MMGiles

Bond Factors for CT Pretrial Detainees

Overview

Connecticut intended 2016's "An Act Concerning Pretrial Justice Reform" to reform bail requirements and lengths of stay for some non-violent misdemeanors. Did Connecticut's pre-Act bail requirements also show bias involving immutable characteristics? Are Connecticut Bail Fund's concerns about "An Act Concerning Pretrial Justice Reform" valid? Are Connecticut residents still being held by the state without a trial because they can't afford bail?

My analysis is checking for biases and testing the post-bill versus the pre-bill data to provide some insight into the effectiveness of the bill, primarily by determining whether a discernible change in the bond amounts for misdemeanor charges exists.

Preliminary Data Sourcing

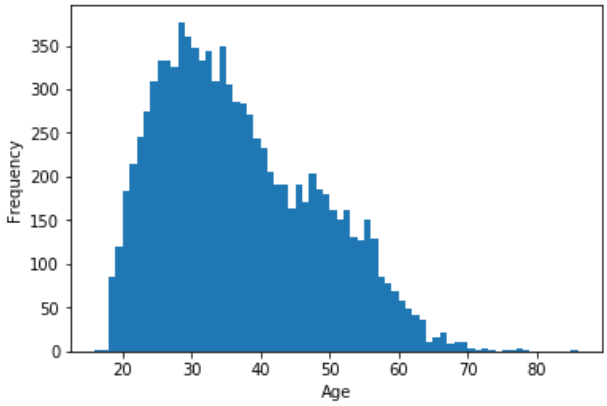
I have obtained the data directly from the state of Connecticut. The dataset begins on 1 July 2016, and the analysis period has been split into two sections: before (1 July 2016 - 30 June 2017) and after (1 July 2017 - 30 June 2018) enactment. Fields utilized were the randomized identifier, latest admission date, race, gender, age, bond amount, and offense. In 20% of the rows, the offense did not include a classifier for felony or misdemeanor. These rows have been removed and the misdemeanor charges separated out as the feature of interest.

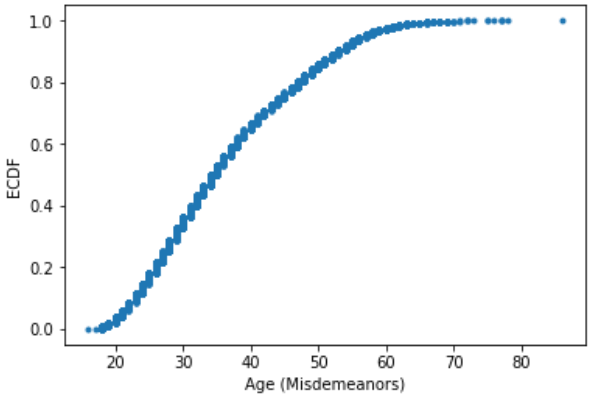
While the state says that they upload the information every day, not every day was represented in the dataset, and the state admits that not all "latest admission dates" are correct. This presented issues with length-of-stay determinations. Two methods were tried - counting the times per arrest that the unique identifier appears in the rows, and performing time math between the latest admission date and the last date the detainee appears in the rows, capped at

1,095 (maximum days plus 365 based on the state's recommendation that admission dates over a year before should not be assumed to be correct). All statistics in this report have been performed with integers sourced from the time math calculation.

