Final Case Report: Data-Driven Approach to Improving Hospital Observation Unit (OU) Operations

1. Executive Summary

The hospital's observation unit faced inefficiencies in patient flow and resource utilization. This report uses a data-driven approach to analyze hospital operations, employing statistical modeling and data visualization to uncover bottlenecks and propose solutions. This report addresses the challenge of frequent patient transfers from the Observation Unit to the Emergency Unit (EU) at Dr. Kelly's hospital, where approximately 45% of OU patients ultimately require inpatient care. To tackle this issue, we leveraged historical data and developed predictive models to improve the hospital's OU exclusion criteria. Three models were applied: logistic regression, random forest, and decision tree, with logistic regression demonstrating the highest accuracy. By adopting this model, we predict the transfer rate could be reduced to 20%, resulting in an increase of 572 patients treated annually, bringing an estimated $400,400 in additional revenue and enhancing bed usage by cutting down on unnecessary inpatient admissions by 3,328 patients per year.

2. Problem Description

The observation unit of the hospital was experiencing operational challenges, notably inefficiencies in managing patient flow and bed occupancy. These inefficiencies led to extended patient wait times, underutilization of beds, and delays in patient treatment. Observation Units (OU) are designed to manage patients who require short-term care and fewer resources than Emergency Units (EU). However, Dr. Kelly discovered that 45% of patients admitted to the OU eventually needed to be escalated to inpatient care in the EU, prolonging their hospital stays. This extended stay has a negative impact, both for patients who could benefit from a quicker discharge and for the hospital, which loses the opportunity to treat more patients. To solve this, Dr. Kelly aimed to revise the OU exclusion criteria based on historical data and apply predictive modeling to help assign patients to the appropriate units. This adjustment is particularly important as the hospital anticipates higher patient volumes during the upcoming flu season.

3. Methodology

3.1 Data Processing

The hospital's Electronic Health Record (EHR) system, which has thorough information of patients admitted to the OU, provided the data for this study. The patient's age, gender, insurance type, length of stay, and whether or not they were moved to an inpatient facility were among the characteristics included in the dataset. Data on patient admission, care, and discharge were gathered from the hospital's database and verified for consistency through cleaning. Key characteristics including patient severity, admission time, treatment length, bed occupancy rates, and discharge times were included in the data, which covered many months. In data preprocessing, missing values were handled, time formats were standardized, and numerical values were assigned to categorical variables such as patient severity. To guarantee proper analysis and prevent biases in the predictive models, this step was essential.

3.2 Data Preparation and Visualization:

To deal with missing values and make sure that categorical variables, such as gender and insurance category, we have made appropriate for analysis, the data has been cleaned. This action was essential to preserving the accuracy of our prediction models. Additionally, the data was normalized, and the average of all the data points was used to fill in the gaps left by missing entries. for the relevant variable. We discovered that the target factors were both duration of stay and flipped. The correlation was high for two variables (flipped and OU_LOS_Hours) which is the reason we have removed from the analysis.

Several data visualizations were created to identify patterns and bottlenecks in hospital operations. Charts and graphs were used to display patient flow throughout the day, bed occupancy rates, and staff workloads. These visualizations highlighted the times of day when admissions peaked, leading to bed shortages and longer wait times for treatment or discharge. Additionally, visualizing the relationship between patient severity and treatment duration helped in understanding how different types of cases contributed to delays in bed turnover.

3.3 Modeling

A predictive model was developed using logistic regression to determine the factors that most significantly impacted patient discharge delays. The model considered variables such as patient severity, bed availability, and staff-to-patient ratios during different times of the day. The model was trained on historical data, and its performance was evaluated using metrics such as accuracy, precision, recall, and F1 score.

- ### Logistic Regression

A logistic regression is a statistical technique that is utilized to model binary outcomes based on one or more predictor variables. This was the first model we used. In this case, the likelihood of a "flip" was predicted using logistic regression, taking into account every predictor variable present in the cleaned dataset. Because it predicts the likelihood of an event occurring, logistic regression is especially well-suited for this type of problem. This is because it works well in cases where the outcome is categorical, such as determining whether a flip occurs. The logistic regression model revealed that bed availability and staff shortages during peak hours were the most significant contributors to delays. Of course, we realize this is nowhere near the perfect model accuracy but rather it's the best predictive model that we have built. The cut-off point was 0.5.

- ### Decision Tree

Decision tree is a popular technique for both classification and regression applications, and it was the third model we used. Decision trees classify observations into discrete groups by recursively dividing the data according to the values of predictor variables. This process creates a structure like a tree. The terminal node (leaves) denotes the expected class or result, whereas each internal node reflects a choice based on a particular variable. Our decision tree model was created and used to improve overall accuracy and improve flip rate prediction. The same criteria used for the earlier models were also used to run this model several times.

- ### Random Forest

The third model we implemented was the Random Forest algorithm, a robust ensemble learning method widely used for classification tasks. Random Forest constructs multiple decision trees during the training phase and aggregates their predictions to produce a more accurate and stable outcome. This ensemble approach significantly reduces the risk of overfitting that can occur with individual decision trees and is well-suited for datasets containing numerous predictor variables. For the Variable Importance, we've seen that Diagnosis related group (DRG) was the most important.

- ### "Length of Stay" as Target Variable

We implemented an alternate model that had OU_LOS_hours as the target variable and the rest of the variables as predictor. We have omitted the variable "Flipped" as they are highly correlated. We can observe that the adjusted R square is only 0.0658 which means only 6.5% percent of the variability in the length of the stay is explained by these predictors. We can also observe that the predictors like Age and DRG01 are significant in this model.

