# Prison sentence discrimination in US Courts

Fernando Voticky
Ramón Galindo

# Problem and hypothesis

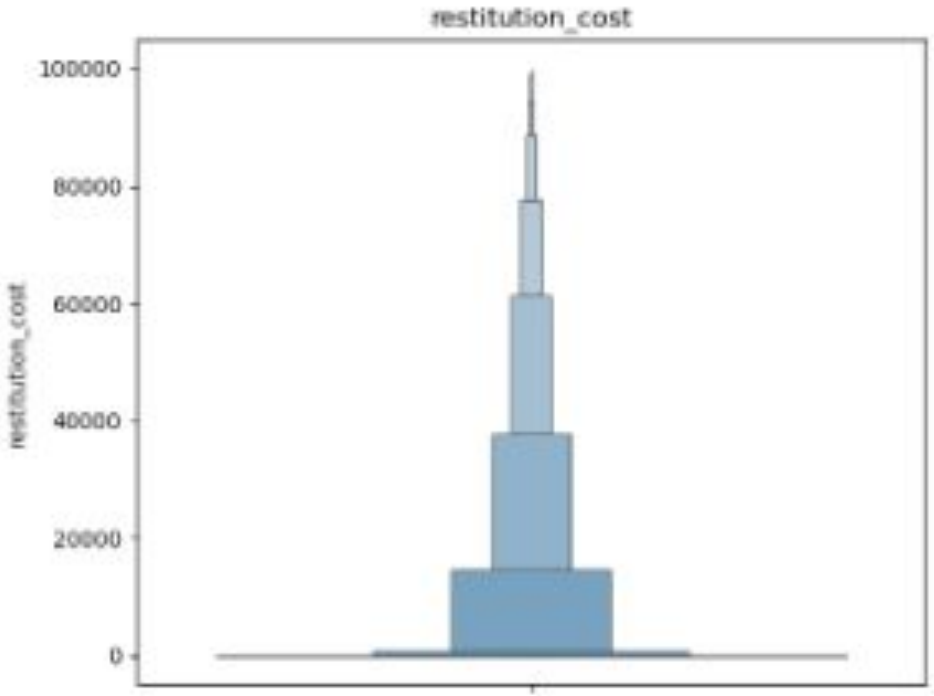
Judges give sentences based on the law, but depending on personal criteria.

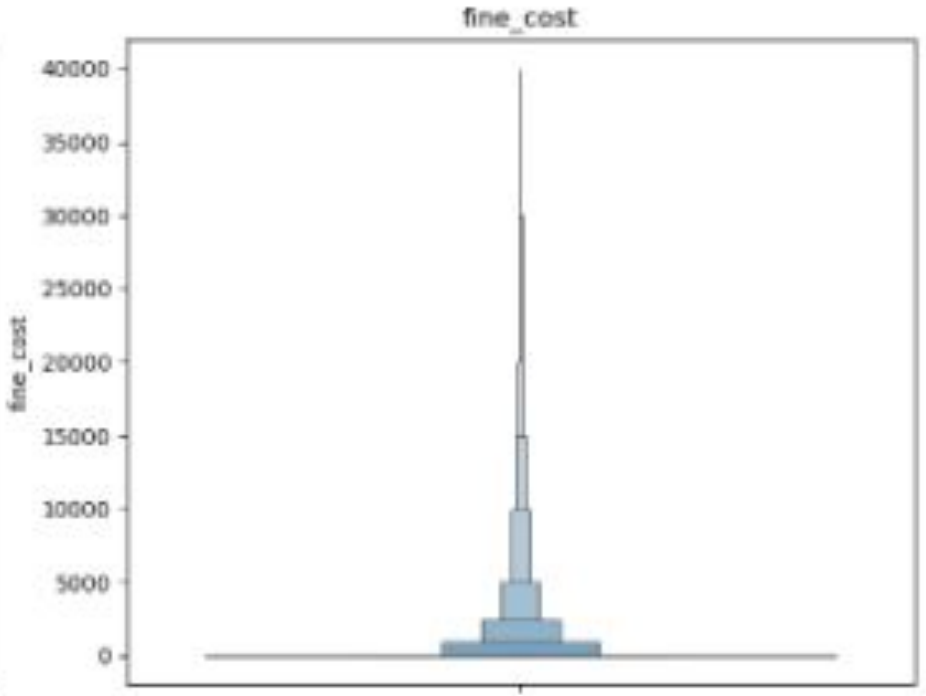
Enormous historic evidence of racial and sociodemographic discrimination.

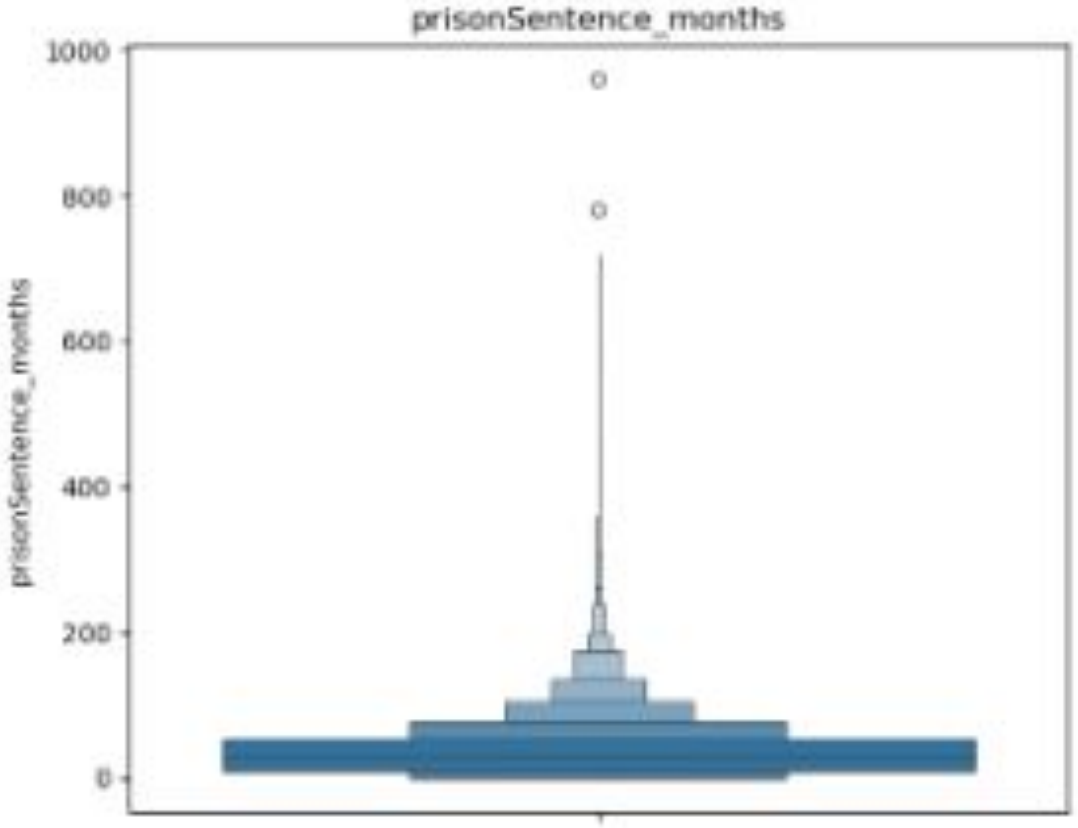
**Can we spot evidence of systemic injustice towards people of certain groups?**

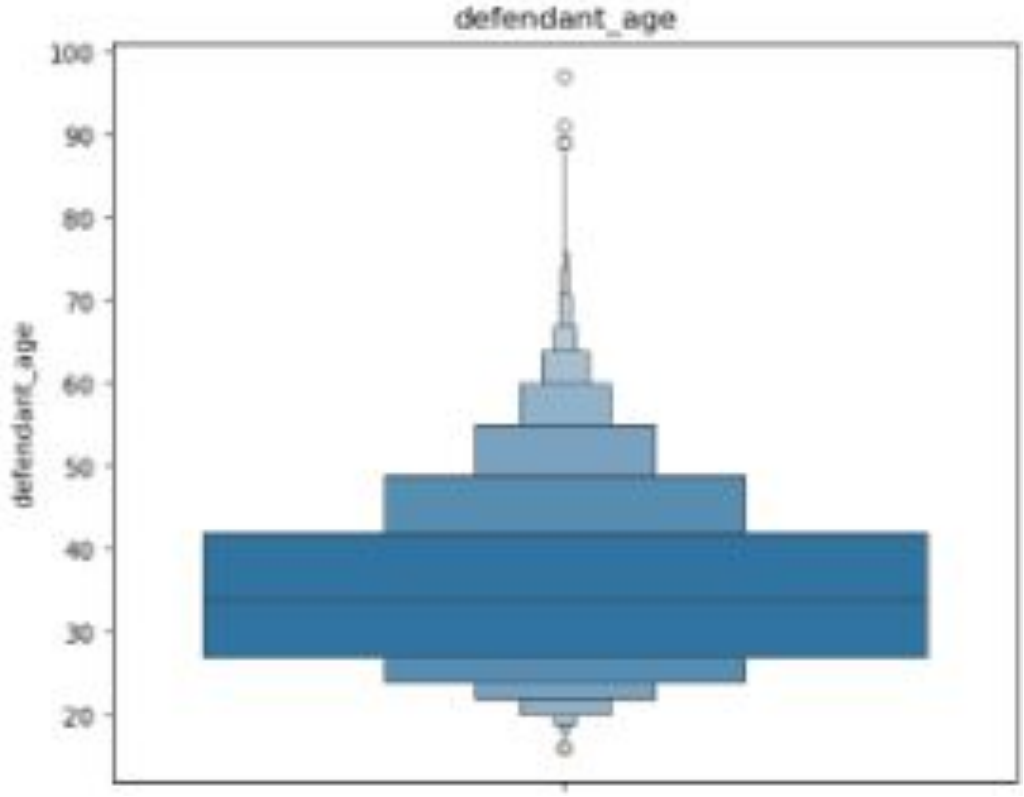
**Hypothesis: there is a slight difference between prison sentences served to people of the US that belong to marginalized sociodemographic groups.**

