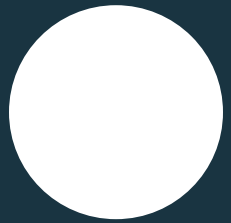
Georgia State Recidivism: Binary Classification Forecasting

Project 5

Project 5:
Frank Novak, Preet Sekhon, Debbie Sim
DSI1010

RECIDIVISM

Recidivism is defined as a criminal act that results in rearrest, reconviction or return to prison with or without a new sentence during a three-year supervision period following the person's release.



PROBLEM STATEMENT

The National Institute of Justice and the US Office of Justice Programs are looking to address recidivism rates by predicting persons most at risk. The focus of this project is to identify key variables that increase the probability a person will recidivate. Using a binary classification, can predictive attributes be identified and used to create solutions that can reduce prison reentry rates?

Background Information

- Data compiled from Federal Government and State of Georgia
- January 1, 2013 - December 31, 2015
- 25,000 persons
- Features included, but not limited to:
    - demographics
    - prior criminal history
    - conditions of current supervision
    - actual recidivism data
- External data from the US census was also used to relate persons to general geographic areas in Georgia

Data Cleaning & Pre-Processing

Nulls

- Applied simple imputer (most frequent & mean) & knn imputer on nulls
- dropped nulls if it was a small % of the entire dataset

Data Type Changes

- converted boolean data types into binary values
- converted certain categorical variables into binary values
- encoded ordinal features into numeric