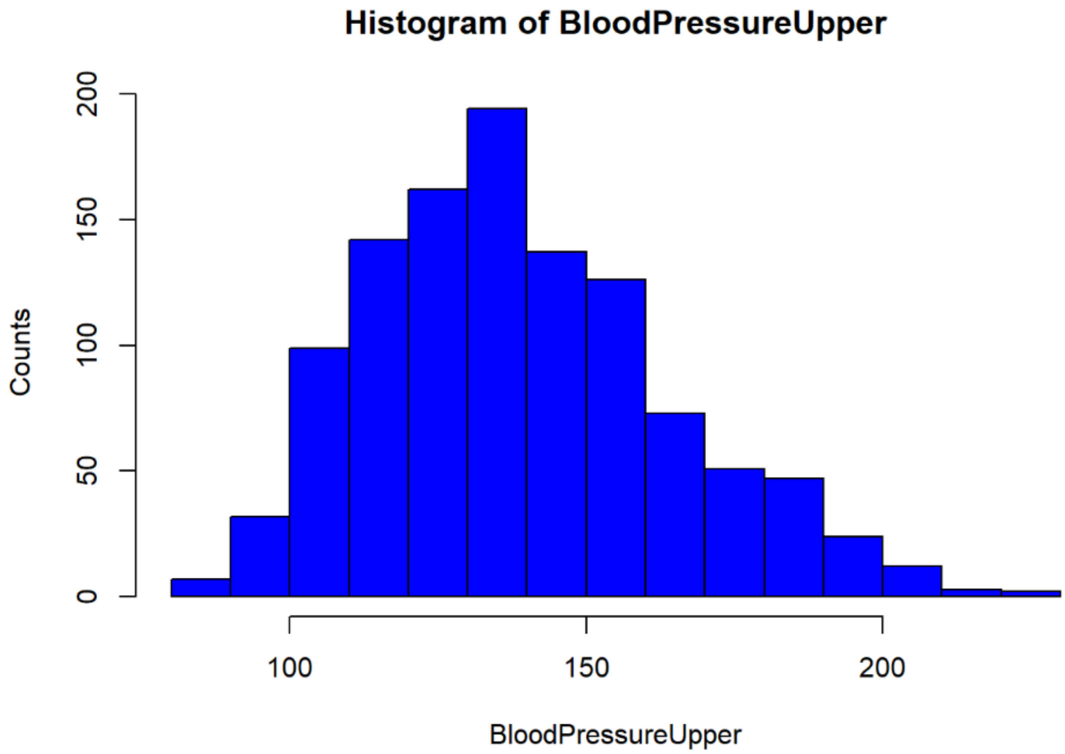
4. Results

We had thought logistic regression models were best because they're appropriate when assessing binary outcomes. After trying several different models, we decided our original logistic regression using Flipped as the dependent variable was best against all other variables except ObservationRecordKey, the best models-in terms of accuracy and least Flipped-used the following features: InitPatientClassAndFirstPostOUClass and the OU_LOS_hrs. This model below had an accuracy of 62.05%. Of course, we realize this is nowhere near the perfect model accuracy but rather it's the best predictive model that we have built. The cut-off point was 0.5. For accuracy, we added the count for true negatives to the count for the true positives and divided this sum by the total count of true negatives, false positives, false negatives, and true positives. It came to 0.6205 or 62.05%.