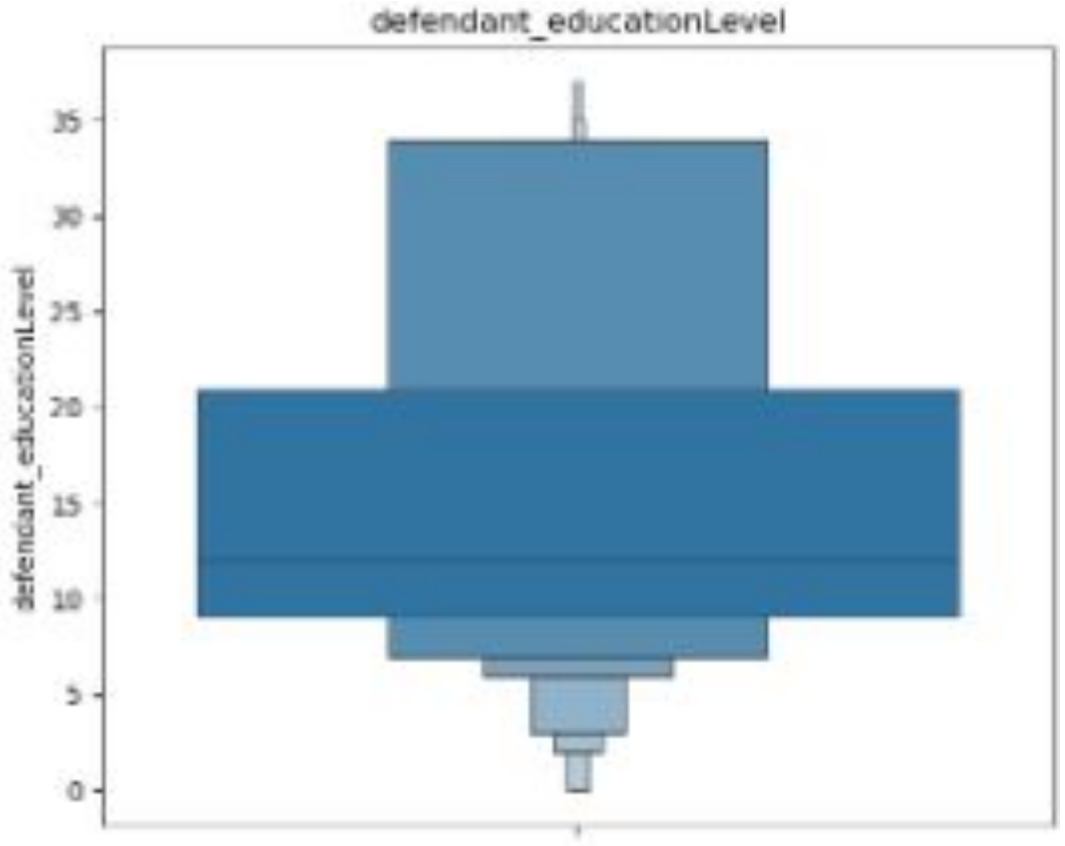
# EDA

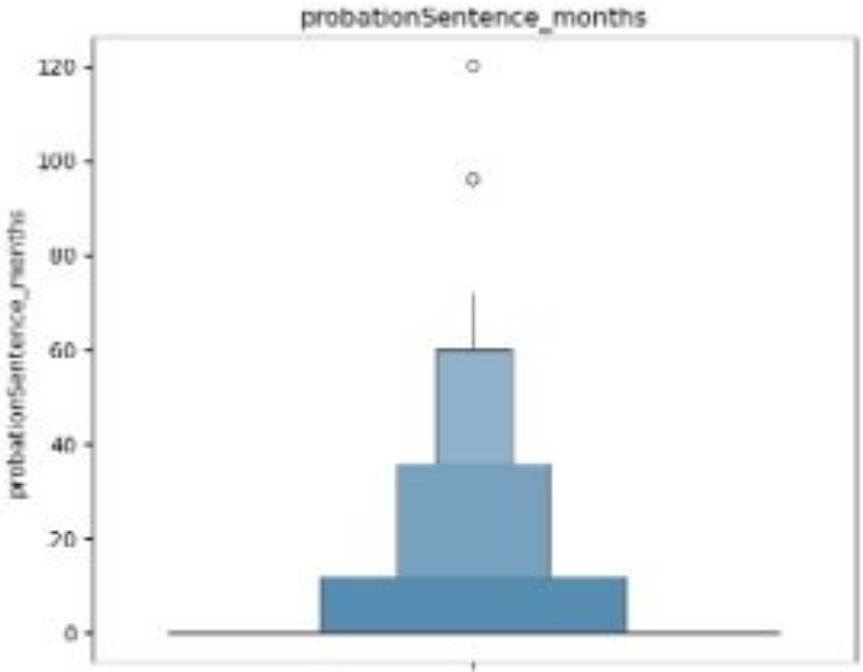

restitution_cost
100000
80000
60000
40000
20000
0
restitution_cost

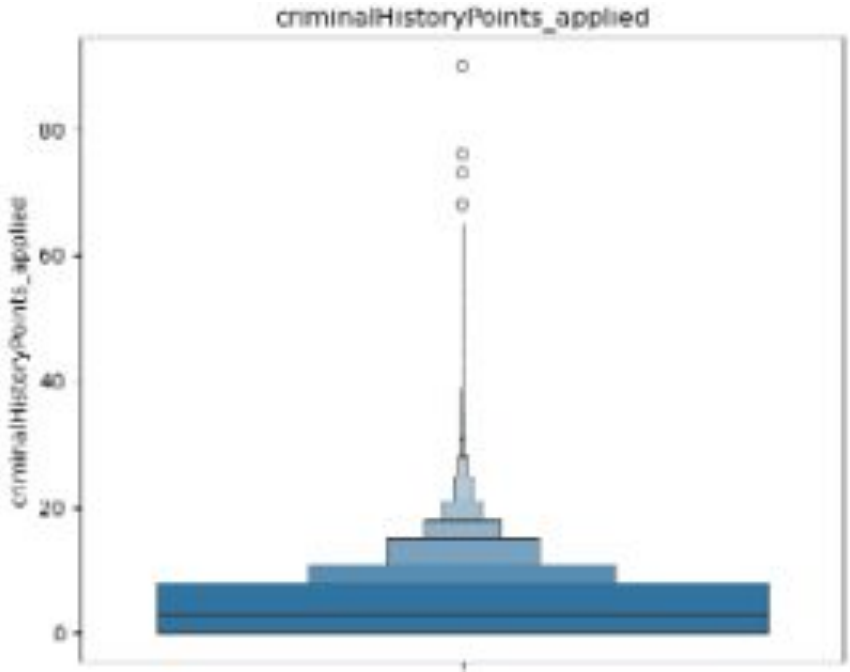

fine_cost
40000
35000
30000
25000
20000
15000
10000
5000
0
fine_cost

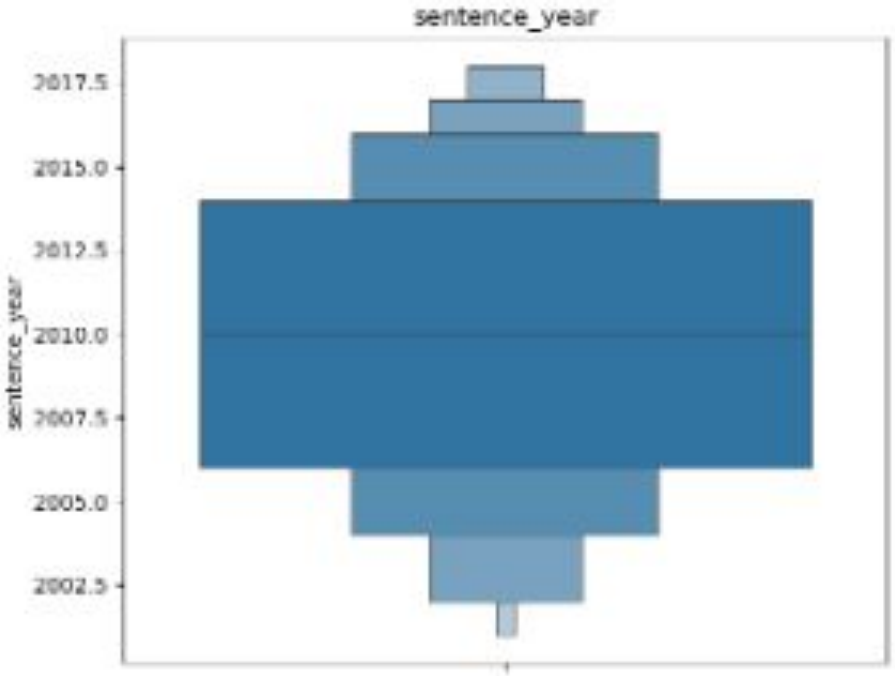

prisonSentence_months
1000
800
600
400
200
0
prisonSentence_months

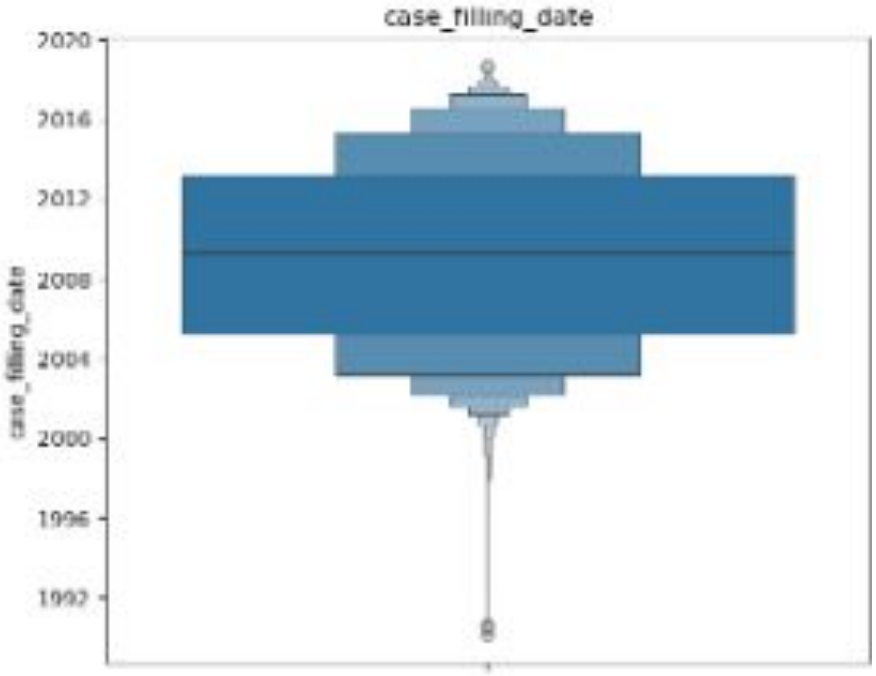

defendant_age
defendant_age
100
90
80
70
60
50
40
30
20

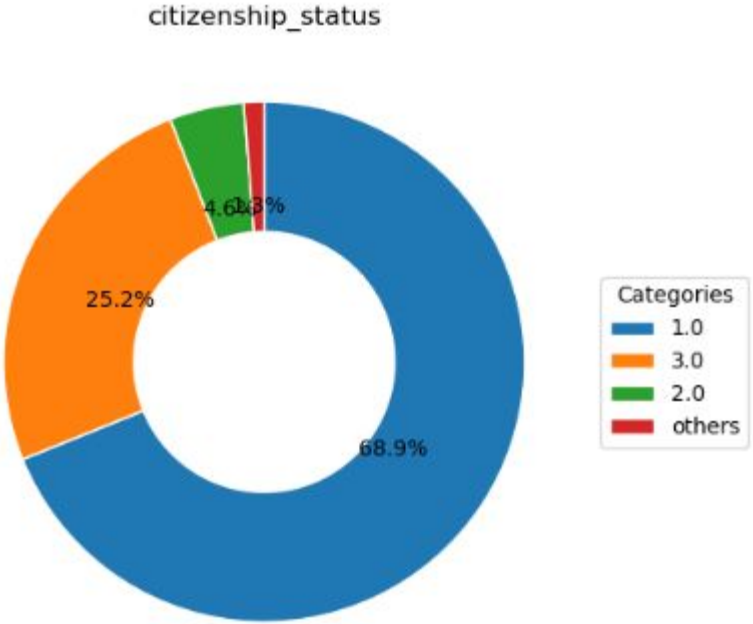

defendant_educationLevel
defendant_educationLevel
35
30
25
20
15
10
5
0

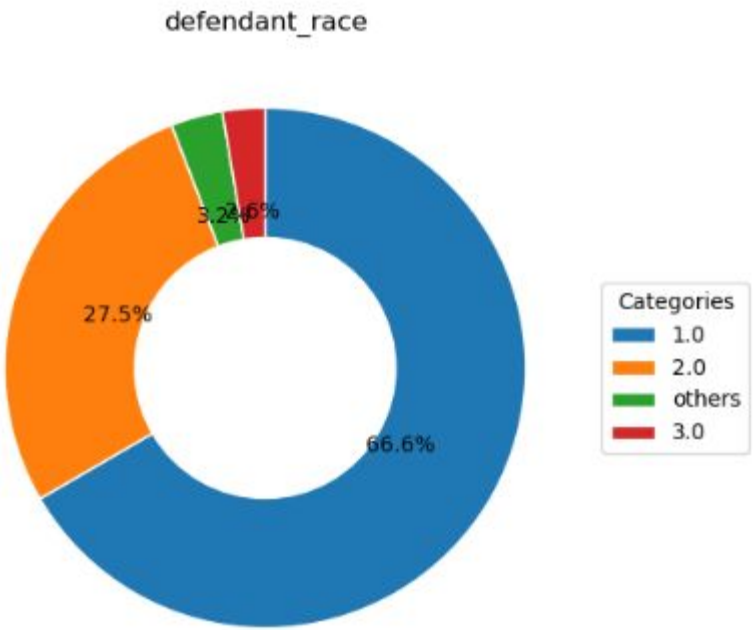

probationSentence_months
probationSentence_months
120
100
80
60
40
20
0

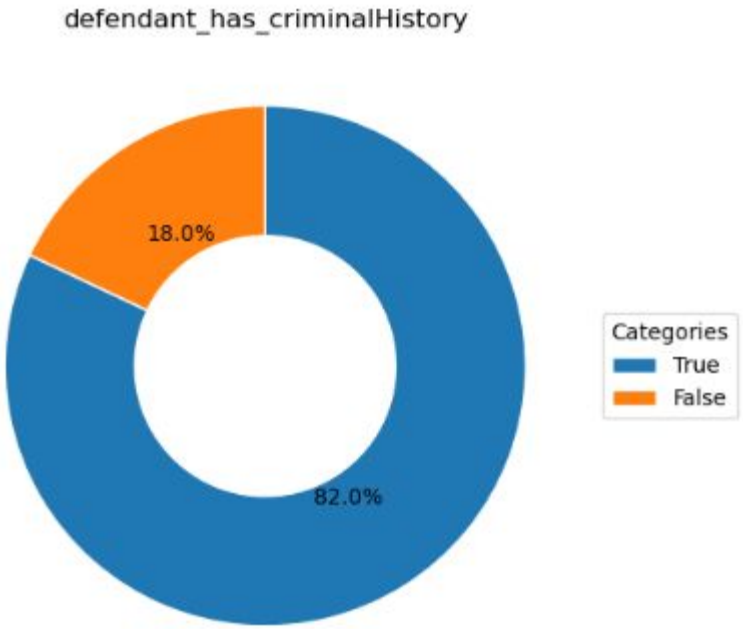

criminalHistoryPoints_applied
criminalHistoryPoints_applied
80
60
40
20
0

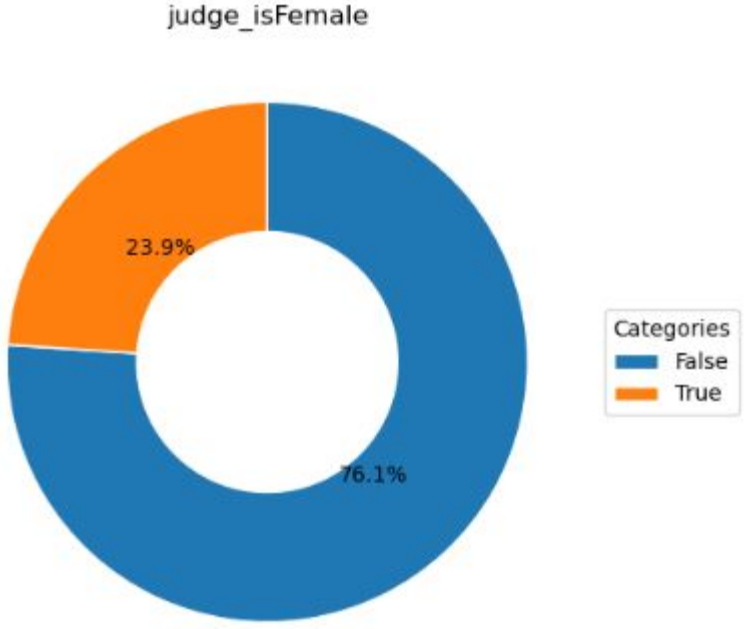

sentence_year
2017.5
2015.0
2012.5
2010.0
2007.5
2005.0
2002.5
sentence_year