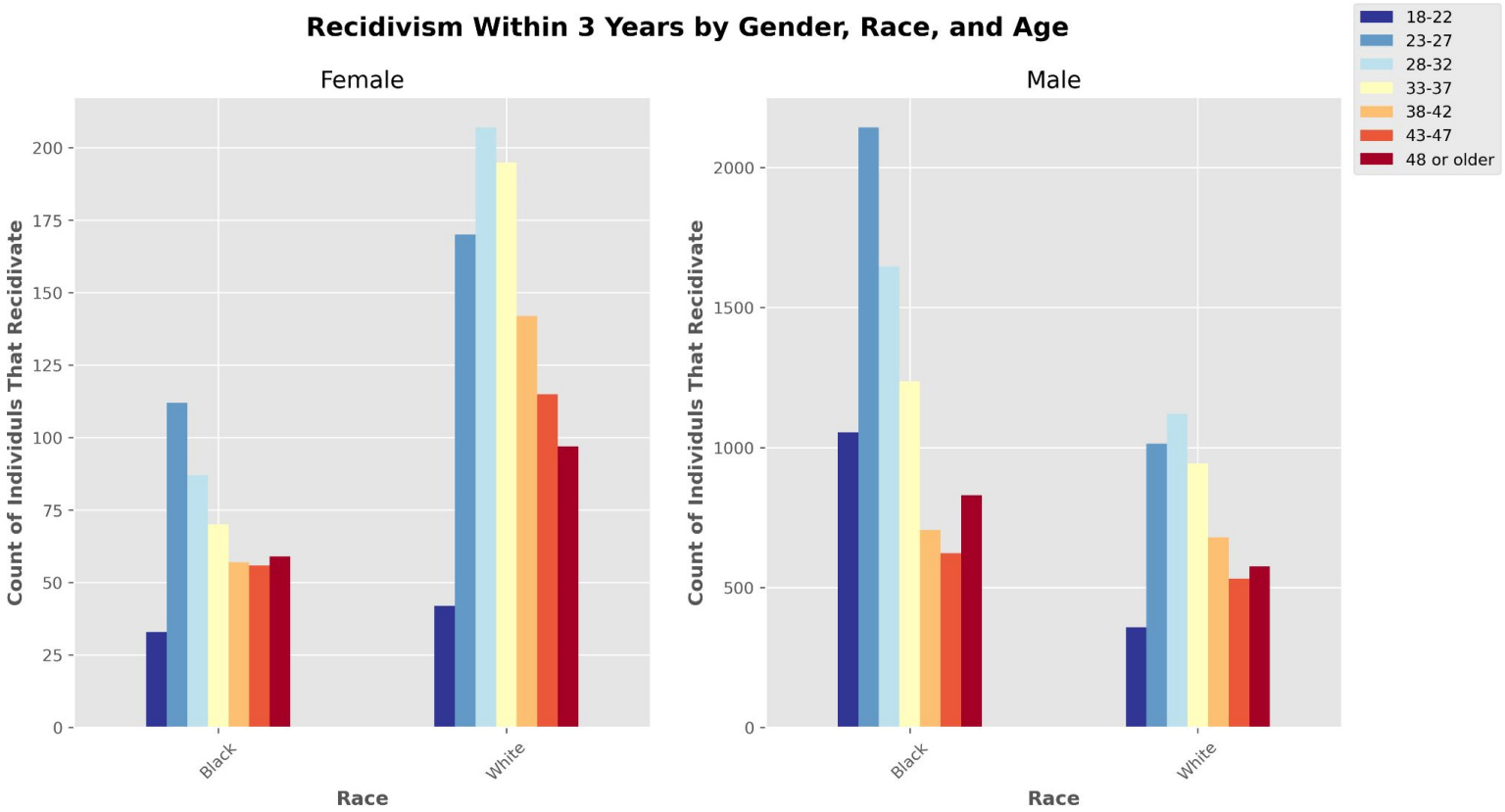
Pre-Processing

- OHE (OneHotEncoder) on categorical variables
- StandardScaler

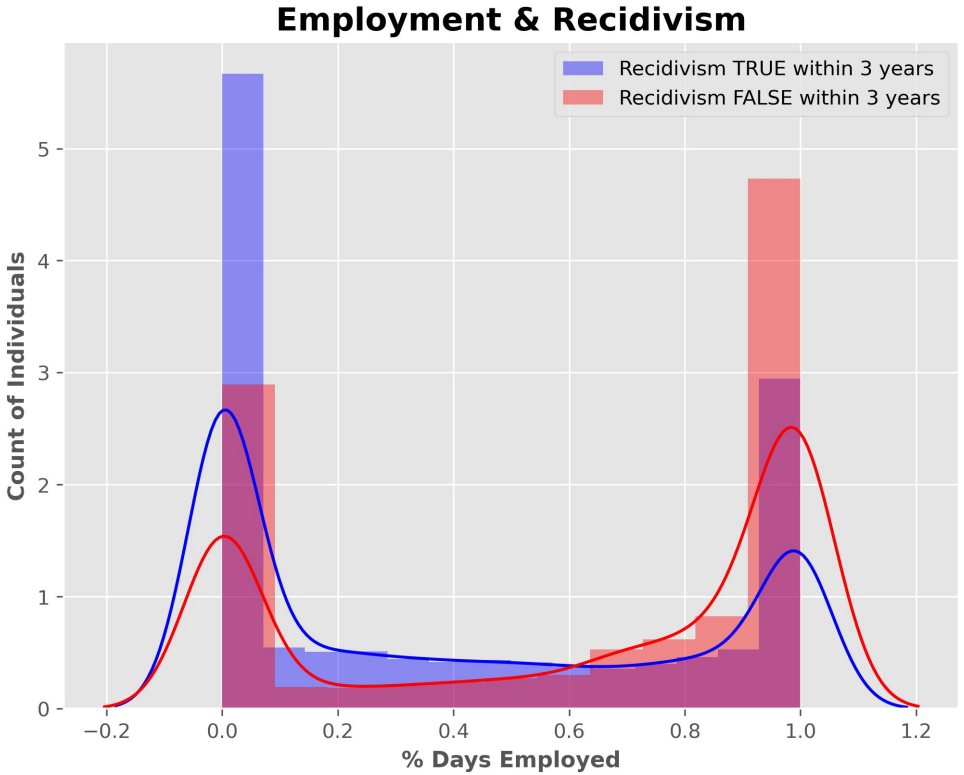
Feature Engineering

- PolynomialFeatures
- Changed ‘Dependents’ to had dependents or not

Recidivism is higher amongst individuals between the ages of 23-32 across race and gender



Individuals employed for at least ~61% of the time are less likely to recidivate within a 3 year period

