

Project Part 1 Report

Minh Ngo, Sri Sai Sushan Gumadavalli, Will Luo

Introduction

Recidivism measures the tendency of released individuals to reoffend after being released from prison. The purpose of this project is to see if there are patterns in recidivism rates within different demographics, such as between genders or age groups. We will do so using data from the New York State prison release records to answer the research questions listed below:

1. Is there a statistically significant difference in the mean age at release between offenders who return to prison within three years and those who do not in the 2008–2020 New York State release cohorts?
2. Is there a statistically significant difference in the proportions of male and female offenders who return to prison within three years of release in the 2008–2020 New York State release cohorts?

Data Summary

The data used comes from the administrative record compiled by the New York State Department of Corrections and Community Supervision (DOCCS). It includes every individual released from a New York State correctional facility between 2008 and 2020, containing a total of 287,139 observations. Each row corresponds to one release event, and the dataset records whether the individual returned to prison within three years of release due to a New Felony Offense or Parole Violation, or if they were Not Returned. The headers and the first two observations from the dataset (without modification) are shown below.

Table 1: First two Observations of the Recidivism Dataset

Release_Year	County_of_Indictment	Gender	Age_at_Release	Return_Status
2008	UNKNOWN	MALE	55	Not Returned
2008	ALBANY	MALE	16	Returned Parole Violation

For our research questions, this dataset is considered a population, as we are only generalizing and drawing conclusions to the scope of the dataset.

Modifications to the dataset are listed below.

1. The header variable names were cleaned to replace spaces with underscores for code compatibility.
2. The categorical variable Return_Status was recoded into a binary variable **Returned** where Not Returned = 0 and both Returned Parole Violation and New Felony Offense = 1 because we are trying to study recidivism rates, and the severity of the crime committed if one were to recidive is not considered.

There are two potential limitations of this dataset. Firstly, each release event is treated as an independent observation, which means that individuals who are released multiple times are going to appear more than once in the dataset. This will lead to overestimation of the recidivism rate since the people who frequently recidivate will have a larger influence on the statistic. Secondly, the return reasons present in the dataset are grouped into broad categories, which prevents us from doing analysis for specific type of offenses. Although these limitations exist and might influence our conclusion drawn for individuals, our dataset is still capable for our goal of examining overall demographic patterns from release events.

This dataset directly measures three-year recidivism outcomes and contains the demographic variables (age and gender) required to answer both research questions. Additionally, the data's wide range and large size allows us to draw broad and strong conclusions.