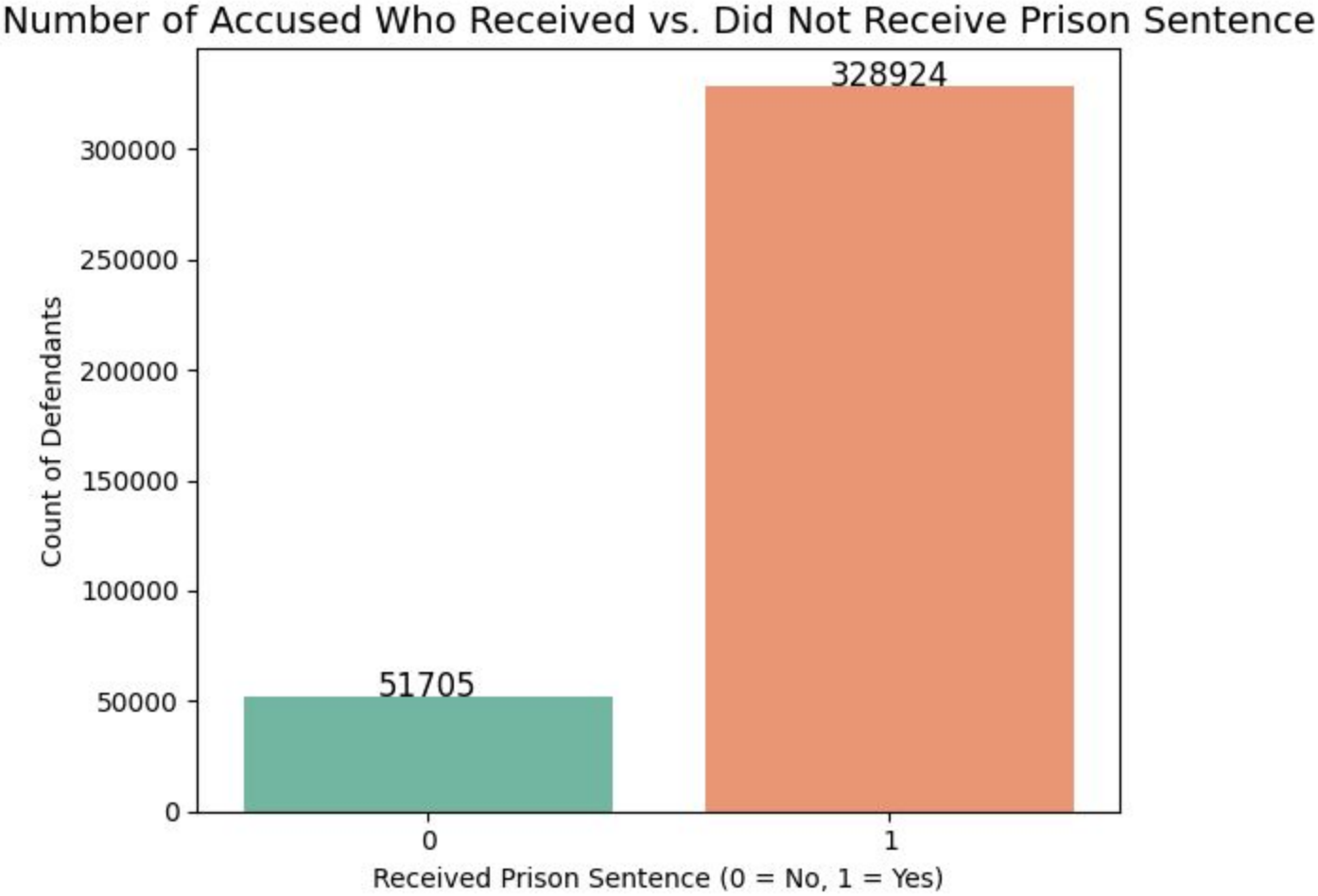

case_filing_date
2020
2016
2012
2008
2004
2000
1996
1992
case_filing_date


citizenship_status
25.2%
68.9%
Categories
1.0
3.0
2.0
others

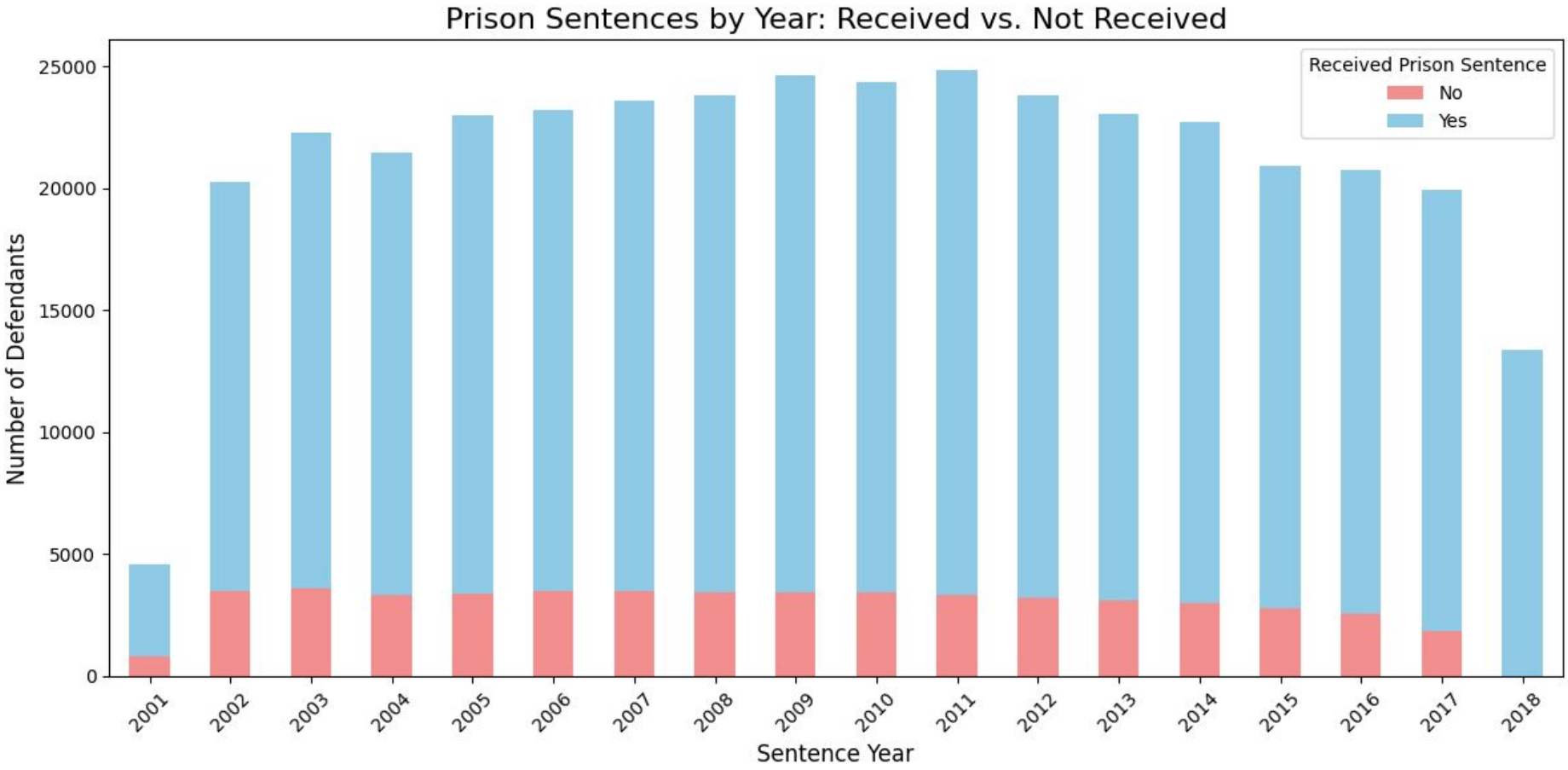

defendant_race
27.5%
66.6%
Categories
1.0
2.0
others
3.0


defendant_has_criminalHistory
18.0%
82.0%
Categories
True
False

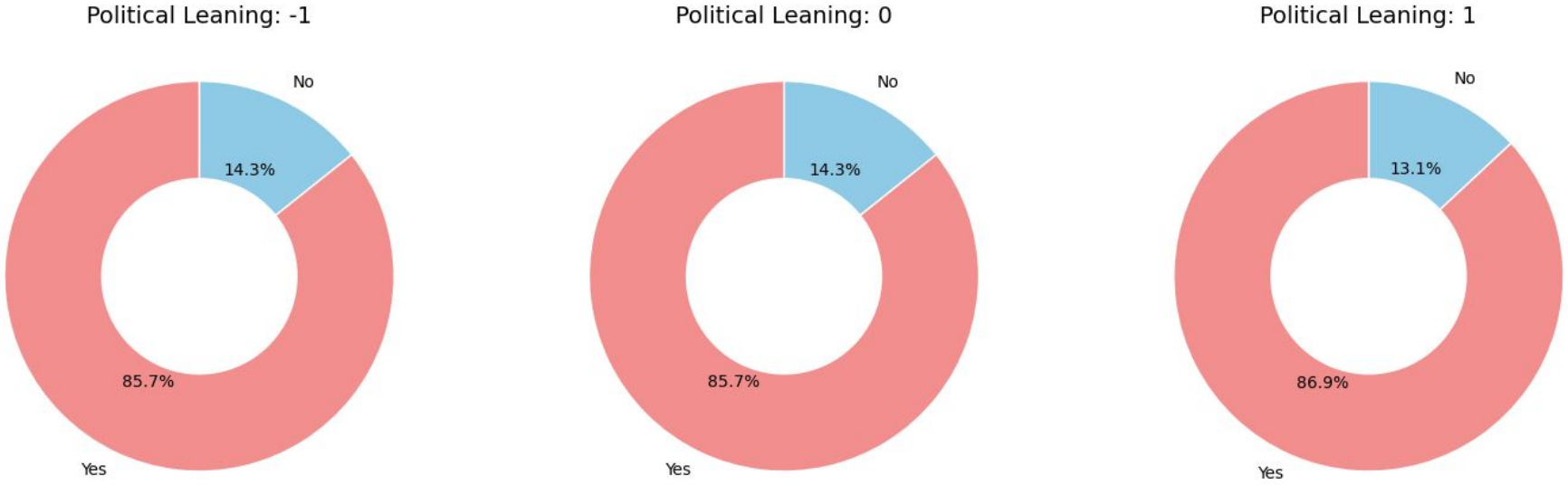

judge_isFemale
23.9%
76.1%
Categories
False
True

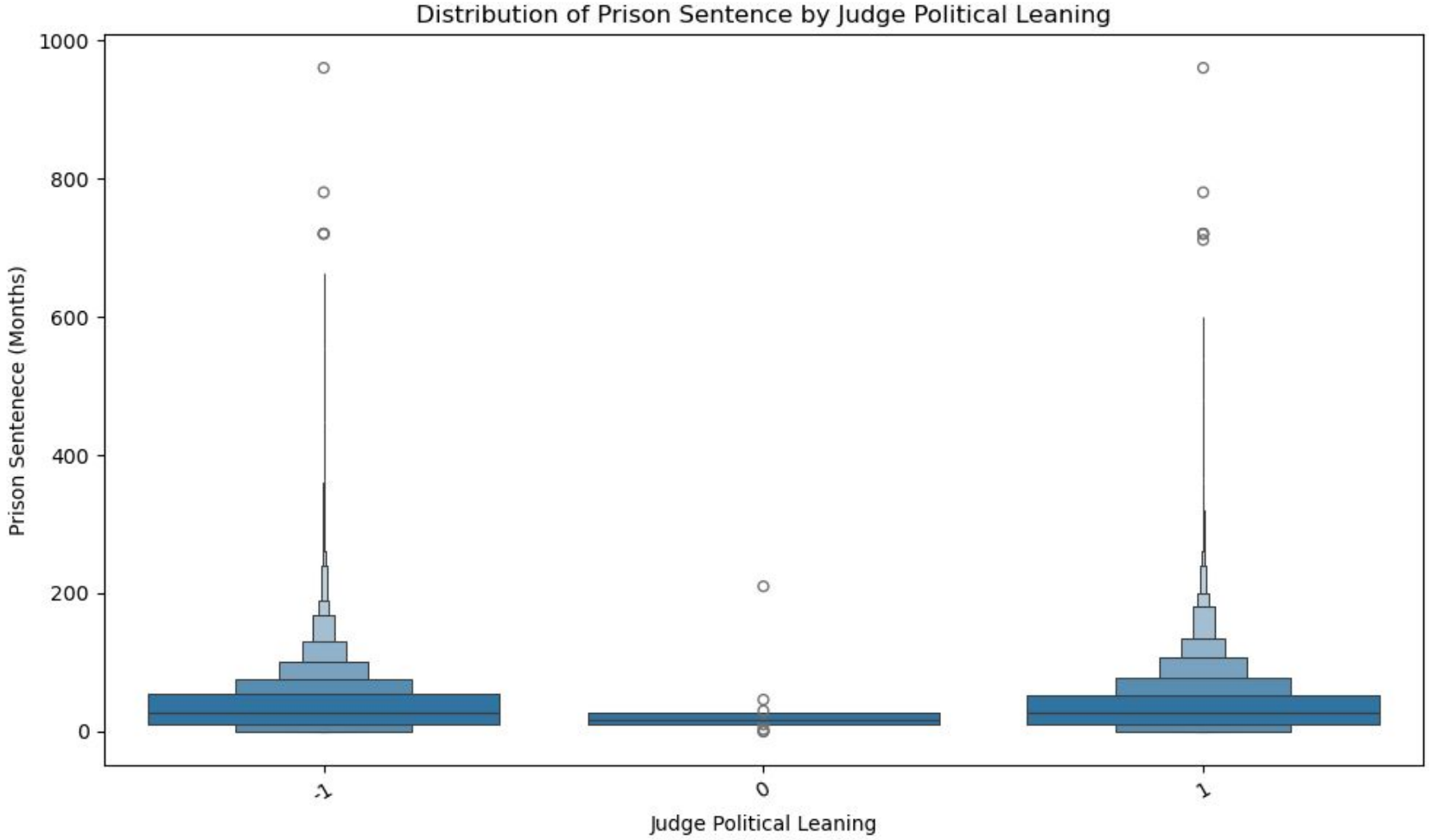

Number of Accused Who Received vs. Did Not Receive Prison Sentence
328924
51705
300000
250000
200000
150000
100000
50000
0
Count of Defendants
0
1
Received Prison Sentence (0 = No, 1 = Yes)

Is there temporality?

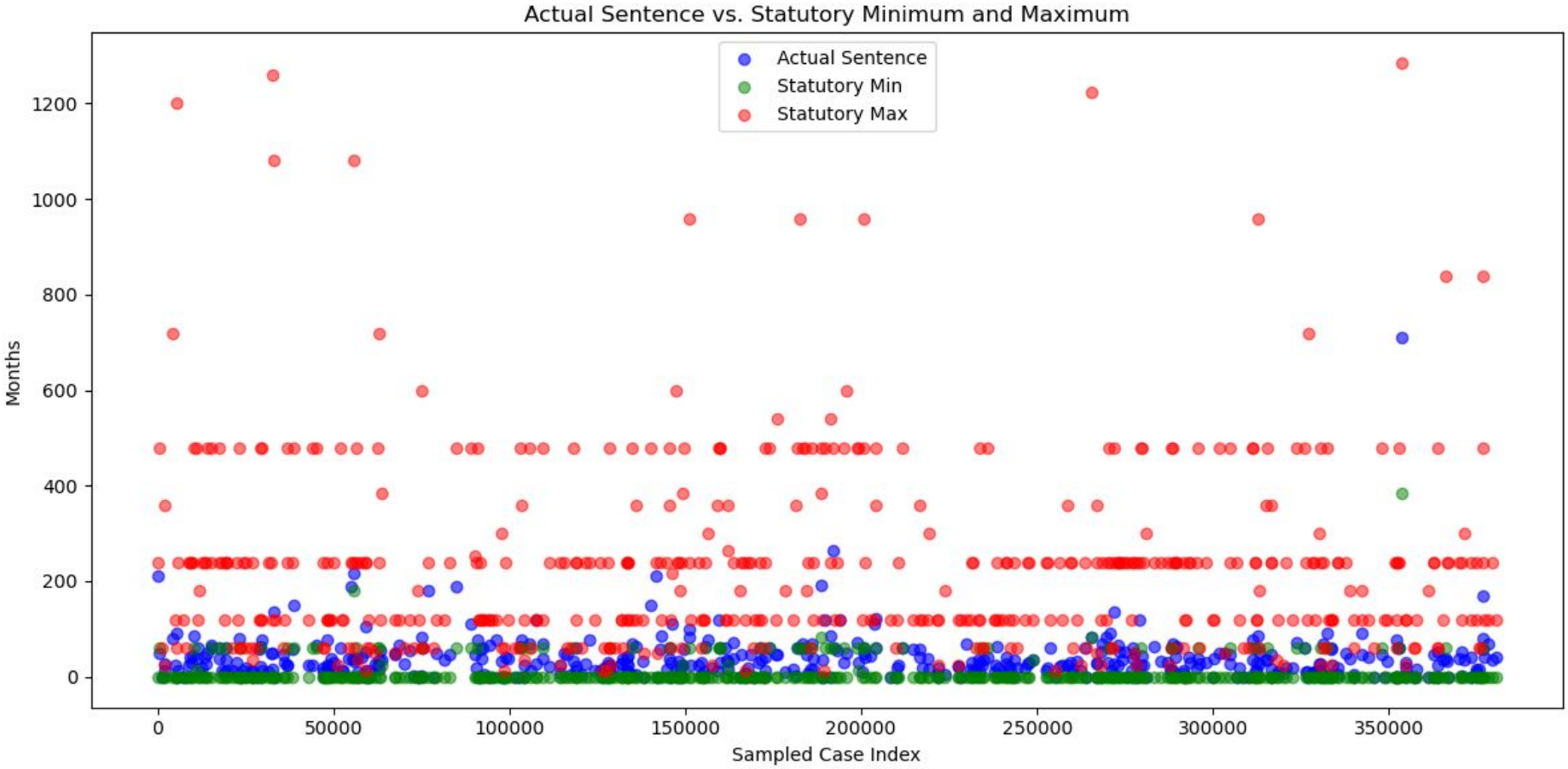

Prison Sentences by Year: Received vs. Not Received
Received Prison Sentence
No
Yes
Number of Defendants
0
5000
10000
15000
20000
25000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Sentence Year

