



# Wrongful Convictions: Are They Predictable?

Eric Geisler



# Wrongful Conviction

/'rōNGfəl/ /kən'vikSH(ə)n/

a person being convicted of a crime that they did not commit, typically due to procedural errors

# **Problem:**

limited resources are available in the  
fight against wrongful convictions

[Ohio Innocence Project](#)

Can we predict how bad  
a wrongful conviction is  
going to be?



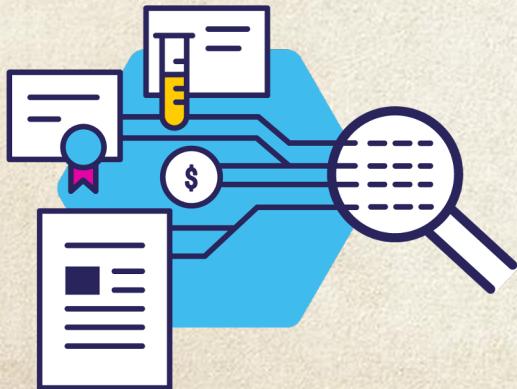
# The Data

[The National Registry of Exonerations](#)

## Contributing Factors:

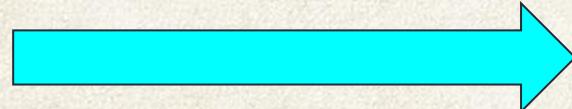
**Official Misconduct, Mistaken Witness Identification, False Confessions,  
Falsified Forensics, False Accusations, Inadequate Legal Defense**

**3353 Rows**  
(after removal of  
missing data)



## **9 Features**

- State
- Race
- Sex
- 6 Contributing Factors



## **2 Labels**

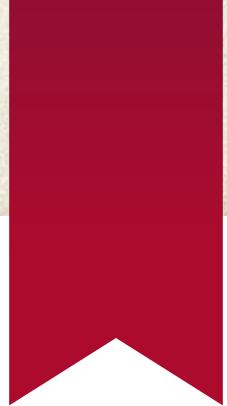
- Sentence
- Time Spent in Prison





## Data Cleaning // Feature Engineering

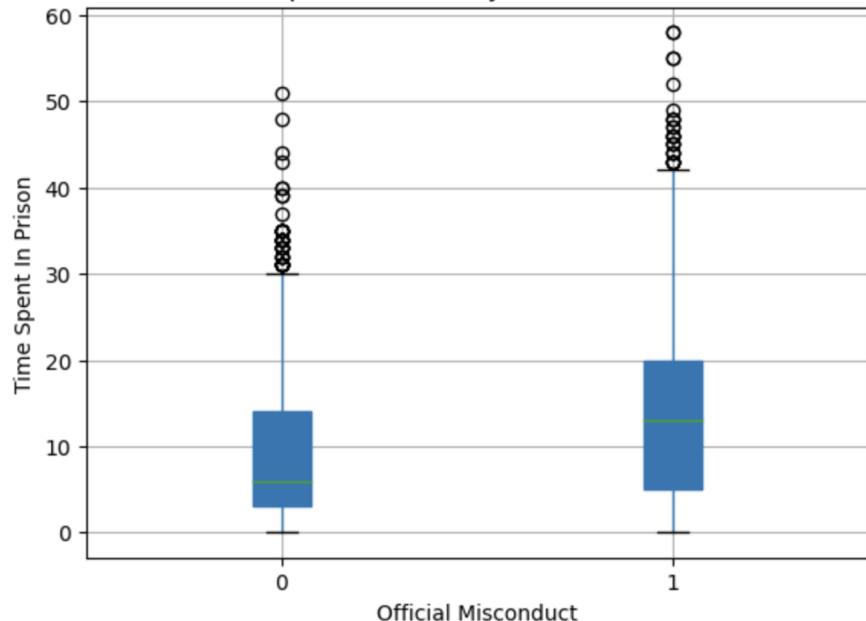
- Contributing factor variables turned to dummy columns
- Label Formation:
  - Calculated “Time Spent In Prison” label column (log)
  - Simplification of “Sentence” label column
    - Death, Possible Life, 40+ Years, 30+ Years, 20+ Years, 10+ Years, < 10 Years, < 1 Year, Other
- Numerical variables for “State”, “Race”, and “Sex”
- Normalization of “State” and “Race”



