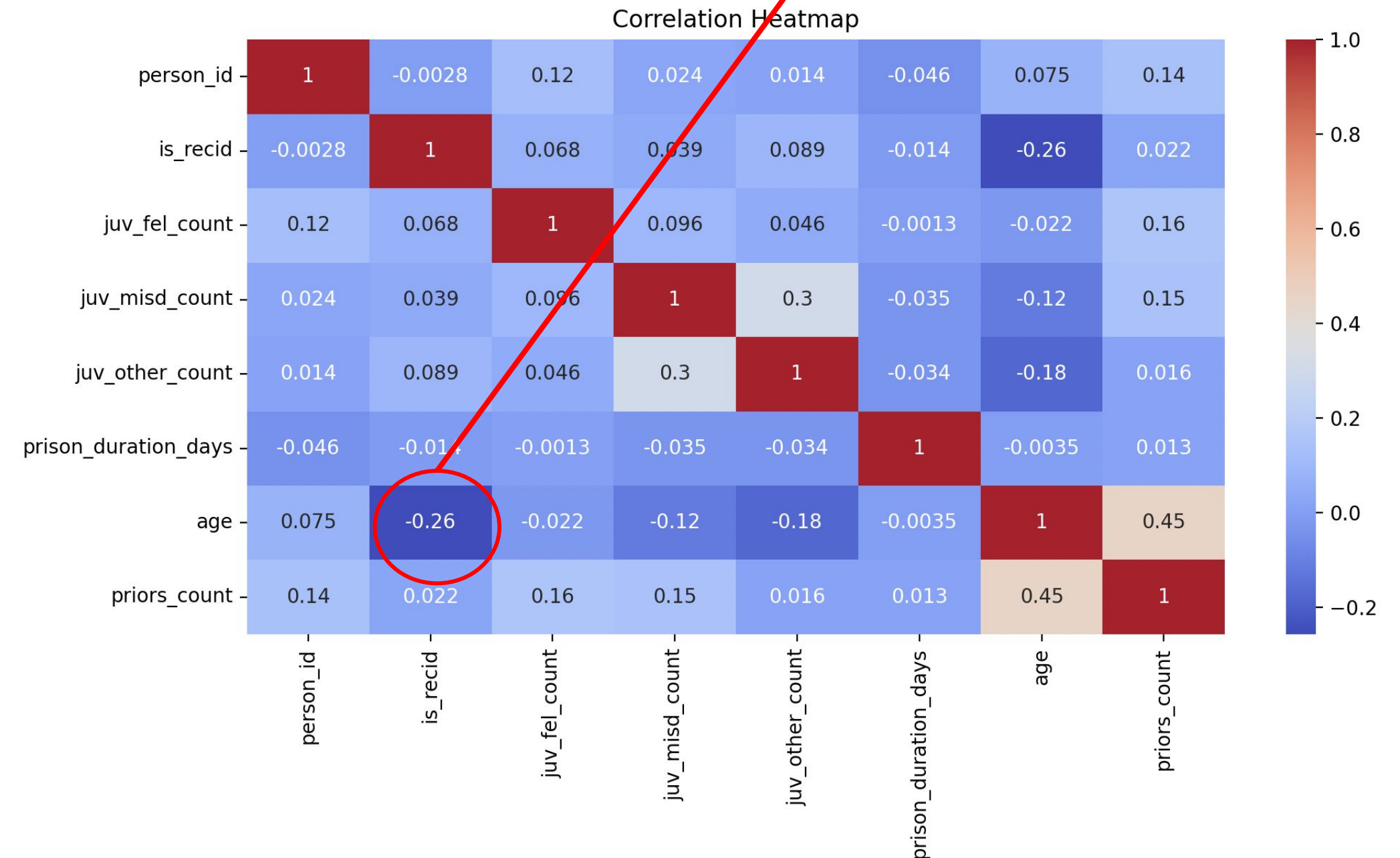
The original dataset also skews left on age distribution, where a lot of younger people are included than older people.

EDA

Young age is correlated with recidivism

Age distribution by race is very different between different race groups. Hispanic group are mostly younger, while native americans are much older.
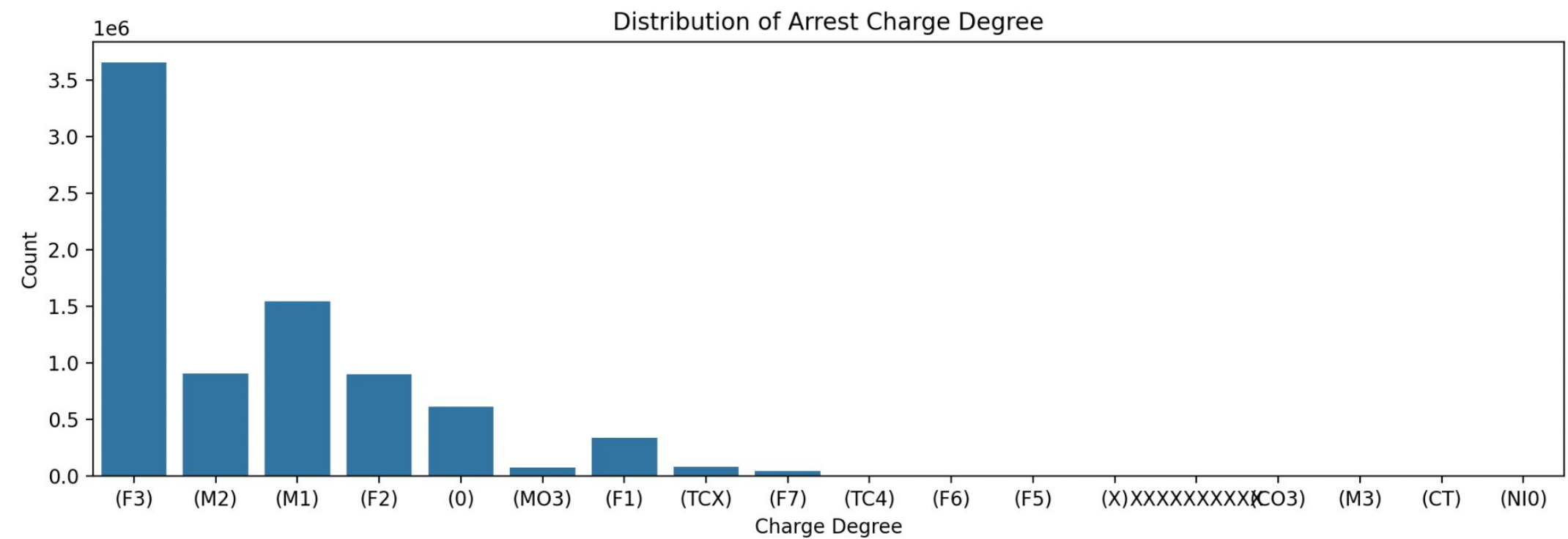
To rule out multicollinearity, the heatmap shows relatively high negative correlation between age and our outcome variable.

12

EDA
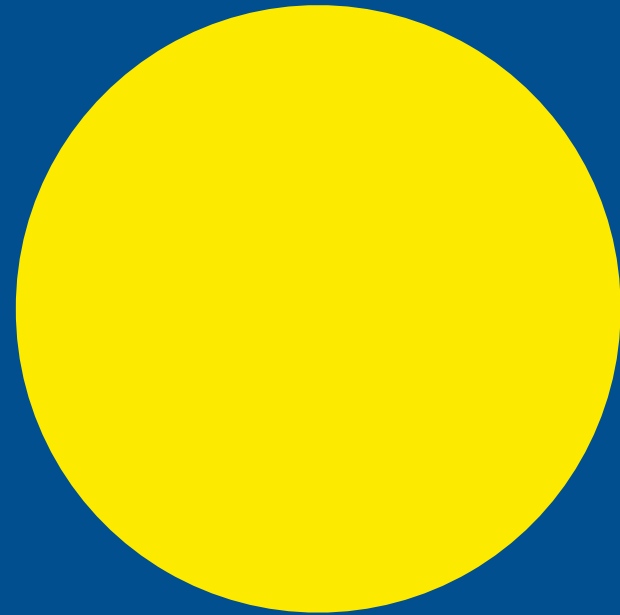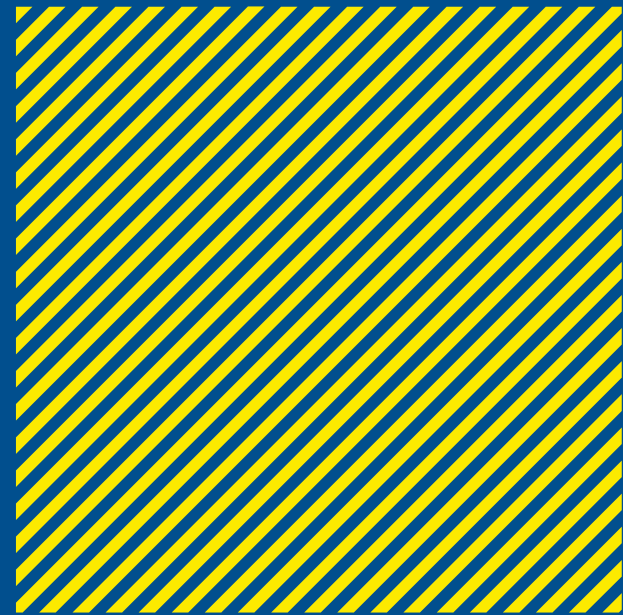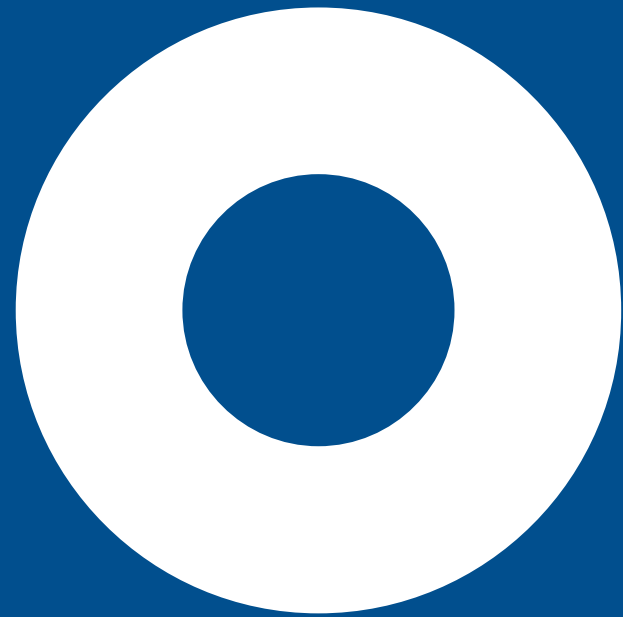