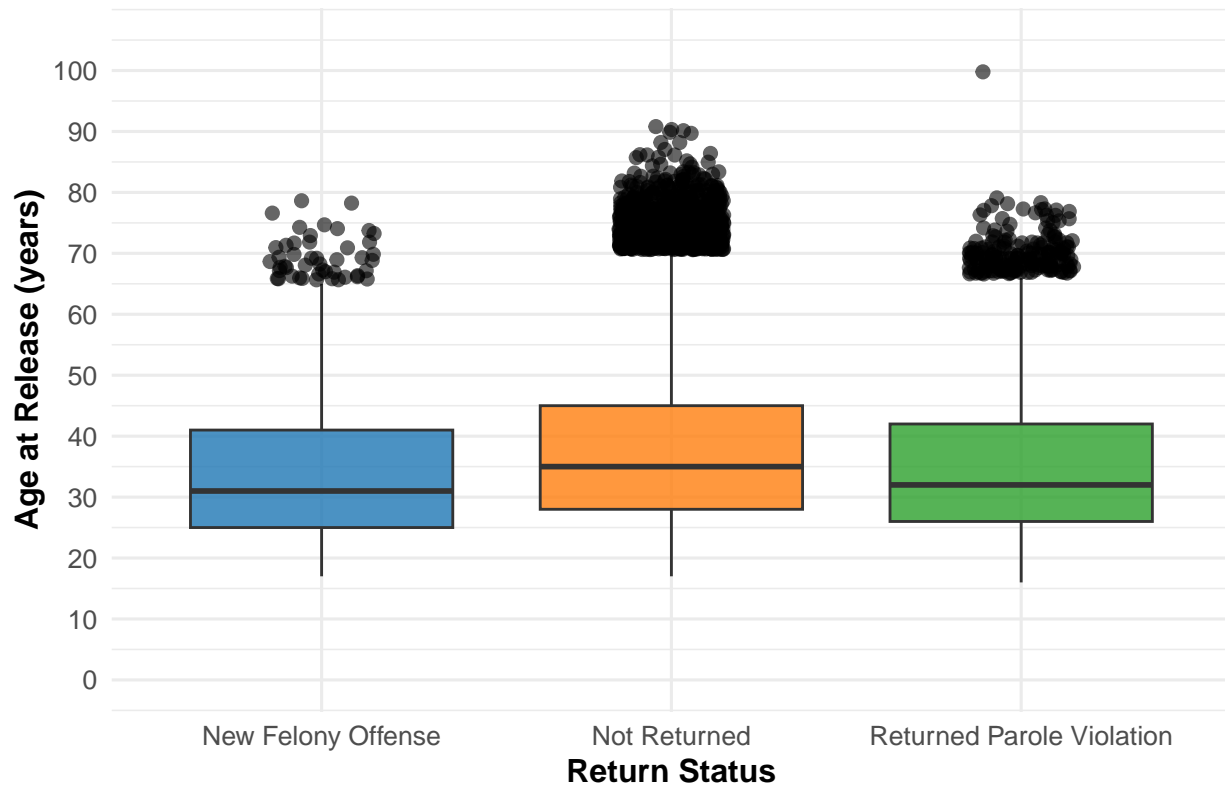
Question 1: Age vs Recidivism

Exploratory Analysis

Table 2: Summary of Age by Recidivism Status

Return Status	Count	Mean Age	Median Age	SD	Min	Max
New Felony Offense	24160	33.34	31	10.35	17	79
Not Returned	177424	37.06	35	11.45	17	91
Returned Parole Violation	85555	34.45	32	10.49	16	100

Figure 1. Distribution of Age at Release by Recidivism Status



Conclusions

From the summary table and boxplot, individuals who were not returned to prison tend to be slightly older on average than those who reoffended. The median age of offenders who were not returned is 35 years, compared to 31 years for those who committed a new felony offense

and 32 years for those returned for a parole violation. The interquartile ranges and whiskers also show that most offenders, regardless of return status, were between their mid-20s and early 40s, though a few older individuals extend into their 60s and 70s. The boxplot also reveals an outlier, which was a 100 year old person who returned due to a parole violation. Statistically, this pattern potentially indicates that offenders who recidivate are, on average, several years younger than those who successfully remain in the community.

This trend suggests that age is inversely related to the likelihood of recidivism. Younger individuals may face greater challenges in re-entry, such as unstable employment, peer influence, or lack of established support networks, which can increase the chance of reoffending or parole violations. In contrast, older offenders likely benefit from greater maturity, stability, and life experience, reducing their risk of returning to custody (Ulmer & Steffensmeier, 2014).

Overall, these findings imply that, within the 2008–2020 New York State release cohorts, younger offenders are a higher-risk group for recidivism, while older individuals are more likely to avoid reoffense. The results reinforce broader criminological evidence that the probability of reoffending decreases with age, highlighting the potential value of age-targeted rehabilitation and reentry programs to address the specific needs of younger offenders.