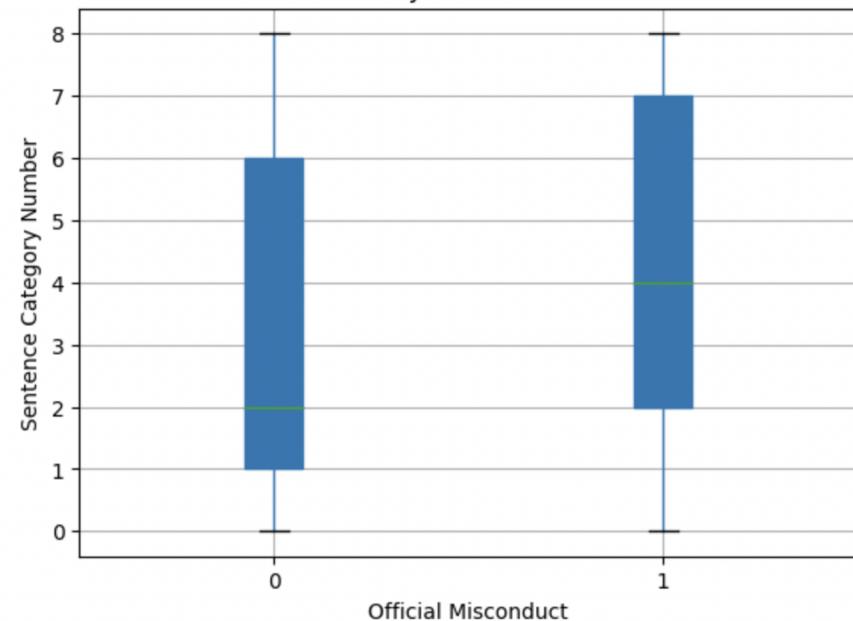
# Examining The Contributing Factors

# Official Misconduct

Time Spent In Prison by Official Misconduct

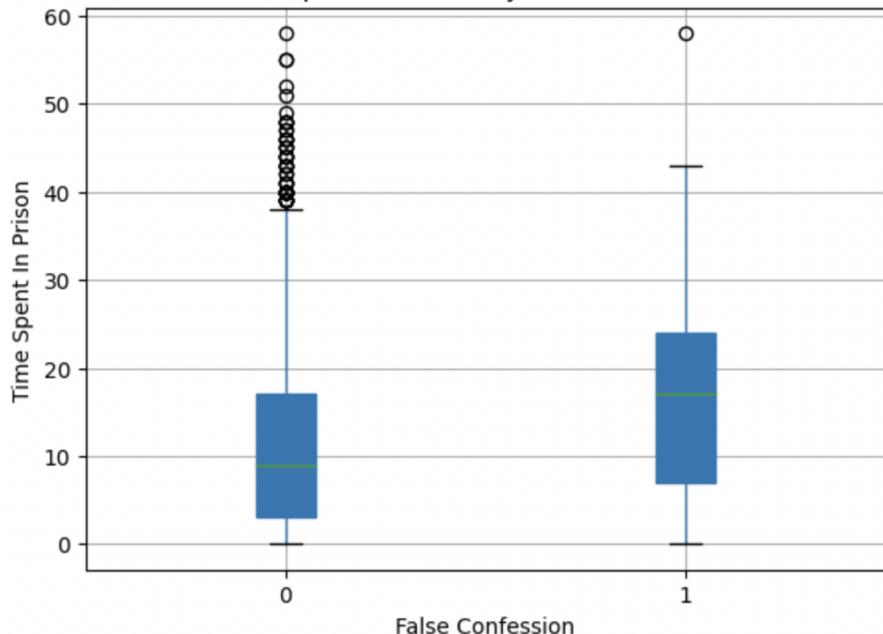


Sentence by Official Misconduct