# Does the political leaning of the judge matter?

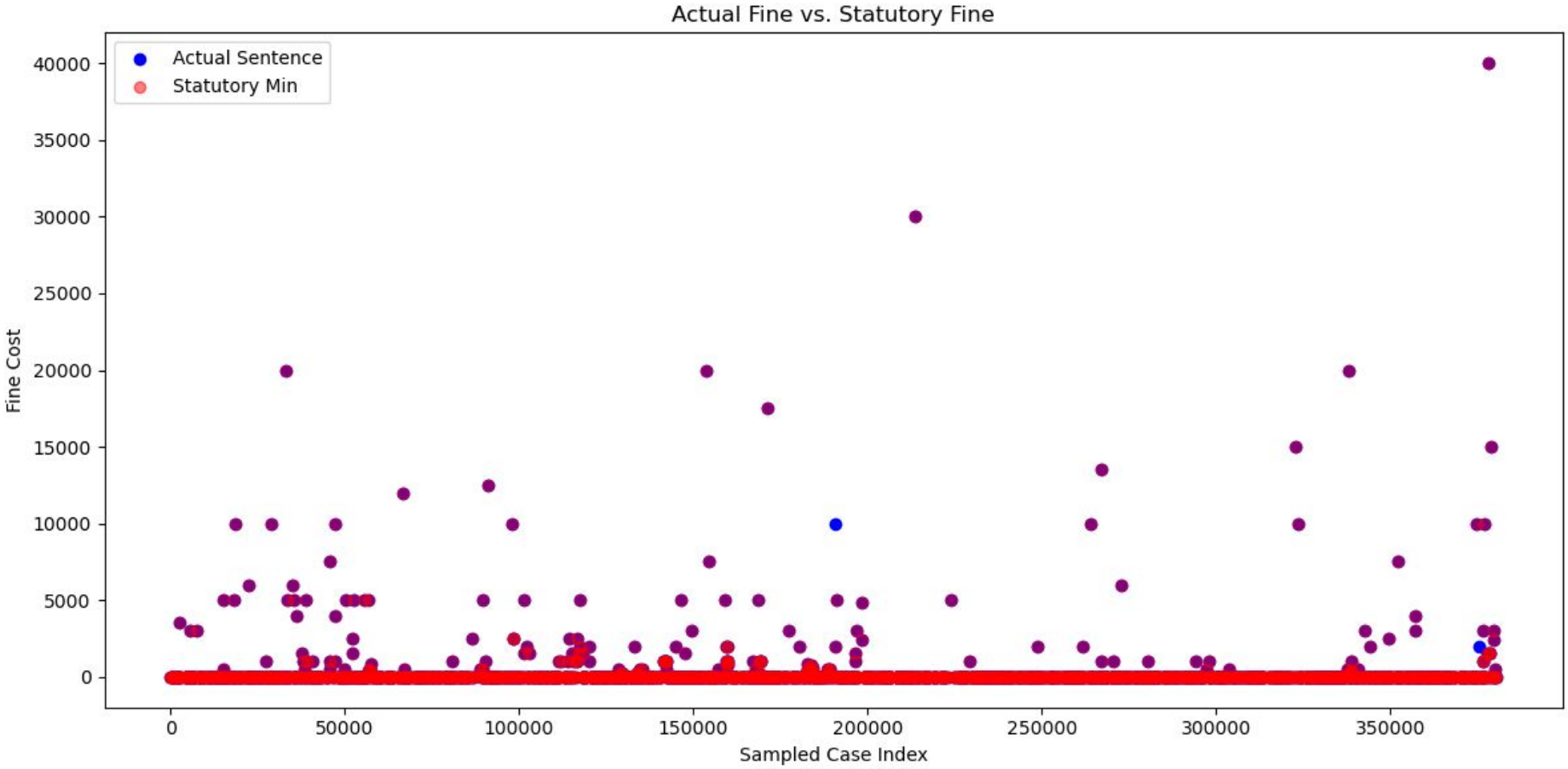

Received Prison Sentence by Judge Political Leaning
Political Leaning: -1
No
14.3%
85.7%
Yes
Political Leaning: 0
No
14.3%
85.7%
Yes
Political Leaning: 1
No
13.1%
86.9%
Yes


Distribution of Prison Sentence by Judge Political Leaning
Prison Sentenece (Months)
1000
800
600
400
200
0
-1
0
1
Judge Political Leaning

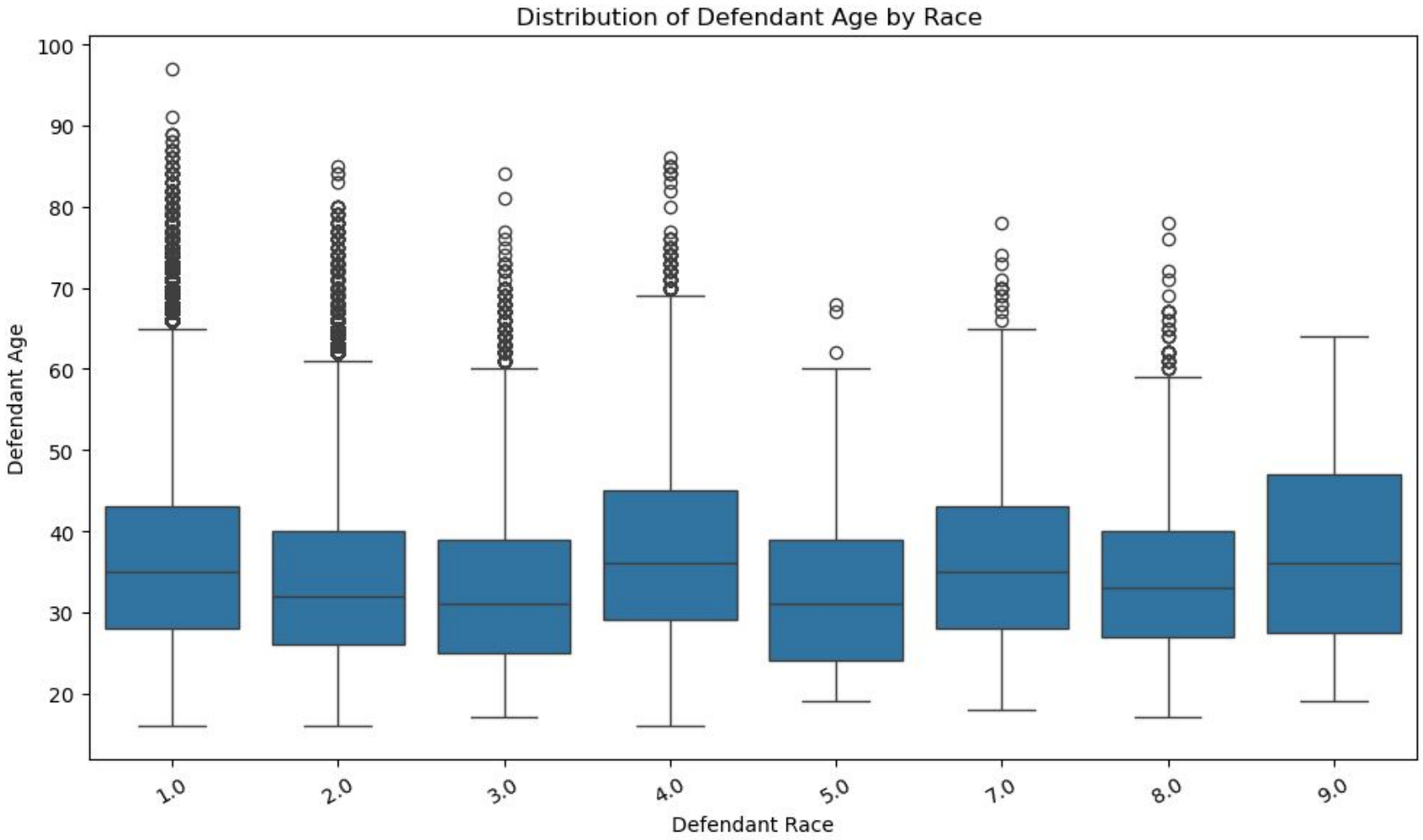
Are sentences on the upper or lower side of the spectrum? Are they Consistent?

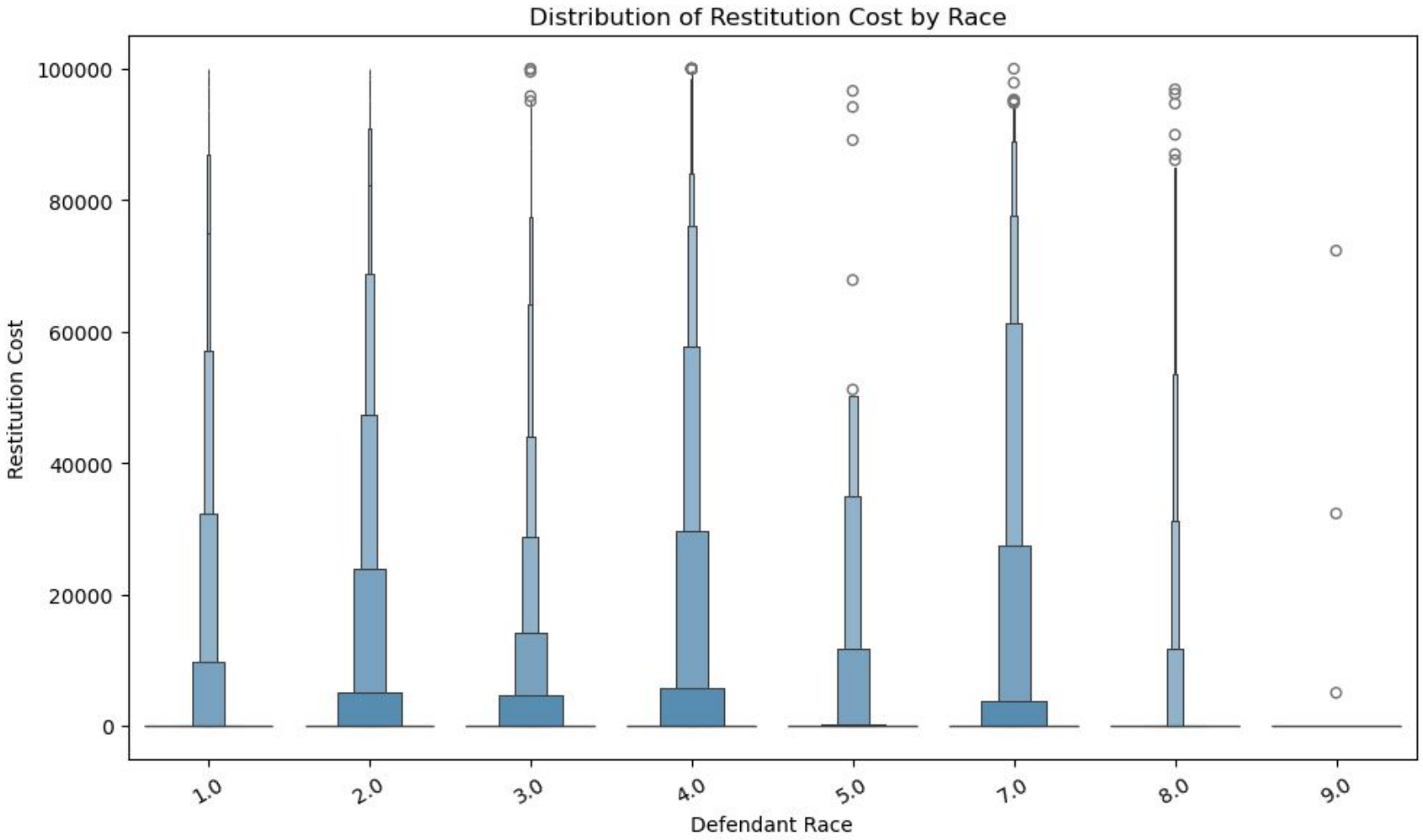

Actual Sentence vs. Statutory Minimum and Maximum
Actual Sentence
Statutory Min
Statutory Max
Months
0
200
400
600
800
1000
1200
0
50000
100000
150000
200000
250000
300000
350000
Sampled Case Index

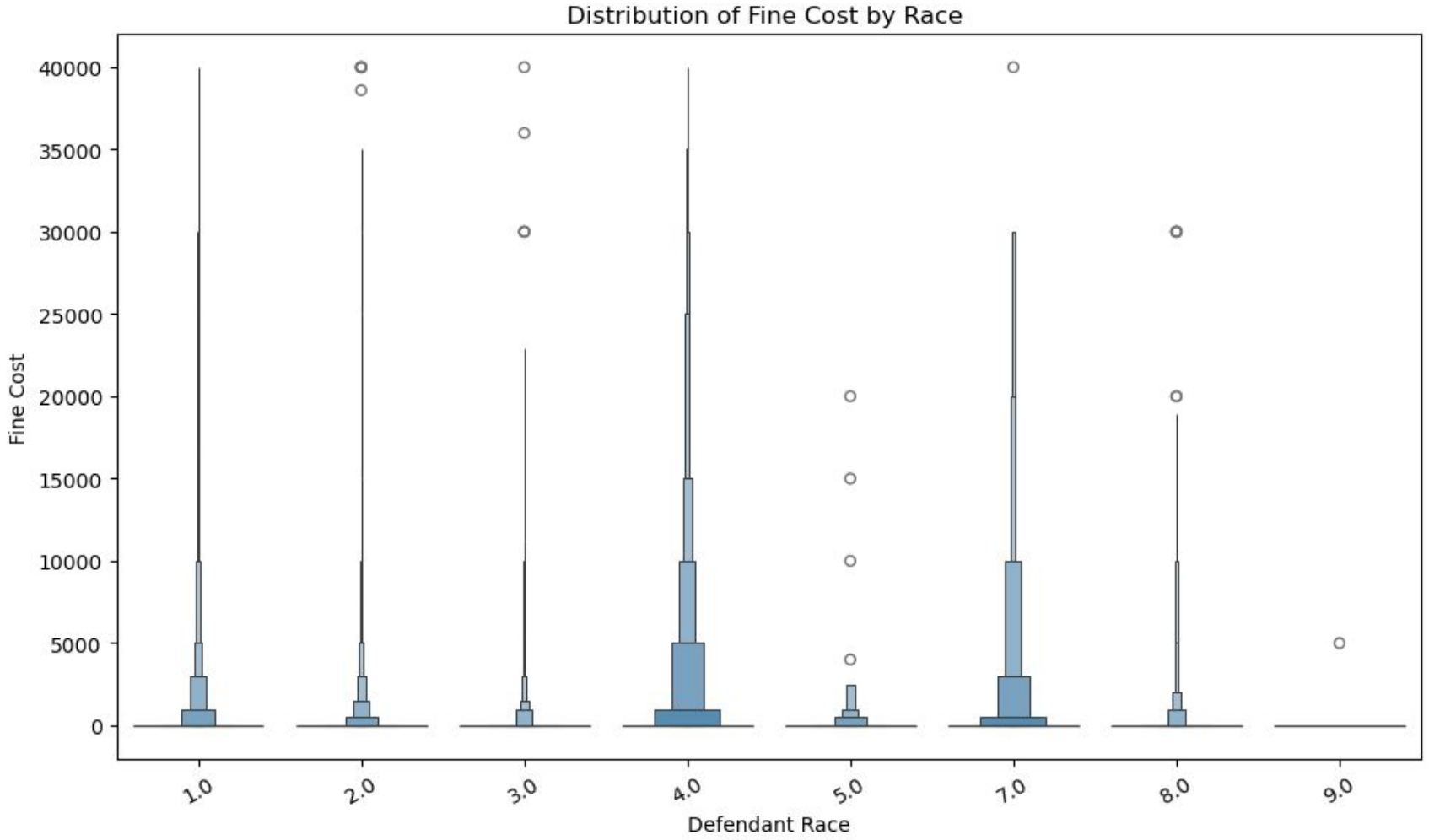
# What about fines?