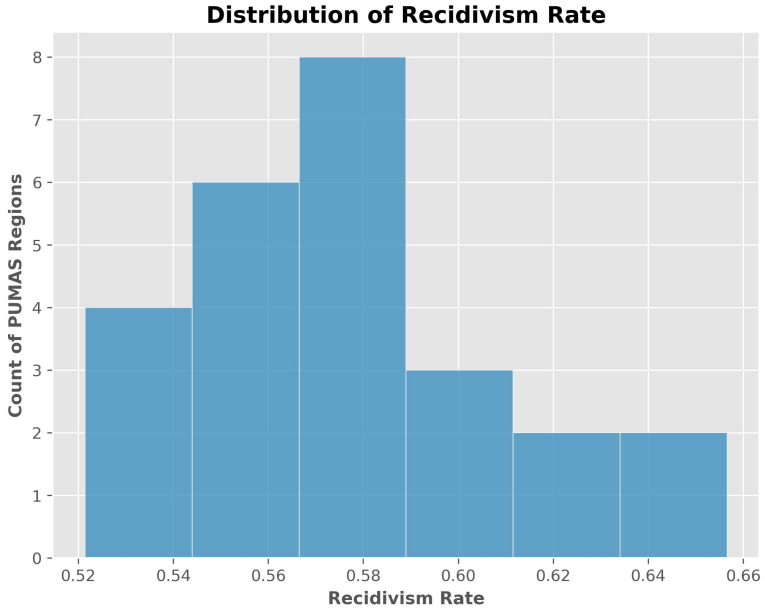


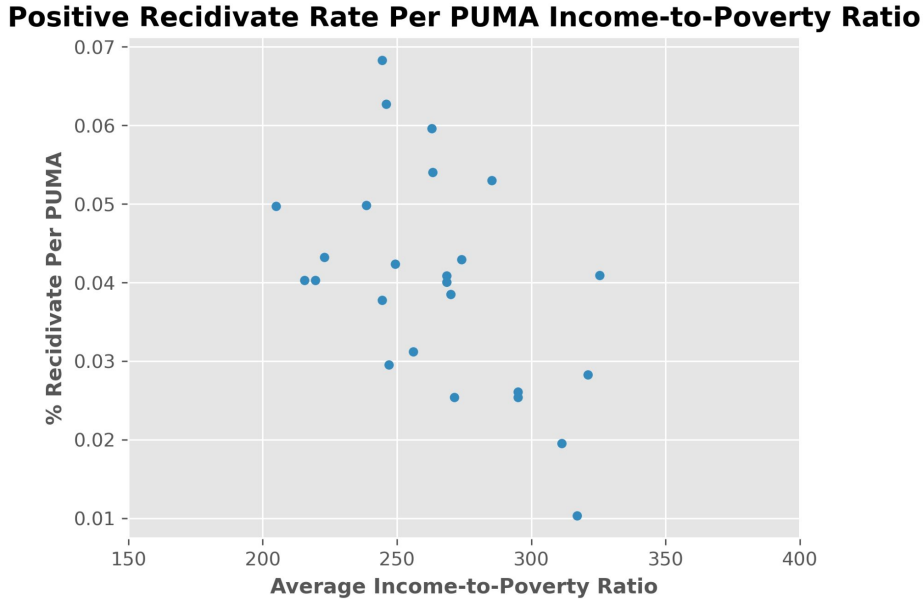
PUMAs
Public Use Microdata Areas

Public Use Microdata Areas (PUMAs) are non-overlapping, statistical geographic areas used by the US Census that partition each state or equivalent entity into geographic areas containing no fewer than 100,000 people each.