Distribution of Sex

Distribution of Arrest Charge Degree
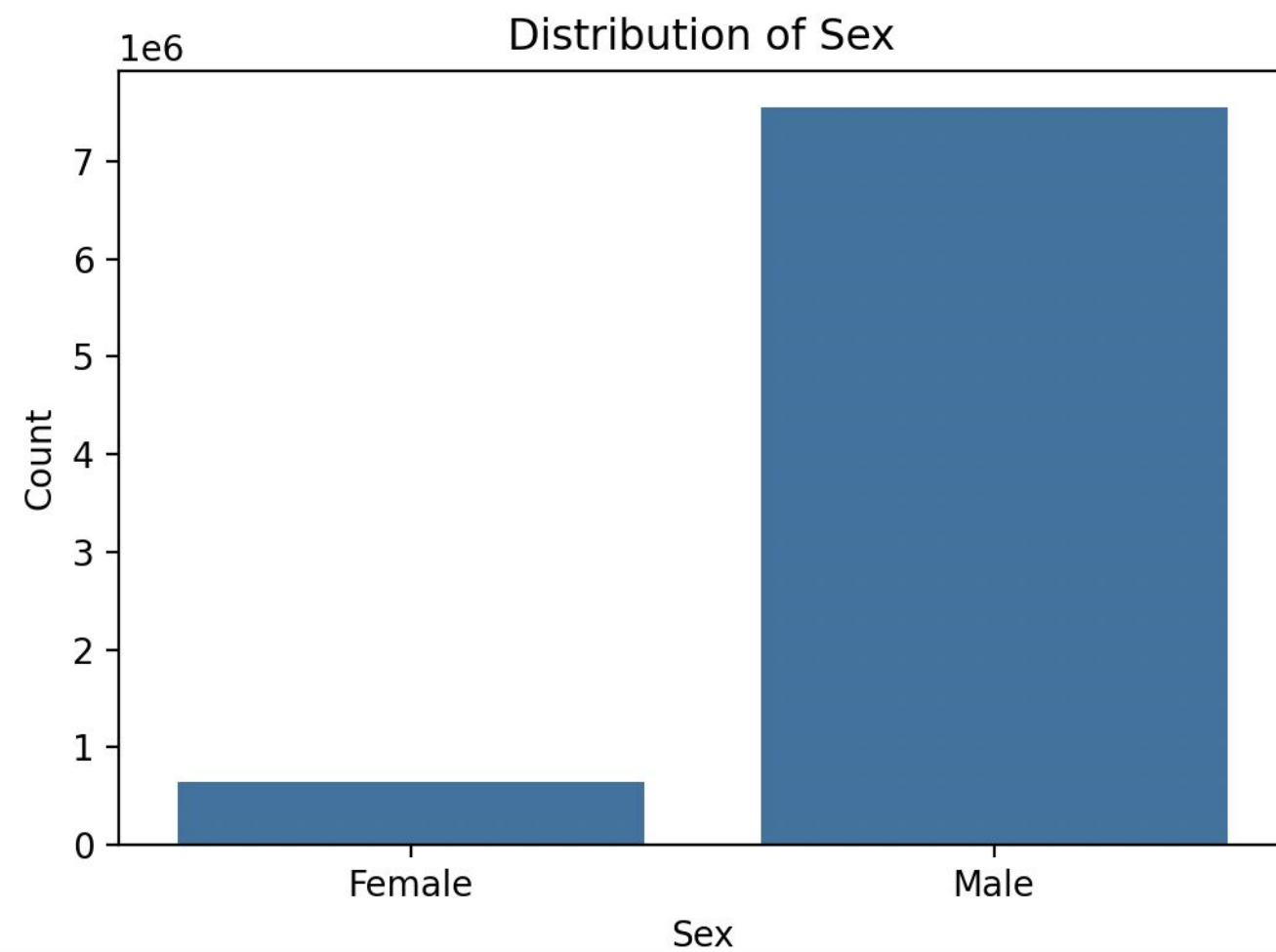
The original dataset includes a lot more males than females.

The original dataset includes a lot of class three felony charges than other types (such as different classes of misdemeanor charges).

Approaches

LOGISTIC REGRESSION

- Simple, interpretable baseline model
- Helps identify key predictive features
- Transparent and audit-ready
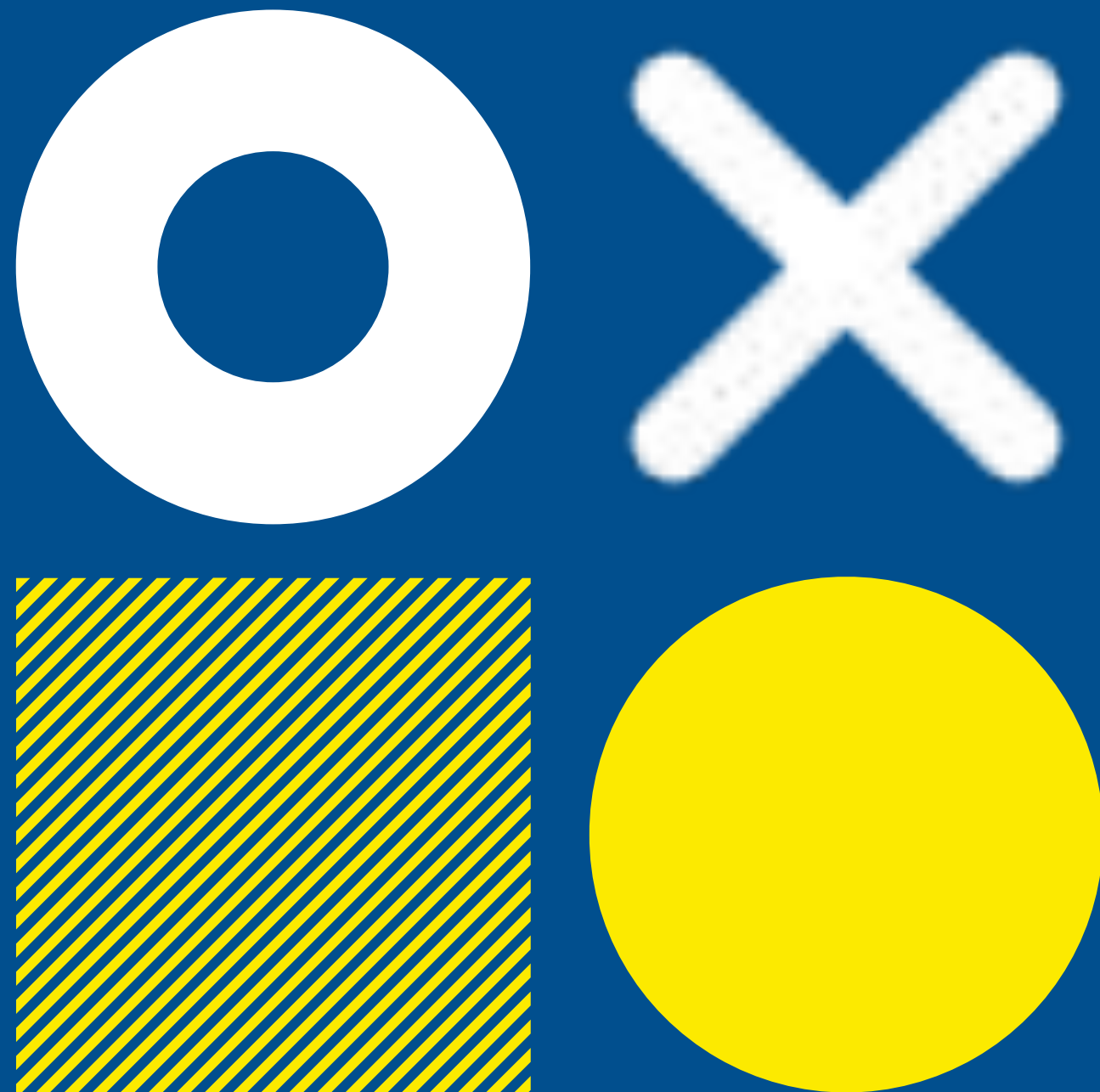
GRADIENT BOOSTING