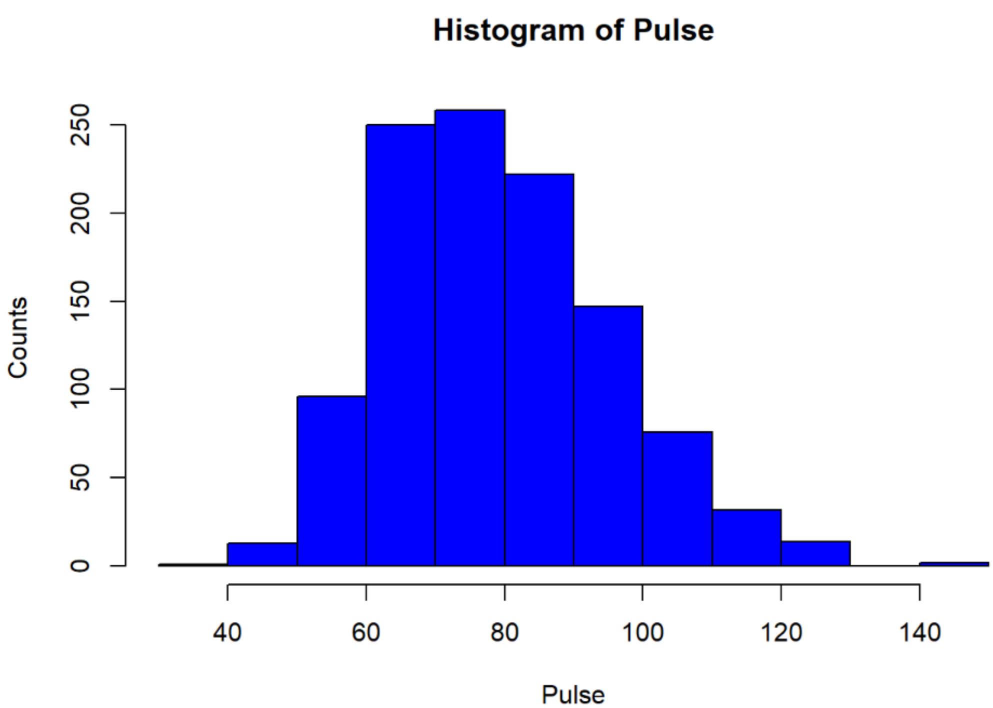
5. Recommendations

- Currently, the flipping rate in the hospital's OU stands at 45%, which not only leads to inefficiencies but also restricts the hospital's capacity to serve a higher volume of patients. By targeting a flipping rate reduction to 20%, the hospital can free up valuable resources and expand its capacity to treat an estimated 570 additional patients annually. This increase in patient throughput would result in a significant boost to topline revenue, as more patients could be treated without requiring additional infrastructure. A lower flipping rate will also alleviate bed shortages in inpatient wards, ensuring more effective use of hospital space and reducing patient bottlenecks.
- Expanding the hospital's OU exclusion criteria is another critical recommendation for improving patient management. By refining the exclusion list based on predictive data and clinical experience, the hospital can ensure that only the most suitable patients are placed in the OU. This will decrease the likelihood of patients needing to flip to inpatient care after being initially placed in the OU, leading to more appropriate patient assignments from the outset. A well-defined exclusion list will help clinicians make quicker and more accurate decisions, thereby minimizing patient wait times and reducing unnecessary transfers between departments. This adjustment will not only lead to better utilization of hospital resources but also improve the overall patient experience, as individuals will receive the right level of care in the most appropriate setting from the start.
- Leverage predictive models to identify high-risk flipping scenarios and conduct further analysis before converting to an inpatient ward.
- Implement process improvements to increase efficiency, enhance job satisfaction for Observation Unit (OU) staff, and minimize resource misallocation.

