Predicting Recidivism

Abhay A., Gabby C., Scott T., and Yadan T.

Problem Statement

Can we use machine learning models to predict recidivism within a period of three years?

Initial Research and Project Path

- We started off wanting to find a correlation between mental health spending and incarceration rates.
- We then reviewed state-level data to find if states could reduce incarceration rates using mental health services.
- During the latter part of discovery, we found a recidivism challenge from the US Department of Justice and that influenced how we approach recidivism, mental health, and incarceration.

Overview: Mental Health and Incarceration

- 1.2 million individuals living with mental illness are in jail/prison.
- Imprisonment typically begins with low-level offenses.
- Studies show a positive correlation between rates of adults in the criminal justice system and lack of access to mental health care.
- States with less access to mental health care have more adults in the criminal justice system.

Rank	State	State Imprisonment (per 100k) Ranked Lowest to Highest	Access to Care Ranking (Sum of Scores)
1	Maine	153	-8.9271
2	Rhode Island	178	-5.74207
3	Massachusetts	188	-13.2939
4	Minnesota	194	-8.34278
5	North Dakota	214	-2.1477
6	New Hampshire	219	-8.12996
7	Utah	237	1.967375
8	Vermont	241	-16.4588
9	New Jersey	241	-3.39434
10	Washington	254	-1.32436

Source: https://www.apa.org/monitor/2019/03/mental-heath-inmates

Predicting Recidivism

- COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is a widely used criminal risk assessment tool.
- This software does not include race, but other data correlated to race could lead to racial disparities.
- COMPAS appeared to favor white defendants over black defendants by underpredicting recidivism for whites and overpredicting for blacks

	HUMANS	COMPAS
Accuracy (overall)	67.0%	65.2%
False positive (black defendants)	37.1%	40.4%
False positive (white defendants)	27.2%	25.4%
False negative (black defendants)	29.2%	30.9%
False negative (white defendants)	40.3%	47.9%

J. DRESSEL *ET AL.*, *SCIENCE ADVANCES*, EAAO5580, 2018, ADAPTED BY C. AYCOCK/*SCIENCE*

Source: https://www.science.org/doi/10.1126/sciadv.aao5580

Our Goal

We hope to understand the **risks associated with recidivism**.

We're also hoping to find possible **correlations between mental health spending and incarceration rates**, which might provide insight into which states should invest more in mental healthcare services.

The Importance

- About 37% of prisoners have a history of mental health disorders.
- Prisoners can lose their access to mental health services during imprisonment.
- Prisoner Jamycheal Mitchell, 24, lost access to his schizophrenia medication when no beds were available for his “competency restoration.” He was found dead 4 months after his arrest covered in urine and feces. He died of cardiac arrhythmia related to wasting syndrome.

“We lock up people with mental health problems when we should really be treating these people in the community”

- Thomas Fagan, PhD, professor emeritus at Nova Southeastern University and former administrator for the Federal Bureau of Prisons.

Source: https://www.apa.org/monitor/2019/03/mental-heath-inmates

The Data

- Census_incarceration_data:

  https://observablehq.com/@themarshallproject/adults-in-correctional-facilities-from-decennial-census

- Public_health_spending_data:

  https://knoema.com/SHPCPHF2020/per-capita-public-health-funding-in-u-s-states

- Mental_health_spending_data:

  https://rehabs.com/explore/mental-health-spending-by-state-across-the-us/

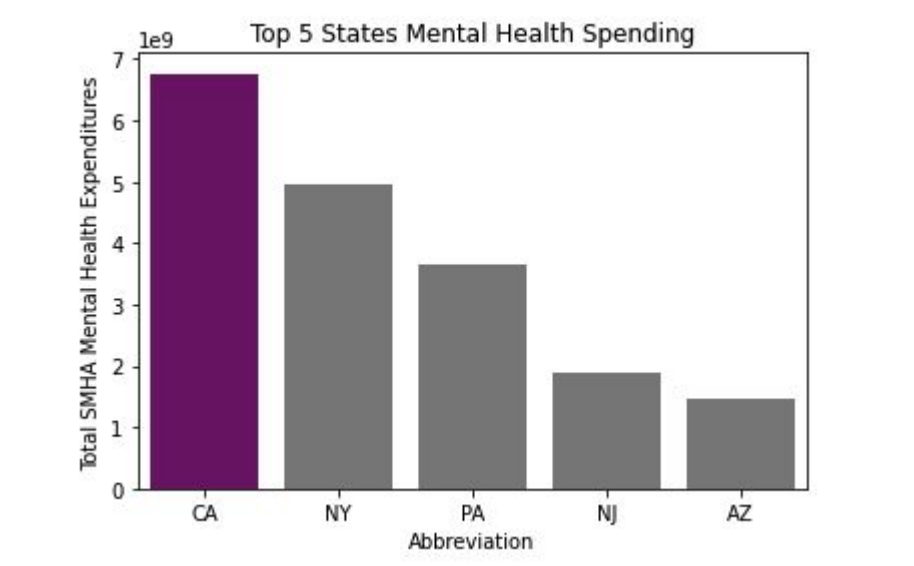
- Recidivism_data:

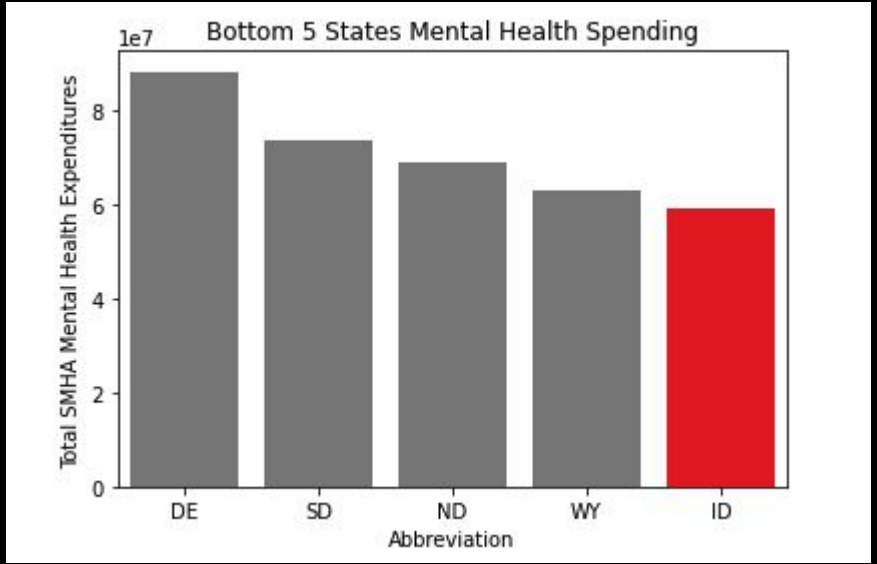
  https://data.ojp.usdoj.gov/stories/s/daxx-hznc

Assumptions

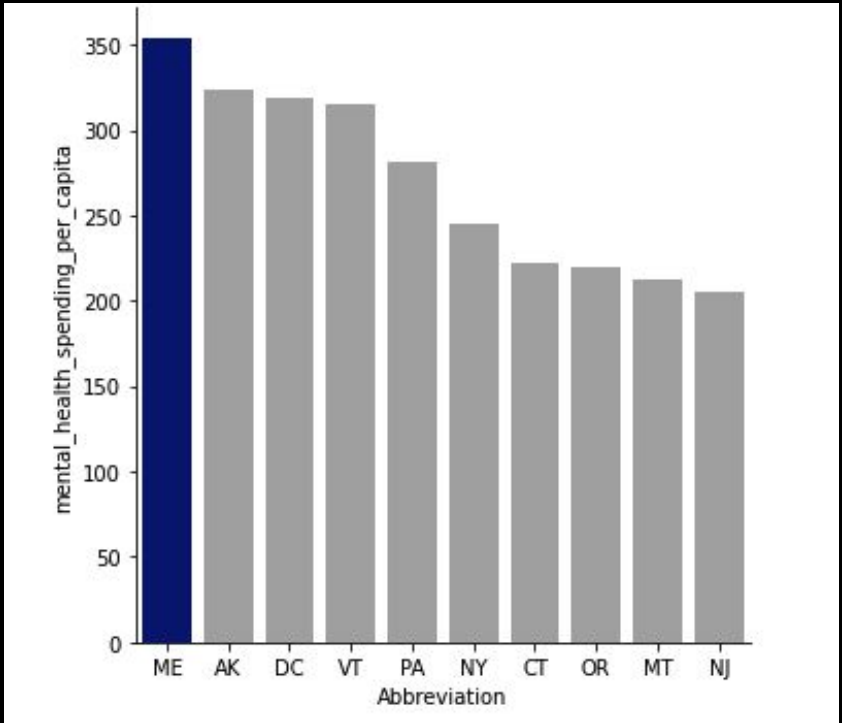
- Over 10% of the data in gang affiliation and prison offenses is missing.
- There are columns that are unreasonable to impute.
    - We cannot assign gang affiliation and/or a prison offense to individuals.
- In this study we are using recidivism score as a measurement of successful and effective state mental health spending.