- High performance on tabular data
- Captures complex feature interactions
- Provides feature importance

MULTI-LAYER PERCEPTRON

- Learns non-linear relationships
- Leverages large dataset size
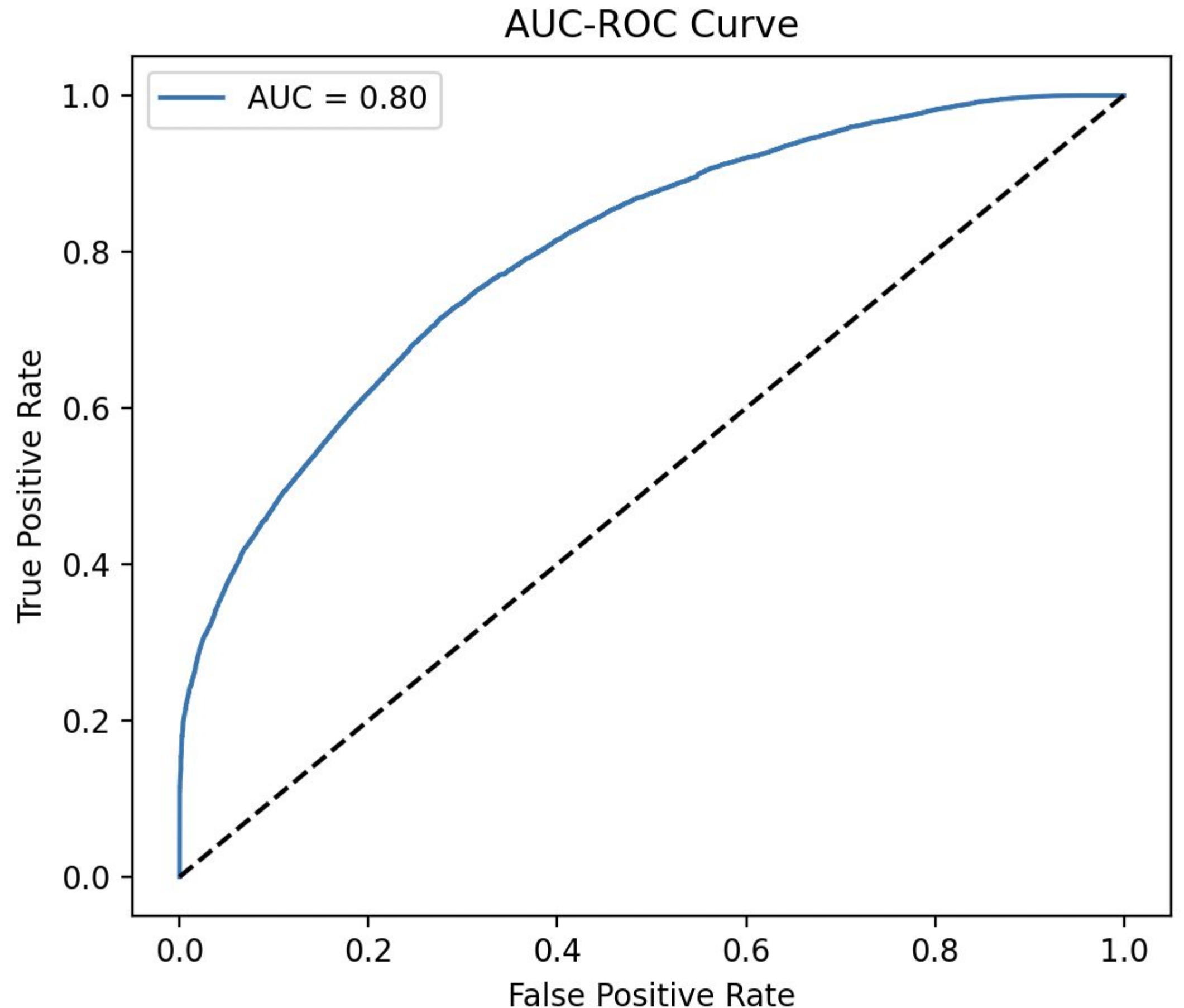- Explores deeper predictive patterns

Modelling Evaluation

How did we apply our models to our data?

Logistic Regression

- Accuracy: 79.0%

- Precision: 0.68 (non-recidivist), 0.80 (recidivist)

- Recall: 0.31 (non-recidivist), 0.95 (recidivist)

- F1-score: 0.43 (non-recividist), 0.87 (recidivist)

- AUC-ROC: 0.80 → Acceptable class separation.

AUC-ROC Curve

Logistic Regression

- Confusion matrix (on the right): more false positive than true negative - bad at predicting no recidivism.

- Good amount of true positives to false negatives - good at predicting true recidivism.

Confusion Matrix
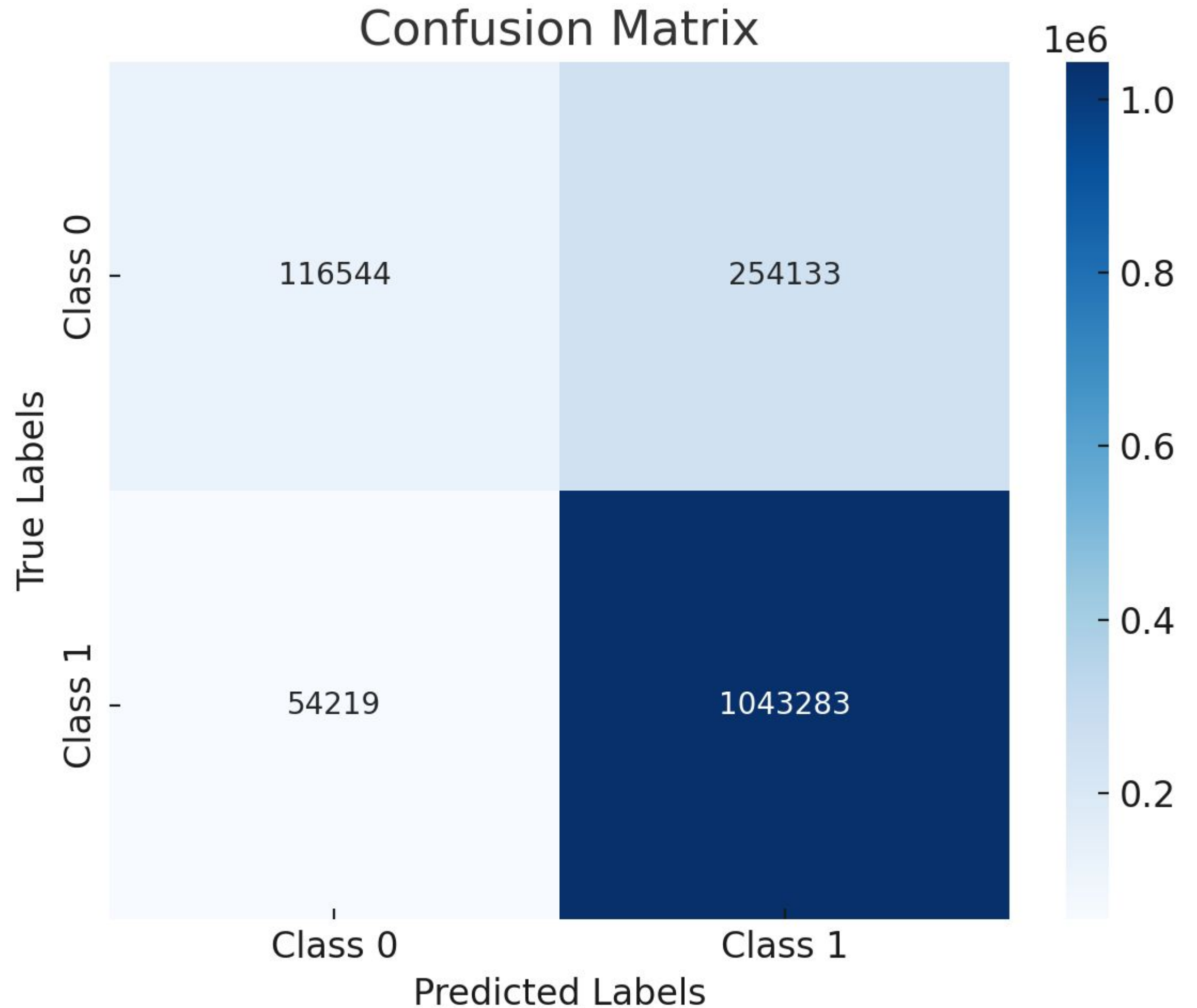
Logistic Regression

- Feature rank shows that this model is still racially biased
- Amongst these top 15 features, the following features have positive coefficients (inversely contribute to recidivism prediction)
  - age
  - initial_statute_316.19
  - initial_charge_Driving Under Influence
  - filing_agency_Broward Sheriff Office
  - race_Hispanic