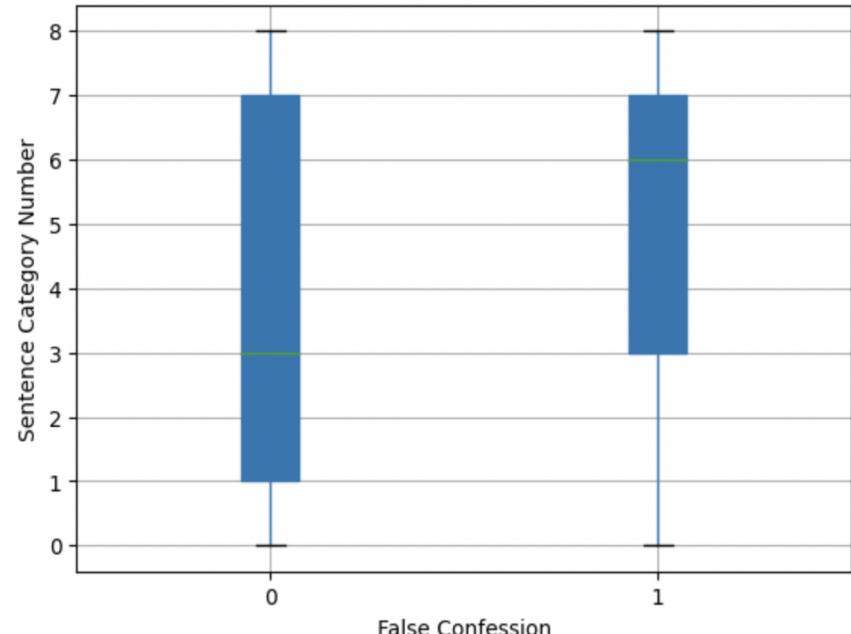


# False Confessions

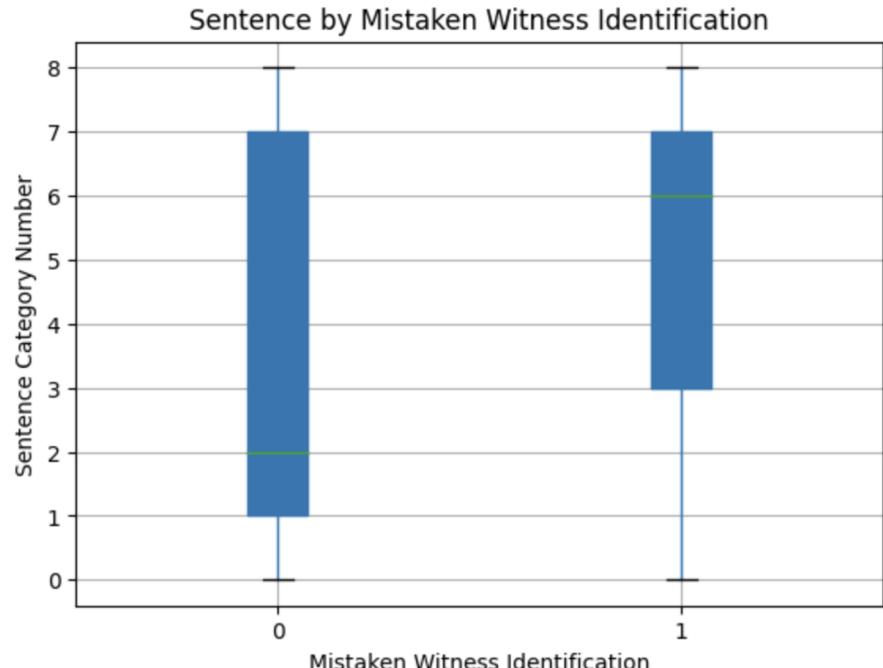
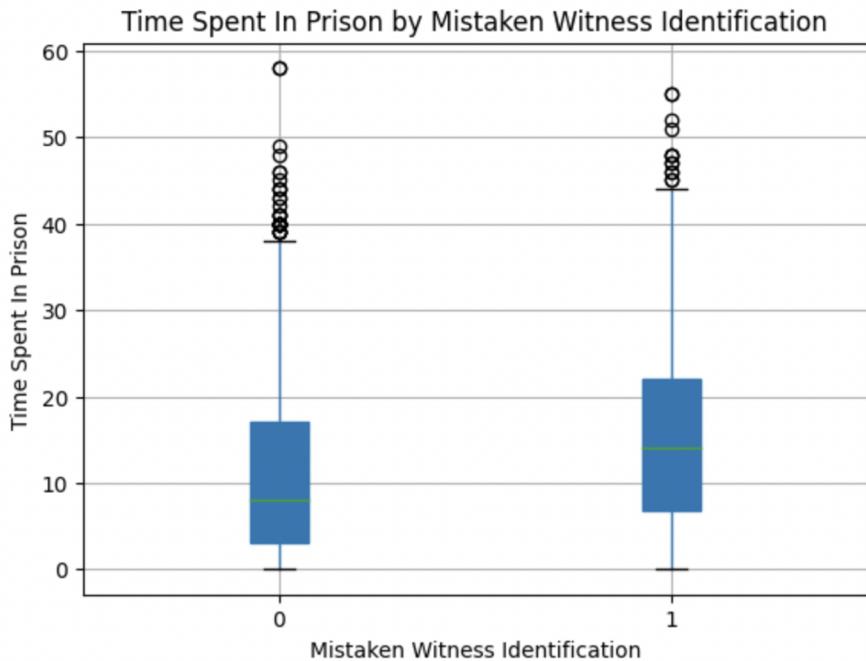
Time Spent In Prison by False Confessions