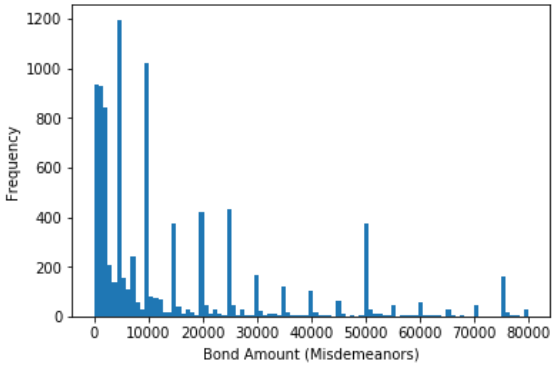
Early Analysis

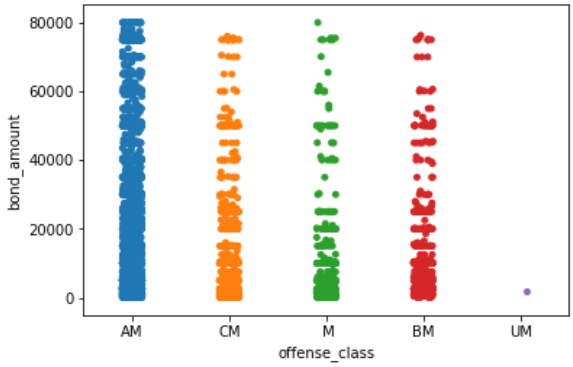
After performing cleaning steps, I examined the age, bond amounts, gender, and offense class visually. First, a look at the age distribution in the sample and its cumulative distribution function (CDF). Most detainees are under 60, with a median age of 34, and with a possible second peak in the late 40s.



