MERCER UNIVERSITY

SCHOOL OF BUSINESS

Data Driven Decision Making

Report On

Improving Hospital Observation Unit Operations

Prepared by

Sowndarya Saini

Goutham Yallapu

Table of Contents

1. Executive Summary:

At Montanaro Hospital, Dr. Erin Kelly identified significant inefficiencies within the 23-bed Observation Unit (OU), including long patient stays and a high proportion of inpatient status from OU. To address these challenges, Dr. Kelly led a data-driven project to improve the OU's exclusion criteria. The approach involved assembling an interdisciplinary team, using predictive analytics to anticipate inpatient care requirements, and revising the exclusion list accordingly. The model utilized previously collected patient data, including demographics, preliminary diagnoses, and clinical values. The analysis indicated that a more precise exclusion list could significantly decrease patient stay durations in the OU, increasing unit throughput and enhancing hospital capacity, especially during the flu season. This strategic initiative is expected to not only streamline patient flow but also to boost financial performance by minimizing the loss incurred from patients leaving without being seen.

2. Problem Description:

Montanaro Hospital's Observation Unit (OU) faced a substantial operational inefficiency. The management reports from the OU revealed a concerning trend that patients were staying longer than necessary on average, and there was a high rate of patients being "flipped" from observation to inpatient status (Refer 8.2). These problems not only impacted patient flow and bed availability but also had financial implications due to different compensation rates for observation and inpatient care. The OU, which was designed to provide less resource-intensive care to patients who did not require full inpatient services, faced the challenge of optimizing its patient placement rules to ensure efficient and effective use of hospital resources, especially in anticipation of the high-demand flu season. The problem demanded a systematic review of the OU's exclusion criteria to better identify patient suitability for observation care and minimize unnecessary transitions to inpatient care.

3. Exploratory Data Analysis:

3.1 Data cleaning and Preprocessing:

This study used the case and dataset "A Data-Driven Approach to Improving Hospital Observation Unit Operations" which had up to 1,111 observations and 15 variables. After modifying the data types of the variables, we have identified 20 observations in the dataset, 3 of which were part of "BloodPressureUpper", 4 were part of "BloodPressureDiff", "Pulse", "Respirations", and "Temperature" each and 1 was part of "PulseOximetry". After individually interpreting each variable, we discovered that every distribution is normally distributed, and we have replaced the null values with mean or median based on skewness. Furthermore, we thoroughly examined the dataset to ensure its accuracy and dependability. During this process, we discovered the presence of extreme outliers or data points that differed significantly from the rest of the distribution. To overcome this issue, we have normalized these extreme outliers using min max scaling which would enable us to research further and run the models.

Correlation Heatmap of Health Metrics:

From the correlation heatmap provided (Refer Figure 1) visualizes the relationship between various health metrics. It reveals binary variable “Flipped” has a moderate positive correlation with Length of stay (Hrs), indicating that when the “Flipped” variable is 1, there tends to be a longer length of stay in hours and no significant linear correlation between age and binary variable 'Flipped', indicating that age may not be a strong predictor in this context. Blood pressure readings, including upper, lower, and differential, are positively correlated with each other, which is expected due to their inherent relationships which may cause multicollinearity. Vital signs including pulse, pulse oximetry, and respirations show moderate positive correlations among themselves, suggesting they tend to increase and decrease together. The binary variable 'Flipped' does not show a strong correlation with the continuous health metrics, implying that these do not strongly predict the likelihood of being 'Flipped'. This heatmap is a preliminary step in understanding the data and suggests that more nuanced statistical methods may be needed to uncover deeper insights, especially for binary outcomes.

Boxplot of Diagnosis-Related Group by various Health Metrics:

The boxplots illustrate the distribution of various health metrics between two patient groups (Figure 2): those who were not “flipped” (0) and those who were “flipped” (1). Both groups show similar distributions for age, blood pressure (upper and lower), and pulse oximetry, with only slight variations in median values. The 'Flipped' group exhibits a slightly higher median age and pulse rate, and a slightly lower median for diastolic blood pressure (BloodPressureLower). Notably, there is a clear difference in the 'Respirations' metric, where the 'Flipped' group has a significantly higher median, indicating more respiratory activity which could suggest distress or a different clinical profile.