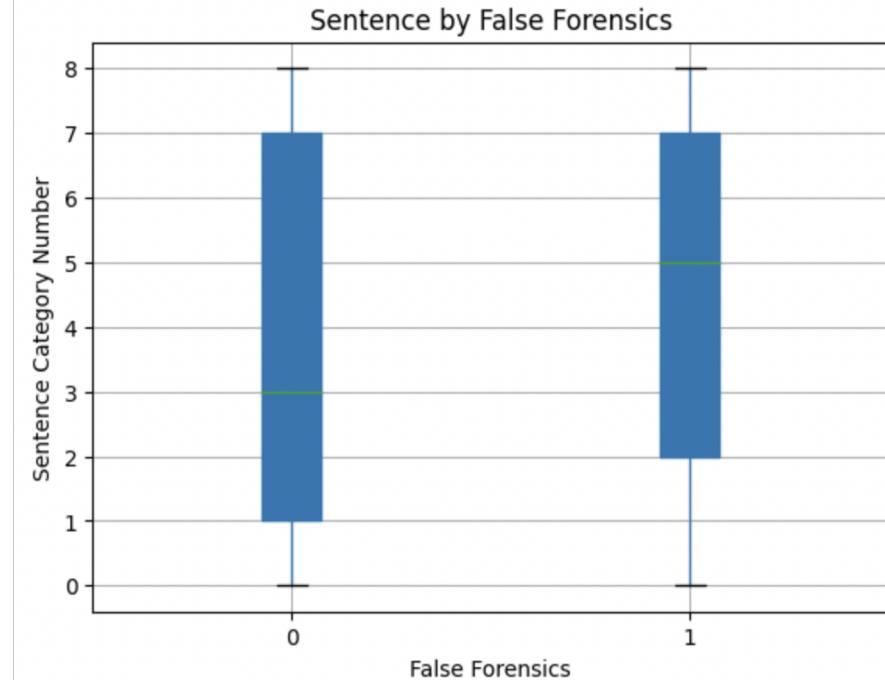
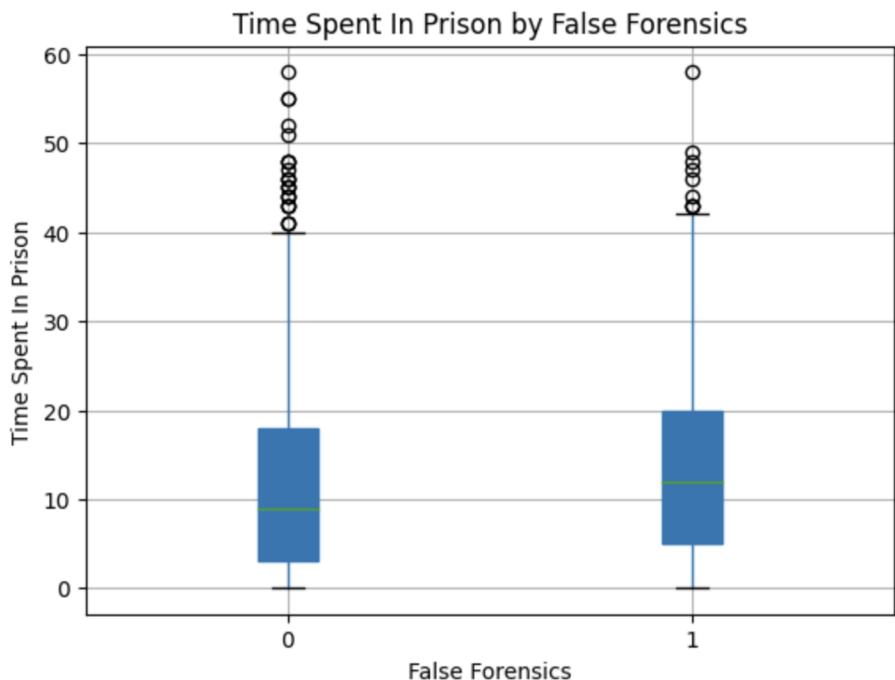
Sentence by False Confessions