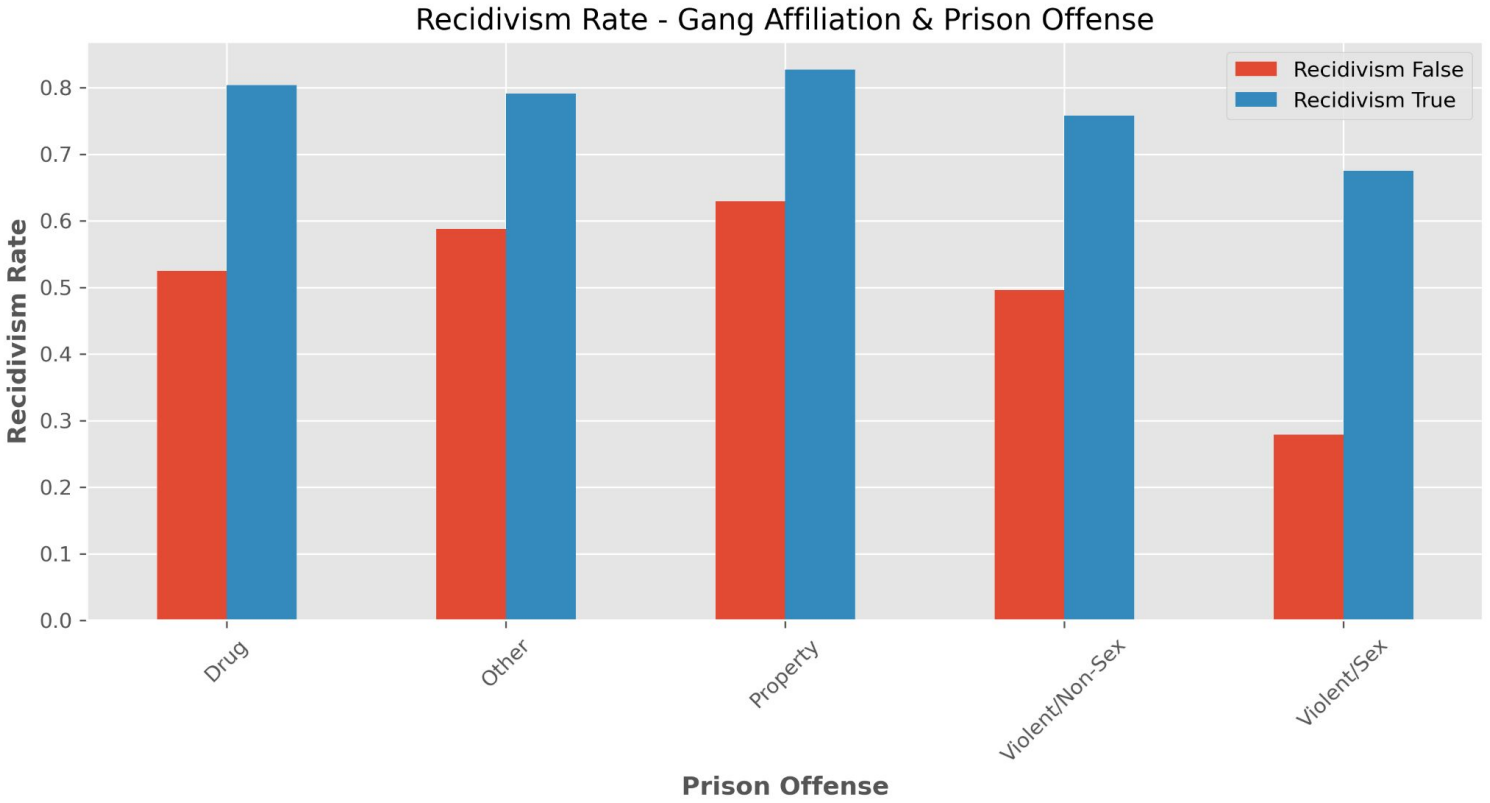
For this dataset, PUMA regions were collapsed to reduce risk of identification. 72 total regions were collapsed into 24.

Average recidivism rate across all regions: ~58%
Lower income-to-poverty ratio shows higher recidivism





Gang Affiliated individuals are more likely to recidivate.