Across all metrics, there are outliers present, indicating individual patient values that fall outside the typical range. These outliers are especially pronounced in the 'Flipped' group's 'Respirations' metric. The presence of outliers suggests that while the central tendency (median) may be similar between groups for some metrics, there are individuals with atypical values that could be clinically significant.

Scatterplot of Age and Length of Stay:

The scatterplot illustrates the relationship between Age and Length of Stay in hours (Refer Figure 3), with age spanning from approximately 20 to 80 years and the Length of Stay ranging widely up to 400 hours. Most data points cluster at the lower end of the Length of Stay, indicating shorter stays are more common across all ages. Notably, there are outliers, particularly in the Length of Stay metric, where a few individuals have much longer stayed than the majority, with these outliers present across various ages.

3.2 Data Partition:

We have parted data into 70% Training with 778 observations and 30% Validation with 333 observations with 15 Variables each. We ensure that our models have a comprehensive learning phase by using most of the data for training, capturing the underlying trends and patterns across various variables, while the testing data serves as new, unseen data for the models, allowing us to assess how well our predictions generalize beyond the data on which the model was trained.

3.3 Variable Selection: After developing both a null model, which does not consider any variables, and a full model incorporating all available variables, we employed a stepwise method for variable selection. Our analysis revealed that only five variables significantly contribute to predicting flipping based on their p-values. These variables are Age, Respirations, PulseOximetry, Diagnosis-Related Group (DRG01), and Length of stay in the Observation Unit (Hours). Therefore, for subsequent model building, only these five variables are considered.

4. Methodology:

Dependent variable: Flipped (Binary outcome)

Independent variable: Age, Respirations, PulseOximetry, Diagnosis-Related Group (DRG01), and Length of stay in OU(Hrs).

4.1 Logistic Regression:

Logistic regression is a statistical method used for analysing a dataset in which there are one or more independent variables that determine a binary outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). It is widely used in numerous fields, including the medical sciences, social sciences, and machine learning. Logistic regression is robust to non-linear relationships between predictors and the outcome and is versatile enough to handle both continuous and categorical independent variables.

The Logistic regression model suggests that higher respiration rates and pulse oximetry readings are associated with an increased likelihood of “Flipped”, while increasing age is associated with a decreased likelihood. Diagnosis categories serve as categorical predictors, with conditions like syncope and chest pain significantly decreasing the odds of “Flipped” when compared to a reference diagnosis category. Notably, the length of stay in the observation unit (OU_LOS_hrs) has a strong positive association with the likelihood of 'Flipped', indicating that longer stays increase this likelihood. The model's coefficients are statistically significant at various levels. Overall, the model suggests that both clinical signs and diagnosis categories, along with the length of hospital stay, are important predictors of the “Flipped” outcome. (Refer 8.2.1)

The area under the ROC curve (AUC) is a measure of how well the model distinguishes between the two outcomes (often 'positive' and 'negative' for a binary classification). In this case, the AUC is 0.83, which indicates a good discriminative ability of the model. Generally, an AUC of 0.5 suggests no discriminative ability (equivalent to random guessing), while an AUC of 1.0 indicates perfect discrimination. An AUC in the range of 0.8 to 0.9 is considered very good, so the model associated with this ROC curve is likely to be useful for whatever prediction task it was designed for. (Refer 8.2.3)

4.2 Random Forest Model:

Random Forest model is a kind of machine learning algorithm that excels at managing complex data sets for tasks such as classification (determining which group something belongs to) and regression (predicting a number).

The model explains 50.78% of the variance in the outcome, which suggests that over half of the outcome variability is accounted for by the model. (Refer 8.2.4) The mean of squared residuals, which measures prediction error, is 0.123; lower numbers are preferable, indicating more accurate predictions. This level of variance explanation is good for such models, provided the complexity and predictive challenges of the dataset. The ROC curve represents the performance of a Random Forest classification model with an AUC of 0.78, indicating good predictive ability. While the model is better than random chance at distinguishing between positive and negative classes, there is still potential for improvement to achieve higher sensitivity and specificity. (Refer 8.2.7)