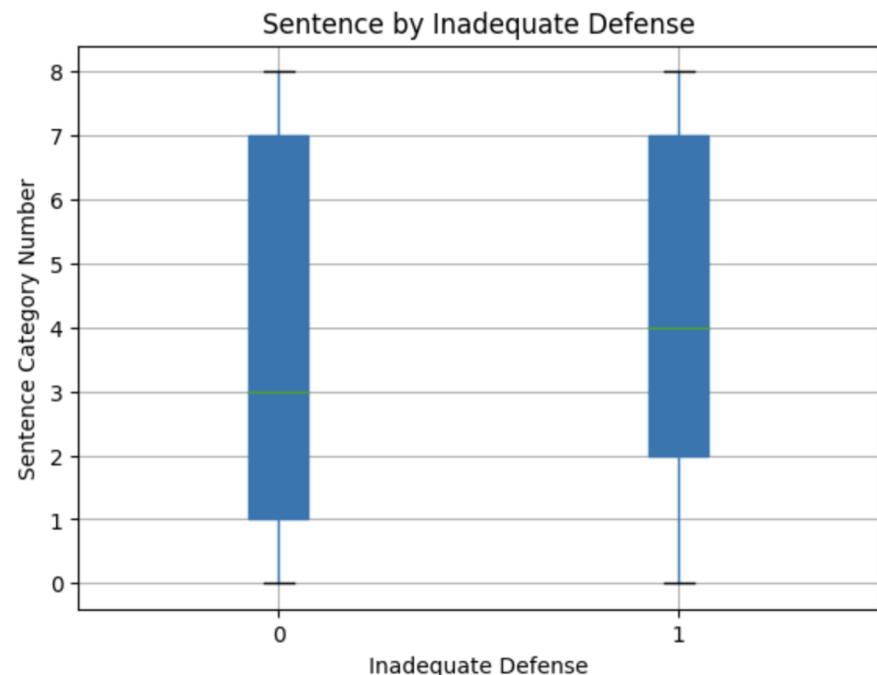
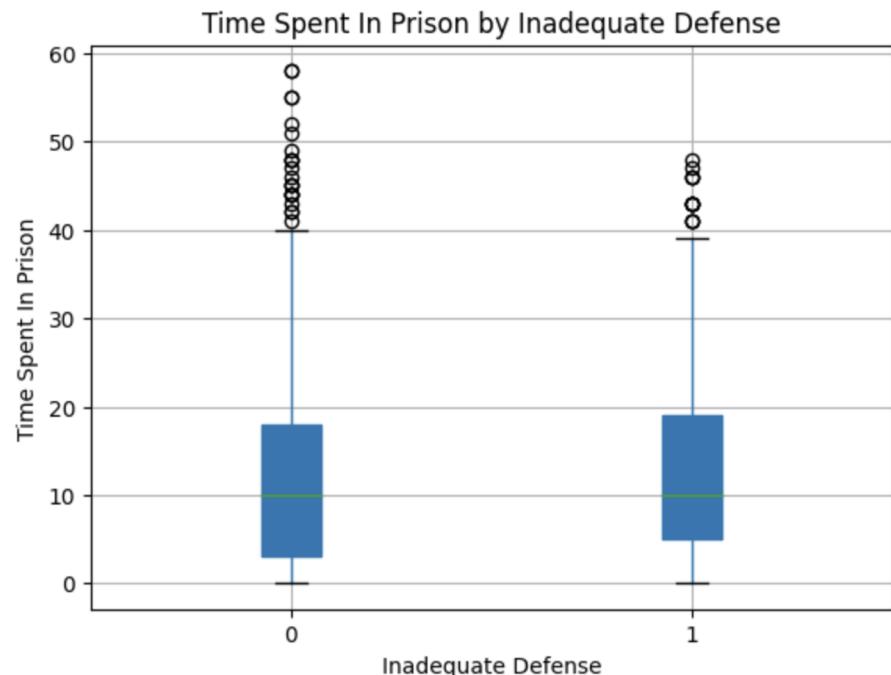
# Mistaken Eyewitness Identification