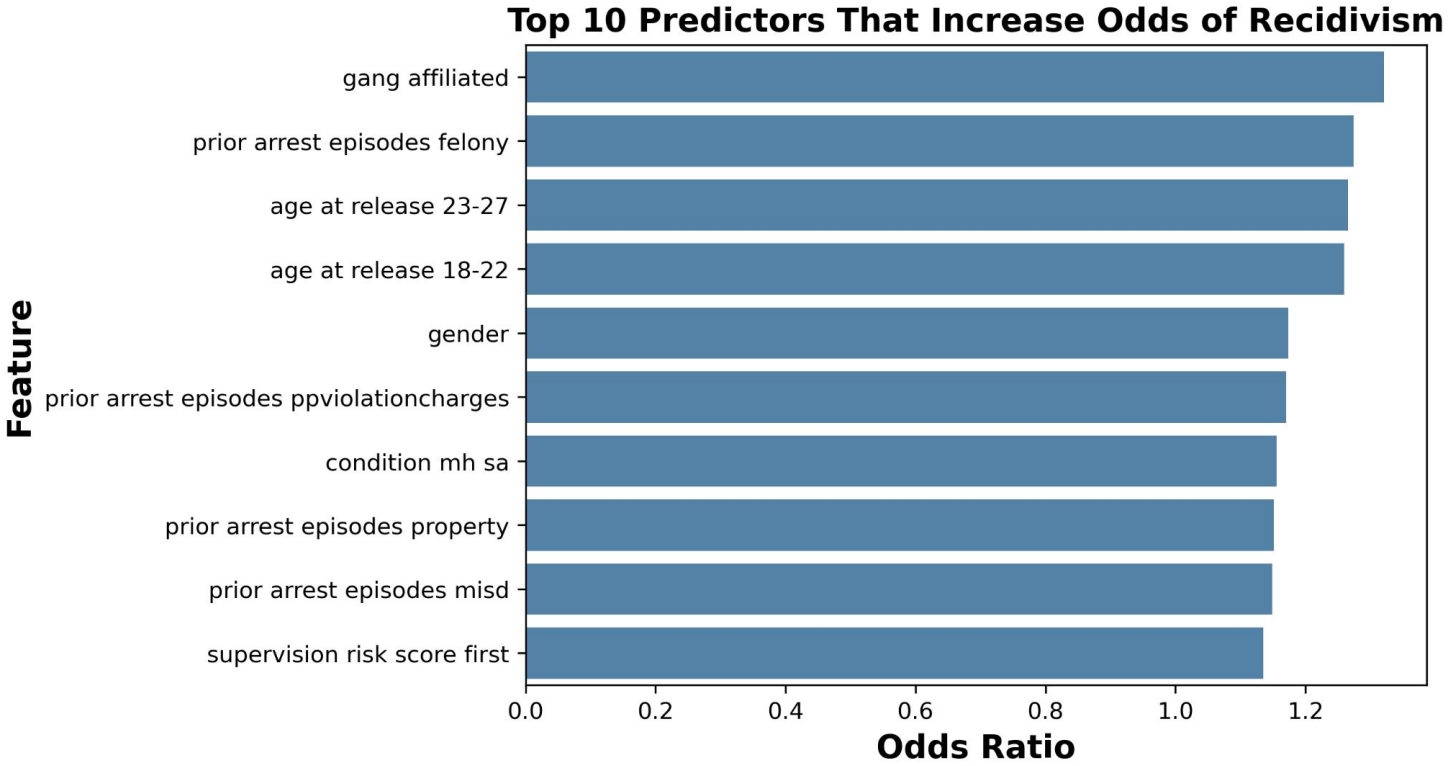
Metrics and Model Comparison

Brier Score is an accuracy benchmark of a model's predicted probability.
A lower brier score is a more accurate model.