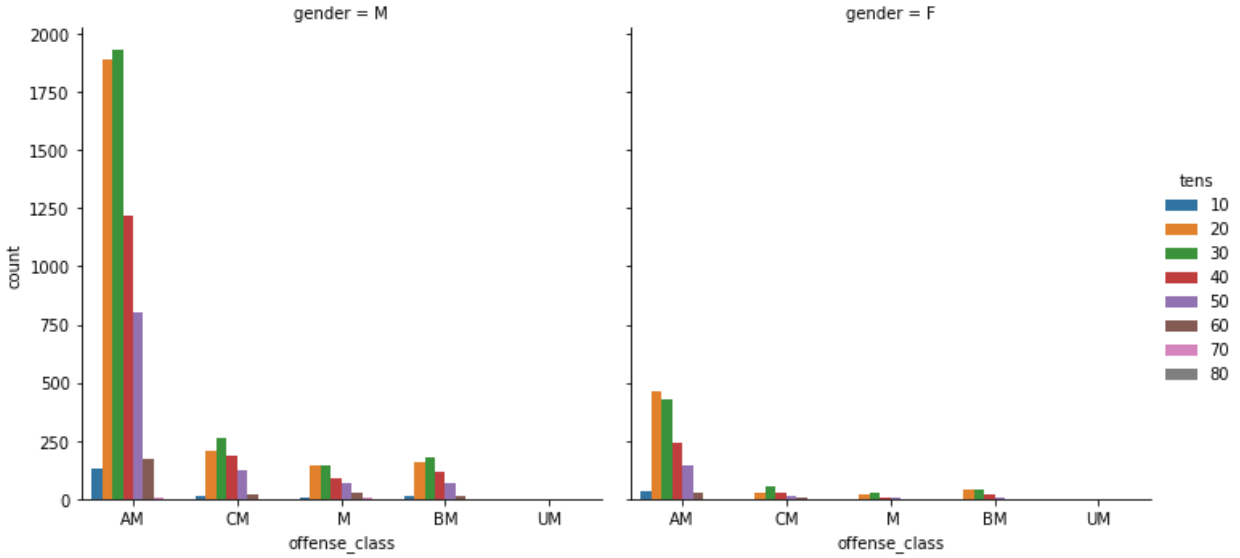


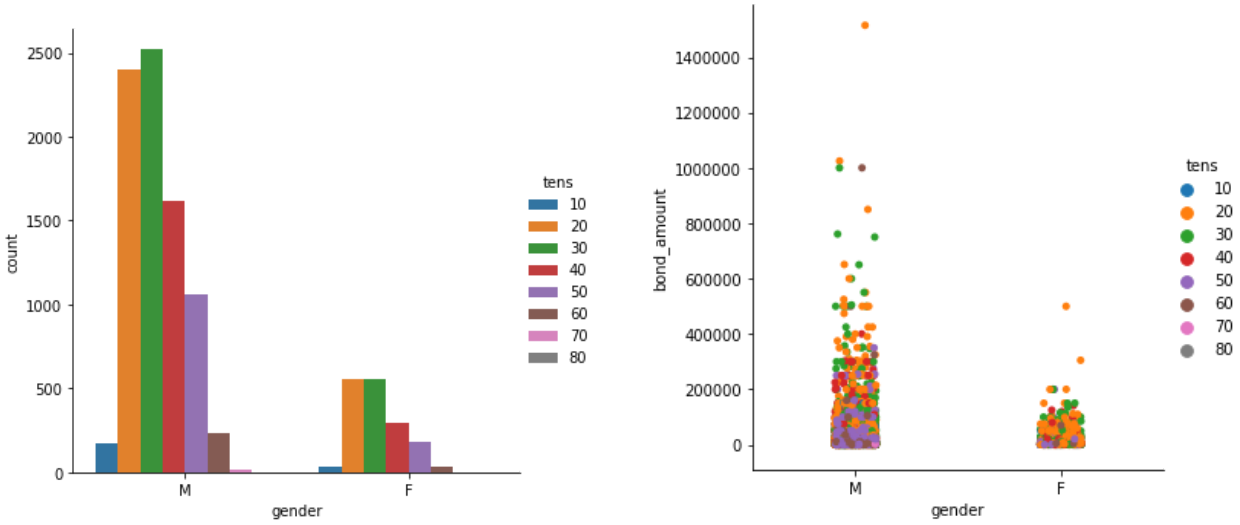
The bond distribution is skewed by a few high bonds. The below plots are the 94% of bonds that are $80,000 or less. Bias is shown toward bond amounts that contain 5s and 0s ($3,000 and $3,500, for example), and bond amounts are bottom-heavy (higher frequency of lower amounts in each offense class).





How are offenses distributed between the genders by age? The first plot below implies that there might be a different age skewing for the genders, but underneath that is a plot of male and female detainees by age. The distributions are fairly similar. Next to it is a plot of the bail amounts by age, also separated by gender.





Now that these factors have been examined in our plots, I'll check the for some statistical comparisons if gender, race, and offense class for their potential impact on bond amounts and days detained. All tests were conducted with significance/alpha = 0.05.

Gender -- For the first set of tests, a chi-squared test determined that gender is a contributing factor to both bond amounts and the number of days detained over the entire two-year sampling period. (Female median bond amount: $5,000. Female median days detained: 17. Male median bond amount: $10,000. Male median days detained: 24.)

The second set of tests checked on the proportion of inmates by gender before and after the enactment. With a p-value of 0.83, we fail to reject our H0, that pre-enactment == post-enactment.

Race -- Tests of the minority percentage of the CT system demonstrate that H0 (no difference between pre-enactment minority and post-enactment minority) came back with a p-value of about 0.7, high enough that we fail to reject H0.

While five races are used in the CT data, only three were used for this analysis. American Indian and Asian were dropped from the dataframe before the next two pieces of analysis since the number of detainees was too small to come to any conclusion.

Testing the mean bond amounts for the remaining races (white, black, and hispanic) from the pre-enactment period to the post-enactment period offered high chi-squared values and p-values below the 0.05 significance value set prior to the test. H0 (white == black == hispanic) should be rejected in favor of differences existing for mean bond amounts.

Bond Amounts	mean before	mean after	*median before*	*median after*
white	18,373	22,007	*7,500*	*7,500*
black	22,818	23,169	*7,500*	*9,000*
hispanic	25,372	24,668	*8,500*	*10,000*

Testing the mean of the days detained (H0: white == black == hispanic) for the three races offered a p-value of 0.62. We do not reject that the mean of the days detained could be statistically equivalent before and after enactment.