Top 15 Feature Importances (|Coefficient|)

Gradient Boosting
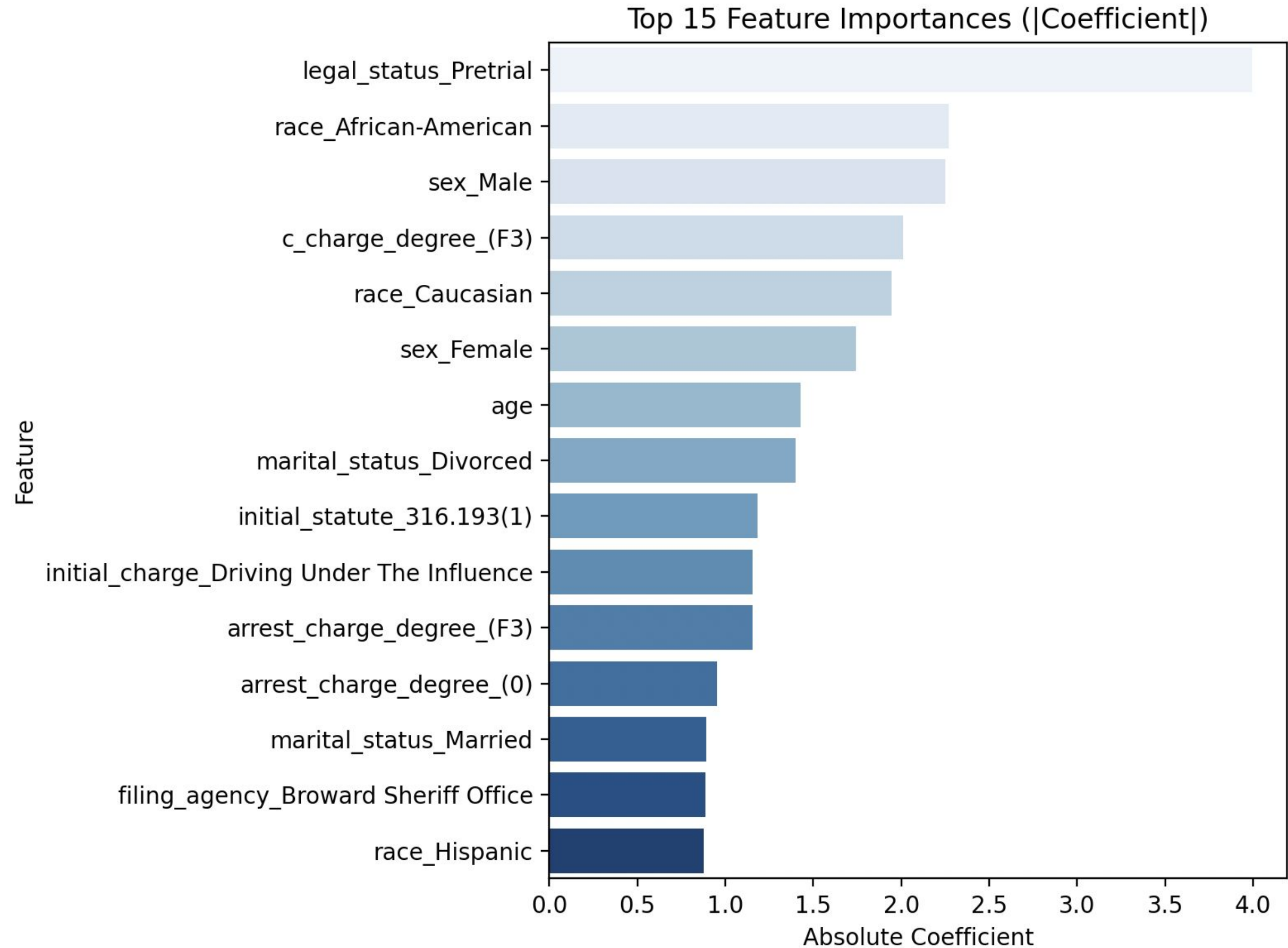
- Accuracy: 89.4%

- Precision: 0.92 (class 0), 0.89 (class 1)

- Recall: 0.64 (non-recidivist), 0.98 (recidivist)

- F1-score: 0.75 (class 0), 0.93 (class 1)

- AUC-ROC: 0.95 → Excellent class separation

ROC Curve
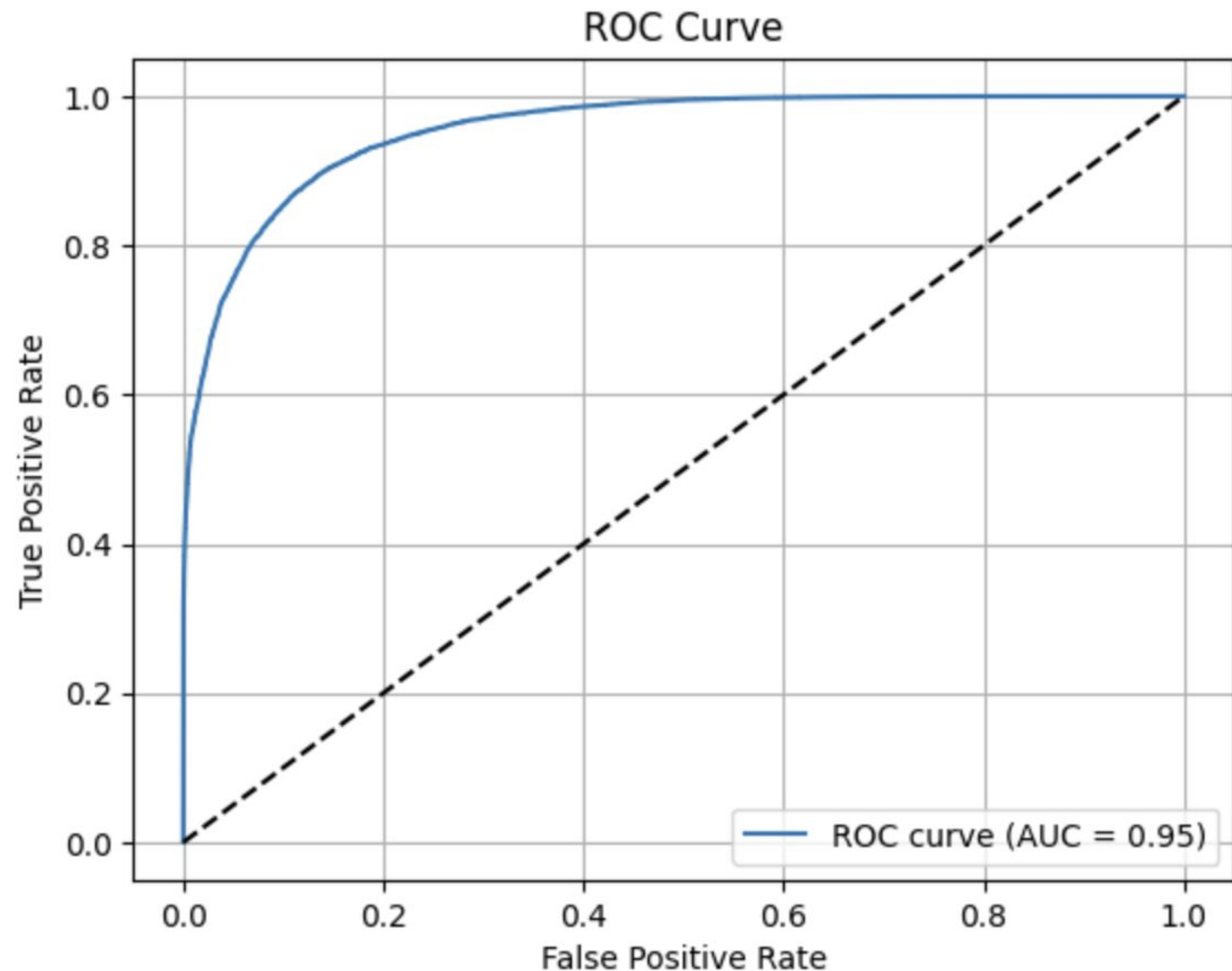
Gradient Boosting

Top drivers of recidivism predictions:

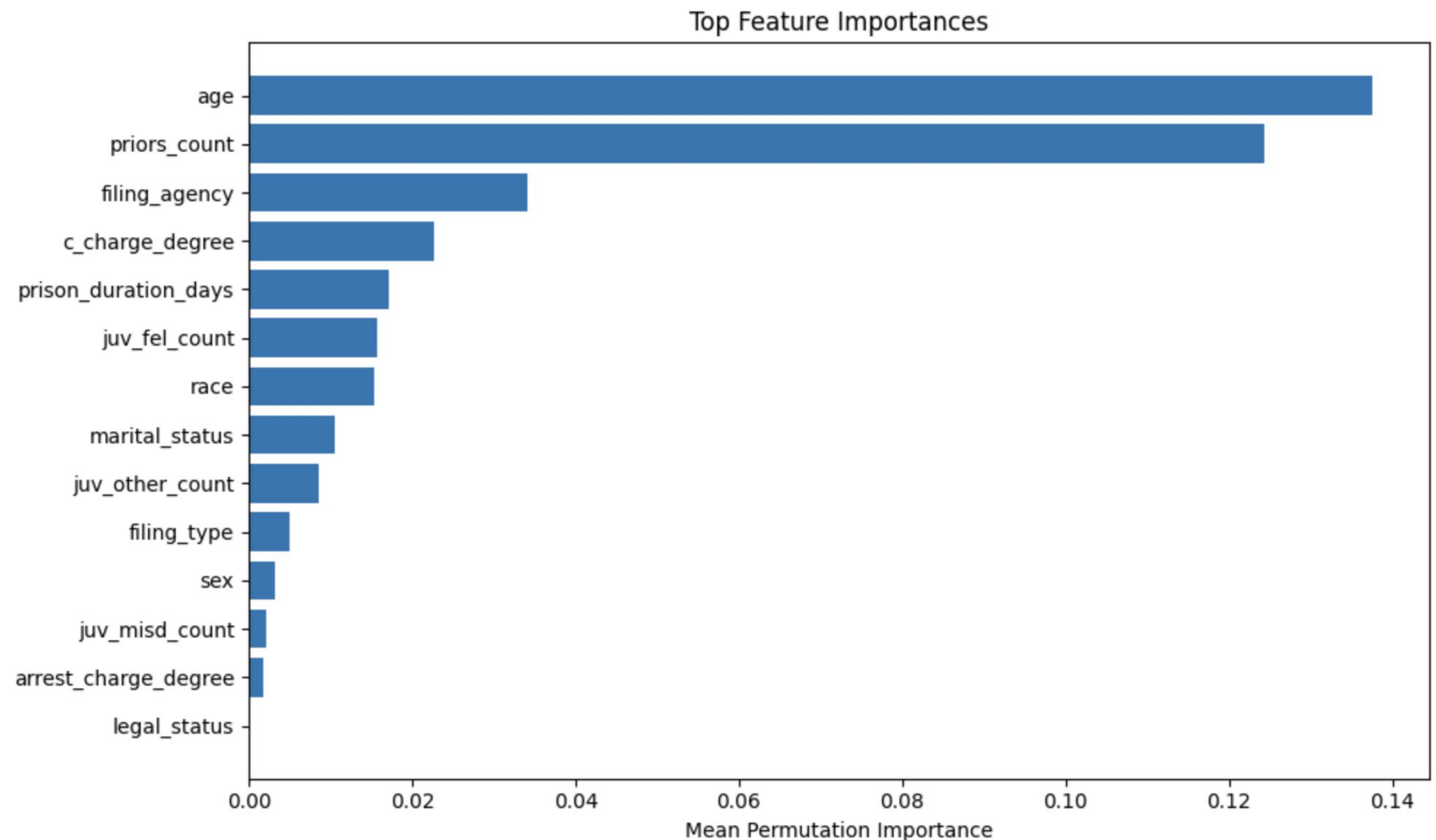- Age and prior convictions are the strongest predictors.
- Filing agency and charge degree also carry strong influence.

Sensitive features like race and sex appear lower in importance — but still require ethical consideration.

Top Feature Importances

Multi-layer Perceptron

- **Consistent decline:** training loss decreases across epochs → effective learning and convergence of model
- **Diminishing returns:** most significant loss reduction occurs during first few epochs → early convergence
- **Low final loss:** final training loss falls below 0.09 → good fit on training dataset

Training Loss Over Epochs
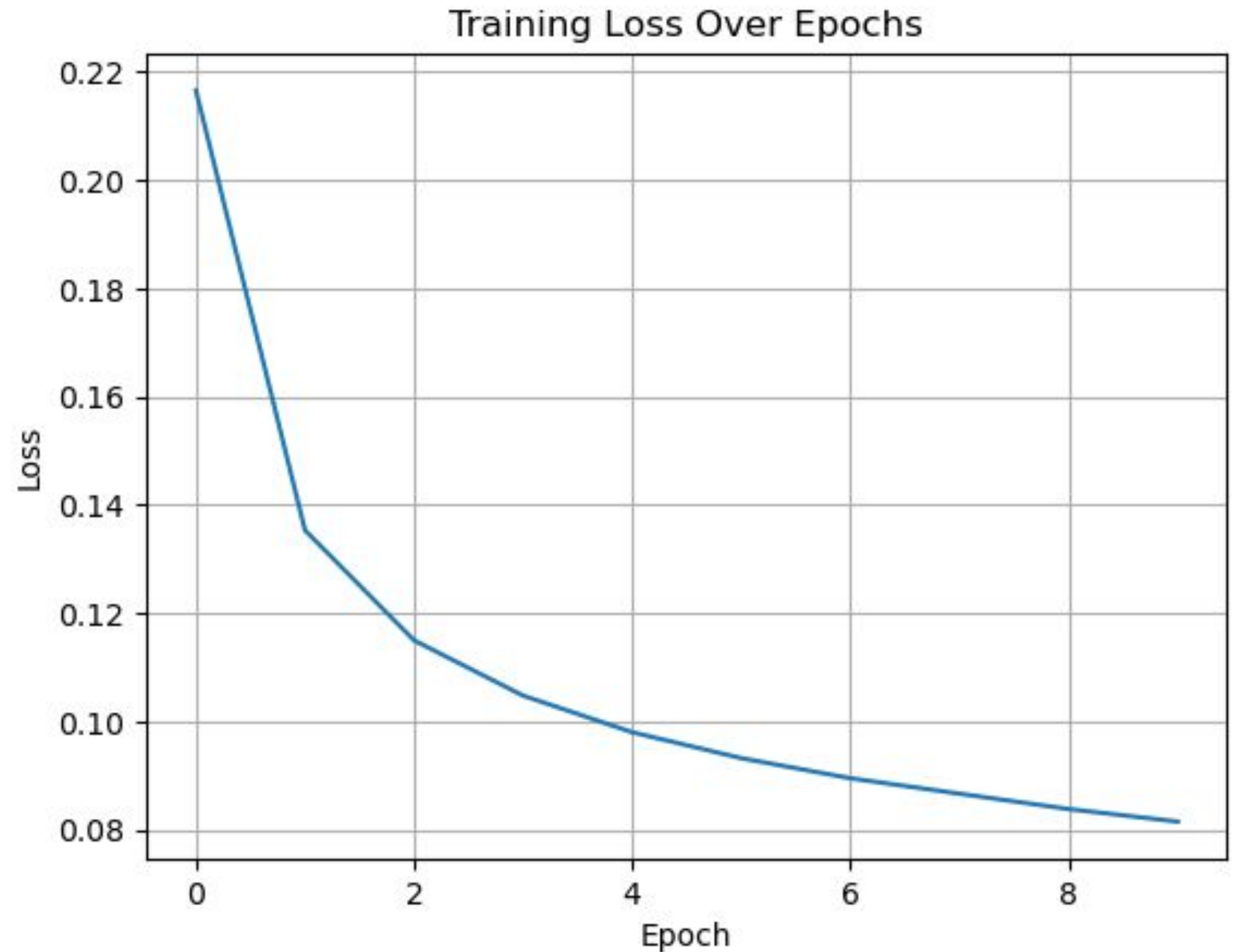
Multi-layer Perceptron

- **High true positives:** Excellent recall for the positive class
- **Moderate false positives:** Acceptable percentage of misclassified positives
- **Balanced performance:** Robust classification across both classes