5. Empirical Results:

The confusion matrix and accompanying statistics indicate the performance of a binary classification model. Logistic Regression model demonstrates an accuracy of 76.28%, which exceeds the baseline accuracy (No Information Rate) of 53.75%, implying that the model's predictions are better than random chance. Sensitivity is high for class 0 (87.15%), indicating the model is good at detecting this class, but specificity is lower for class 1 (63.64%), suggesting less effectiveness in identifying class 1. Precision for class 0 is 73.58%, and for class 1 it is 80.99%, indicating how often the model's predictions are correct when it predicts each class. Balanced accuracy stands at 75.39%, reflecting the model's average effectiveness across both classes. McNemar's test has a very low p-value, pointing to a significant difference in misclassification rates between classes, hinting at potential prediction bias toward one class. (Refer 8.2.2)

The accuracy of the Random Forest model is 78.08%, which is better than the baseline accuracy, with a statistically significant p-value indicating this improvement is unlikely due to random chance. Sensitivity and specificity rates are 75.98% and 80.52%, respectively, showing a relatively balanced ability to predict both classes. The positive and negative predictive values are also relatively high (81.93% and 74.25%, respectively), indicating that most predictions are correct. The balanced accuracy is 78.25%, showing that the model performs well on both classes, with the 'positive' class being class 0. (Refer 8.2.6)

6. Conclusions and Recommendations

Selection of technique:

The Random Forest model was chosen over the Logistic Regression model based on its superior performance metrics as shown in the confusion matrices and the summary table (Refer Figure 8.2.8). With a higher overall accuracy of 78.08%, the Random Forest model outperforms the Logistic Regression model, which has a 76.28% accuracy. Despite a lower baseline rate, the Random Forest model still maintains a more favourable balance between sensitivity (80.52%) and specificity (75.98%), indicating a consistent performance across different classes. The Logistic Regression model, while having a higher AUC, suggests better discriminative power; however, the Random Forest model's advantages in accuracy, balance between sensitivity and specificity, and likely better generalization capabilities make it the more suitable choice for the predictive analysis at hand.

Recommendations:

Risk Mitigation Before Conversion: Prioritize patient safety by implementing necessary safeguards and conducting additional studies if a predictive model suggests a high risk of patient flipping (from observation to inpatient status). This proactive approach reduces adverse events and ensures appropriate patient care transitions.

Reducing Flipped Percentage: Implement measures to decrease the percentage of patients being flipped from 45% to 20%. This reduction allows the hospital to efficiently serve more patients, leading to increased revenue (topline) and capacity for treating an additional 570 patients annually.

Process Improvement and Staff Satisfaction: Utilize predictive models to enhance process efficiency, which in turn improves job satisfaction among observation unit (OU) staff. By streamlining operations and reducing unnecessary workload, staff members can focus on delivering quality care, leading to higher morale and retention rates.

Resource Allocation Optimization: By reducing flipping instances and enhancing patient flow, hospitals can mitigate resource misallocation and minimize extended stays in the observation unit. This optimization improves resource utilization, enhances patient throughput, and ultimately leads to better overall outcomes.

Expanding Exclusion Criteria: Expand the criteria for excluding certain patients from observation status, thereby decreasing the likelihood of flipping. This strategic approach helps in better patient classification and ensures that resources are allocated appropriately based on patient needs, reducing unnecessary transitions and improving overall efficiency.

Enhanced Compensation from Insurance Companies: If we can lower the flipping percentage, it will decrease the underutilization of inpatient ward resources, leading to more effective resource allocation. Additionally, patients can be served within the observation unit itself, resulting in treatment compensation based on actual costs, thereby improving reimbursement from insurance companies.