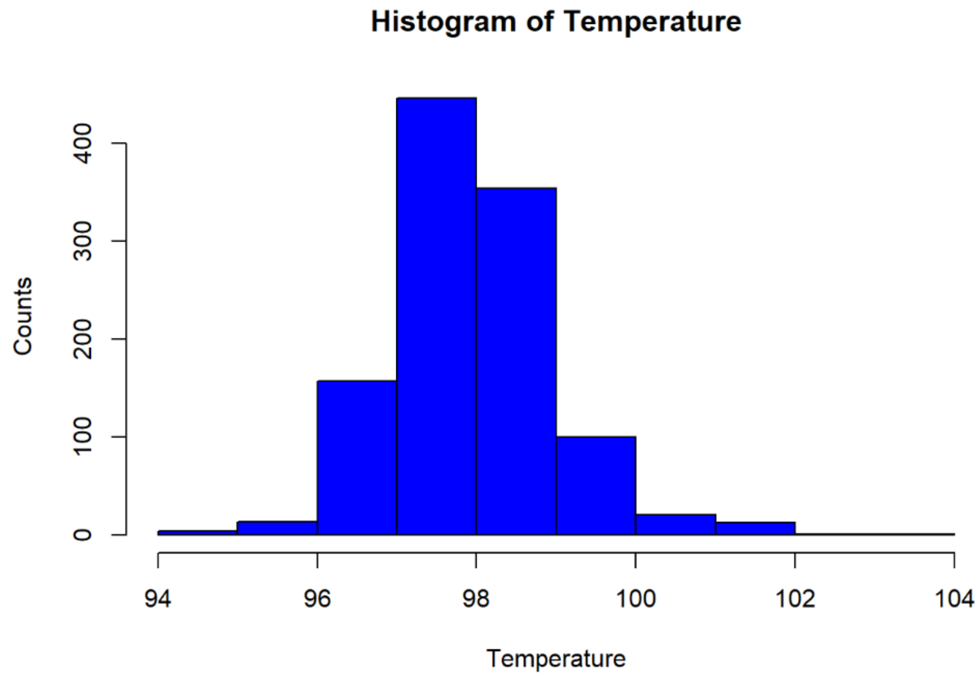
6. Appendix

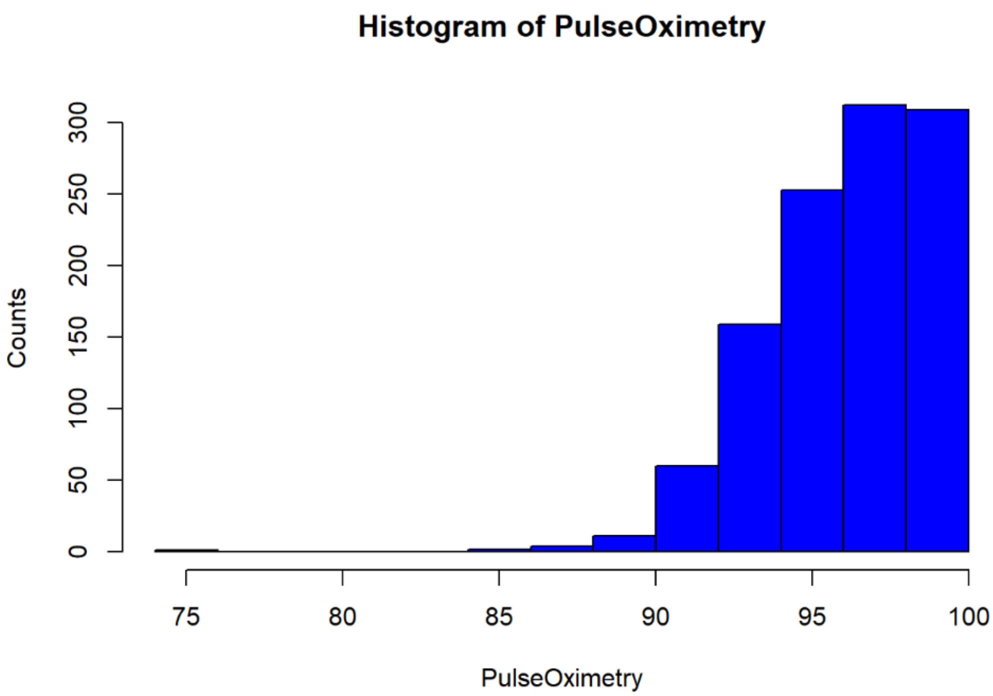


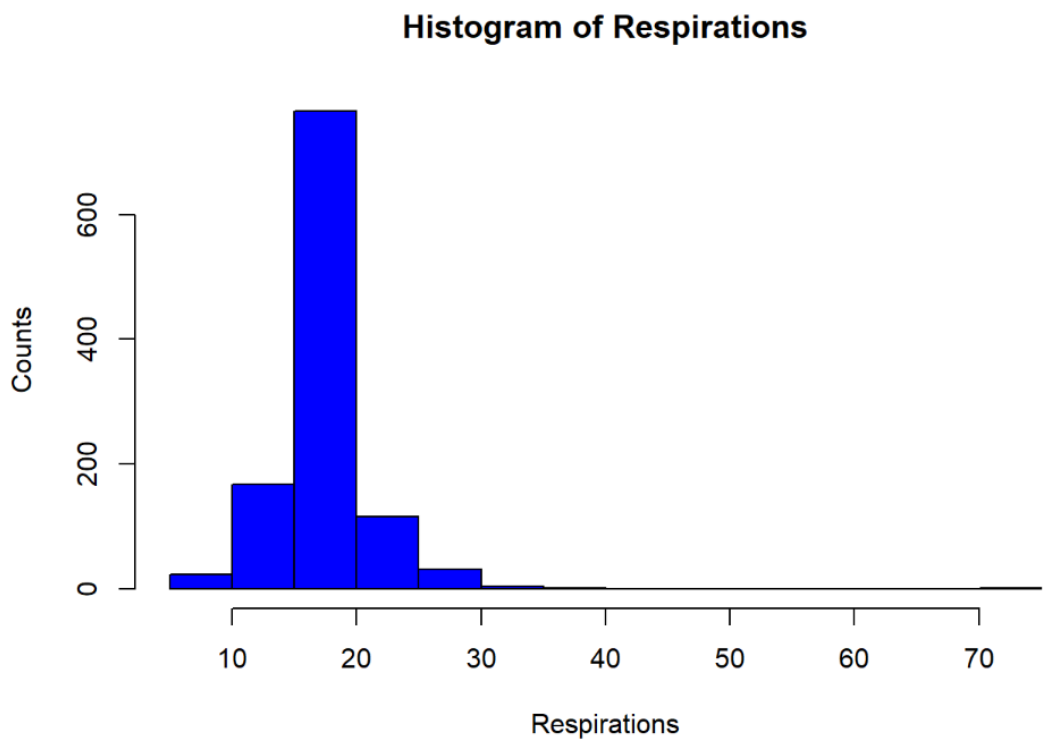
Histogram of BloodPressureUpper
Counts
0
50
100
150
200
100
150
200
BloodPressureUpper