Confusion Matrix

	Predicted: No	Predicted: Yes
Actual: No	344850	26211
Actual: Yes	15709	1081412

Actual
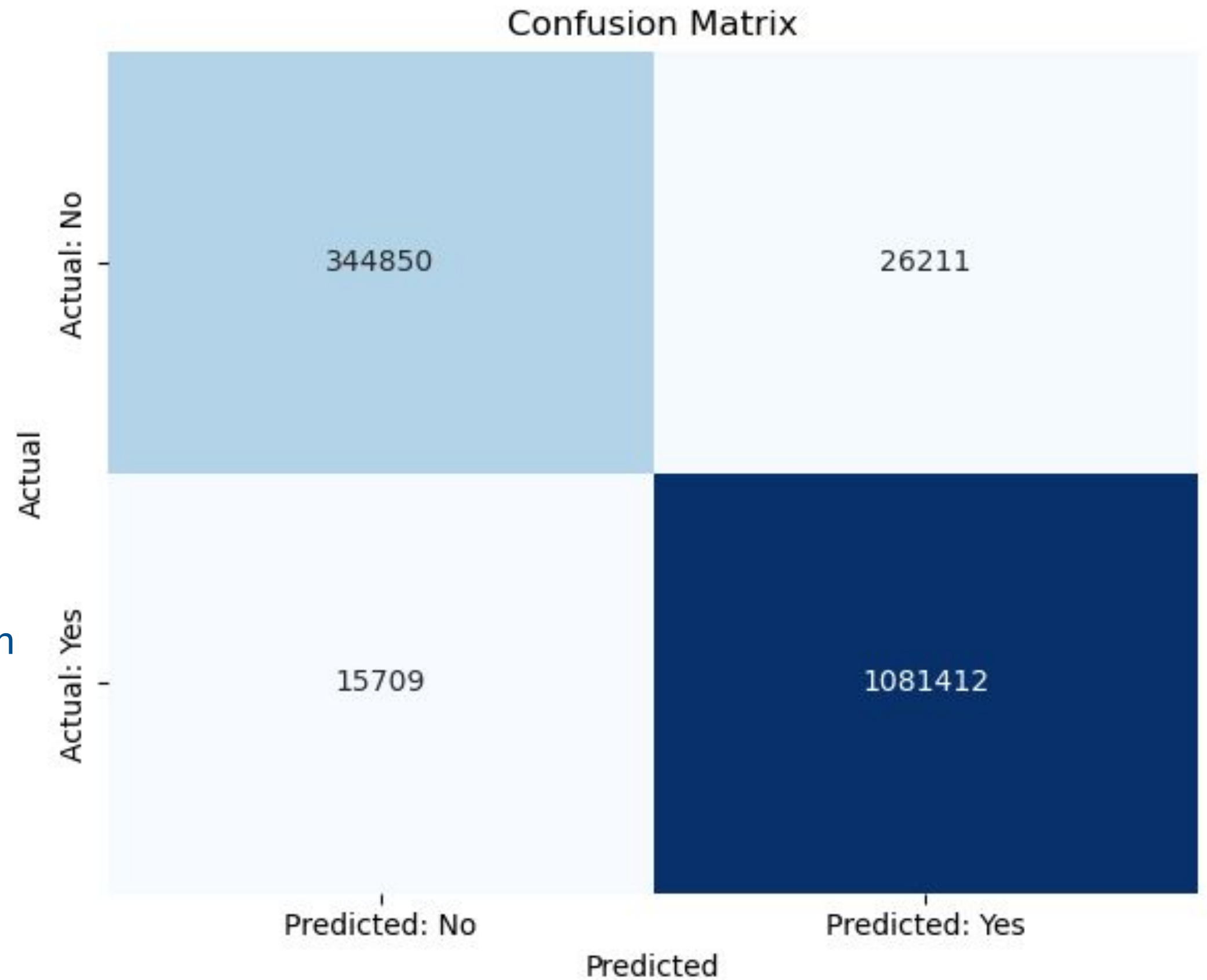
Predicted

Multi-layer Perceptron

- **priors_count** and **age** are the most influential features, causing the largest drops in model accuracy when permuted
- **Juvenile history** also shows notable predictive value → relevance of early offences
- Sociodemographic variables (**race, marital status**) contribute to model accuracy but with smaller impact

Top 15 Most Important Features (Permutation Importance)
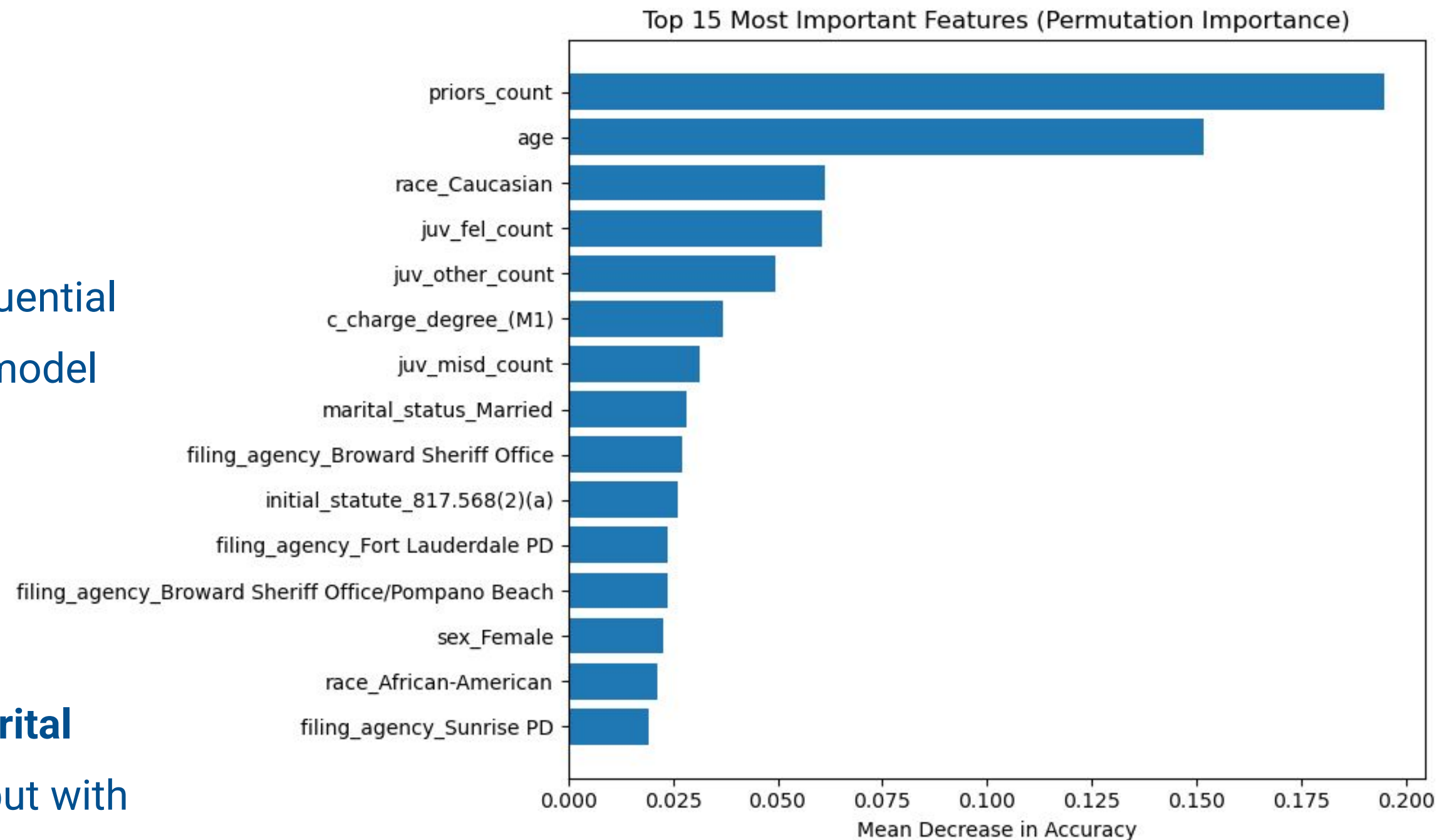
	LOGISTIC REGRESSION	GRADIENT BOOSTING	MULTI-LAYER PERCEPTRON
TEST ACCURACY	0.79	0.89	0.97
TEST PRECISION	0.77	0.90	0.93
TEST RECALL	0.79	0.89	0.99
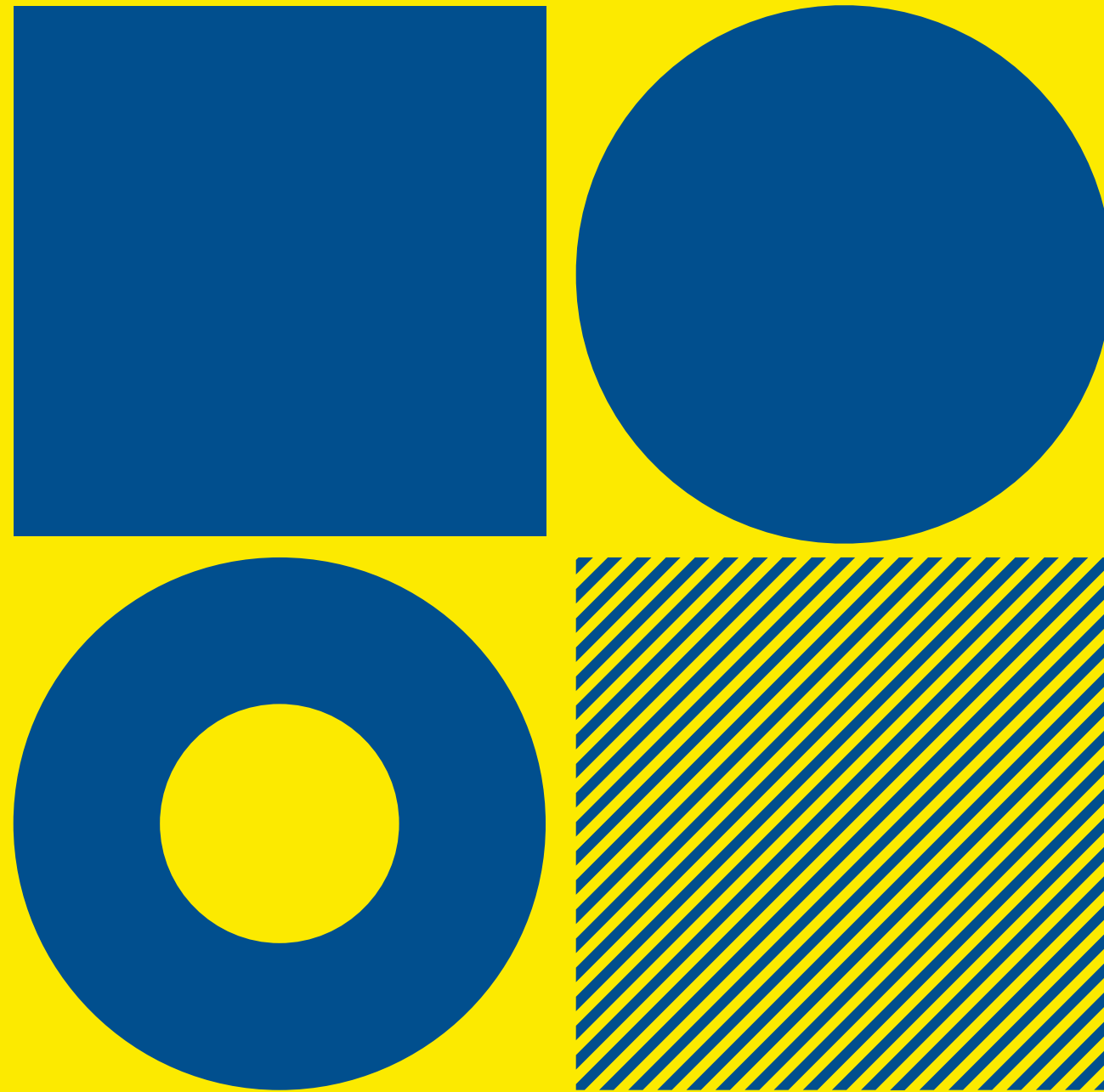
Conclusion

