Section 14 Springboard Submission
Written and submitted by Saleha Bakht.
NY Daily Inmates in Custody - Proposal
The following is a proposal for the Springboard Capstone 2.

1. What is the problem you want to solve?

    a. Can top charges be predicted?
    b. Does race or gender have overly large influence on custody level?
    c. Is there a pattern in infractions amongst certain groups of inmates?
    d. Are there correlations between top charge numbers and inmates being under mental observation?

2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

    The dataset is hosted by the City of New York. They have not provided a list of questions to be answered. Their main concern would be about ensuring that there is no bias in their charges in sentencing. They are giving people the opportunity to look at anonymous prison information so that they can check against racism, sexism, mental bias, and gang bias themselves. They would be very interested to know if there is bias in their system.

    Algorithms have been made to determine the sentencing of an offender. Due to possible accusations of racial and sexual discrimination, can an effective algorithm be made to remove such biases? How would the algorithm perform?

3. What data are you going to use for this? How will you acquire this data?

    The data for this would come from NYC OpenData and Kaggle. The dataset is updated daily on both platforms. It includes data on things like inmate ID, admission dates, discharge dates, custody level, whether the inmate is under mental observation, race, gender, age, inmate status code, whether the inmate's case is sealed, whether the inmate is affiliated with a gang, the inmate's top charge, and if they have committed any infractions.

    It should be noted that all discharge dates in the dataset are marked as unavailable.

4. In brief, outline your approach to solving this problem (knowing that this might change later).

    e. Can top charges be predicted?
        i. Use a label encoder to predict what the top charges are using 1700 values to find a missing 100.
    f. Does race or gender have overly large influence on custody level?
        i. Use graphs to see if there are any patterns in race and gender or otherwise amongst cusody levels.
    g. Is there a pattern in infractions amongst certain groups of inmates?

        i. Use graphs to see if there are any patterns in race, gender, gang affiliations, mental observation, inmate status, top charges, and the duration of inmate incarceration amongst whether or not the inmates are written up for infractions.
    h. Are there correlations between top charge numbers and inmates being under mental observation?
        i. Use graphs and perhaps a train and test set to see if top charges can be predicted whether or not the inmate is places under mental observation.

5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

    i. Solutions would presented with the Python code file, a slidedeck for the graphs to answer exploration questions. The comments in the Python file would detail the approach, the methods used, and the answers found.
    j. Attempt a predictive model to see how the top charge values can be affected by whether the inmate is under mental observation or not, also to see what other factors may be involved.