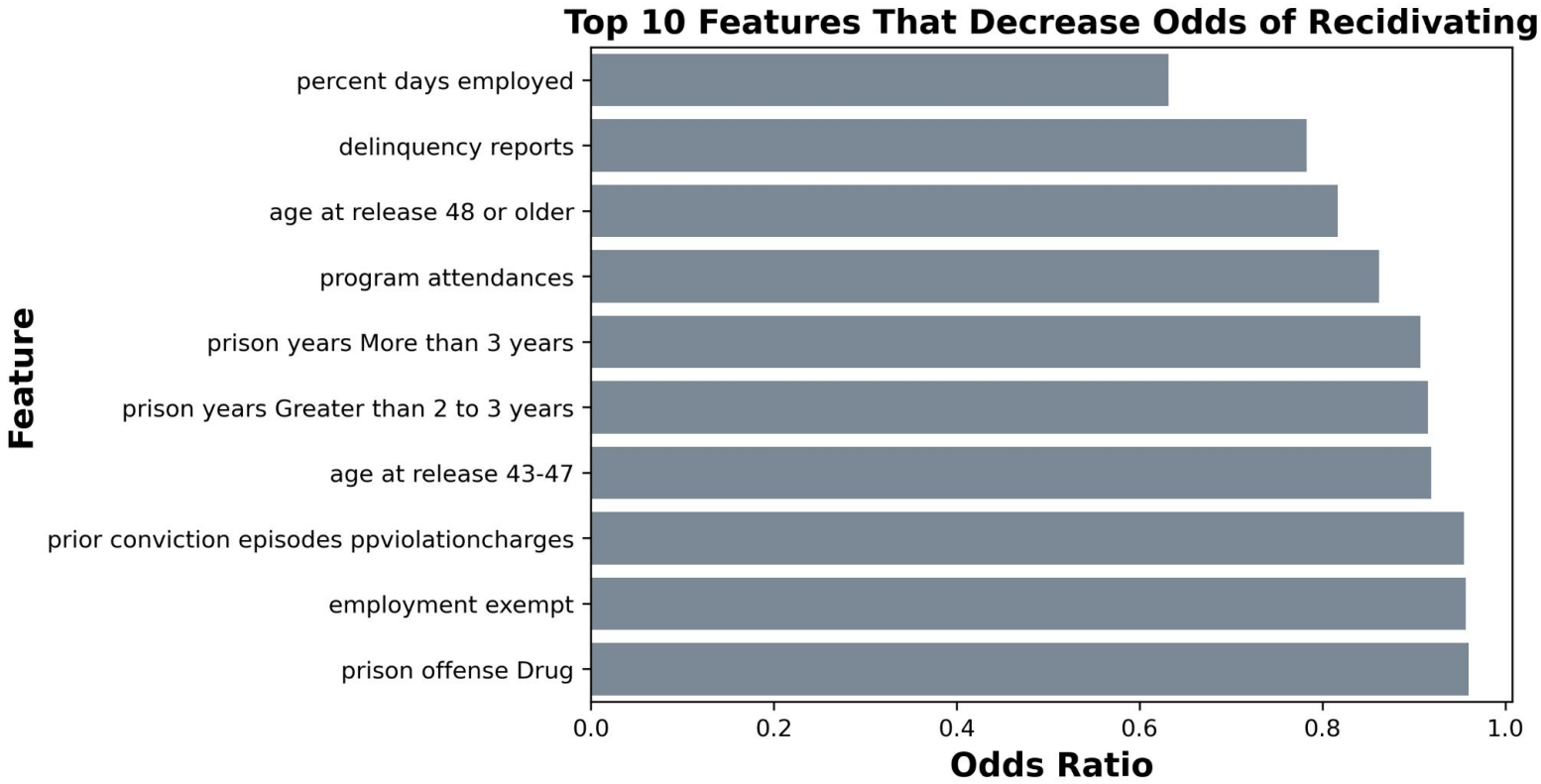
Models	Brier Score (Train / Test)	Precision (Train / Test)
Baseline	.25*	N/A
Logistic Regression	0.188 / 0.192	0.7287 / 0.7264
Random Forest Classifier	0.2045 / 0.2049	0.6943 / 0.6988
XGBoost Classifier	***0.1720 / 0.1822***	***0.7543 / 0.7453***

*source: https://www.ojp.gov/pdffiles1/nij/304110.pdf

Top logistic regression features that increase odds of recidivating.



Top logistic regression features that decrease odds of recidivating.



Insights & Recommendations

1. Employment & program attendances decrease odds of recidivism. Barriers to employment need to be addressed. Individuals need guaranteed employment opportunities rather than finding jobs independently.
2. Increased support for young individuals between the ages of 18-27 is needed to decrease recidivism rates.
3. Poverty is shown to factor into recidivism. Individuals need wrap-around support including, but not limited to stable income and housing

Next Steps

1. PUMAs data are grouped together to hide identity. It may be better if PUMAs were analyzed on a more granular level for better geographic insights
2. Compare models split on race and gender to compare bias and help improve false positive rate.
3. Test out additional models such as Neural Networks to improve predictive ability.

Thank you!

Questions?

Sources

- "NIJ's Recidivism Challenge Full Dataset." Office of Justice Programs, U.S. Department of Justice, data.ojp.usdoj.gov/Courts/NIJ-s-Recidivism-Challenge-Full-Dataset/ynf5-u8nk.
- "Recidivism Forecasting Challenge." National Institute of Justice, U.S. Department of Justice, nij.ojp.gov/funding/recidivism-forecasting-challenge#g0jtto.
- "Evaluating Reentry Programs: A User's Guide to Measuring Recidivism." Office of Justice Programs, U.S. Department of Justice, www.ojp.gov/pdffiles1/nij/304110.pdf