The multi-layer perceptron model is
the best predictor of recidivism

Pros and Cons

The multi-layer perceptron model was the strongest across all metrics:

- Model accuracy exceeds COMPAS (61%)
- Strong and balanced precision and recall
- Not overly reliant on sociodemographic features
- Only 7.06% of actual non-reoffenders are classified as reoffenders

1 **Model Accuracy**
97% √

**2** **Precision & Recall**
0.93 & 0.99 √

3 **Features**
Demographic predictors are less influential

27

Race Remains a Primary Feature

- **race_Caucasian** is the 3rd most influential feature
- **race_African_American** is the 14th most influential feature
- The model isn't specifically reliant on sociodemographic features over other features such as **priors_count** and **juv_fel_count**, but they are still high-ranking enough to be impactful

However:
- Low misclassification rate
- Higher-ranking non-sociodemographic features

Top 15 Most Important Features (Permutation Importance)
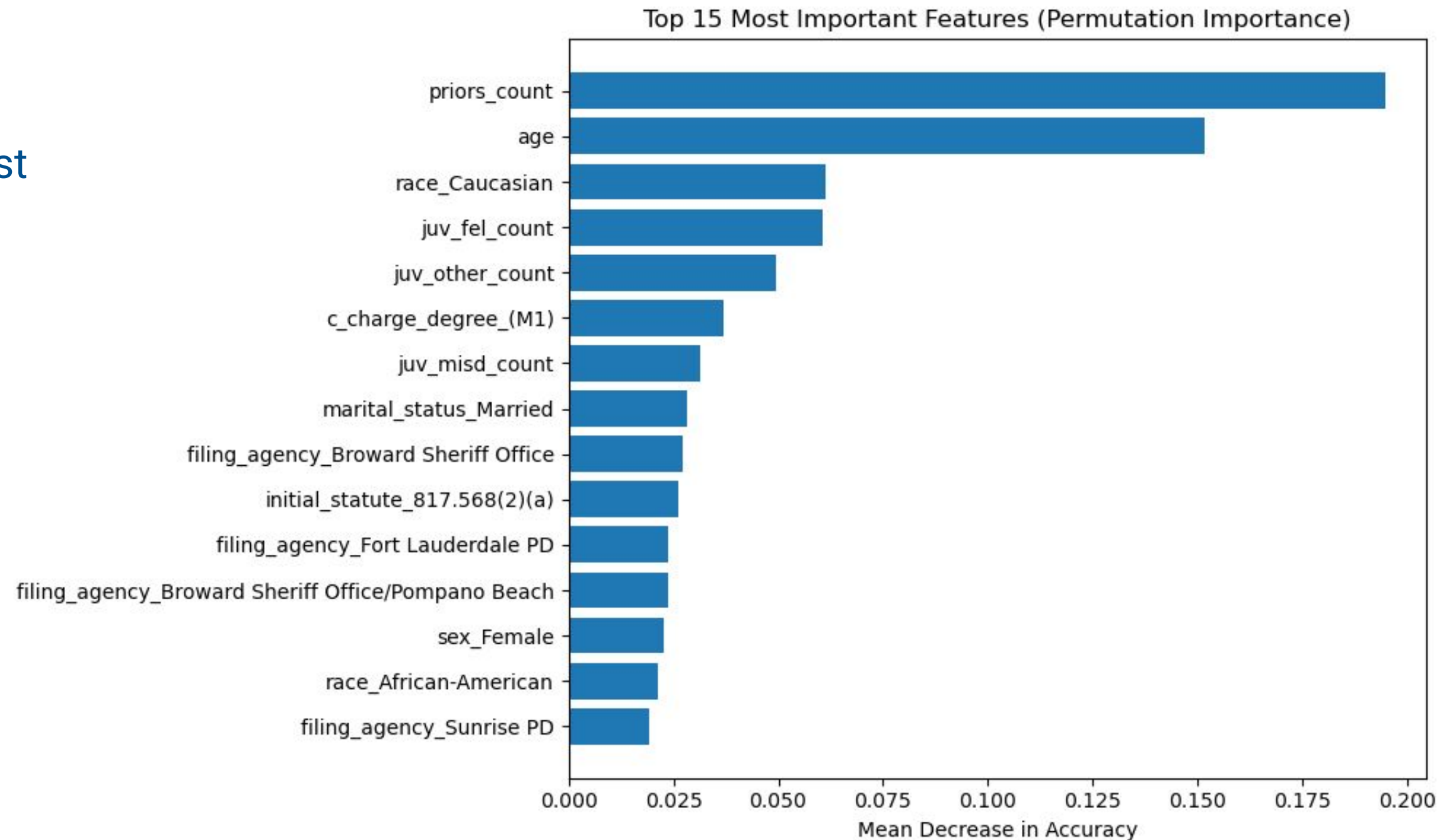
Our best model is less discriminatory than existing justice system tool COMPAS