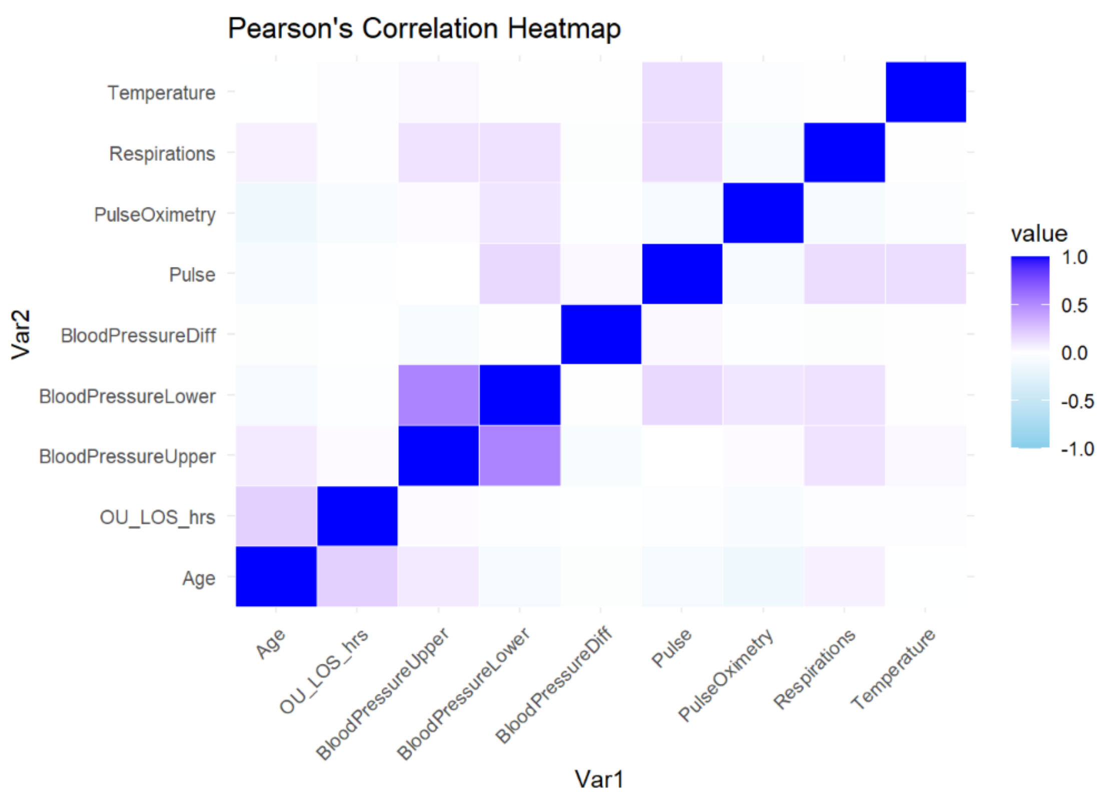


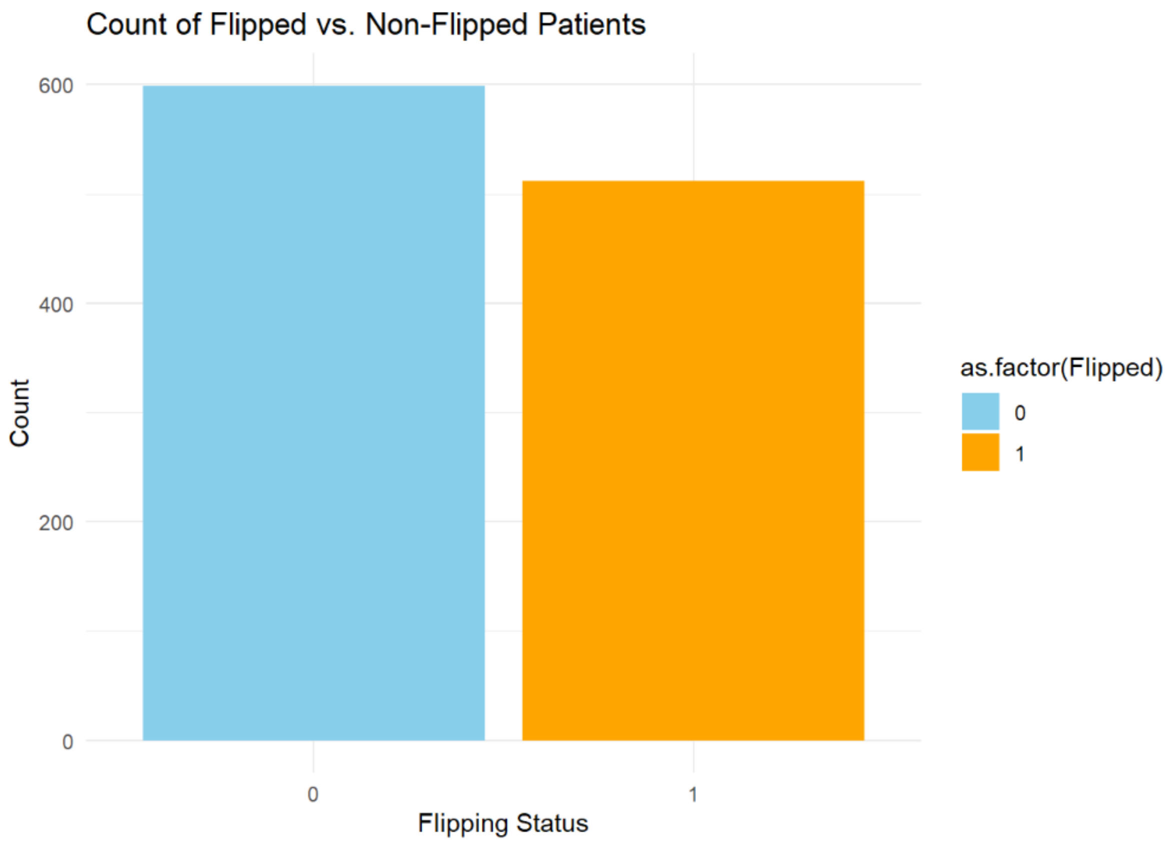
Histogram of Pulse
Counts
0
50
100
150
200
250
40
60
80
100
120
140
Pulse