7. References:

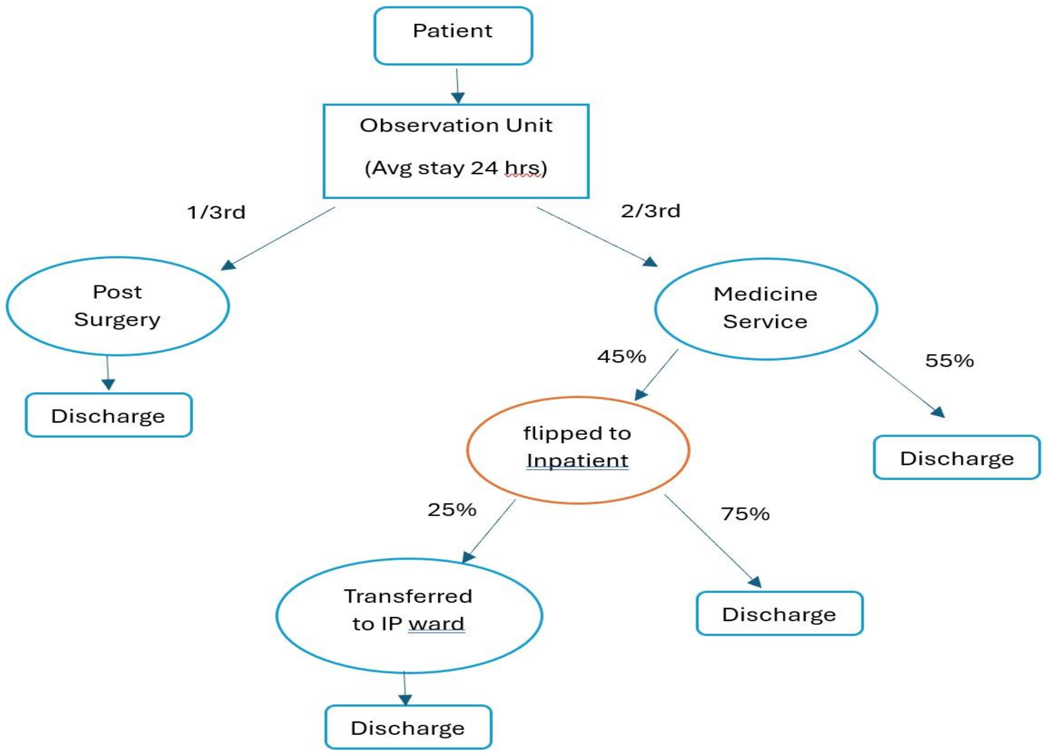
Pachamanova, D., Tilson, V., & Dwyer-Matzky, K. (2022). Case—A Data-Driven Approach to Improving Hospital Observation Unit Operations. INFORMS Transactions on Education, 22(3), 188-194.

8. Appendix:

8.1 Data Dictionary

Variable Name	Description
Age	Age of Patient (in Years)
Gender	Patient gender (Male/Female)
Primaryinsurancecategory	Insurance Provider for Patient
Flipped	Binary Variable that is 1 if the patient "flipped"' from observation status to Inpatient status and 0 if the patient stayed in observation status and was discharged from the OU
OU_LOS_hrs	Length of stay in the OU in hours
DRG01	Initial diagnosis-related group(Code) corresponding to the patient's primary complaint
BloodPressureUpper	Systolic, or upper, blood pressure numbere in mm Hg
BloodPressureLower	Diastolic, or lower, blood pressure numbere in mm Hg
BloodPressureDiff	Difference between systolic and diastolic blood pressure
Pulse	Patient Pulse
PulseOximetry	Measure of Level of Oxygen in patient's blood
Respirations	Number of breaths patient takes per minute
Temperature	Patient's temperature in Fahrenheit