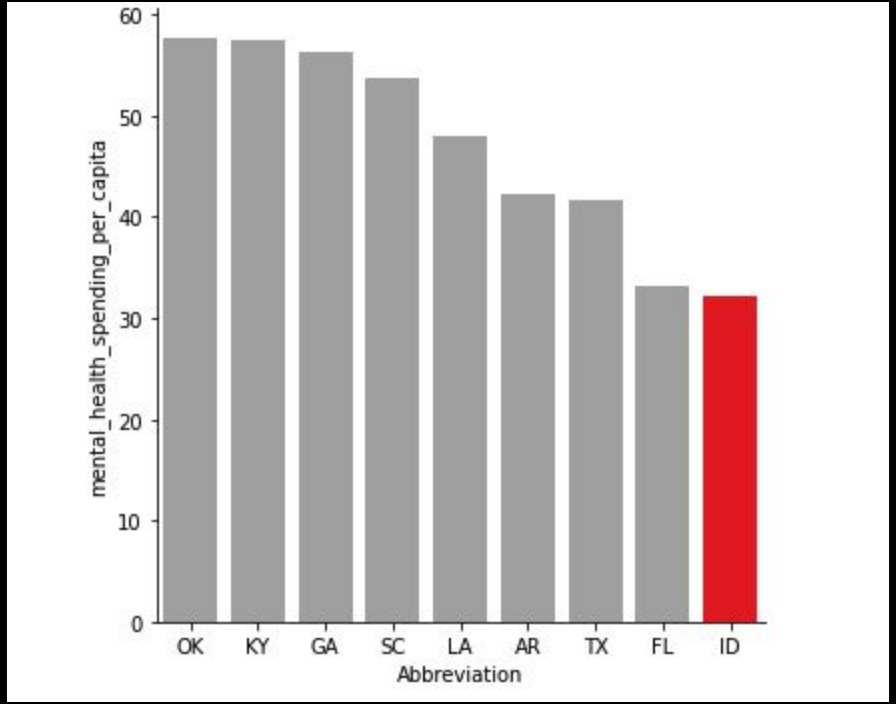
Data Visualizations for Total SMHA Spending



