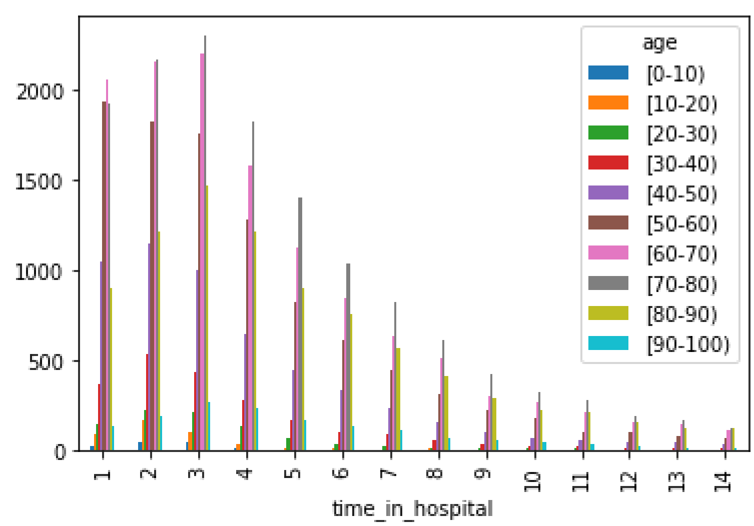
Kelci Mensah, Madiha Abdul Maajid
Fairness Analysis Project: Algorithms in the Wild
12 December 2022

## Action Item 1: Data Analysis

- **Analysis**

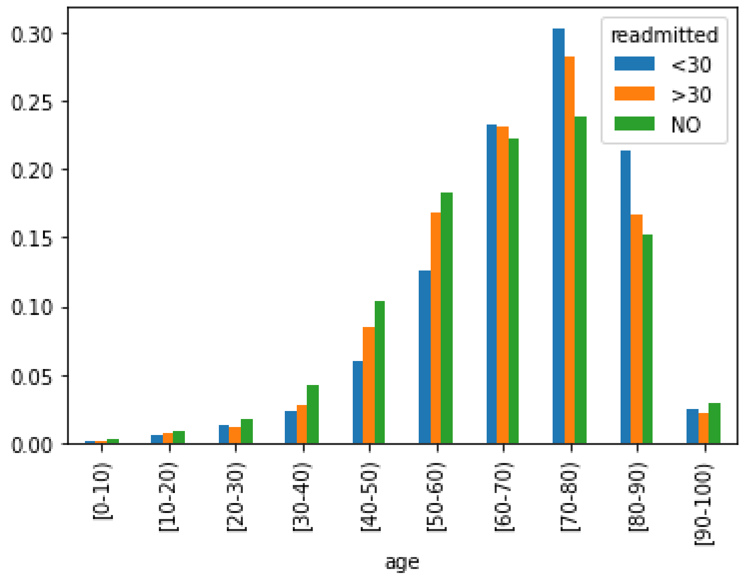
  With the main factors being age, race, gender, and time spent at the hospital in regards to being readmitted, the strategy employed was to find correlations between each of the factors (age, race, gender) and how long the stays were. For example, some questions to answer were: did older or younger people have a longer or shorter stay? Which racial groups had longer or shorter stays? Were there any differences with gender? Then, by establishing these correlations, a comparison was made between time of stay and rate of readmission. Therefore, by determining which factors led to a longer hospital stay, the patterns which caused a longer hospital stay, which then would be a factor for a greater chance of readmission could be determined

- **Correlation Between Age and Time of Stay**
  - A larger amount of older patients (starting from 40+)
  - Sample Size: 30-40 age group has a only 500 max patients, 40+ goes up to 2000, 50-80 goes to 2500, then goes down to 2000 at 80 and 250 to 90+
  - There is a direct relationship between age and time of stay, the older a person is, the longer their time of stay is going to be. In terms of count, there is a clear direct correlation but with frequency, past a 4 day stay, generally showing that older patients tend to have a longer stay rather than a shorter one