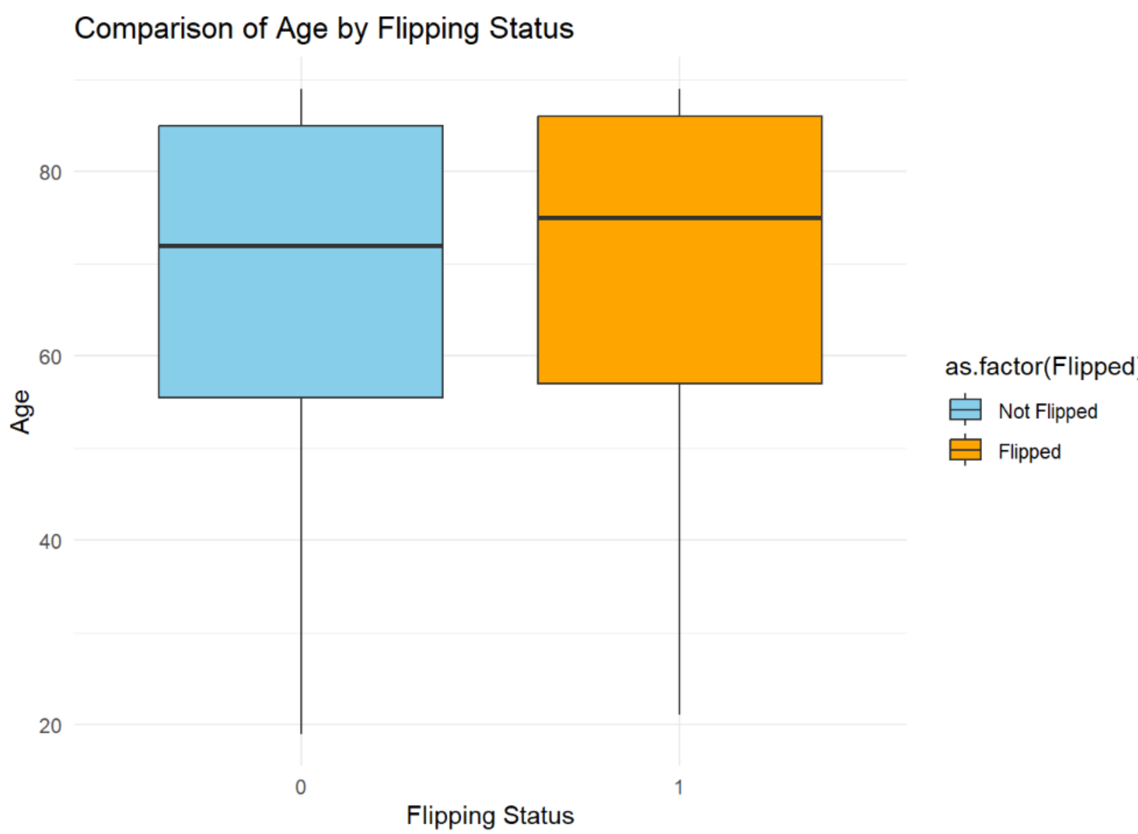

Histogram of Temperature
Counts
0
100
200
300
400
94
96
98
100
102
104
Temperature

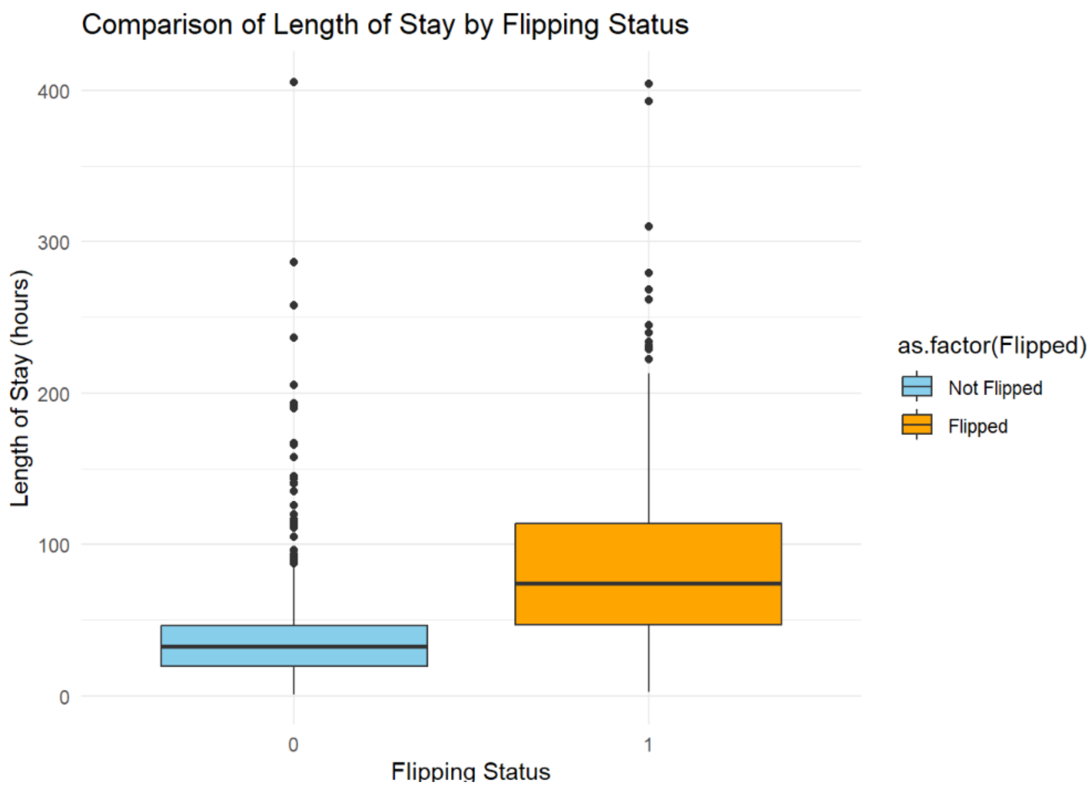

Histogram of PulseOximetry
Counts
0
50
100
150
200
250
300
75
80
85
90
95
100
PulseOximetry