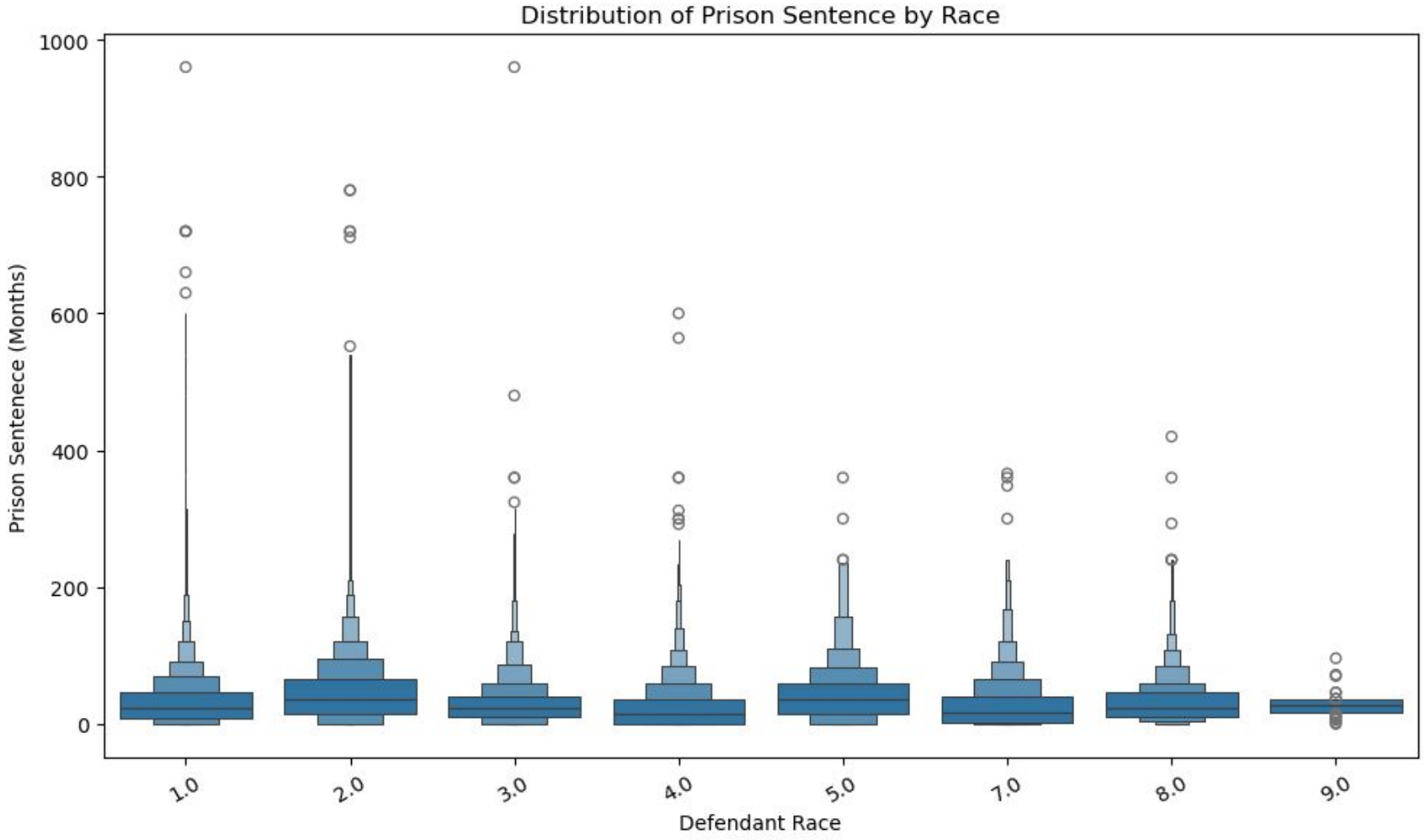

Actual Fine vs. Statutory Fine
Actual Sentence
Statutory Min
Fine Cost
40000
35000
30000
25000
20000
15000
10000
5000
0
0
50000
100000
150000
200000
250000
300000
350000
Sampled Case Index

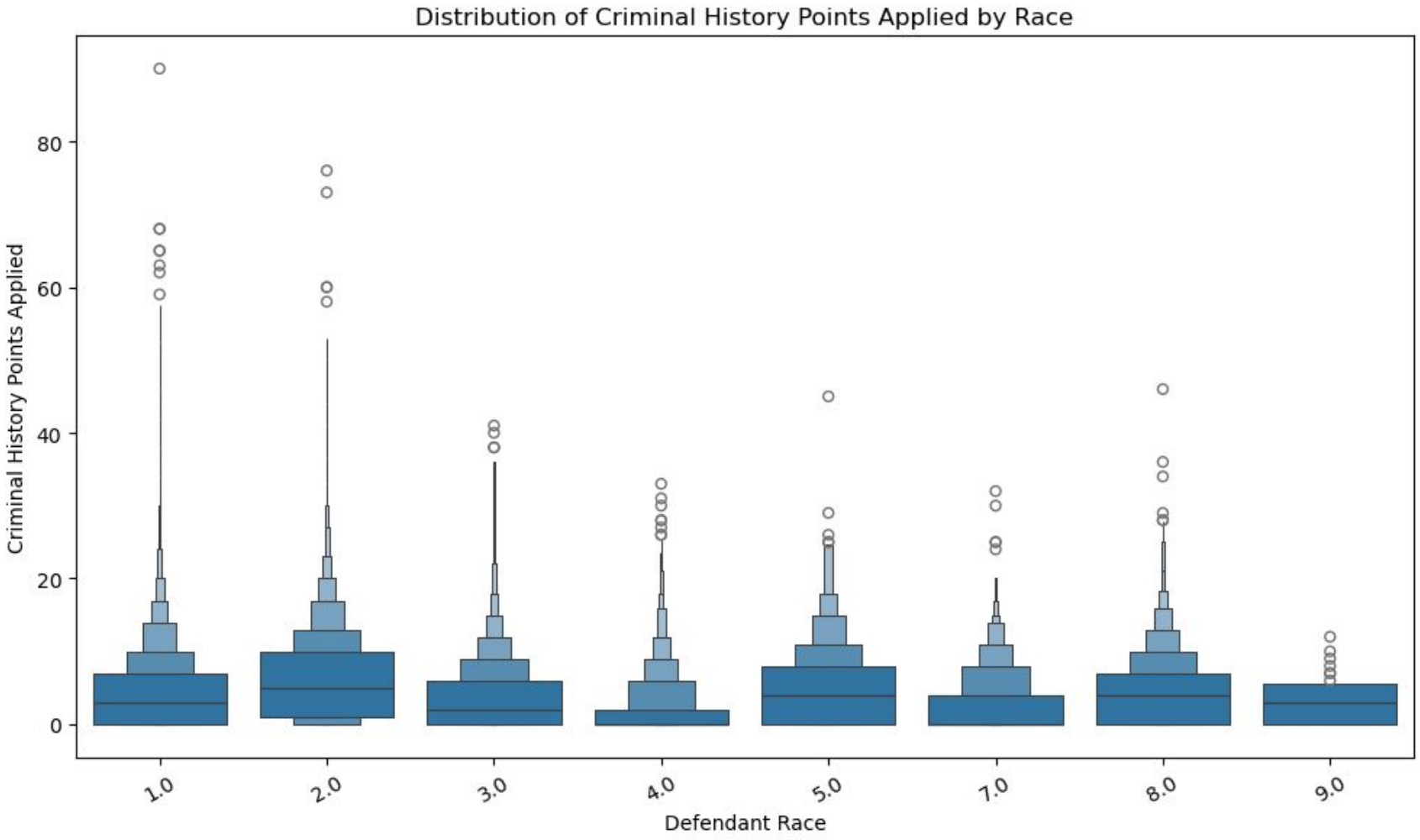
Does race have a noticeable impact in the involvement with the judiciary system?


Distribution of Defendant Age by Race
Defendant Age
100
90
80
70
60
50
40
30
20
1.0
2.0
3.0
4.0
5.0
7.0
8.0
9.0
Defendant Race


Distribution of Restitution Cost by Race
Restitution Cost
100000
80000
60000
40000
20000
0
1.0
2.0
3.0
4.0
5.0
7.0
8.0
9.0
Defendant Race


Distribution of Fine Cost by Race
Fine Cost
40000
35000
30000
25000
20000
15000
10000
5000
0
1.0
2.0
3.0
4.0
5.0
7.0
8.0
9.0
Defendant Race


Distribution of Prison Sentence by Race
Prison Sentenece (Months)
1000
800
600
400
200
0
1.0
2.0
3.0
4.0
5.0
7.0
8.0
9.0
Defendant Race


Distribution of Criminal History Points Applied by Race
Criminal History Points Applied
0
20
40
60
80
1.0
2.0
3.0
4.0
5.0
7.0
8.0
9.0
Defendant Race

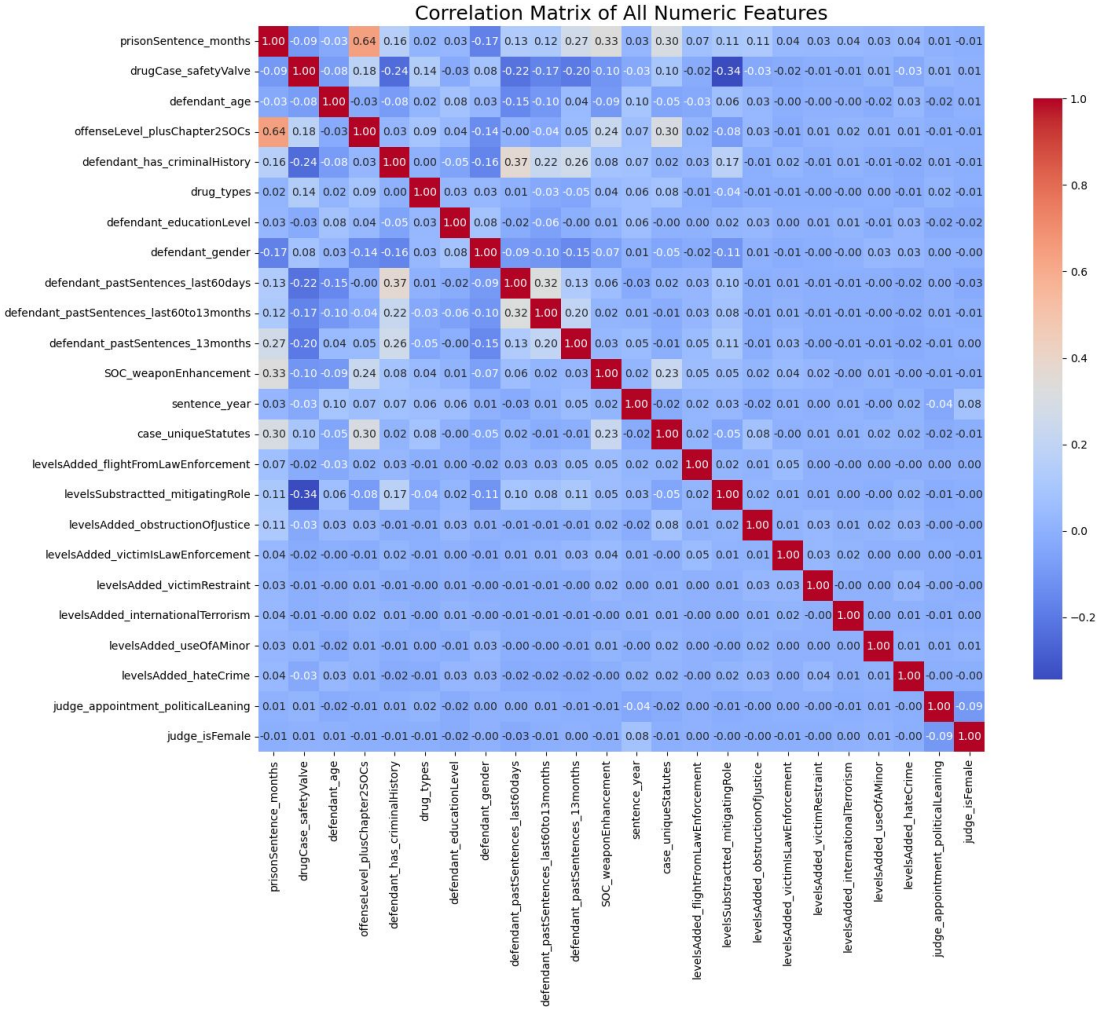
# Prediction problem

What are the factors that predict how long will a sentence be? Are they all lawful or are there also demographic ones?