8.2 Process Flow



8.3 Results:

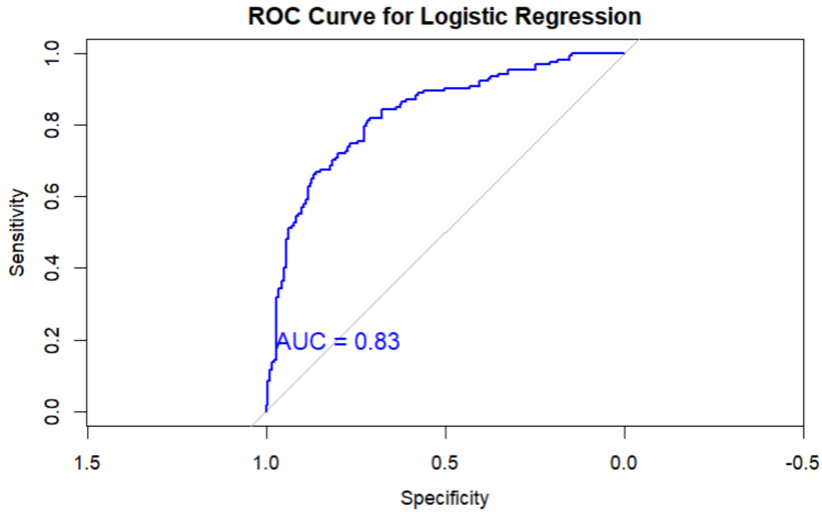
8.3.1 Logistic Regression Model output

	Dependent variable:
	Flipped
Respirations	2.978** (1.479)
PulseOximetry	1.155** (0.454)
Age	-1.223*** (0.393)
DRG01Congestive Heart Failure	0.123 (0.420)
DRG01Pneumonia	0.921* (0.524)
DRG01Colitis	-0.431 (0.537)
DRG01Pancreatitis	-0.321 (0.937)
DRG01GI Bleeding	-0.509 (0.577)
DRG01Urinary Tract Infection	0.207 (0.449)
DRG01Syncope	-1.628*** (0.334)
DRG01Edema	-1.188 (0.857)
DRG01Chest pain	-1.456*** (0.366)
DRG01Nausea	-0.735* (0.435)
DRG01Abdominal pain	-1.779*** (0.378)
OU_LOS_hrs	11.629*** (1.017)
Constant	-1.500*** (0.569)
Observations	778
Log Likelihood	-375.088
Akaike Inf. Crit.	782.176
Note:	*p<0.1; **p<0.05; ***p<0.01

8.3.2 Confusion Matrix for Logistic Regression Model

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 156  56
         1  23  98

               Accuracy : 0.7628
                 95% CI : (0.7133, 0.8074)
    No Information Rate : 0.5375
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.5156

 Mcnemar's Test P-Value : 0.0003179

            Sensitivity : 0.8715
            Specificity : 0.6364
         Pos Pred Value : 0.7358
         Neg Pred Value : 0.8099
             Prevalence : 0.5375
         Detection Rate : 0.4685
   Detection Prevalence : 0.6366
      Balanced Accuracy : 0.7539

       'Positive' Class : 0
```

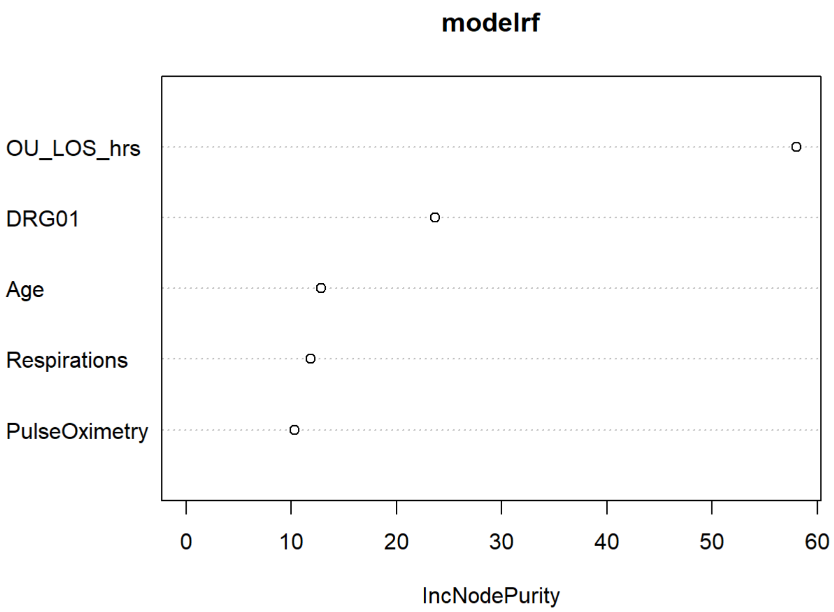
8.3.3 ROC Curve for Logistic Regression Model



8.3.4 Random Forest Model output