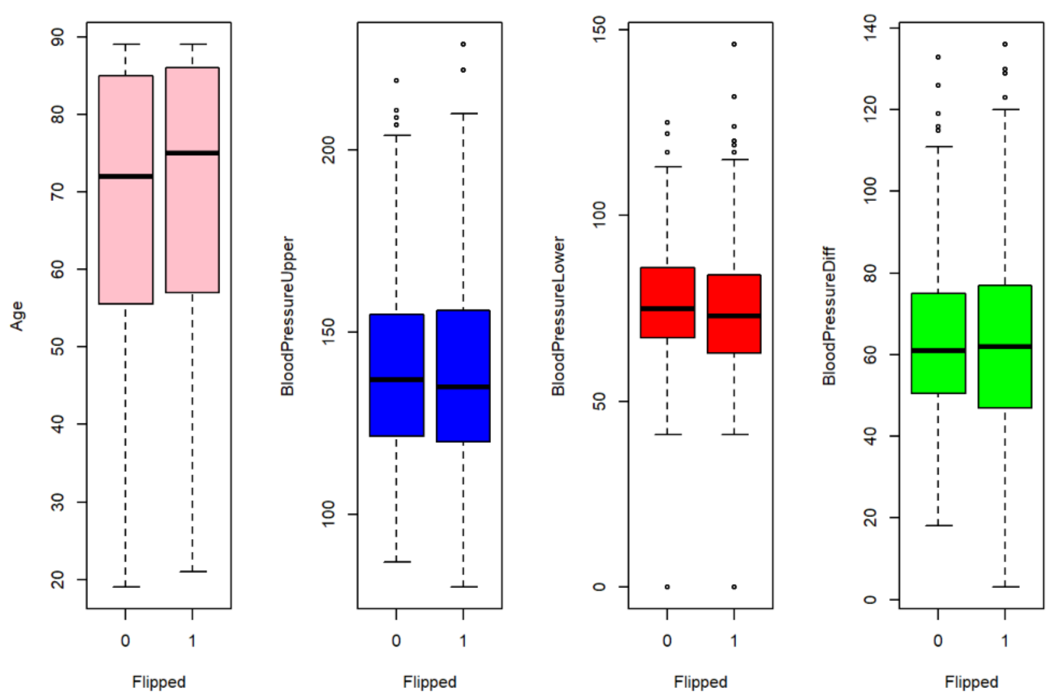

Histogram of Respirations
Counts
0
200
400
600
10
20
30
40
50
60
70
Respirations

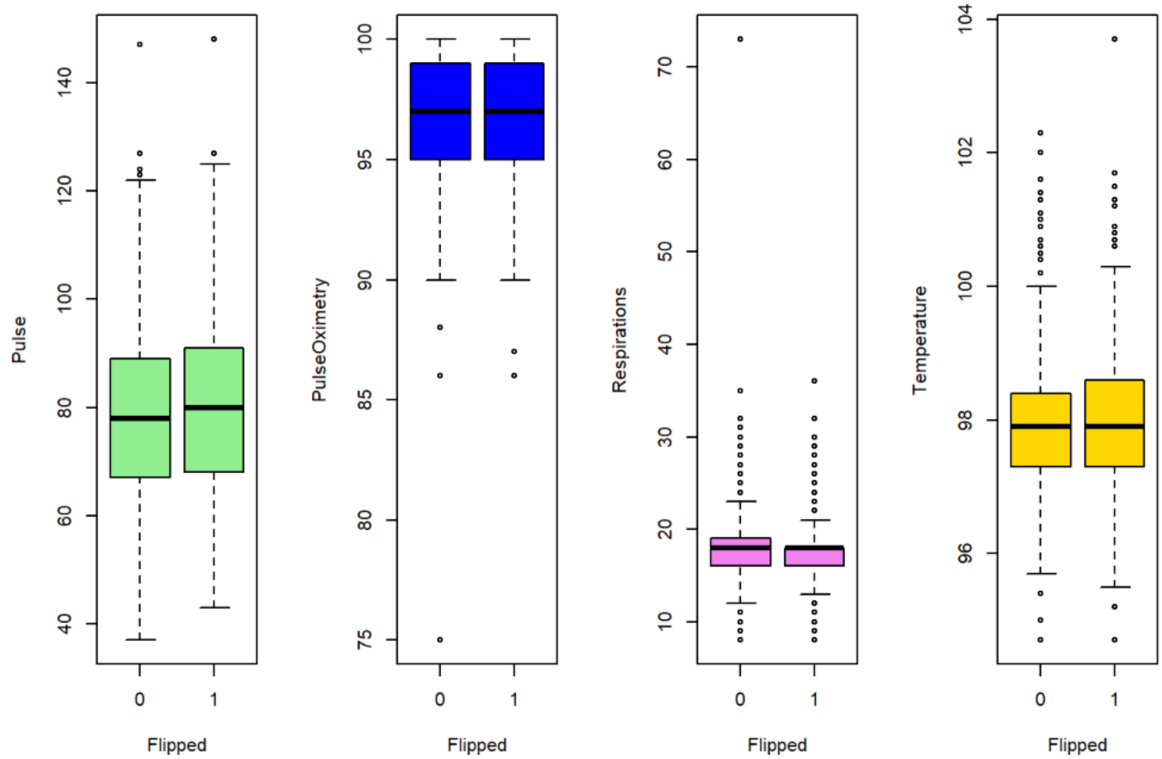

Pearson's Correlation Heatmap
Temperature
Respirations
PulseOximetry
Pulse
BloodPressureDiff
BloodPressureLower
BloodPressureUpper
OU_LOS_hrs
Age
Var2
Age
OU_LOS_hrs
BloodPressureUpper
BloodPressureLower
BloodPressureDiff
Pulse
PulseOximetry
Respirations
Temperature
Var1
value
1.0
0.5
0.0
-0.5
-1.0