The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) approach is a risk assessment tool used in the U.S. criminal justice system to help predict a defendant's likelihood of reoffending.
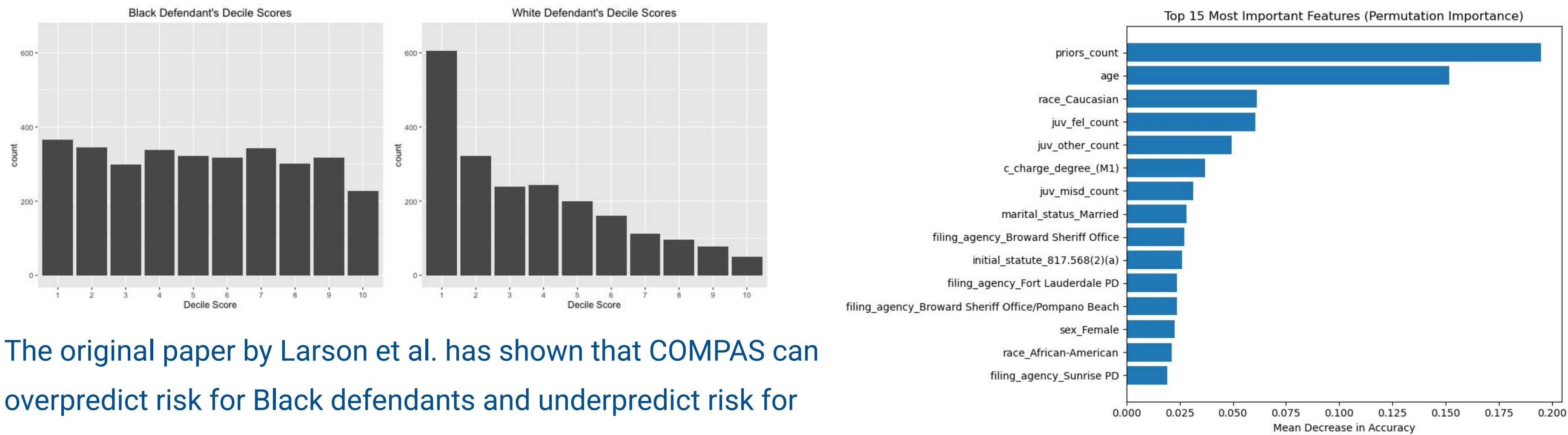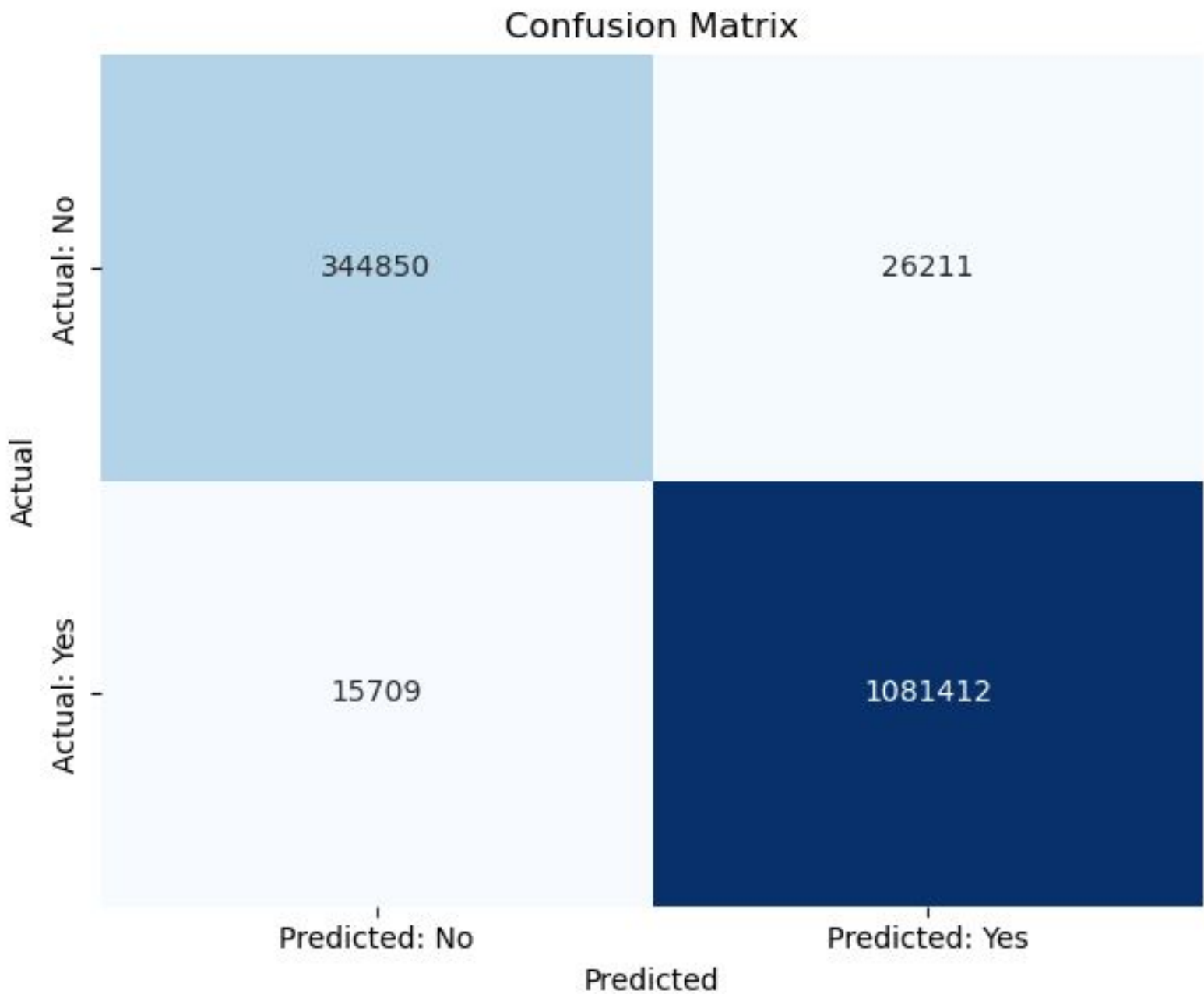

The original paper by Larson et al. has shown that COMPAS can overpredict risk for Black defendants and underpredict risk for white defendants. Our MLP result uses racial predictor differently and assigns higher risk to Caucasian offendants.
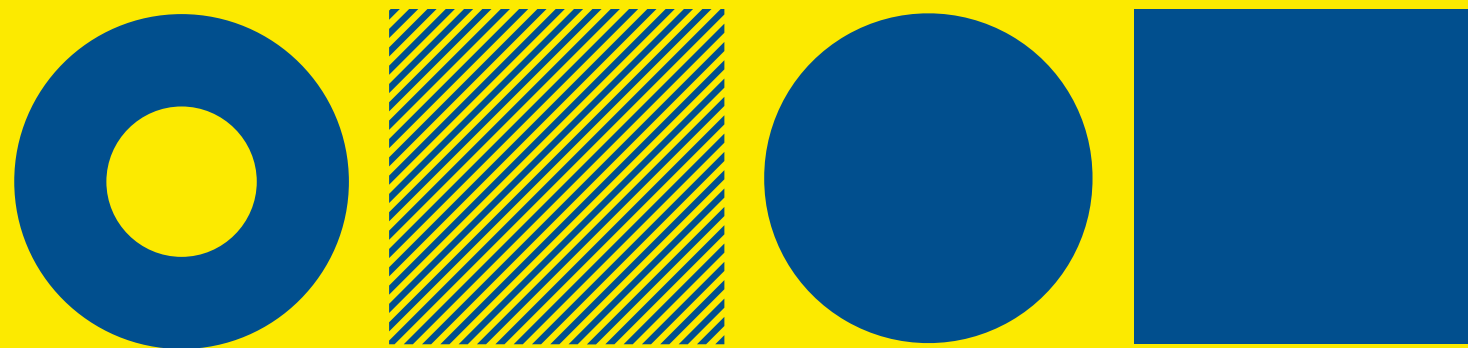
Our best model gives less misclassification results than existing justice system tool COMPAS

The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) approach is a risk assessment tool used in the U.S. criminal justice system to help predict a defendant's likelihood of reoffending.

All Defendants

	Low	High
Survived	4121	1597
Recidivated	347	389

FP rate: 27.93
FN rate: 47.15
PPV: 0.20
NPV: 0.92
LR+: 1.89
LR-: 0.65

The COMPAS tool has false positive (FP) rate of 0.28 and false negative (FN) rate of 0.47. Our best model lowered FP rate to 0.07 and FN rate to 0.01.

Confusion Matrix

	Predicted: No	Predicted: Yes
Actual: No	344850	26211
Actual: Yes	15709	1081412

Thank You!

—

Full code:

https://github.com/ursulaGUO/compas-analysis

Sources

Blomberg, T. (2010, September). *VALIDATION OF THE COMPAS RISK ASSESSMENT CLASSIFICATION INSTRUMENT*. College of Criminology & Criminal Justice. https://criminology.fsu.edu/sites/g/files/upcbnu3076/files/2021-03/Validation-of-the-COMPAS-Risk-Assessment-Classification-Instrument.pdf

Innocence Project New Orleans. (n.d.). https://ip-no.org/what-we-do/advocate-for-change/mass-incarceration-and-racial-oppression/?gad_source=1&gad_campaignid=11715335968&gbraid=0AAAAADOiIuigtqWW7nN_k-dEOaTtJWWCJ&gclid=CjwKCAjw6NrBBhB6EiwAvnT_rqE4UuU3h9eGqRfKYJ3_R7jRcU1kkm0uMj7L5ugvamRBfGfNr-nCuBoCBy8QAvD_BwE

Jeff Larson,Julia Angwin,Lauren Kirchner,Surya Mattu. (2016, May 23). *How we analyzed the COMPAS recidivism algorithm*. ProPublica. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Partners for justice. (2025, May 25). Partners for Justice. https://www.partnersforjustice.org/?gad_source=1&gad_campaignid=16172307588&gbraid=0AAAAAoL1I_JB6u-A-qalrgeDw-W906tFD&gclid=CjwKCAjw6NrBBhB6EiwAvnT_rrwRKrUQ0w4kFEq1SMchGxabq__s20LhvstMBD-SwszP-BVLtwMIwhoC1bUQAvD_BwE

Prison Policy Initiative. (n.d.). *Racial and ethnic disparities*. https://www.prisonpolicy.org/research/racial_and_ethnic_disparities/

Propublica/compas-analysis: Data and analysis for 'Machine bias'. (n.d.). GitHub. https://github.com/propublica/compas-analysis

Racial disparities persist in many U.S. jails. (2023, May 16). The Pew Charitable Trusts | The Pew Charitable Trusts. https://www.pew.org/en/research-and-analysis/issue-briefs/2023/05/racial-disparities-persist-in-many-us-jails