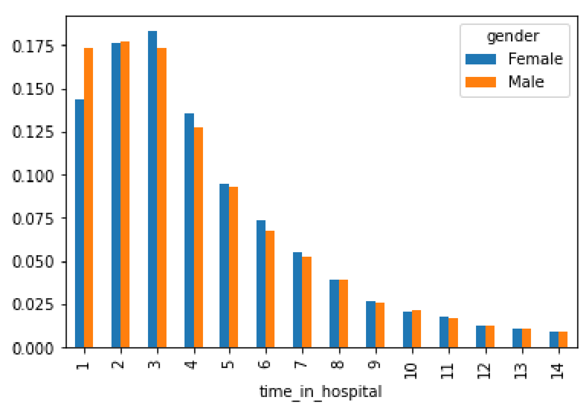



- **Correlation Between Age and Readmission:**

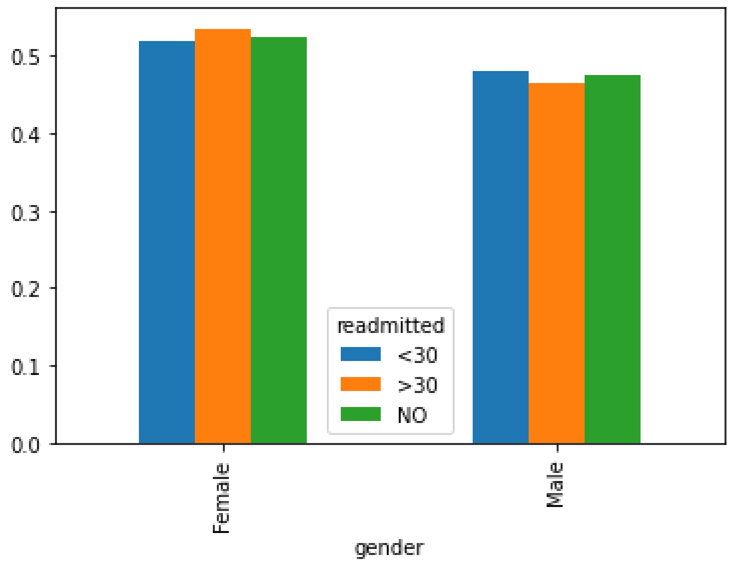
From 60+ onwards, the rate that a person would be readmitted in less than 30 days and more than 30 days becomes higher than not being readmitted at all. The younger someone is from 60, the chance they are not readmitted is higher than being readmitted, and only goes down with age, with the exception of 0-10 and 20-40, but with only a small difference.




- **Correlation Between Gender and Time of Stay:**
  More females tend to stay longer in the hospital than males, starting at 2 days:

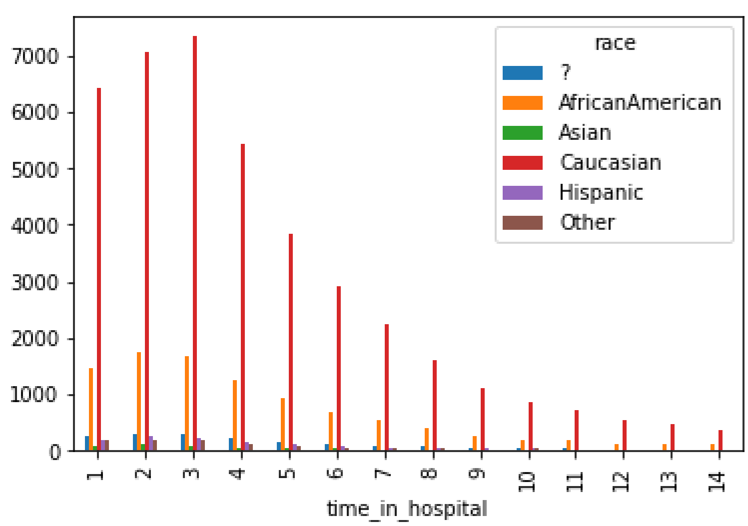


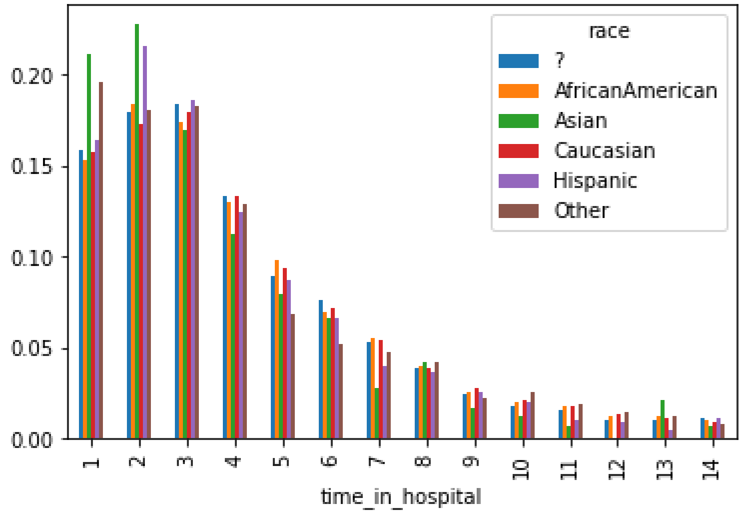

- **Correlation Between Gender and Readmission:**
  - Females are slightly more likely to readmitted than males after 30 days and before 30 days
  - In terms of the rate of readmission with either gender, males are slightly more likely to be readmitted under 30 days, but overall seem to have an equal chance of being readmitted, and so do females when compared to their own statistics. However, the comparison between rates of female and male readmission is that females are more likely to be readmitted.




- **Correlation Between Race and Time of Stay:**

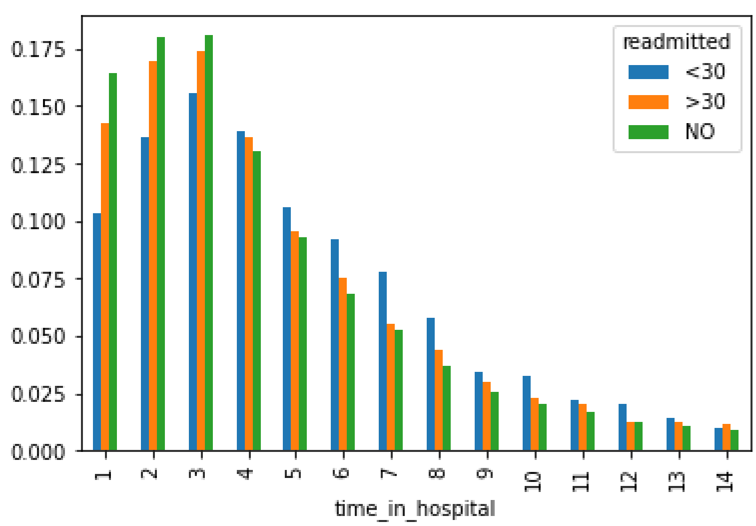
  Caucasians both have a significantly larger count than Africans Americans, Hispanics, Other, then Asians. However, when considering the frequency, Asians have a much higher frequency in being admitted, especially in stays that are 1-2 days long. Following Asians, the races that tend to have longer stays are Caucasians, African Americans, and Other. The significance of high Asian frequency in hospital stays, despite having the smallest count in the population count is significant. Other race groups that also have a higher frequency in staying at the hospital are African Americans, Caucasians, with Hispanic and Other being among the lower frequency groups, generally.






- **Correlation Between Race and Readmission:**
  Caucasians were the most likely to be readmitted after 30 days, and the least likely to not be readmitted. Asians were the most likely to be readmitted, followed by Caucasian within 30 days. After 30 days, Hispanic, Other, African American, then Asian patients were likely to be readmitted. However, it should also be noted that the groups least likely to be readmitted were African Americans.

- **Correlation Between Time of Stay and Readmission:**
  There is a direct correlation between the length of the stay, and the likelihood that someone will be readmitted. In the graph below, someone is less likely to be readmitted within 30 days if their stay is 3 days or shorter. However, after the 3 day mark, the likelihood of a patient not being readmitted decreases dramatically, with the risk of being readmitted within 30 days remaining high and the risk of being readmitted after 30 days still being higher than not being readmitted if the patient has a stay longer than 3 days.