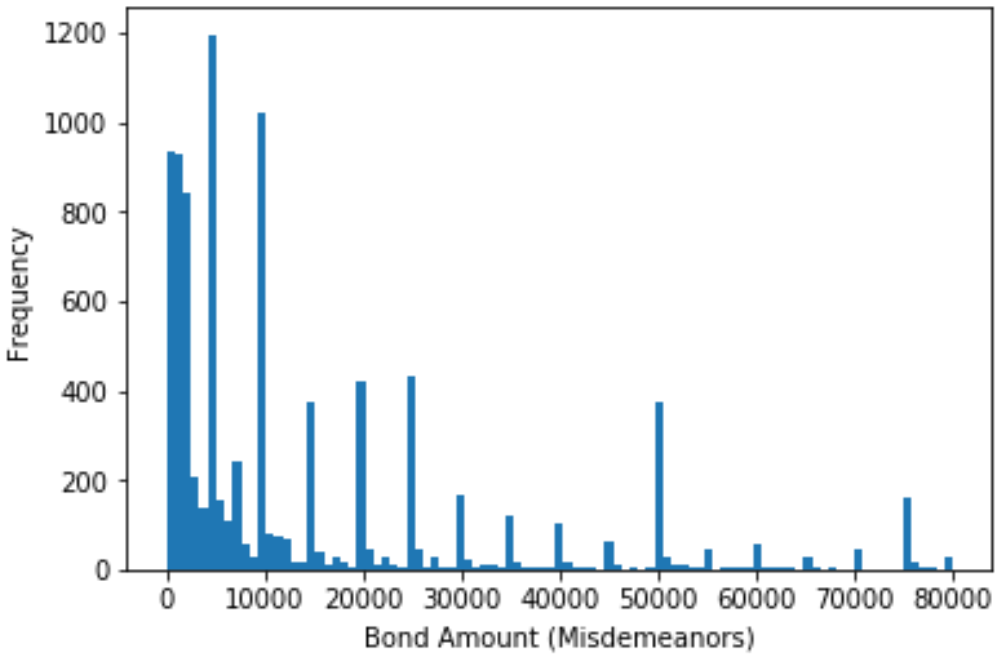
Days detained	mean before	mean after	*median before*	*median after*
white	51	49	*23*	*15*
black	70	76	*22*	*15*
hispanic	58	62	*21*	*14*

Offense class -- Testing for offense class only offers suspicion that the distribution of offense classes for detainees remained relatively stable over the two years for offenses without a letter classification ("M" offenses) provided by the state. For A, B, and C misdemeanors, we reject H0 (that the classification distribution remained stable) for the post-enactment period.

Initial Modeling

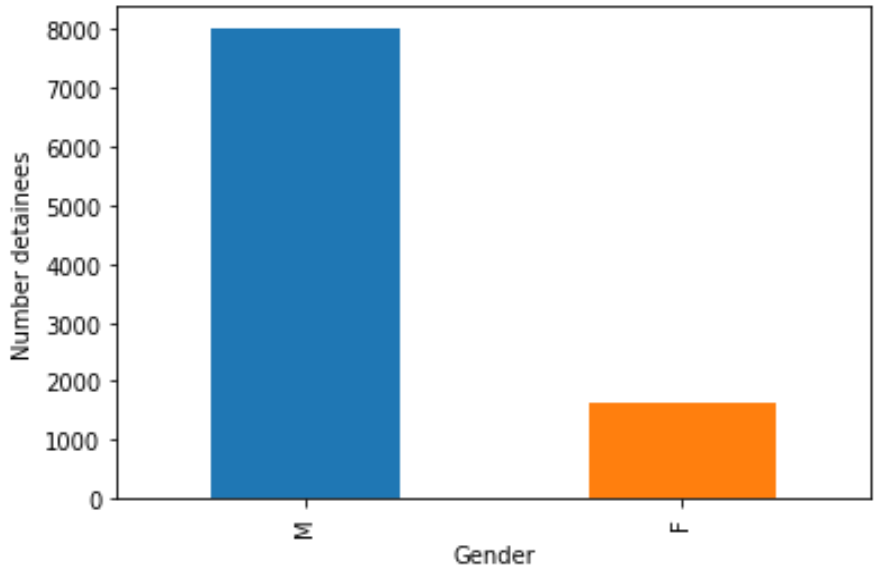
The original plan for this project was to predict bond amounts from a mixture of criminal and immutable characteristics. Criminal characteristics include the offense classification, bond, and days detained, and immutable characteristics are those of the individual that cannot be changed: age, race, gender (author acknowledges that age itself changes and gender can *be changed* but must place limitations on playing devil's advocate).

Linear regression proved difficult to perform on this dataset because of the distribution of the bonds as pictured to the right.