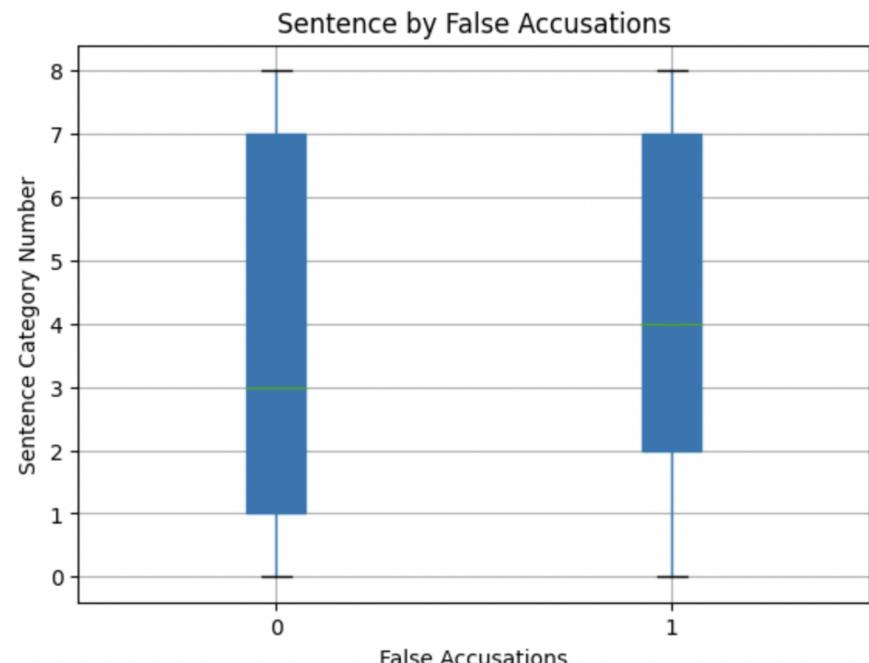
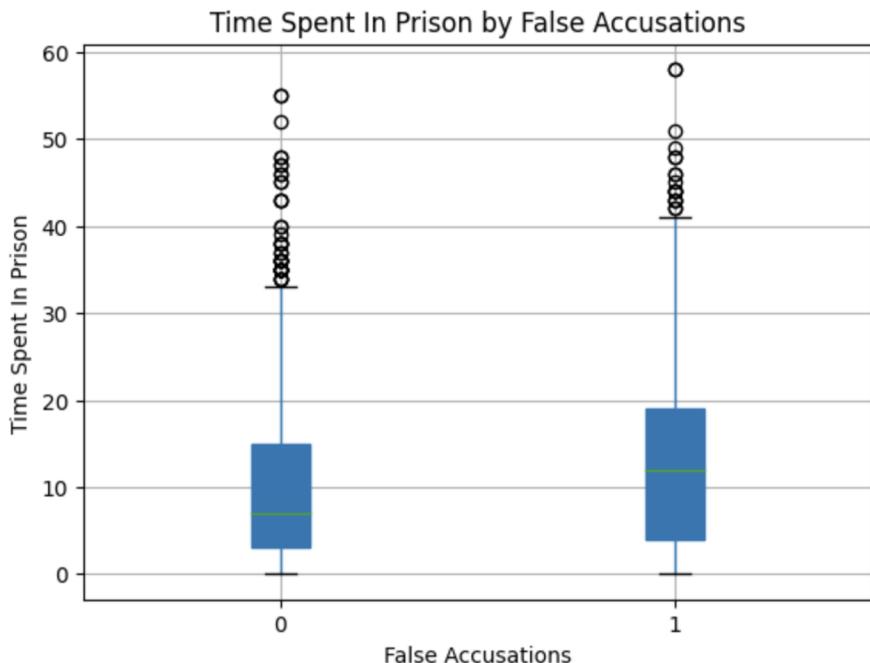
# False Forensics