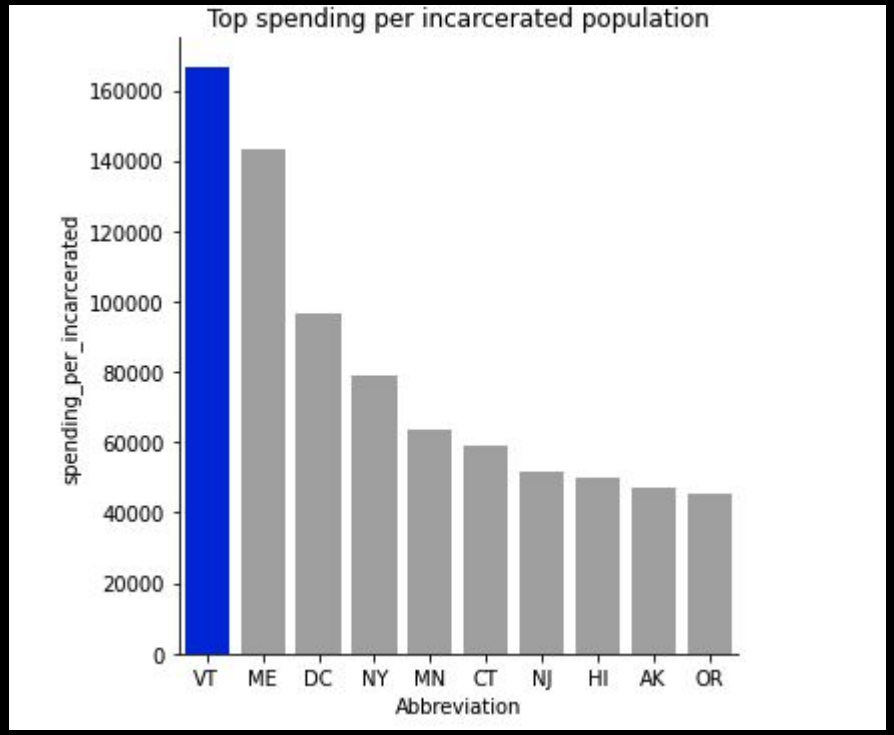


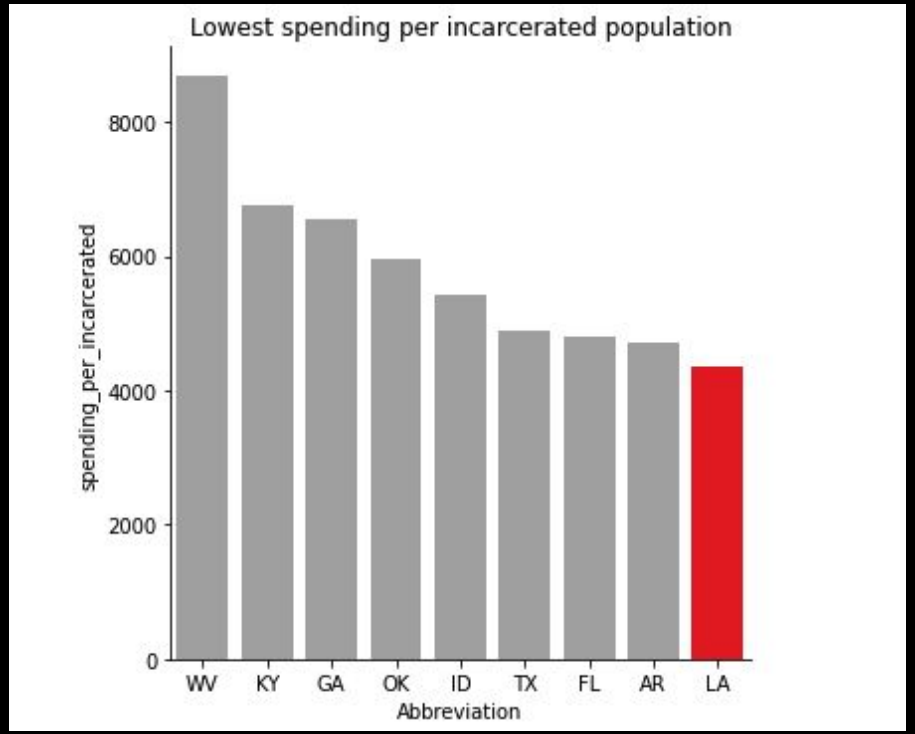
Mental Health Spending Per Capita



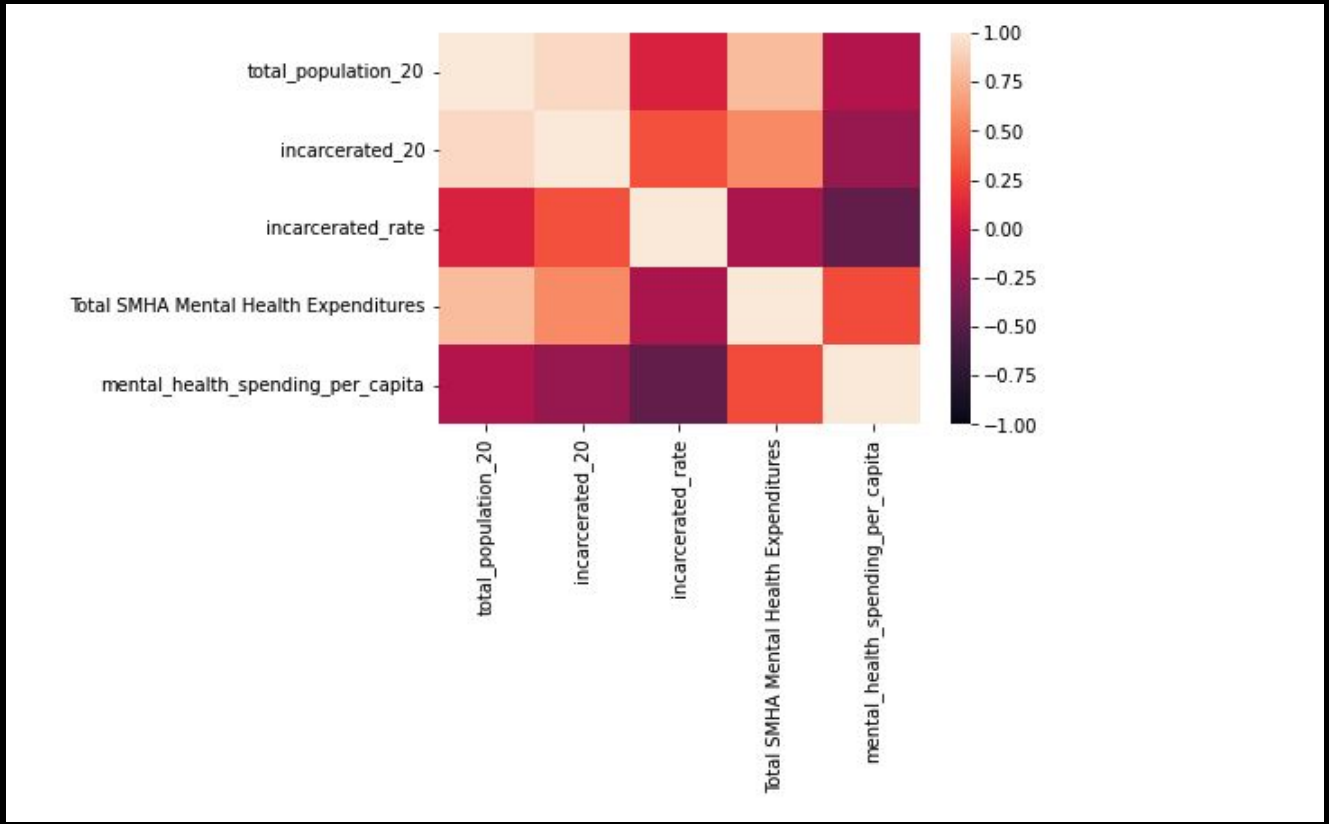


Spending By Incarcerated Population

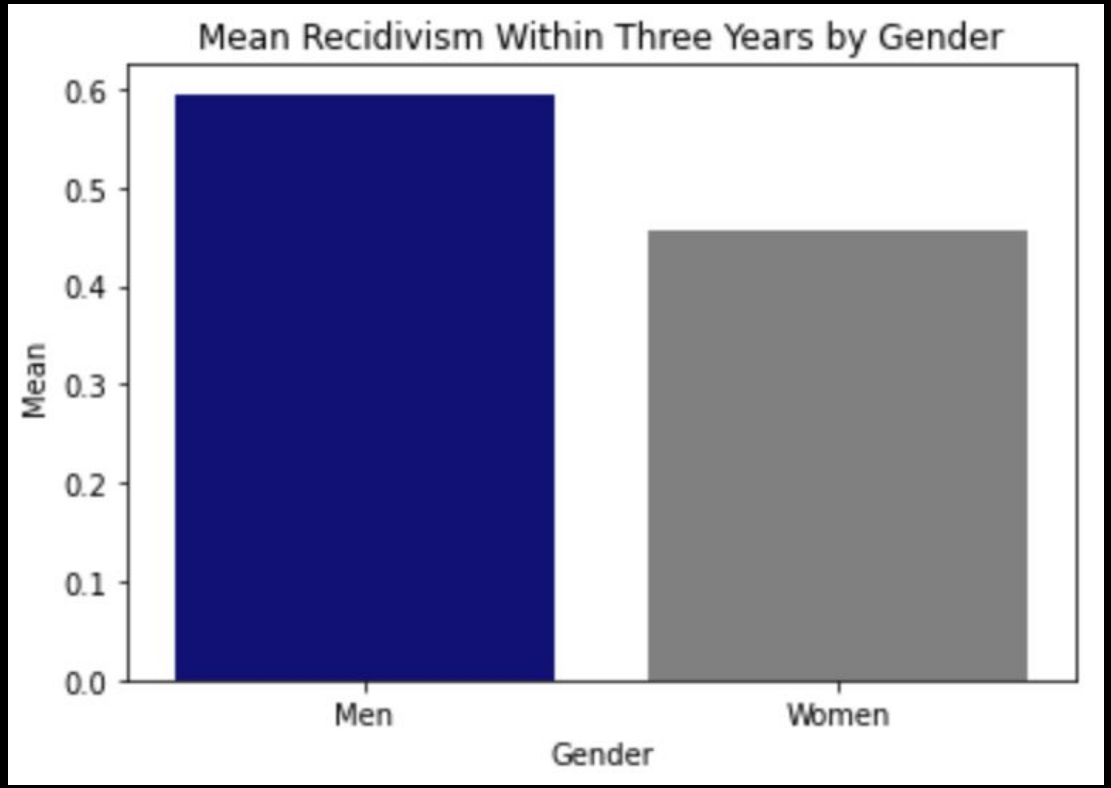




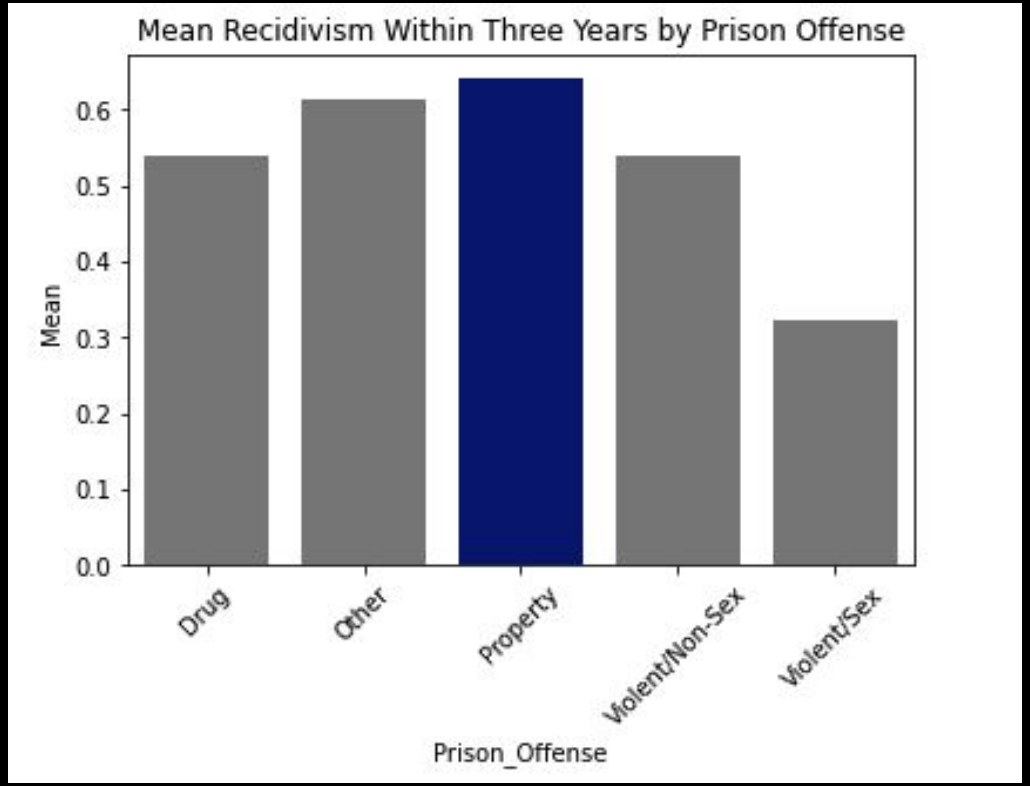
Correlations of Variables



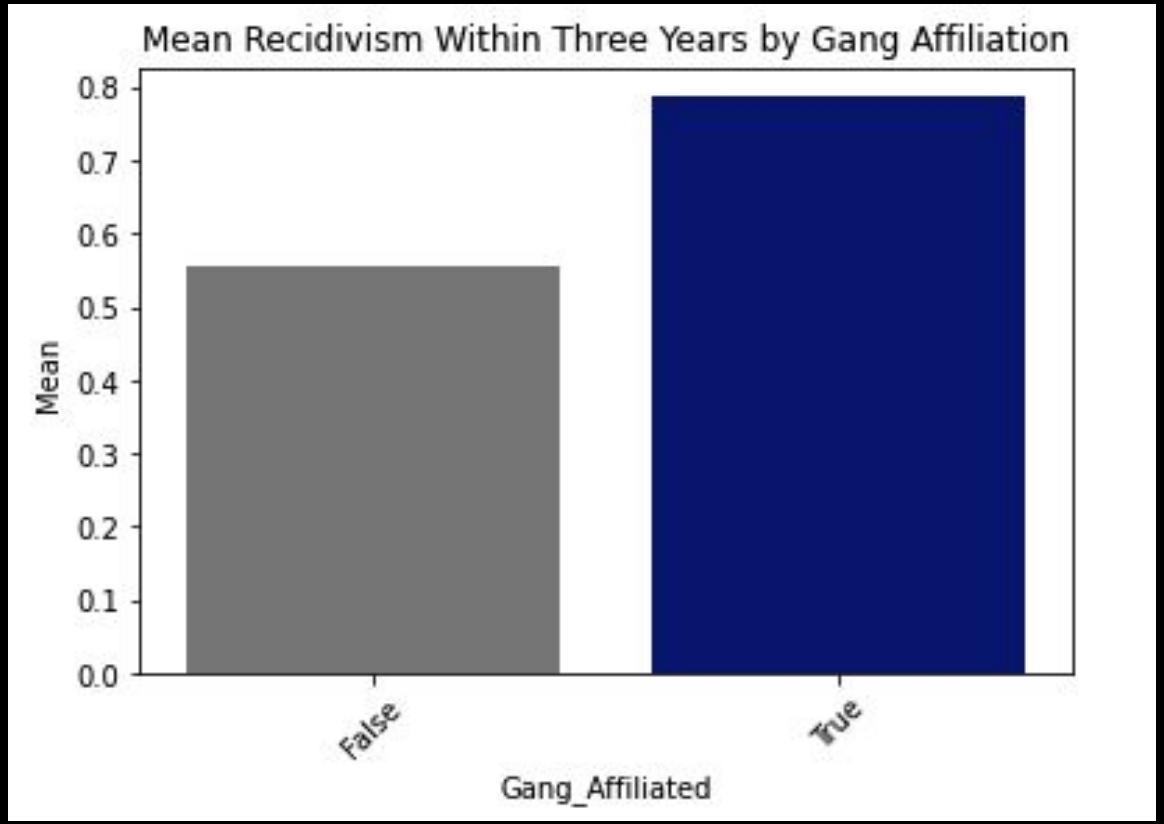
Recidivism by Gender



Recidivism by Prison Offense