Progression and Final Modeling

While the mean bond amounts by race were statistically significant enough to fail to reject the hypothesis that they differed, they weren't practically significant. White/non-white populations also didn't shift much during the two periods. I needed to look for a more significant variable and a different way of looking at the data. I chose to predict gender from a combination of other features: bond amount, days detained, age, offense class, and race.

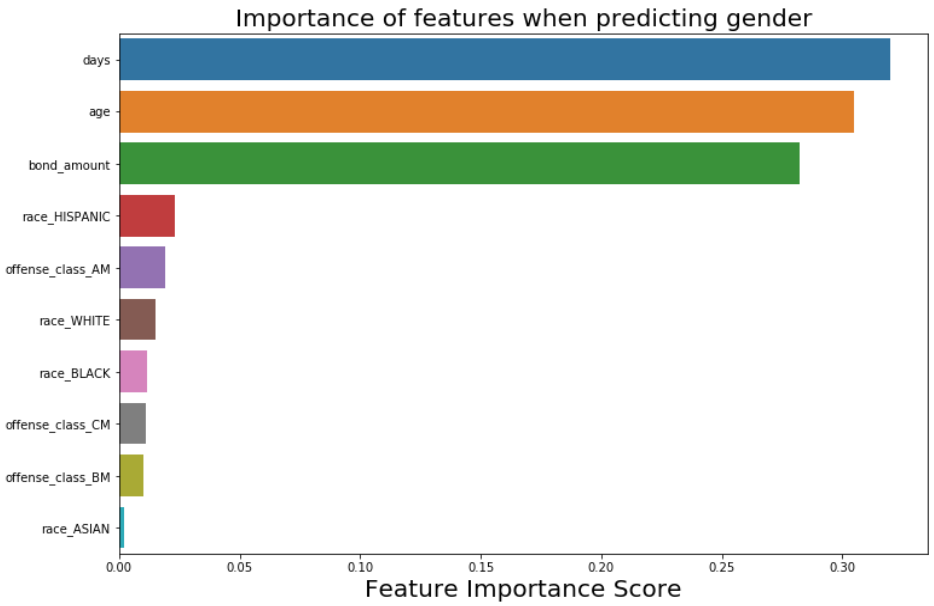


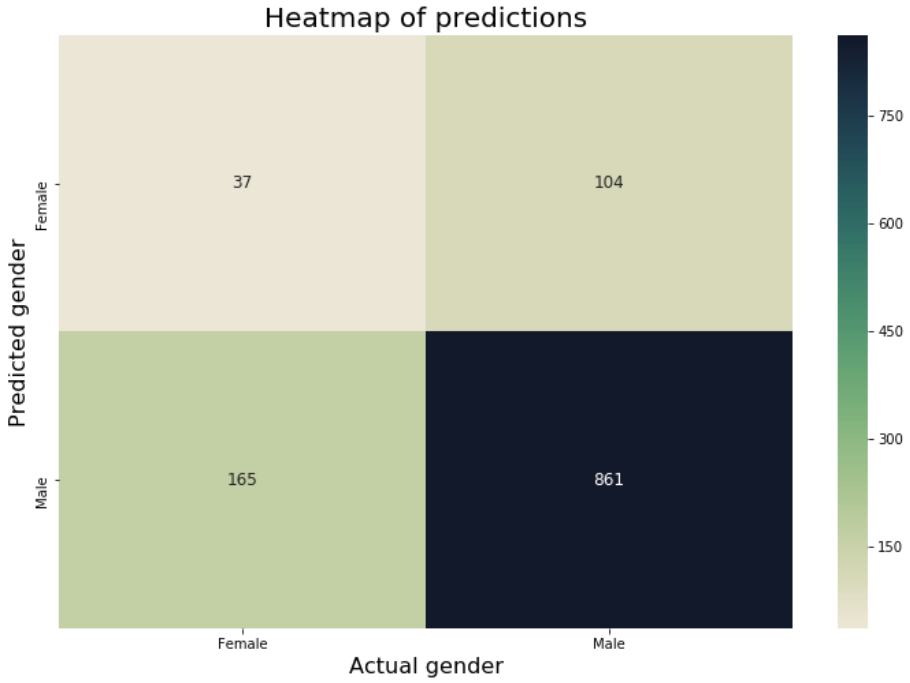
Female detainees make up about 17% of the population and have a mean bond amount of 60% of the male detainees'. They also stay behind bars while awaiting trial half the mean days of their male counterparts. These differences could matter in modeling.