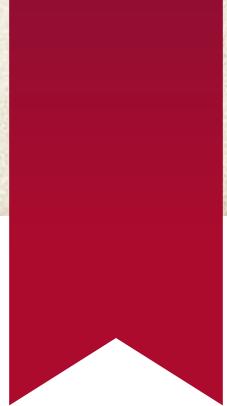


# Inadequate Defense



# False Accusations





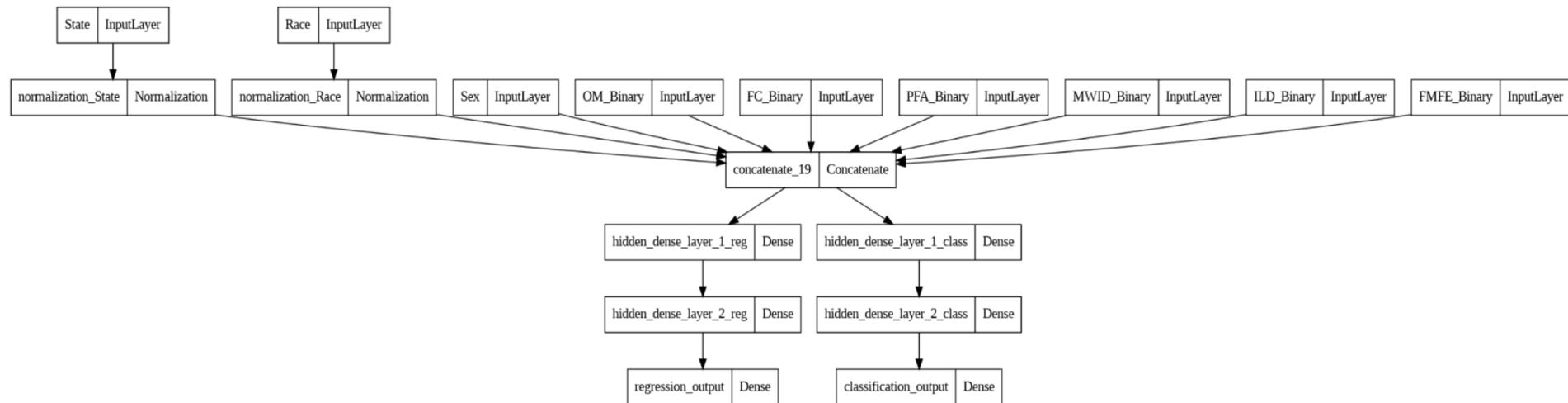
# The Model

Feed Forward Neural Network

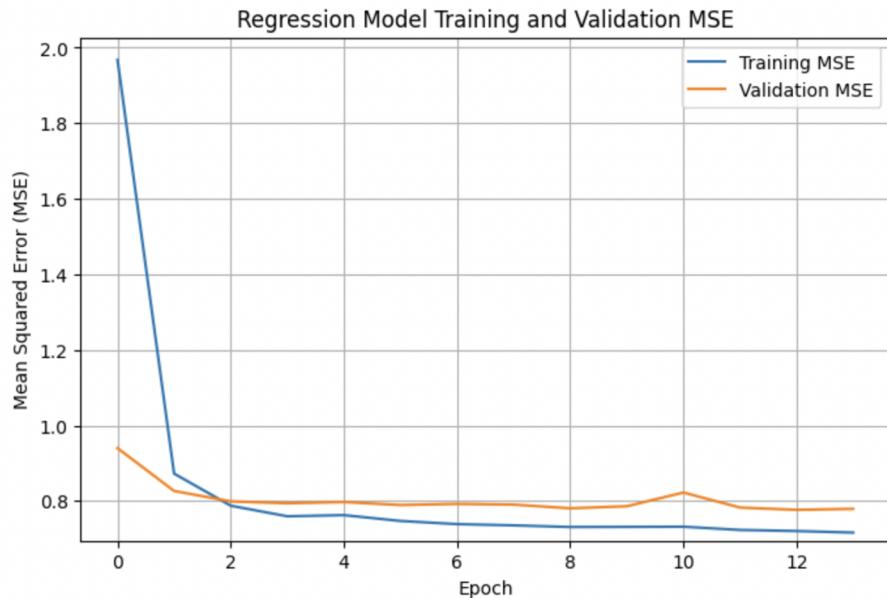


# Model Specifics

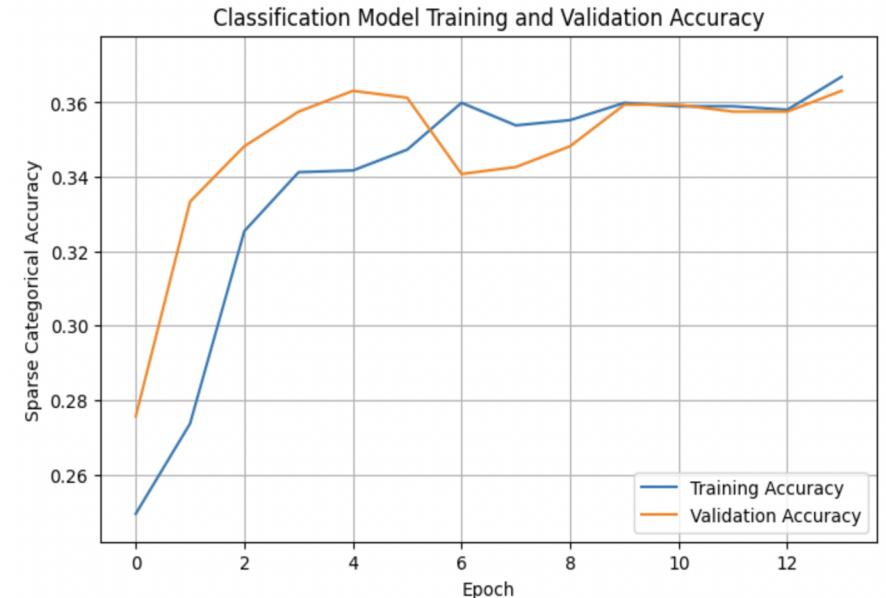
- Again, 9 inputs
- 2 Outputs:
  - Regression: “Time Spent In Prison”
  - Classification: “Sentence Category Number”
- 3 layers per output (2 hidden, 1 output)
- 14 epochs