- **Patterns:**
  - Older patients have a likelihood for having a longer stay as opposed to younger patients, and this chance increases once a patient is past the age of 30
  - Older patients are also more likely to be readmitted, with the significant jump starting at the age of 40, and the chance that being readmitted is higher than not being readmitted starting at the age of 60
  - Females are more likely to be readmitted than males, although females are slightly more likely to be readmitted after 30 days and males are more likely to be readmitted under 30 days
  - Asian, Caucasian, and African American patients seem to have the highest frequency of longer stays, with the other race groups either being lower or matching the rate of stays
  - There is a direct correlation between time of stay and readmission, the more likely that a person is going to have a longer stay, the more likely it is that they will be readmitted within 30 days or after 30 days.
  - Race, age, and gender groups that are associated with longer stays in the hospital have a larger chance for readmission

## Action Item 2: Analysis of Prediction Metrics

**Strengths:**
The prediction has an accuracy rate of 0.89, which means that for 89% of cases, the predicted labels gave the correct outcome. The strengths associated with this accuracy rate include a low false positive rate. False positive, or screening that someone is more likely to be readmitted, therefore giving this patient more medical attention when not necessary, would waste medical resources. The false positive rate for this prediction model is 0.075, meaning that a very low percentage of people will be screened for being readmitted when they are not supposed to be.

**Weaknesses:**
On the other hand, the weaknesses with the predictive model include a high false negative rate, meaning that patients that could have used extra care did not receive it. The false negative rate is 0.74, which is a significant number of patients not receiving the care they need. This could indicate that the predictive model could have a bias towards certain groups, or the predictive model does not have enough data on certain groups or ignores groups of people when considering who might need extra help.

Next, The model has a recall rate of 0.25, and a precision rate of 0.12. With the recall rate being higher than the precision rate, this could indicate that the model returns many results but the predicted labels are incorrect when compared to the training labels. Despite the low FPR, the precision rate is not high, indicating that even though the model returns many outputs, the prediction labels can be incorrect. With both being low, it can indicate that the predictive model is not producing accurate results.

Finally, the ROC AUC score indicates the trade off between detecting false positives in order to make sure that all true positive screenings are detected. For this model, the AUC score is a 0.58, indicating that the accuracy to determine the difference between false positives and true positives is not as adequate, or only correctly indicates the difference between a false positive and a true positive at only 58%. Unfortunately, given the low AUC score, the difference between false positive and true positive is almost like random guessing with this predictive model.

**Advice for Hospitals:**
Some advice for hospitals using this predictive model would be to do a second check done by actual doctors following the predictive model. For example, if the predictive model says that a patient is not likely to be readmitted, it should be further ruled out that the patient is not a false negative given the high false negative rating. On the other hand, given the low false positive rating, it can be believed that a positive rating is accurate, but should still be checked considering that many patients get listed as a false negative and deserve the care. Also considering that the AUC score is only at 58%, the patients ranked as positive for readmission should also be reconsidered. The hospital should look at the rate of patients that are being predicted for negative for readmission only for it to be false, and see which demographics are being impacted by false negatives the most.

## Action Item 3: Analysis of Group Fairness

- [ ] Is the prediction fair for different groups? Are there any differences in the performance that stand out to you? Any case of Simpson's paradox?

I do not believe that the initial predictions created by Fairlearn's MetricFrame are necessarily fair for each of the different groups. The **group_by_gender** table shows similar values for both subgroups, so that might be the only case that we can consider fair with very little difference between values. However, the **group_by_age** and **group_by_race** subgroupings show varying selection rates for instance, meaning readmittance per group varies, and not by a pattern that makes sense.

Let's take a look at the **group_by_age** table for example:

- In the **group_by_age** table, the age-range with one of the higher selection rates has a sample size of 6107. One would assume that the larger the sample size of a particular age-range, the higher percentage of readmission. This would lead us to assume similar principles for an age-range with a smaller sample size, but we do not see that. The selection rate for the 30-40 year old group is approximately 3.2877%, meaning that only about 3% of those 1095 patients are readmitted. Compared to the selection rate for the 80-90 year old group which shows approximately 10.9608% (nearly 11%) of those 4288 patients are readmitted. The 20-30 year old age range has the lowest selection rate but it is the 10-20 year old age-range that actually has the lowest sample size (excluding the 0-10 year old range). 11% of the 90-100 age-range are readmitted.

