

# NY Daily Inmates in Custody

## Milestone Report

EXPLORER: SALEHA BAKHT

---

### Introduction

This milestone report details the first steps taken to explore the dataset provided by the City of New York on their Daily Inmates in Custody. While the City has not posed any questions with the release of its data, our exploration will look into exploring the patterns presented in the data to see if there are any unexpected findings. The goal is also to use machine learning processes to predict missing values and verify the findings in the dataframe.

After exploring the data, we believe it may be useful to also use demographic population data to study the proportions of races that have and have not been incarcerated in the timeframe provided in this data.

### Data Cleaning

To start us off, we replaced the original column names 'BRADH' and 'SRG\_FLG' with 'MENTAL\_OBSERVATION' and 'GANG\_AFFILIATION' respectively. The original names we did not find intuitive. The new names are longer but easier to remember for the explorer.

Next we dropped the columns 'DISCHARGED\_DT' and 'SEALED'. Since all of the values in the former were null values and all the values in the latter were the same (in that none of the inmate's files were sealed) the columns would have no influence on the outcomes of our exploration. They were removed from the dataframe as considerable features.

For the preliminary purposes of EDA all rows that contained a single null value in the columns of 'CUSTODY\_LEVEL', 'GENDER', 'RACE', and 'AGE' were dropped. The rows that had null values in the 'TOP\_CHARGE' column were kept as they will be attempted to be guessed

---

in the future machine learning portion of this exploration. The total percentage of dropped values is 1.6%, leaving us with 98.4% of our original dataframe.

The last thing we did to clean the data was to break up the time stamps in 'ADMITTED\_DT' into columns for year, month, day, and hour. The 'ADMITTED\_DT' column was then dropped.

## **Exploratory Data Analysis**

A preliminary exploration into gender distinctions first, where we found that:

- There were 14.45 times as many male inmates as there were women.
- Low numbers of gang involvement.
- Proportionally higher numbers of women being under mental observation than men.
- The highest proportion of men and women are at medium security custody level and that the least proportion is in maximum security. This holds true for both genders.
- More men had unrecorded values in their top charge but the proportion disparity is getting smaller and smaller by the year.

We found almost 100 records belong to inmate 16-17 years of age. Most of them are given maximum security custody levels. A higher relative proportion of them among their own set have gang affiliations than the adults. More than half of their top ten most common charges involve robbery or burglary.

In relevance to race, there was nothing noteworthy when comparing the genders in incarceration. Race had no bearing on the existence of infractions either. There are proportionally less of each race under mental observation except for the White and Islander populations. Those races have more people under mental observation than not.