# Training/Validation: 80/20 Split



Corrected RMSE: 2.34



Accuracy: ~37%

**Good News:  
The Model is Learning!**



## Testing Evaluation (The Bad News)

**55%**

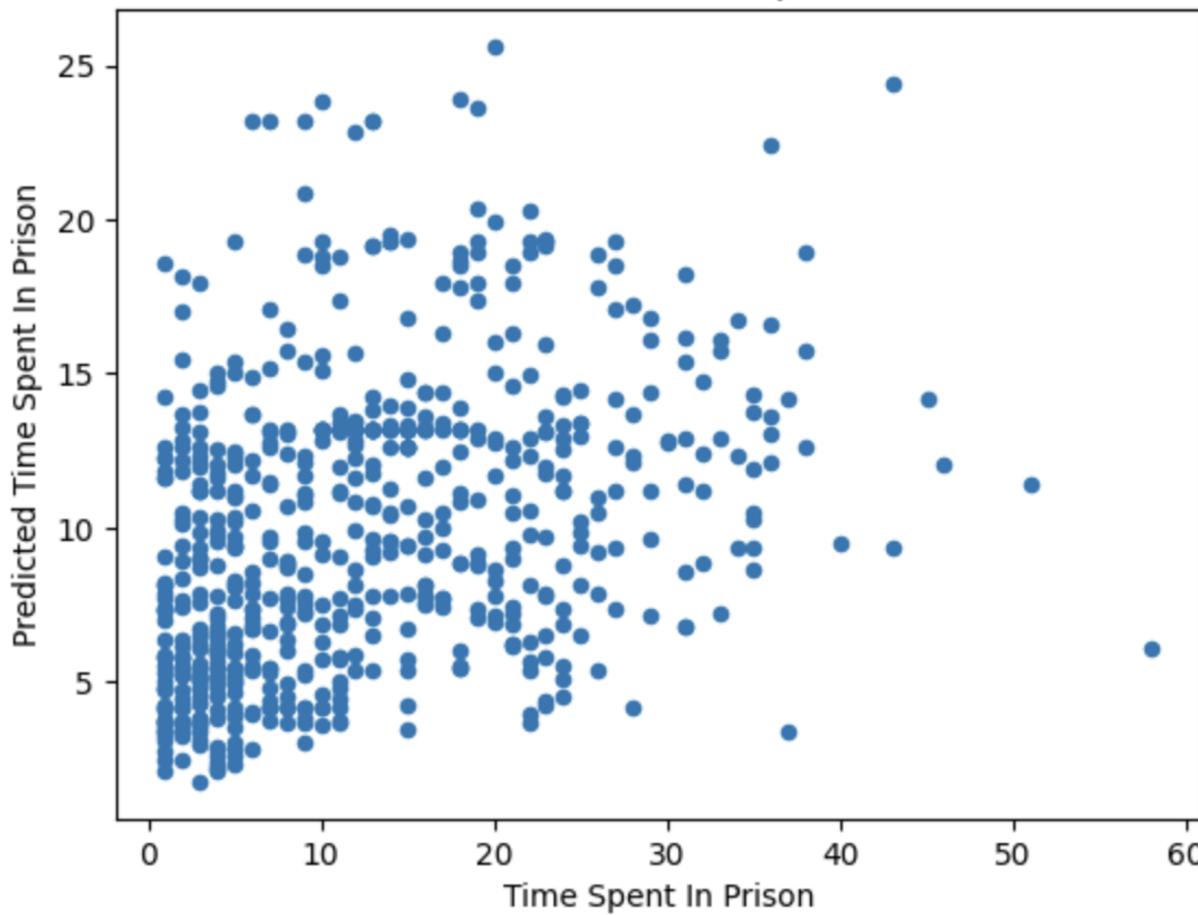
Classification Accuracy

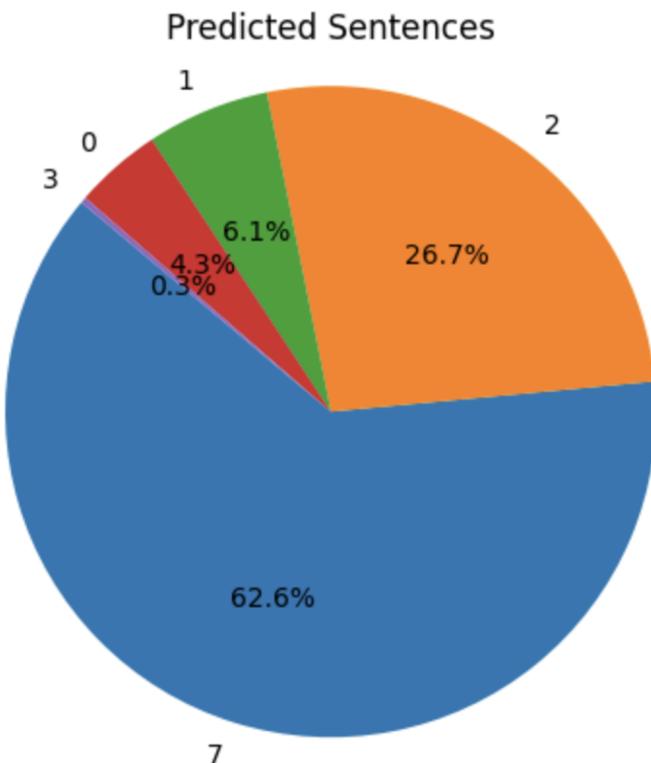
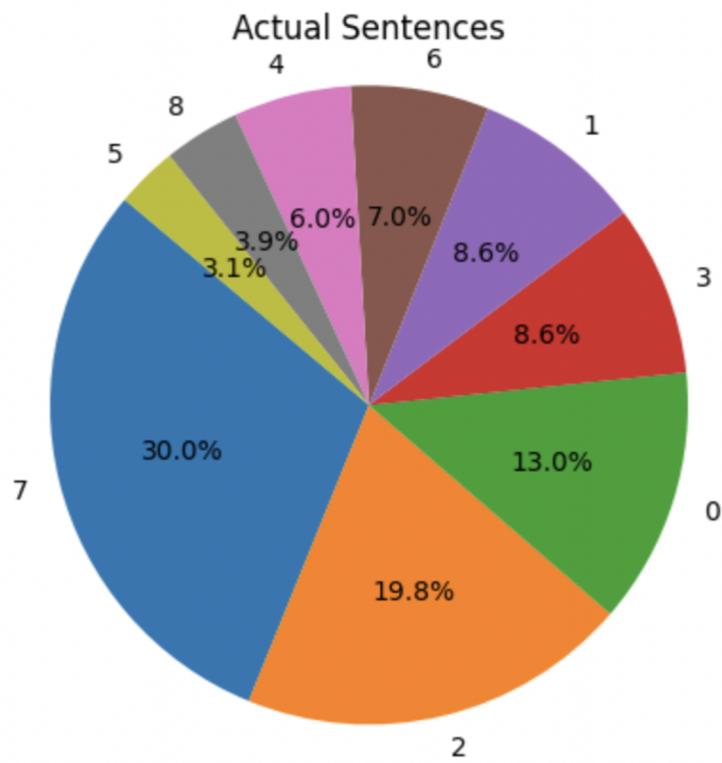
(22% random chance)

**6.83**

Regression MAE

### Predicted vs Actual Time Spent In Prison







# Takeaways

- The Model has some beneficial predictive ability regarding criminal sentences
- Predicting time in prison using basic demographics and contributing factors in this way is impractical
- Regarding potential policy change, the contributing factors of Official Misconduct, False Confessions, Mistaken Eyewitness Identification, and False Forensics are associated with longer time in prison and sentences



# Limitations / Future Work

- Some practical difficulty in prospective application of predictions and retrospective data
- Making the “Sentence Category” label from the initial “Sentence” column was not perfect
  - “Time Spent In Prison” label inherently over-represents past data
  - Further analyzing predictive power of variables to make a more efficient model
  - Using case data that does not exclusively contain wrongful convictions



# GitHub: Eric-Geisler