Predicting sentence length with:

- Demographic data
- Aggravants
- Amount and severity of statutes
- Political ideology / leaning
- Drugs / weapons implications
- Previous and recent defendant sentences

# Feature selection


Correlation Matrix of All Numeric Features
prisonSentence_months
drugCase_safetyValve
defendant_age
offenseLevel_plusChapter2SOCs
defendant_has_criminalHistory
drug_types
defendant_educationLevel
defendant_gender
defendant_pastSentences_last60days
defendant_pastSentences_last60to13months
defendant_pastSentences_13months
SOC_weaponEnhancement
sentence_year
case_uniqueStatutes
levelsAdded_flightFromLawEnforcement
levelsSubstractted_mitigatingRole
levelsAdded_obstructionOfJustice
levelsAdded_victimIsLawEnforcement
levelsAdded_victimRestraint
levelsAdded_internationalTerrorism
levelsAdded_useOfAMinor
levelsAdded_hateCrime
judge_appointment_politicalLeaning
judge_isFemale
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

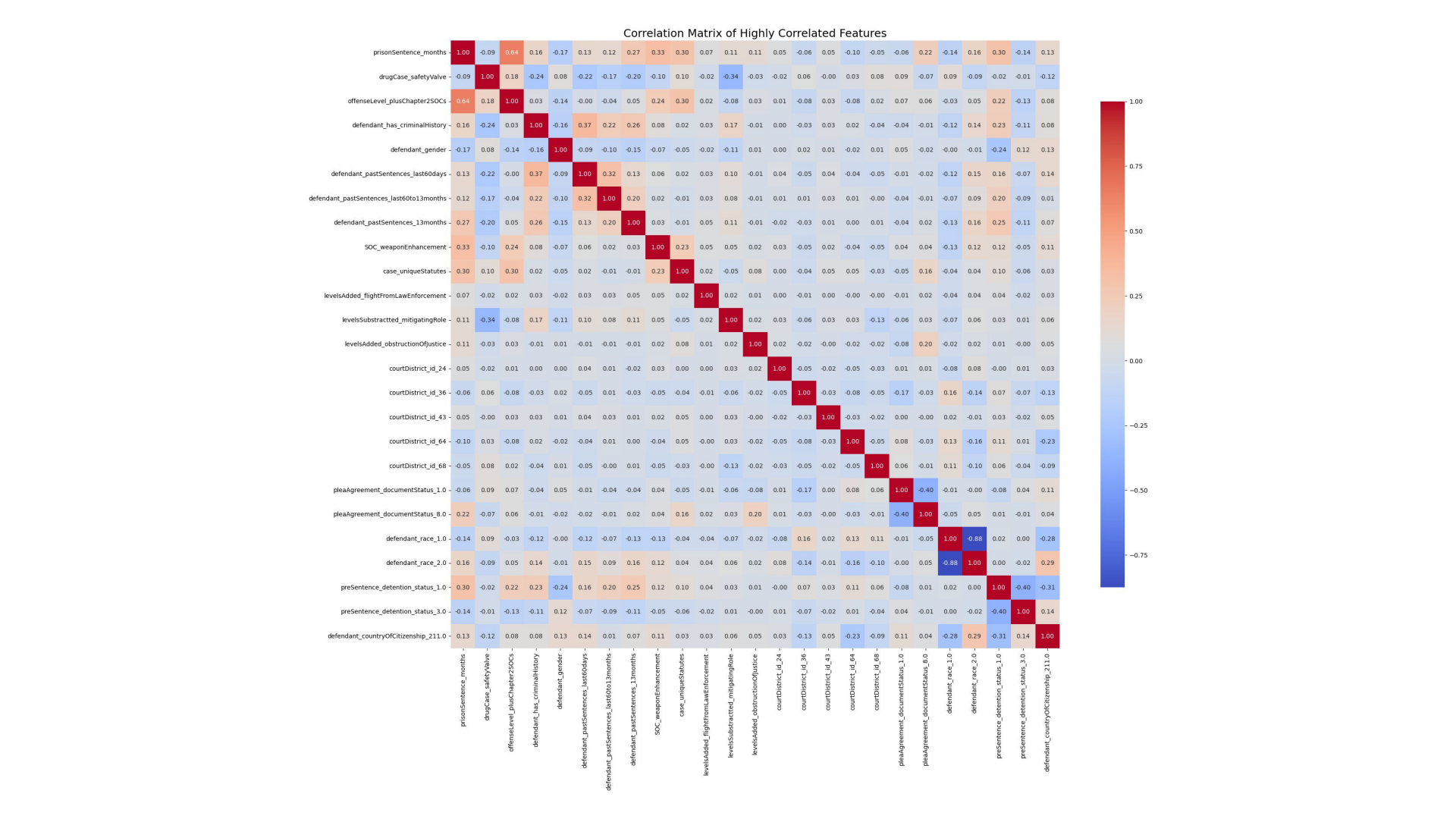
```
Correlation between prisonSentence_months and courtDistrict_id_24: 0.0517
Correlation between prisonSentence_months and courtDistrict_id_36: -0.0627
Correlation between prisonSentence_months and courtDistrict_id_43: 0.0503
c:\Users\ragal\.conda\envs\up_ml\Lib\site-packages\numpy\lib\function_base.py:2897: RuntimeWarning: invalid value encountered in divide
  c /= stddev[:, None]
c:\Users\ragal\.conda\envs\up_ml\Lib\site-packages\numpy\lib\function_base.py:2898: RuntimeWarning: invalid value encountered in divide
  c /= stddev[None, :]
Correlation between prisonSentence_months and courtDistrict_id_64: -0.0961
Correlation between prisonSentence_months and courtDistrict_id_68: -0.0532
Correlation between prisonSentence_months and pleaAgreement_documentStatus_1.0: -0.0625
Correlation between prisonSentence_months and pleaAgreement_documentStatus_8.0: 0.2220
Correlation between prisonSentence_months and defendant_race_1.0: -0.1364
Correlation between prisonSentence_months and defendant_race_2.0: 0.1632
Correlation between prisonSentence_months and preSentence_detention_status_1.0: 0.3041
Correlation between prisonSentence_months and preSentence_detention_status_2.0: -0.2404
Correlation between prisonSentence_months and preSentence_detention_status_3.0: -0.1430
Correlation between prisonSentence_months and defendant_countryOfCitizenship_49.0: -0.1140
Correlation between prisonSentence_months and defendant_countryOfCitizenship_211.0: 0.1318
```

**There appears to be sustantially larger prison sentences for people of two specific countries, two races and in a couple of specific courts and districts, among others.**

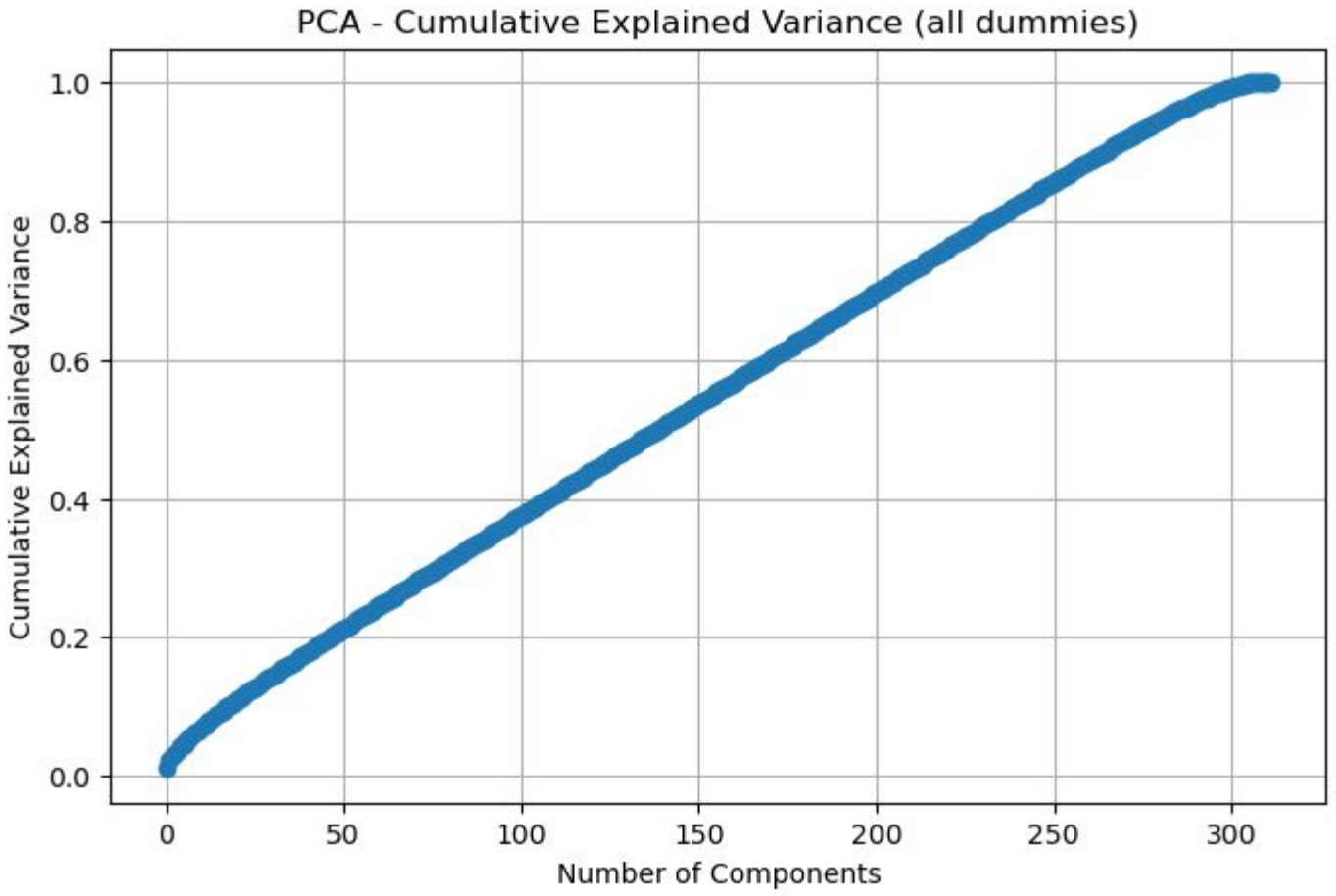
```
Correlation between prisonSentence_months and drugCase_safetyValve: -0.0930
Correlation between prisonSentence_months and defendant_has_criminalHistory: 0.1582
Correlation between prisonSentence_months and defendant_gender: -0.1748
Correlation between prisonSentence_months and SOC_weaponEnhancement: 0.3308
Correlation between prisonSentence_months and judge_isFemale: -0.0093
```

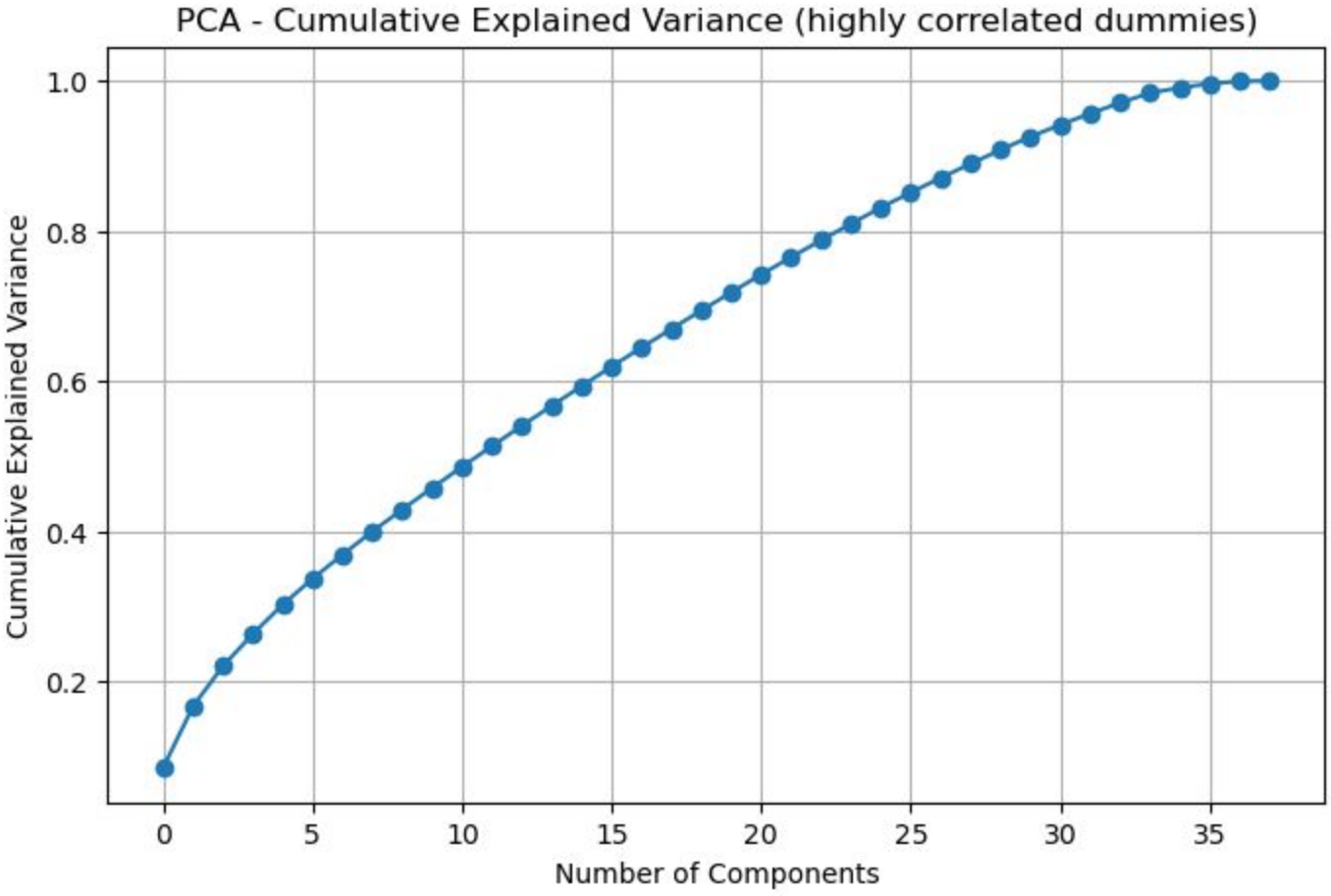
**Modified weapons seem to be very important when it comes to longer sentences.**

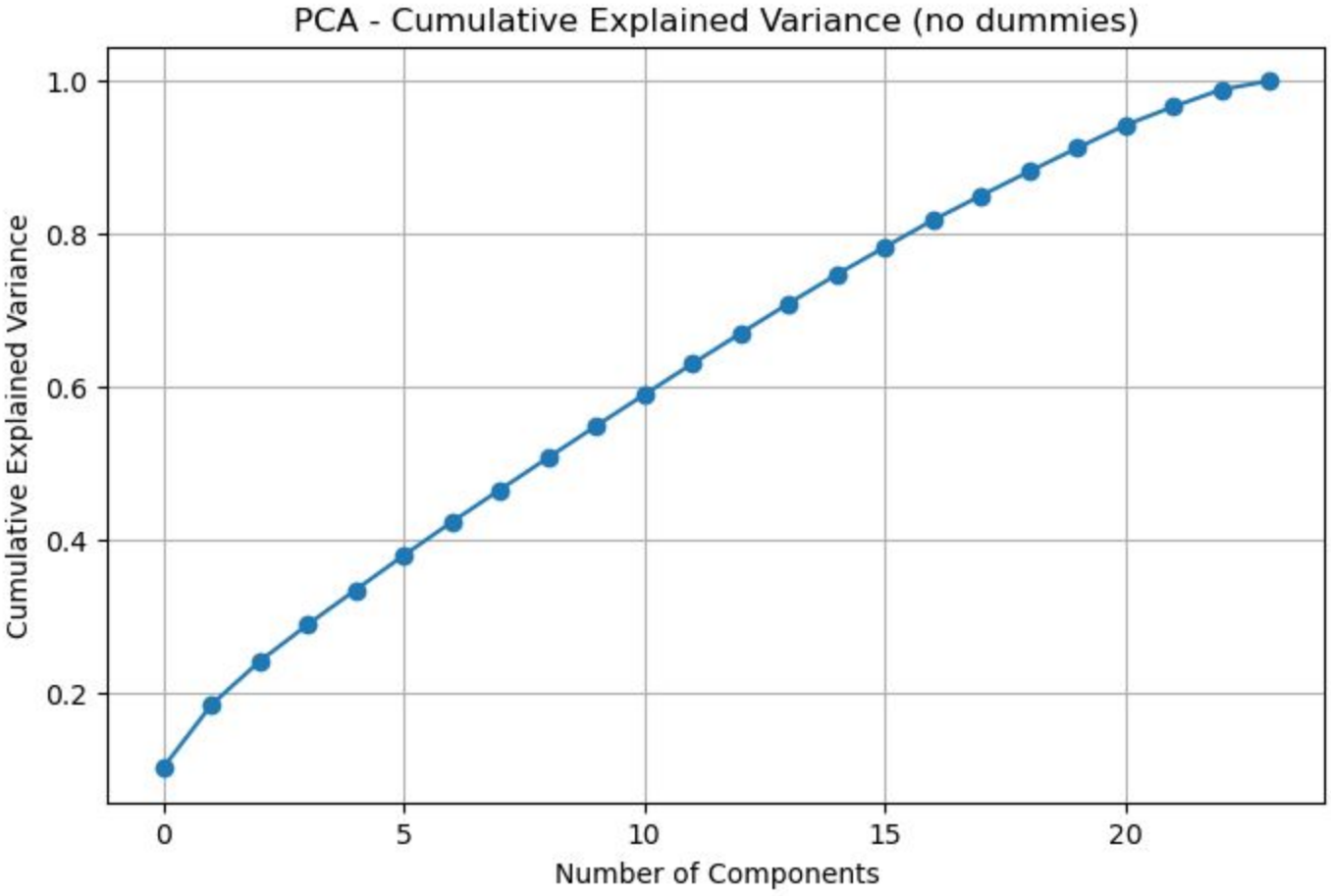
**Judge gender may be completely irrelevant, but the one of the defendant is definitely not.**