Let's also take a look at the **group_by_race** table:

- The subgroup with the highest sample size is the Caucasian grouping, and they are not in the lead by a little. With a size of 20368, about 8.8% of these patients are likely to be readmitted. Now in comparison let's check the African-American grouping with a sample size of 4845. This grouping has a 6.6% selection rate. This does *not* make much sense as explained in the analysis of the **group_by_age** table - if the selection size grew by a factor of nearly 5, then the sample size should reflect that. This data shows us that African-American patients are more likely to be readmitted in this instance.

We believe there is unfairness in the prediction and that can only occur as a product of the people in charge of building prediction models: there is subconscious stereotyping that can occur during the development and implementation phases.

Additionally, in this dataset there was at least one case of Simpson's paradox which occurred: There was a trend amongst the FPR and FNR which disappeared when the data was aggregated together versus separated as patient age-ranges increased (they got older).

## Action Item 4: Analysis of Fairness Mitigation

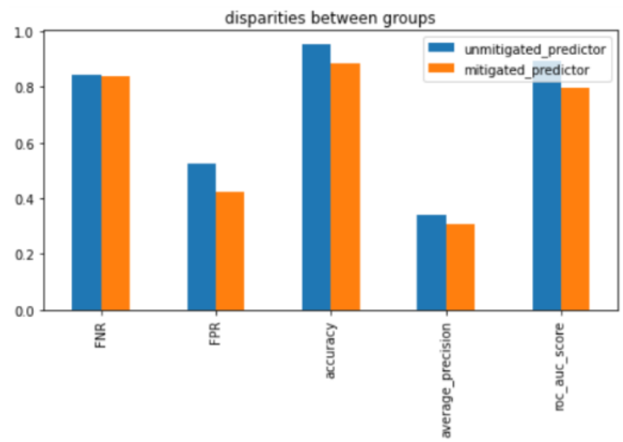
- [ ] How do the unmitigated and mitigated predictors compare?

The mitigated prediction model is an attempt to reduce any bias in the data using the Fairlearn plugin ThresholdOptimizer. The goal here is to make it so that the overall performance is better in the mitigated model as well as have the disparities between groups be reduced.

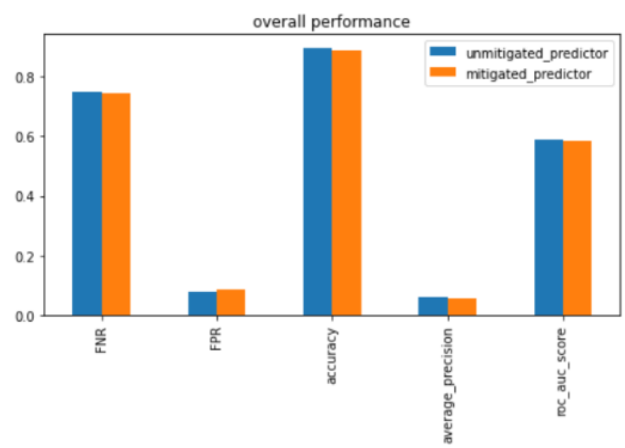
In both model tables, we see that most values fluctuate irregularly by a few fractions of a %. For example, in the mitigated model the African-American subgroup has 6.9% and 7.2% respectively for the FNR and FPR values. In the unmitigated model however, these percentages are actually 7.1% and 6.0%. We actually see a regular pattern of overall decrease in the mitigated model's **accuracy**, **average_precision**, and **roc_auc_score** compared to the unmitigated.

In removing the prediction odds (setting them to zero), the graphs are showing that the mitigated values are reduced.

Let’s take a closer look at the graphs quantifying both the unmitigated and mitigated performance:




- For the disparities between groups graph above, the unmitigated values are higher for each category of data, meaning that there is actually *more* disparity between groups. In this instance, the mitigated model predictions are better.




- In this graph of overall performance, we see that the only category value that doesn’t follow the trend of reduced mitigated value is the FPR category. However in this graph, we must interpret the bars such that performance is actually *better* for the **unmitigated_predictor** overall compared to the **mitigated_predictor**.