I began the classification with logistic regression and quickly discovered the imbalance of the data would be problematic. After rebalancing the data by upsampling the female rows, the basic model could predict the correct gender about 82% of the time. The area under the ROC curve (AUROC) also increased. (This upsampled data will be used in all the other models in this section.) Unfortunately, after tuning the hyperparameters with a randomized search, the model could still only correctly predict gender about 82% of the time with no additional increases in recall, precision, or AUROC.

The final attempt at a working classifier utilized a random forest. Out-of-the-box accuracy was well over 75%, a hopeful sign. Precision and recall were both much stronger for this model, both in the mid-80 percents.

The classifier was then tuned using a random search, and the relative importance of the model features is displayed to the right. Days detained, age of detainee, and bond amount had the most influence.





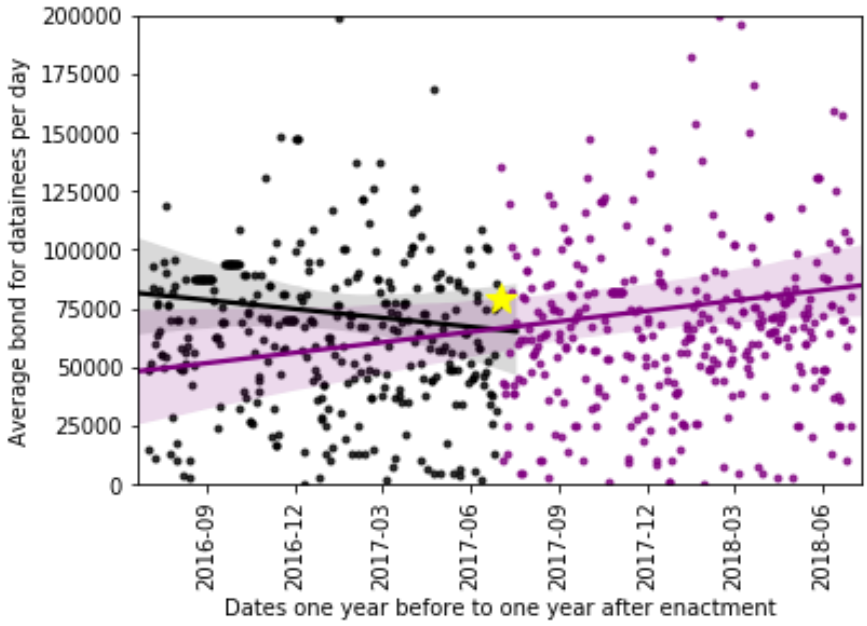
Cross tabulation of the predictions on the test set also showed good accuracy, with 89% of males correctly identified.

At this point, I believe that a tuned random forest classifier can accurately predict from the available data whether a detainee is male 90% of the time.

Final Thoughts

While I have not yet devised a model for predicting bond amounts based on an individual's characteristics, I have demonstrated that some bias does seem to exist in CT's bond system. The differences are most extreme for gender but appear in a statistically significant amount in the racial data. The statistical testing also showed that the gender and racial makeup of the pretrial population remained at similar levels before and after the enactment -- so neither improvement nor worsening conditions can be determined from this analysis.

I stand behind the other conclusions I formed during during earlier reports. Part of this work is discouraging. Mean bond amounts since bill enactment (as shown on here) have increased $1,500 after decreasing in the time between bill passage and enactment. Connecticut Bail Fund is right to be concerned that the act isn't doing enough to stop the debtors' prison effect that originally concerned them.



Whether through the efforts of groups like CBF or effectiveness of the act for the lower end of the bond spectrum, however, most pretrial detainees appear to be staying behind bars at least seven fewer days per arrest in the year after "An act concerning pretrial justice reform" was enacted. The mixed results bear watching.