Correlation Matrix of Highly Correlated Features
prisonSentence_months
drugCase_safetyValve
offenseLevel_plusChapter2SOCs
defendant_has_criminalHistory
defendant_gender
defendant_pastSentences_last60days
defendant_pastSentences_last60to13months
defendant_pastSentences_13months
SOC_weaponEnhancement
case_uniqueStatutes
levelsAdded_flightFromLawEnforcement
levelsSubstractted_mitigatingRole
levelsAdded_obstructionOfJustice
courtDistrict_id_24
courtDistrict_id_36
courtDistrict_id_43
courtDistrict_id_64
courtDistrict_id_68
pleaAgreement_documentStatus_1.0
pleaAgreement_documentStatus_8.0
defendant_race_1.0
defendant_race_2.0
preSentence_detention_status_1.0
preSentence_detention_status_3.0
defendant_countryOfCitizenship_211.0
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

# PCA


PCA - Cumulative Explained Variance (all dummies)
Cumulative Explained Variance
Number of Components
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
250
300

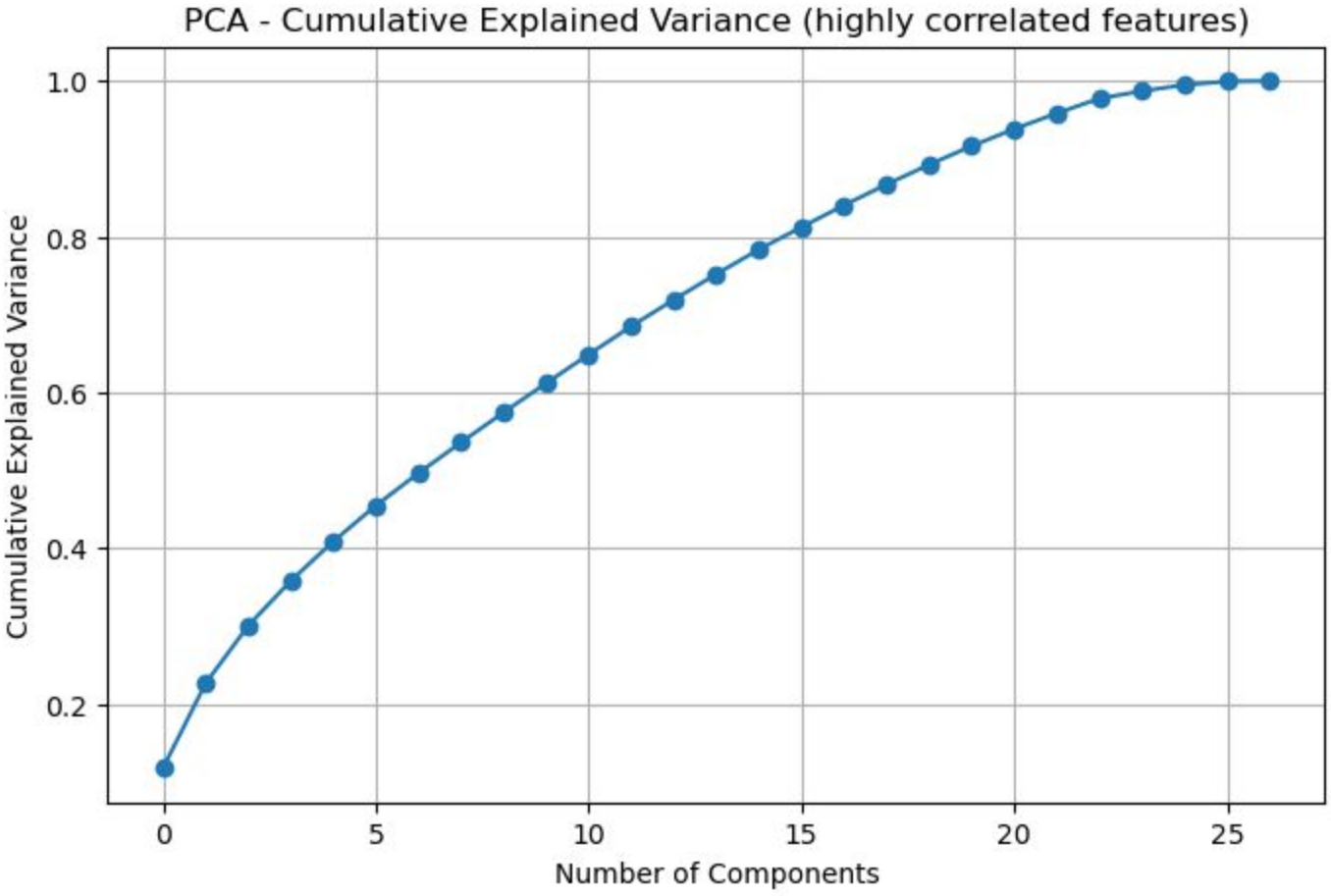

PCA - Cumulative Explained Variance (highly correlated dummies)
Cumulative Explained Variance
1.0
0.8
0.6
0.4
0.2
0
5
10
15
20
25
30
35
Number of Components


PCA - Cumulative Explained Variance (no dummies)
Cumulative Explained Variance
Number of Components
1.0
0.8
0.6
0.4
0.2
0
5
10
15
20

```
Top 10 weights for PC1:
prisonSentence_months: 0.4595
defendant_has_criminalHistory: 0.3595
defendant_pastSentences_last60days: 0.3149
defendant_pastSentences_13months: 0.3116
offenseLevel_plusChapter2SOCs: 0.2896
SOC_weaponEnhancement: 0.2804
defendant_pastSentences_last60to13months: 0.2725
defendant_gender: -0.2505
drugCase_safetyValve: -0.2460
levelsSubstractted_mitigatingRole: 0.2035
```

```
Top 10 weights for PC2:
offenseLevel_plusChapter2SOCs: 0.4911
drugCase_safetyValve: 0.3802
case_uniqueStatutes: 0.3777
prisonSentence_months: 0.3204
levelsSubstractted_mitigatingRole: -0.2576
defendant_pastSentences_last60to13months: -0.2485
defendant_has_criminalHistory: -0.2447
defendant_pastSentences_last60days: -0.2424
SOC_weaponEnhancement: 0.2287
drug_types: 0.1719
```

```
Top 10 weights for PC3:
defendant_age: 0.5384
defendant_educationLevel: 0.3383
levelsSubstractted_mitigatingRole: 0.3275
sentence_year: 0.3199
drugCase_safetyValve: -0.2866
defendant_pastSentences_last60days: -0.2848
defendant_pastSentences_last60to13months: -0.2408
levelsAdded_hateCrime: 0.1933
levelsAdded_obstructionOfJustice: 0.1767
judge_appointment_politicalLeaning: -0.1466
```


PCA - Cumulative Explained Variance (highly correlated features)
Cumulative Explained Variance
Number of Components
1.0
0.8
0.6
0.4
0.2
0
5
10
15
20
25