```
##
## Call:
##  randomForest(formula = Flipped ~ Respirations + PulseOximetry +      Age + DRG01 + OU_LOS_hrs, data = data.train.bal, nt
ree = 200,      nodesize = 25, importance = TRUE)
##                Type of random forest: regression
##                      Number of trees: 200
## No. of variables tried at each split: 1
##
##           Mean of squared residuals: 0.1230371
##                     % Var explained: 50.78
```

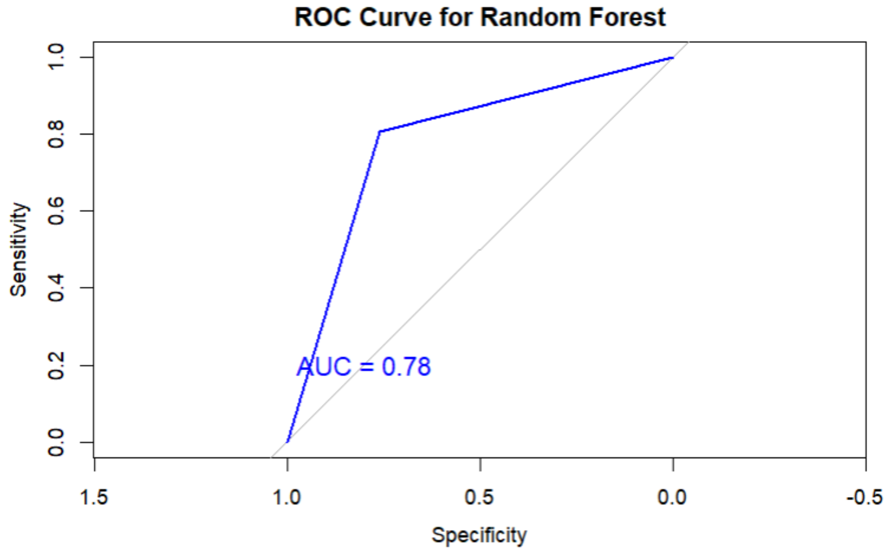
8.3.5 Variable Importance Plot



8.3.6 Confusion Matrix for Random Forest Model

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 136  30
         1  43 124

               Accuracy : 0.7808
                 95% CI : (0.7324, 0.824)
    No Information Rate : 0.5375
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.5617

 Mcnemar's Test P-Value : 0.1602

            Sensitivity : 0.7598
            Specificity : 0.8052
         Pos Pred Value : 0.8193
         Neg Pred Value : 0.7425
             Prevalence : 0.5375
         Detection Rate : 0.4084
   Detection Prevalence : 0.4985
      Balanced Accuracy : 0.7825

       'Positive' Class : 0
```

8.3.7 AUC curve for Random Forest Model



8.3.8 Summary Result table

Model	Accuracy	Baseline	Specificity	Sensitivity	AUC
Logistic Regression Model	76.28%	53.75%	87.15%	63.64%	83%
Random Forest Model	78.08%		75.98%	80.52%	78%

8.4 Data Visualizations:

Figure 1:

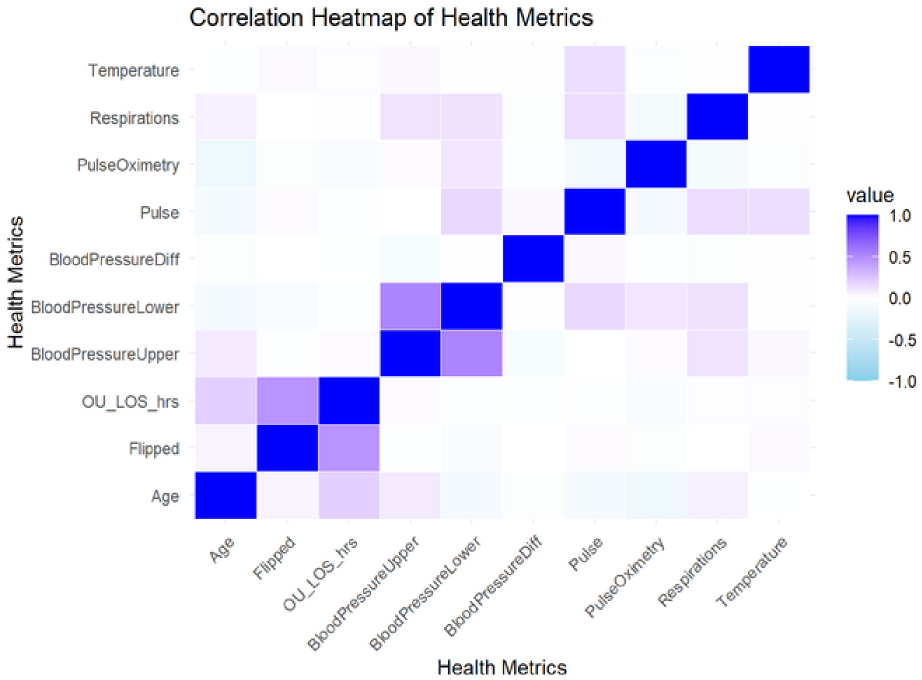
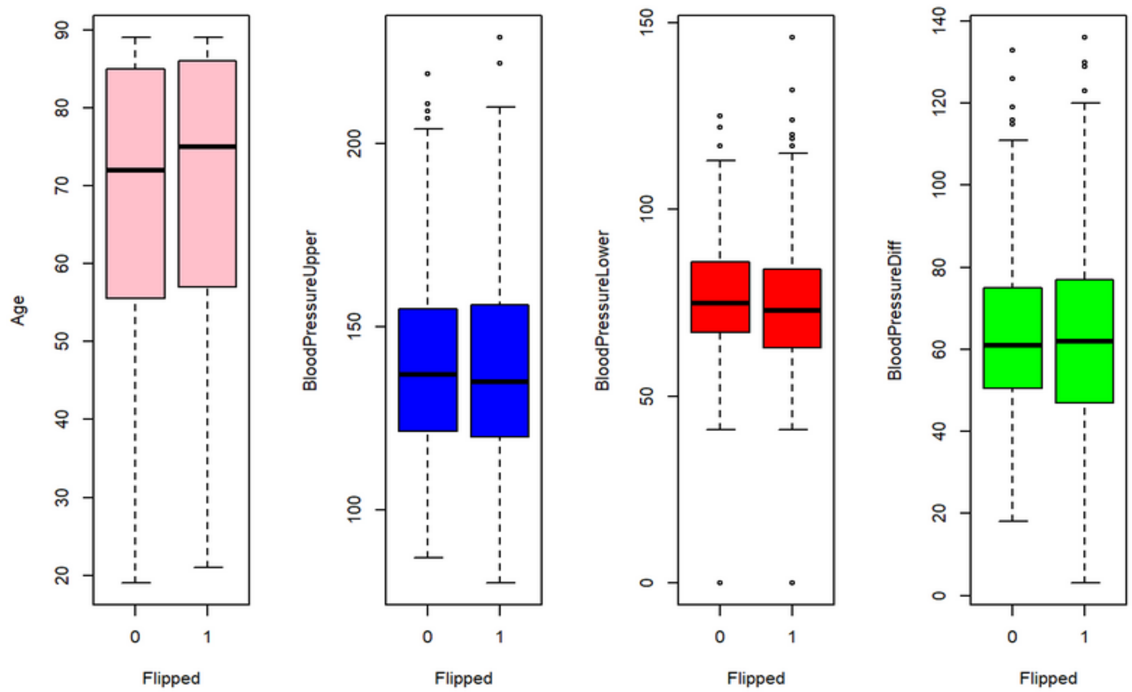


Figure 2:



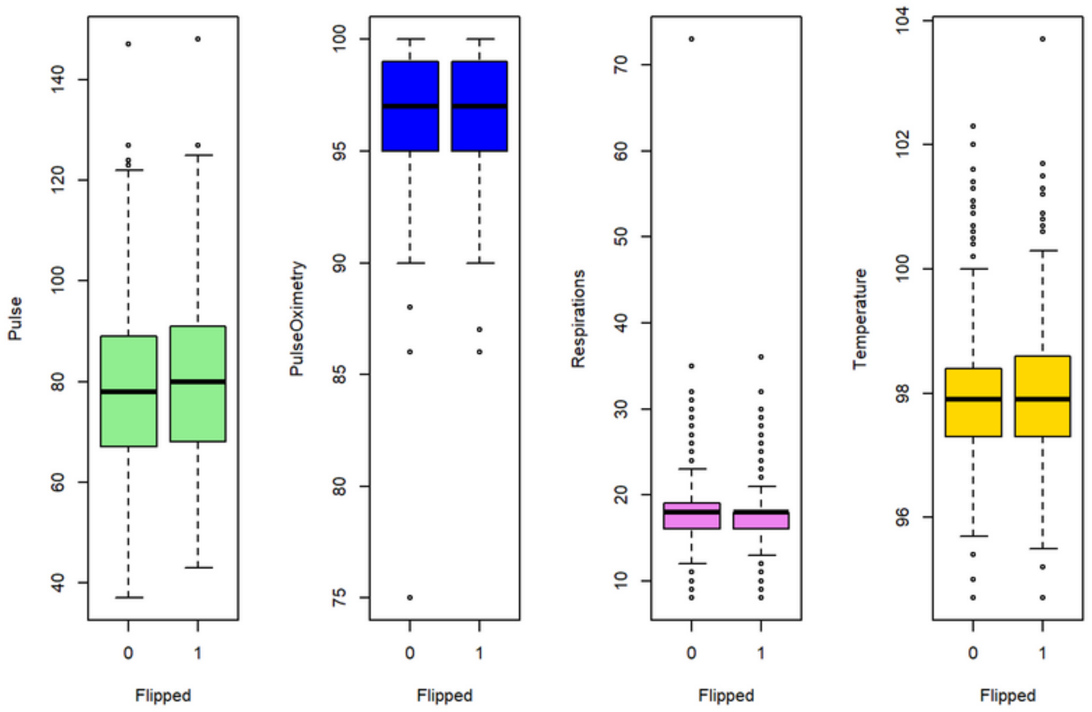
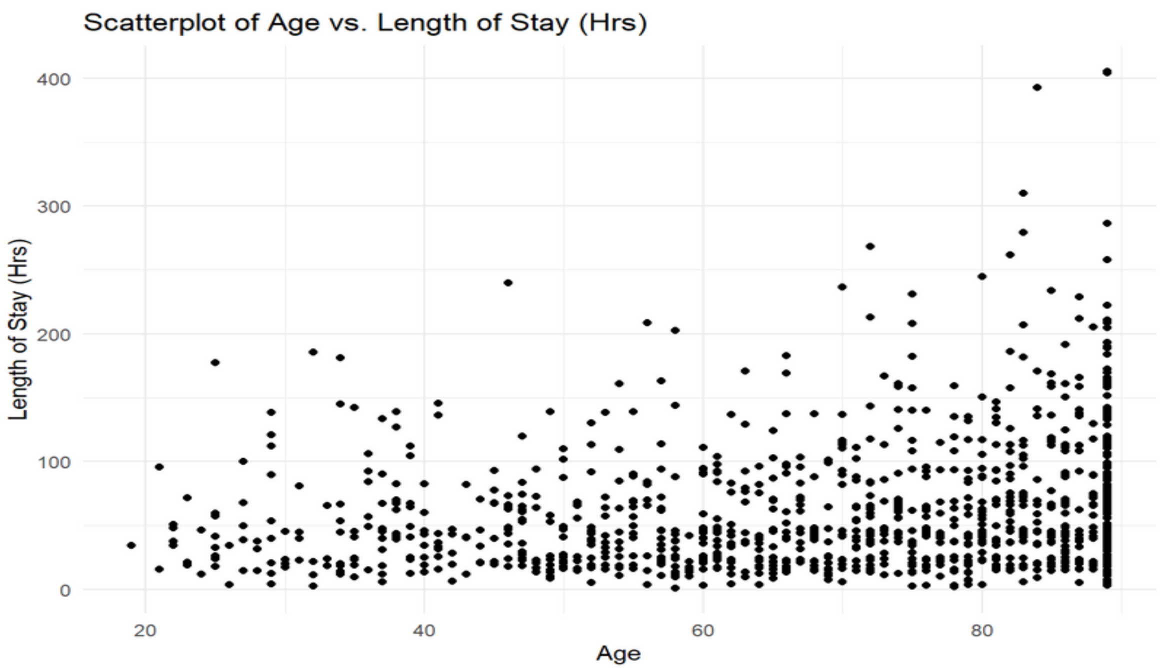

Pulse
PulseOximetry
Respirations
Temperature
Flipped
Flipped
Flipped
Flipped

Figure 3:


Scatterplot of Age vs. Length of Stay (Hrs)
Length of Stay (Hrs)
Age