**A Data-Driven Approach to Improving Monataro Hospital’s Observation Unit Operations**

**Tam Nguyen & Uyen Tran**

**1. Executive Summary**

Montanaro Hospital, under the management of Dr. Erin Kelly, has been facing challenges related to prolonged patient stays and frequent transition from observation to inpatient status, leading to operational inefficiencies and suboptimal resource utilization. In the long-term, this puts a huge constraint on revenue generation of the hospital and employee’s satisfaction. This paper aims to assist Dr. Kelly addressing these issues through a data-driven approach that employs predictive modeling to enhance patient flow and overall operational efficiency. The analysis is arranged in the order of data pre-processing and visualization, modeling and applying both parametric and non-parametric machine learning techniques (Logistic Regression, Classification Tree, Random Forest and Naive Bayes), technique evaluation, implications of the result and recommendations. The results confirm the associations between flipping rate and patient’s age, primary insurance category, length of stay and diagnosis-related group (DRG), with the model demonstrating its ability in reducing flipping rates by 36% and cost per case by 38%. Among the four techniques, Classification Tree is more efficient in predicting flipping cases, with the highest overall specificity of 80.5%. The recommendations include the adoption of the predictive model, modifying the OU exclusion list based on predictors, adjusting allocation of resources based on patient characteristics, integrating technology to enhance data-driven decision-making, and using relevant KPIs to track performance metrics to monitor the model’s impact. The implementation of such recommendations is expected to enhance patient care, streamline operations, and increase overall satisfaction within the OU at Montanaro Hospital.

**2. Problem Description**

The business problem revolves around the operational challenges within the 23-bed observation unit (OU) at Montanaro Hospital, managed by Dr. Erin Kelly. Firstly, the concern over the prolonged average length of stay for patients in the OU. Another problem is the significant number of patients transitioning from observation to inpatient status. The OU, established to efficiently manage lower-acuity cases, is grappling with outdated patient placement protocols and an exclusion list that may not accurately capture appropriate cases. This situation poses inefficiencies, with patients being placed in inpatient beds due to OU capacity limitations. The impact includes suboptimal resource utilization, prolonged hospital stays, and potential medical errors during patient handoffs. Dr. Kelly aims to address these challenges through a data-driven approach, employing a predictive model to enhance the OU exclusion list and ultimately improve patient flow, capacity, and overall operational efficiency.

**3. Methodology**

**Data Pre-Processing**

In preparation for subsequent data analysis, it is essential to conduct thorough data cleaning and preprocessing. We evaluate the dimensions and characteristics of the dataset by assessing the data types of variables and perform summary statistics to gain an initial overview. The summary results reveal the presence of missing values, potential outliers deduced from maximum and minimum values, and several inconsistencies in some variables. Besides, there are many variables at are not in the correct data type (i.e. “BloodPressureUpper” is listed as a character, when it is in fact a numerical variable). Thus, we converted them to the correct form and used some functions such as ‘is.na’ and ‘duplicated’ in R to identify the dataset for missing values and duplicates. Subsequently, we have missing values in ‘BloodPressureUpper’, ‘BloodPressureDiff’, ‘PulseOximetry’, ‘Respirations’ and ‘Temperature’ variables. Since these variables are numeric variables with skewed distribution, we treated these missing values by median imputation.

**Data Visualization**

To identify the patterns, correlation, outliers present in the data and derive meaningful insights into the patient dynamics within the OU, we created a set of visualizations that are instrumental for flipping rate prediction.

1. **Correlation Matrix Heatmap (Appendix 1)**

The heatmap provides a snapshot of the correlation among numerical variables, with darker shades indicating stronger correlation. Overall, there is no significant correlation among most of the variables, except for a relatively weak positive correlation between blood pressure upper and blood pressure lower (0.53), which is to be expected as these measurements are physiologically related.