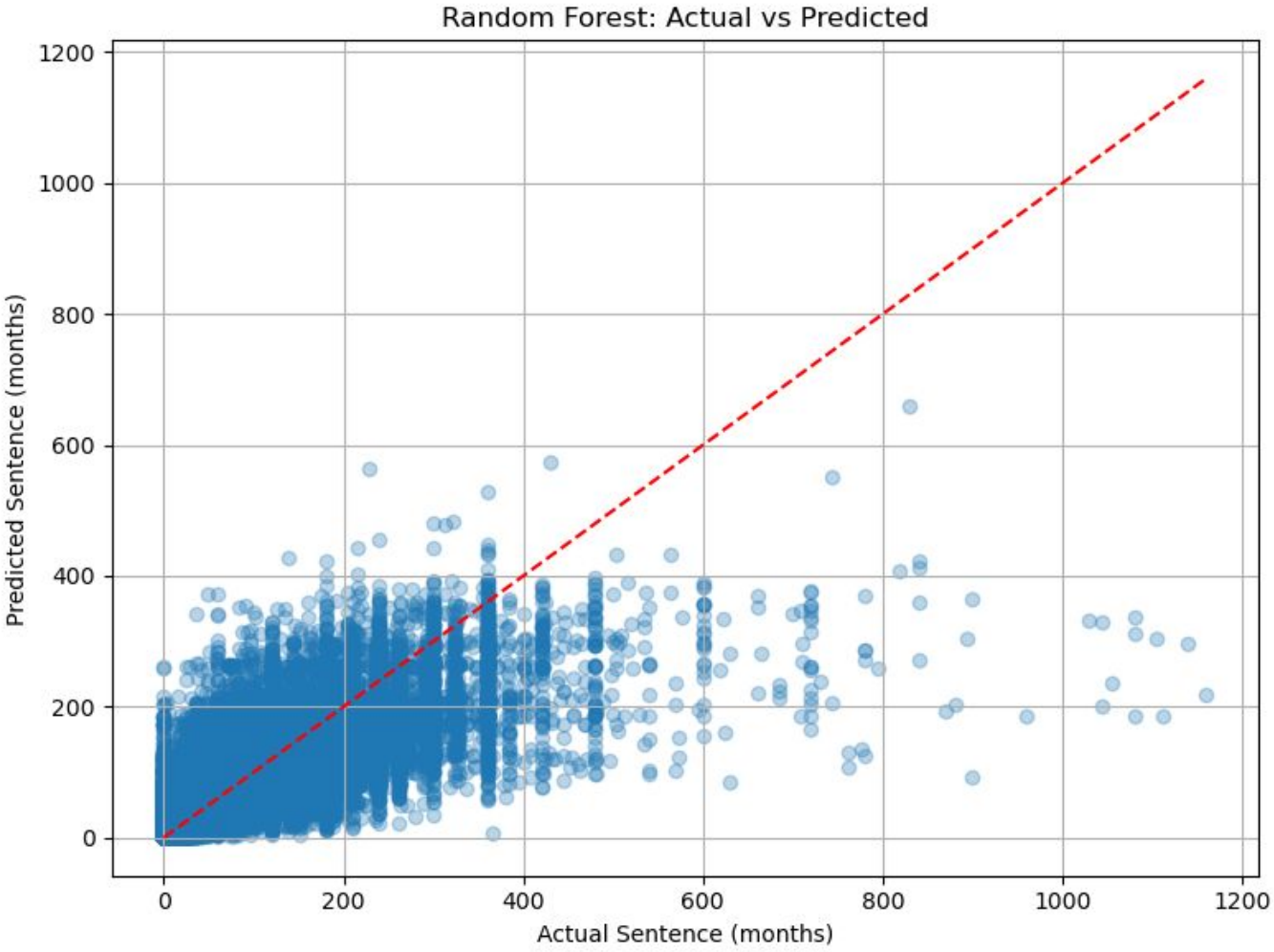
# Model Comparisons & Results

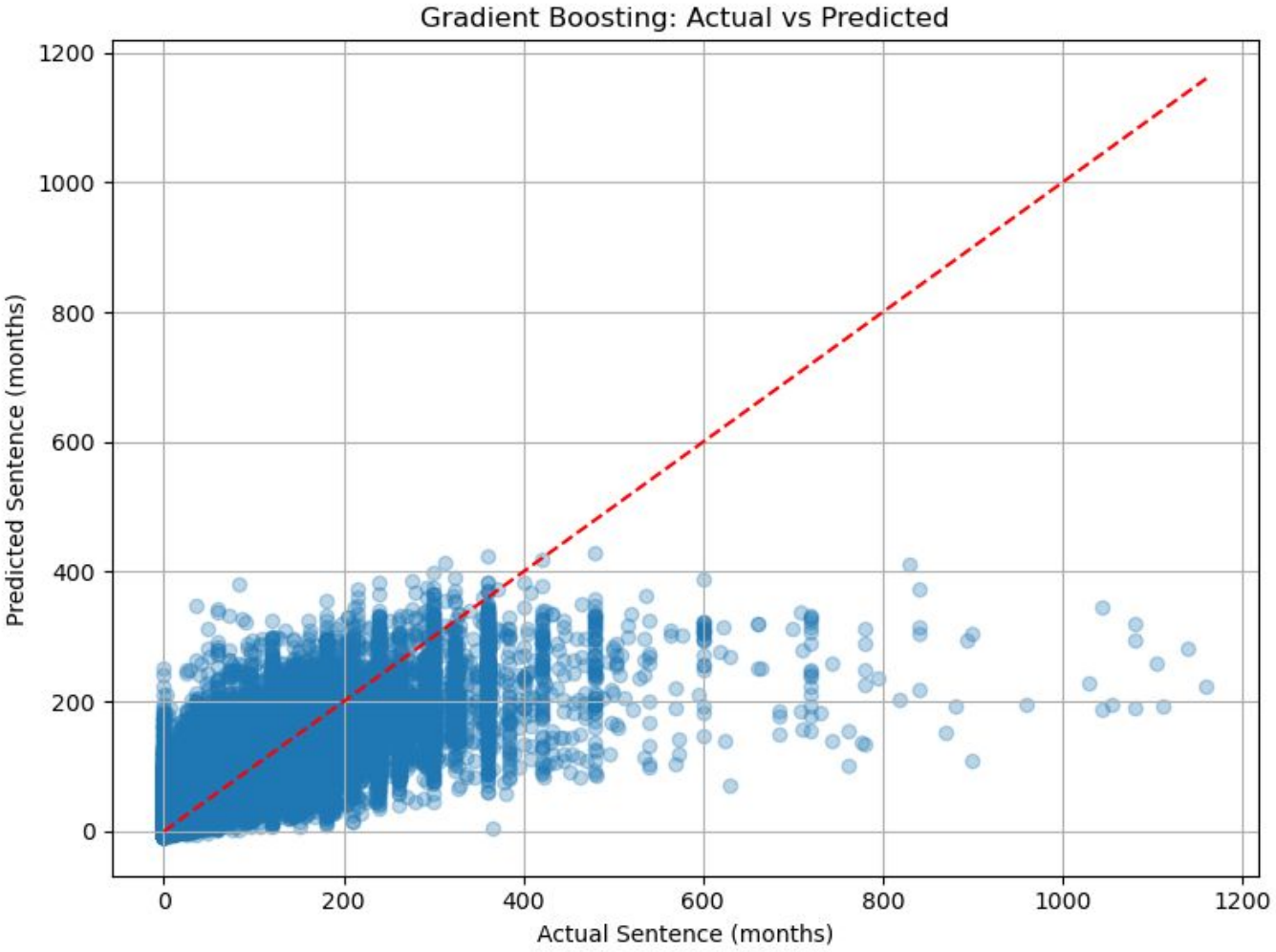
# Performance Metrics

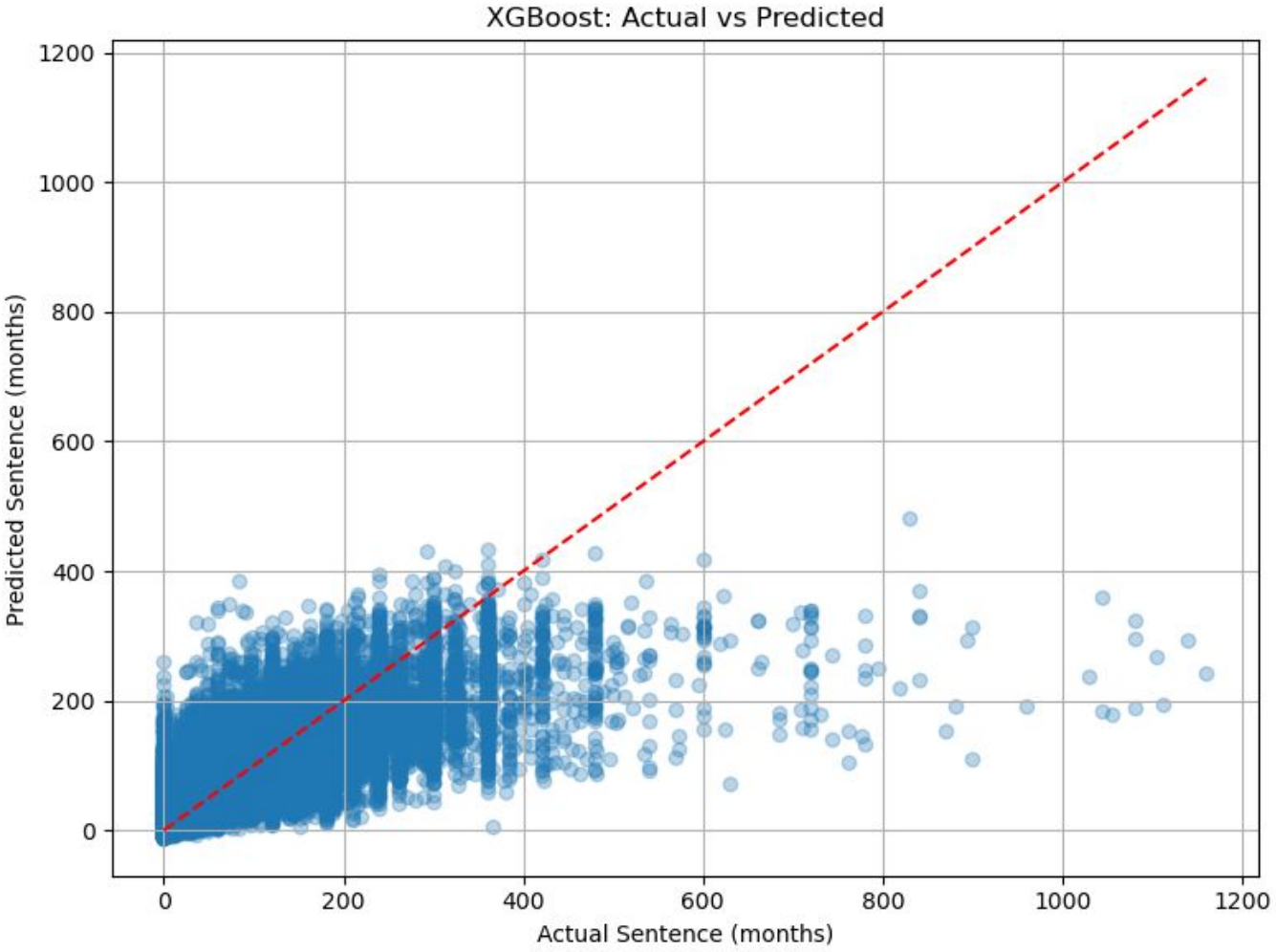
```
Random Forest Performance:
  RMSE: 35.63
  MAE : 19.19
  R²  : 0.691
```

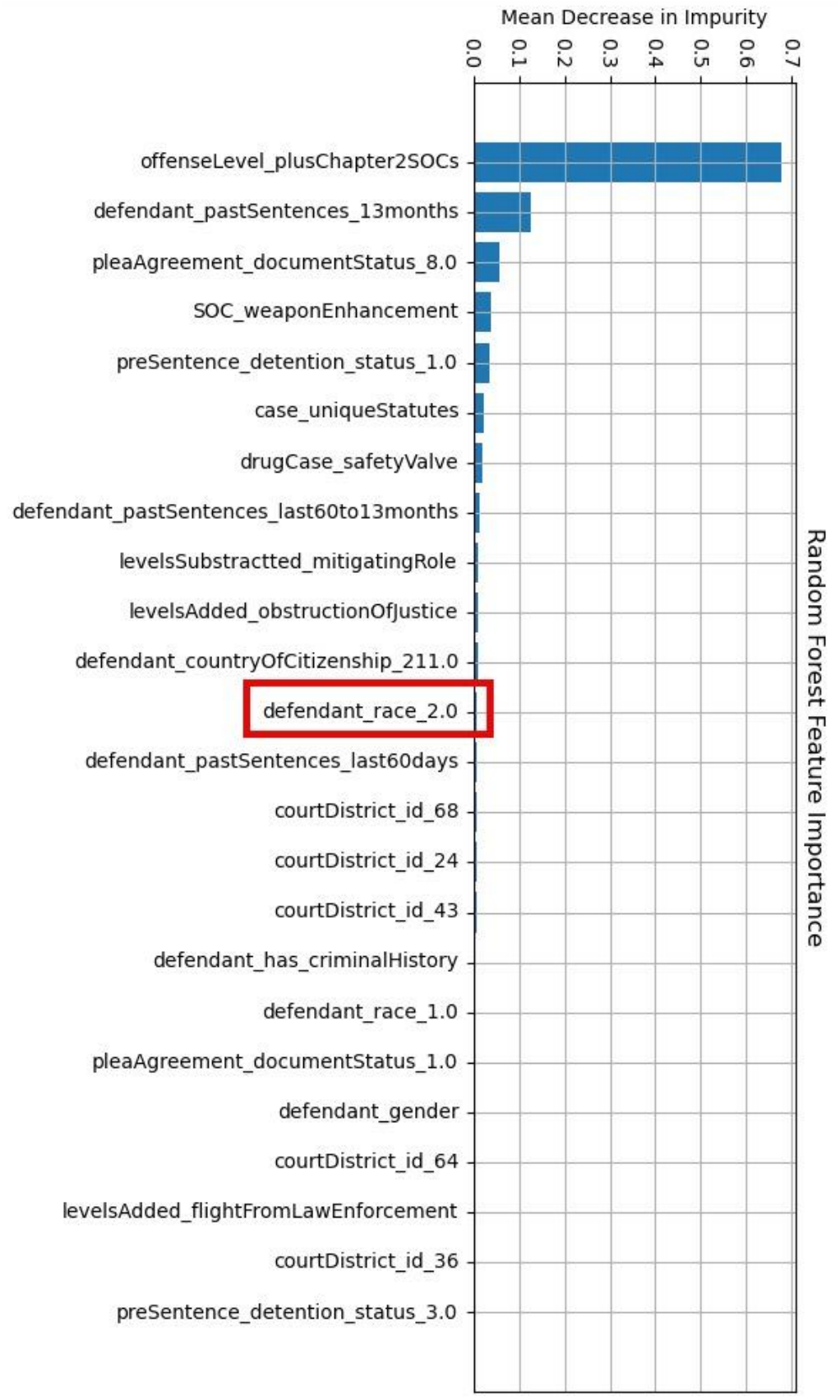
```
Gradient Boosting Performance:
  RMSE: 35.43
  MAE : 19.08
  R²  : 0.695
```

```
XGBoost Performance:
  RMSE: 35.41
  MAE : 19.17
  R²  : 0.695
```

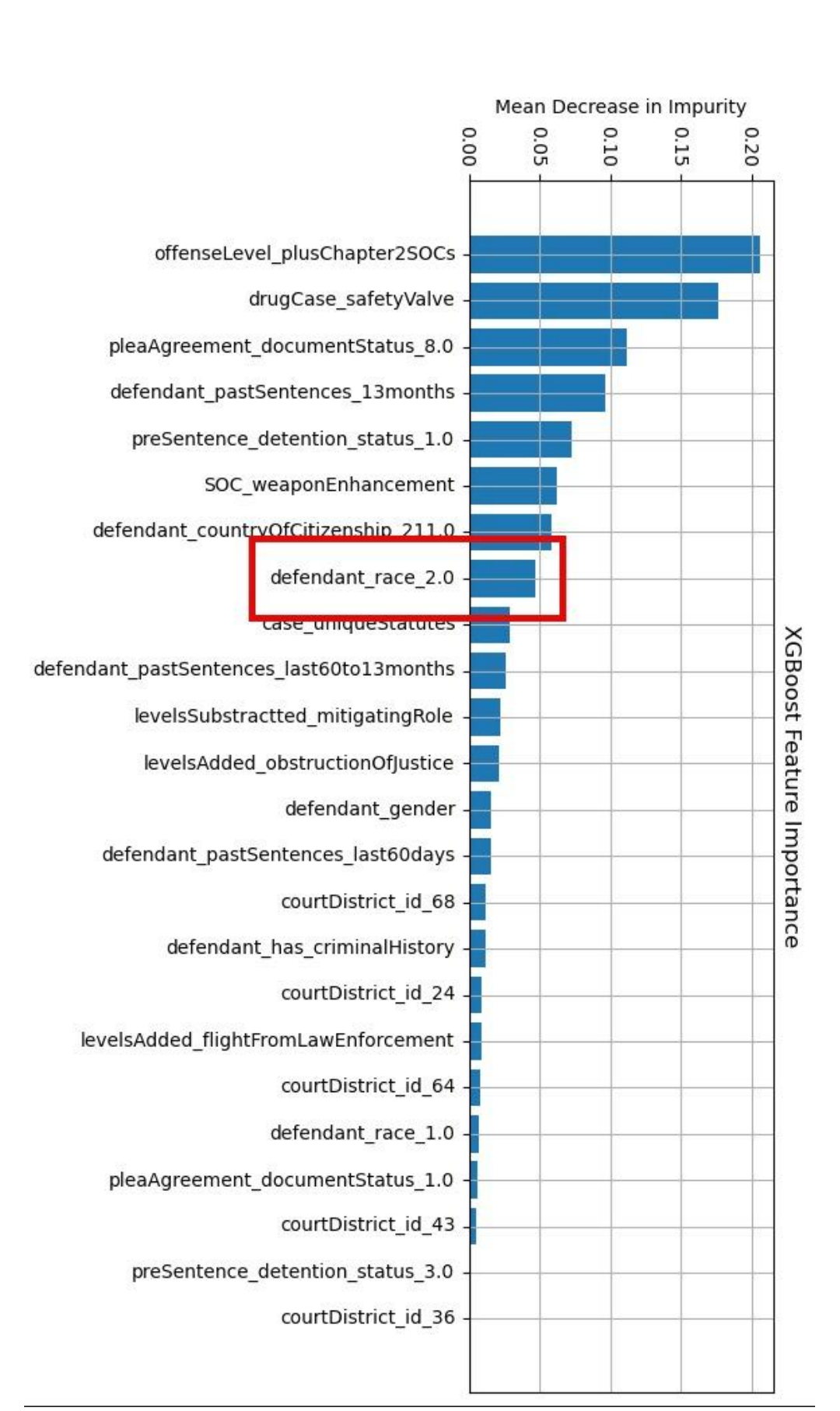

Random Forest: Actual vs Predicted
Predicted Sentence (months)
Actual Sentence (months)
0
200
400
600
800
1000
1200


Gradient Boosting: Actual vs Predicted
Predicted Sentence (months)
Actual Sentence (months)
0
200
400
600
800
1000
1200


XGBoost: Actual vs Predicted
Predicted Sentence (months)
Actual Sentence (months)
0
200
400
600
800
1000
1200


Mean Decrease in Impurity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
offenseLevel_plusChapter2SOCs
defendant_pastSentences_13months
pleaAgreement_documentStatus_8.0
SOC_weaponEnhancement
preSentence_detention_status_1.0
case_uniqueStatutes
drugCase_safetyValve
defendant_pastSentences_last60to13months
levelsSubstractted_mitigatingRole
levelsAdded_obstructionOfJustice
defendant_countryOfCitizenship_211.0
defendant_race_2.0
defendant_pastSentences_last60days
courtDistrict_id_68
courtDistrict_id_24
courtDistrict_id_43
defendant_has_criminalHistory
defendant_race_1.0
pleaAgreement_documentStatus_1.0
defendant_gender
courtDistrict_id_64
levelsAdded_flightFromLawEnforcement
courtDistrict_id_36
preSentence_detention_status_3.0
Random Forest Feature Importance

Mean Decrease in Impurity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
offenseLevel_plusChapter2SOCs
defendant_pastSentences_13months
pleaAgreement_documentStatus_8.0
SOC_weaponEnhancement
preSentence_detention_status_1.0
drugCase_safetyValve
case_uniqueStatutes
defendant_pastSentences_last60to13months
levelsSubstractted_mitigatingRole
defendant_race_2.0
levelsAdded_obstructionOfJustice
defendant_countryOfCitizenship_211.0
defendant_pastSentences_last60days
courtDistrict_id_68
defendant_gender
defendant_has_criminalHistory
courtDistrict_id_24
courtDistrict_id_64
levelsAdded_flightFromLawEnforcement
pleaAgreement_documentStatus_1.0
courtDistrict_id_43
defendant_race_1.0
preSentence_detention_status_3.0
courtDistrict_id_36
Gradient Boosting Feature Importance


Mean Decrease in Impurity
0.00
0.05
0.10
0.15
0.20
offenseLevel_plusChapter2SOCs
drugCase_safetyValve
pleaAgreement_documentStatus_8.0
defendant_pastSentences_13months
preSentence_detention_status_1.0
SOC_weaponEnhancement
defendant_countryOfCitizenship_211.0
defendant_race_2.0
defendant_pastSentences_last60to13months
levelsSubstractted_mitigatingRole
levelsAdded_obstructionOfJustice
defendant_gender
defendant_pastSentences_last60days
courtDistrict_id_68
defendant_has_criminalHistory
courtDistrict_id_24
levelsAdded_flightFromLawEnforcement
courtDistrict_id_64
defendant_race_1.0
pleaAgreement_documentStatus_1.0
courtDistrict_id_43
preSentence_detention_status_3.0
courtDistrict_id_36
XGBoost Feature Importance

# Conclusions