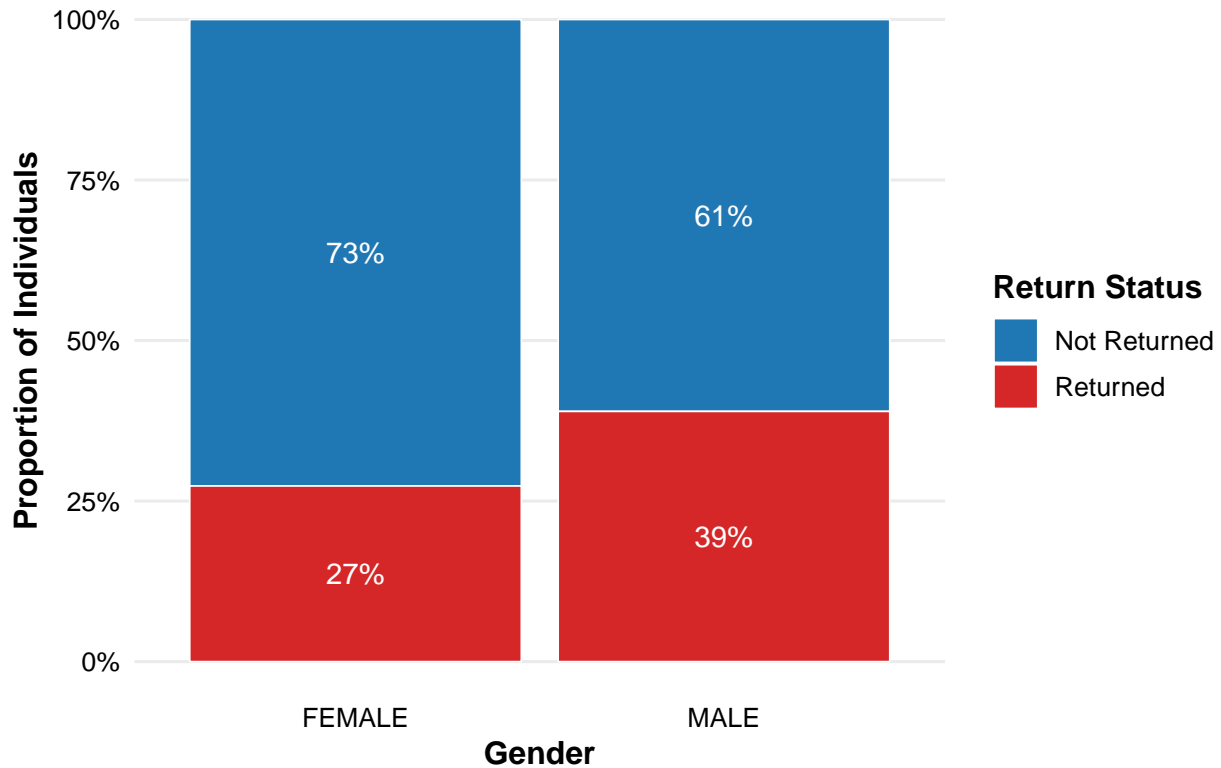
Question 2: Gender vs Recidivism

Exploratory Analysis

Table 3: Summary of Recidivism by Gender

Gender	Count	Mean Age	Recidivism Rate
FEMALE	18728	36.72	0.27
MALE	268411	35.92	0.39

Figure 2. Proportion of Offenders Who Returned by Gender



Conclusions

From the summary table and bar chart, it is clear that there are notable differences in recidivism rates between male and female offenders. Males make up the majority of releases in the dataset (approximately 268,000 compared to 18,700 females), and their recidivism rate is higher at about 39%, compared to 27% for females. While both groups have similar mean ages (around 36 years) gender appears to be a meaningful factor in the likelihood of returning to prison within three years of release.

Statistically, the approximately 12-percentage-point gap in recidivism rates suggests a significant disparity between genders. Given the large sample size, this difference likely reflects a real pattern rather than random variation. The proportional difference indicates that males are roughly 1.5 times more likely to recidivate than females, aligning with established criminological findings that men tend to have higher rates of reoffending.

In context, these trends may imply that female offenders face different post-release circumstances that reduce their risk of recidivism, such as stronger family connections, greater motivation for stability, or lower exposure to criminogenic environments. Conversely, males

may encounter more barriers to reintegration, such as limited employment opportunities or higher supervision intensity, increasing their likelihood of return (Spjeldnes & Goodkind, 2009).

Overall, within the 2008–2020 New York State release cohorts, the results support the conclusion that gender is associated with recidivism outcomes: males are significantly more likely to return to prison within three years than females. This finding suggests that reentry programs and support services could be strengthened through gender-responsive approaches that address the distinct needs and risk factors of each group.

GitHub Repository Link

https://github.com/UVA-STAT3080/project-team_msw

References

Portions of the code, data visualization formatting, and written text in this report were developed with the assistance of OpenAI's ChatGPT (Version 5).

New York State Department of Corrections and Community Supervision. (2025, March 7). Recidivism: Beginning 2008 [Data set]. State of New York: data.ny.gov. <https://catalog.data.gov/dataset/recidivism-beginning-2008>

OpenAI. (2025). ChatGPT (Version 5) [Large language model]. OpenAI. <https://chat.openai.com/>

Spjeldnes, S., & Goodkind, S. (2009). Gender differences and offender reentry: A review of the literature. *Journal of Offender Rehabilitation*, 48(4), 314–335. <https://doi.org/10.1080/10509670902850812>

Ulmer, J., & Steffensmeier, D. (2014). The age and crime relationship: social variation, social explanations. In K. M. Beaver, J. Barnes, B. B. Boutwell (Eds.) *The age and crime relationship: Social variation, social explanations* (pp. 377-396). SAGE Publications Ltd, <https://doi.org/10.4135/9781483349114.n24>