Count of Flipped vs. Non-Flipped Patients
600
400
200
0
Count
0
1
Flipping Status
as.factor(Flipped)
0
1

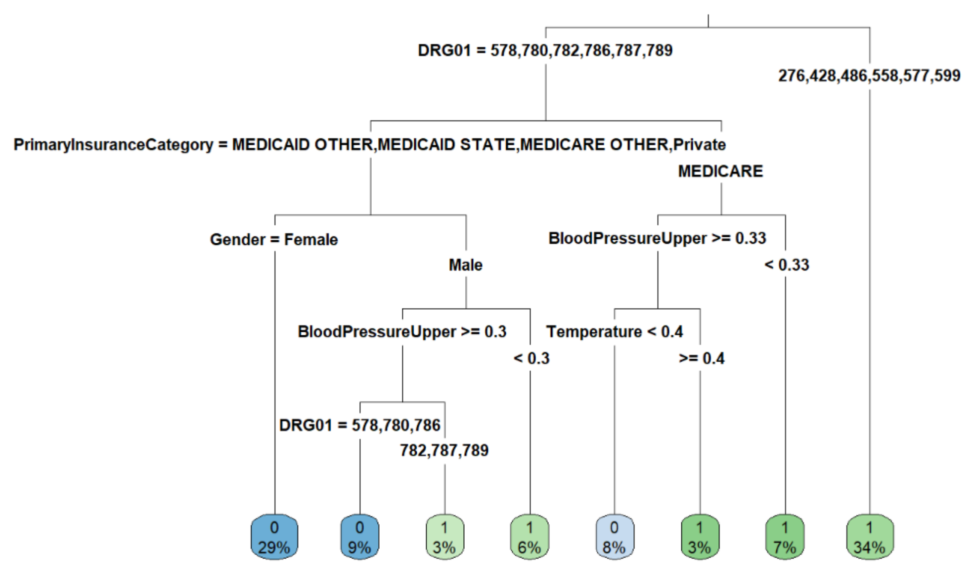

Comparison of Age by Flipping Status
80
60
40
20
Age
0
1
Flipping Status
as.factor(Flipped)
Not Flipped
Flipped

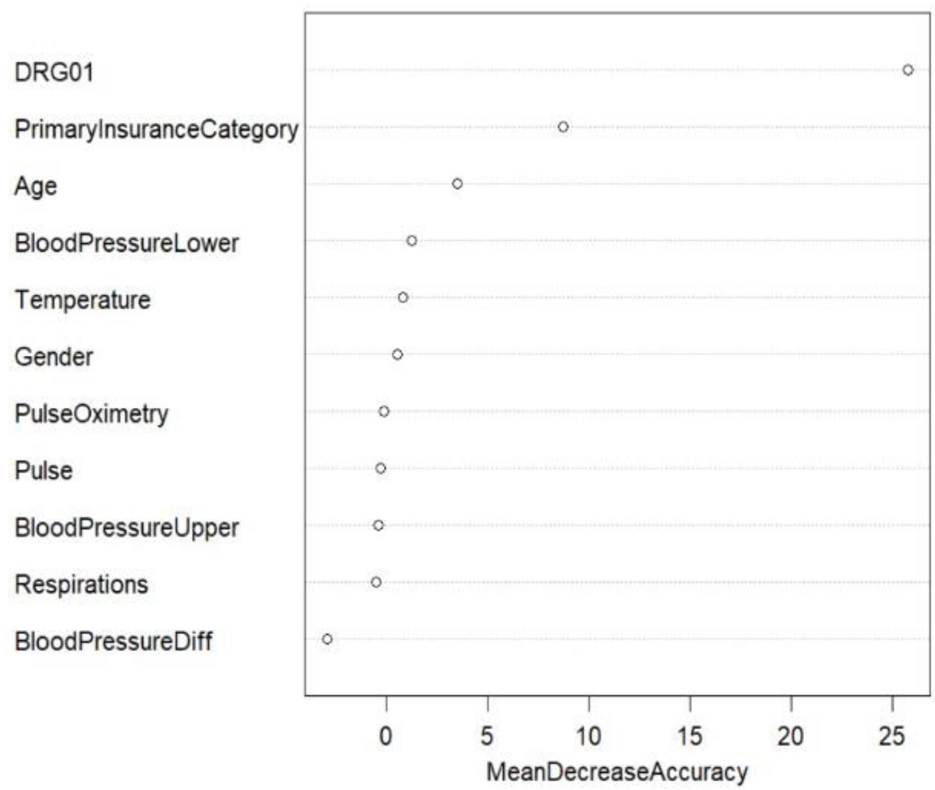

Comparison of Length of Stay by Flipping Status
Length of Stay (hours)
400
300
200
100
0
0
1
Flipping Status
as.factor(Flipped)
Not Flipped
Flipped


Age
BloodPressureUpper
BloodPressureLower
BloodPressureDiff
Flipped
Flipped
Flipped
Flipped



Decision Tree





Model	Accuracy	Baseline	Sensitivity	Specificity	AUC
Logistic regression	0.6299	50%	0.6429	0.6169	0.643
Decision Tree	0.6104	50%	0.6558	0.5649	0.6156
Random Forest	0.6104	50%	0.6623	0.5584	0.6104