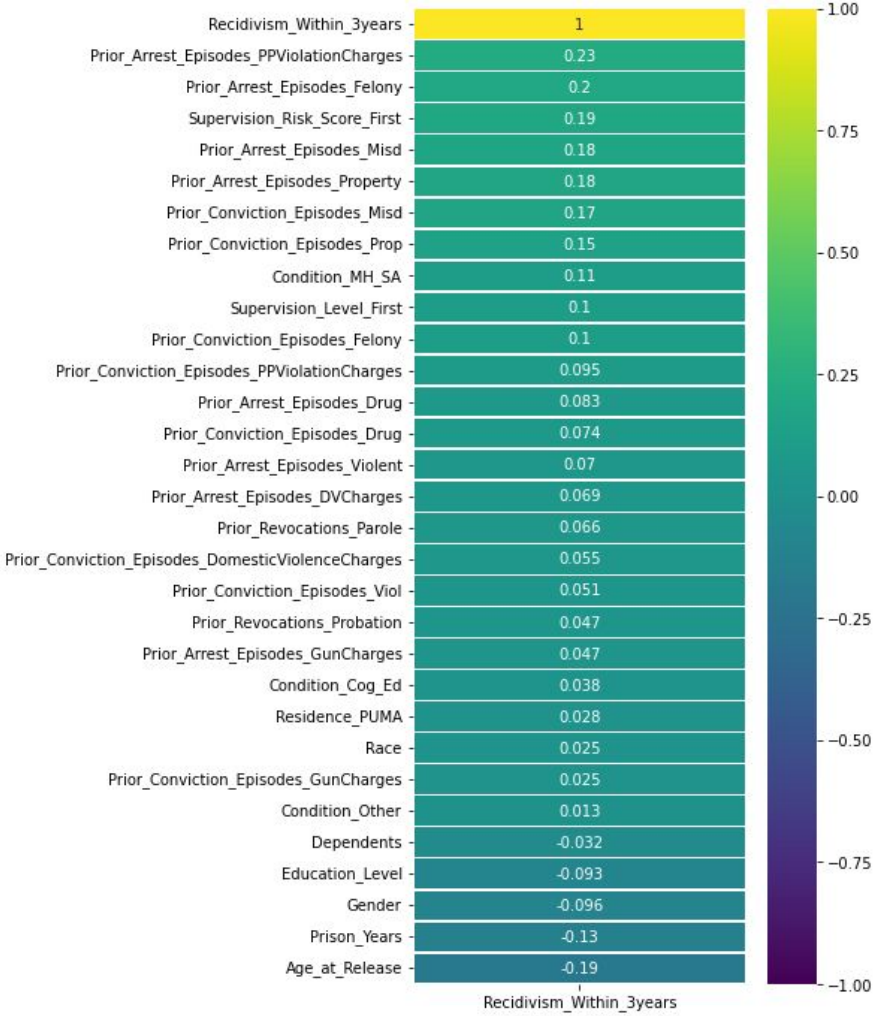


Recidivism by Gang Affiliation



Correlation

Checking to see the extent to which the categories are related.



Base Model

- Our base model han an accuracy of 57.4% for recidivism within three years.

Model: Neural Network

- Neural network modeling performed moderately well.
- KNN analysis yielded a score of ~68.6% with grid searching over neurons and 1 or 2 layers. And Model was slightly overfitted.
- Additional and “deeper” grid searching over parameters and 4 hidden and dropout layers with (drop rate 0.2) achieved a score of ~68.74%, not improvement on accuracy score, but model is not overfitted.
- F1_score: 75.2%
- Recall_score: 81.7%

Model: XGBoost

- While discussing how to solve for missing data, we found that some data scientists recommended XGBoost.
- Oddly enough, the model performed better on datasets that were not our "cleanest."
- We had an accuracy score of **67.2%**, before hyperparameter tuning.
- F1_score: 69%
- Recall_score: 63%
- We tried Grid Search, but ended up overfitting.

Model: RandomForest

- Random Forest was the worst model performed.
- It achieved a score of ~66.70% with no hyperparameter tuning
- After grid searching over parameters, it was able to achieve a score of ~67.71% which only went up 1%
    - F1 score: 74.9%
    - Recall score: 82.4%
- Both models were definitely an overfit.

Conclusions (and interpretations)

- Found negative or low correlation between select features (gender, age, race, etc.) and recidivism within three years.
- Next steps would be to find similar recidivism data for another state in order to compare mental health spending with incarceration rates and recidivism predictions between states.