1. **Bar chart**
2. **Proportion of Patients Flipped to Inpatient Status by DRG Code (Appendix 2)**

From the chart, patients who are initially diagnosed with Pancreatitis (DRG Code 577) and Colitis (DRG Code 558) are most likely to require escalated care, indicating higher flipping rate for such diagnoses. On the other hand, patients with Abdominal Pain are least likely to change status from OU to Inpatient. This suggests that more staff and resources should be allocated for patients with diagnoses that imposes higher chances for flipping status.

1. **Flipped Status Percentage by Primary Insurance Category (Appendix 3)**

The chart indicates that type of insurance can be a significant predictor of flipping rate due to an opposing pattern between a certain insurance company and the rest. In this case, those who have Medicare insurance are associated with greater status change possibility.

1. **Percentage of Flipped Status based on Gender (Appendix 4)**

The recorded proportion of male who have flipped from OU to Inpatient is slightly higher than that of females (50% versus 38%). The difference is unsubstantial between two genders; thus, gender might not be a significant indicator of flipping rate.

1. **Boxplots distribution of Pulse Oximetry, Length of Stay, Blood Pressure Difference and Temperature for “Flipped” and “Non-Flipped” groups. (Appendix 5)**

Among the four box plots that demonstrate the influence of numerical variables on flipping rate, length of stay is the most separated between two groups, indicating it being a potential predictor. Besides, the plots suggest the presence of many outliers. In this context, we decided to not remove them as they are valid patient’s information and can induce meaningful insights for the analysis.

1. **Scatter Plot of Pulse Oximetry and Length of Stay (Appendix 6)**

The scatter plot provides insight into the relationship between pulse oximetry and patient’s length of stay in the OU. The charts suggest that lower length of stay is associated with lower flipping rate. However, there is no relation between pulse oximetry and length of stay that can indicate whether or not a patient will flip from OU to Inpatient status.

**Modeling**

To identify or classify potential patients, we are using a predictive model where we have tried to classify patients who were more likely to flip from observation to inpatient status. To do this, we utilize methodologies such as logistic regression, decision tree, and random forest as our target variable is categorical. Besides, we have considered “Age”, “Gender”, “PrimaryInsuranceCategory”, “DRG01”, “OU\_LOS\_hrs”, “BloodPressureUpper”, “BloodPressureLower”, “BloodPressureDiff”, “Pulse”, “PulseOximetry”, “Respirations”, “Temperature” as a independent variables.

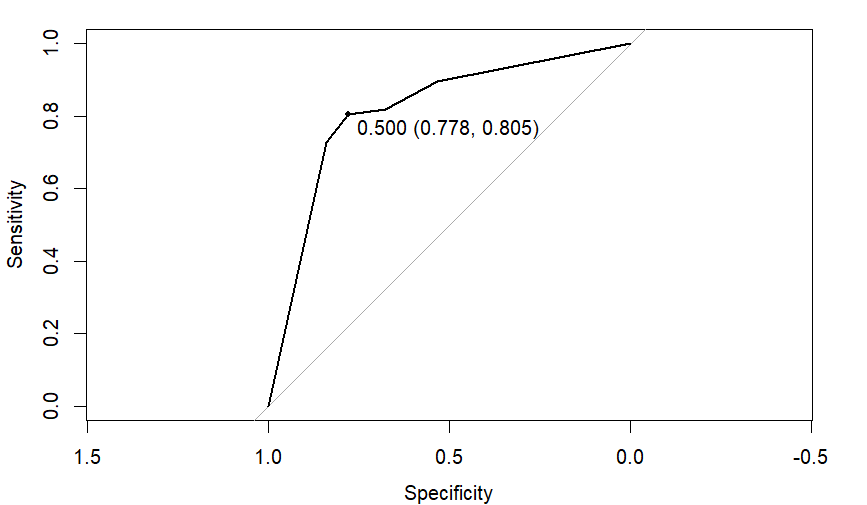
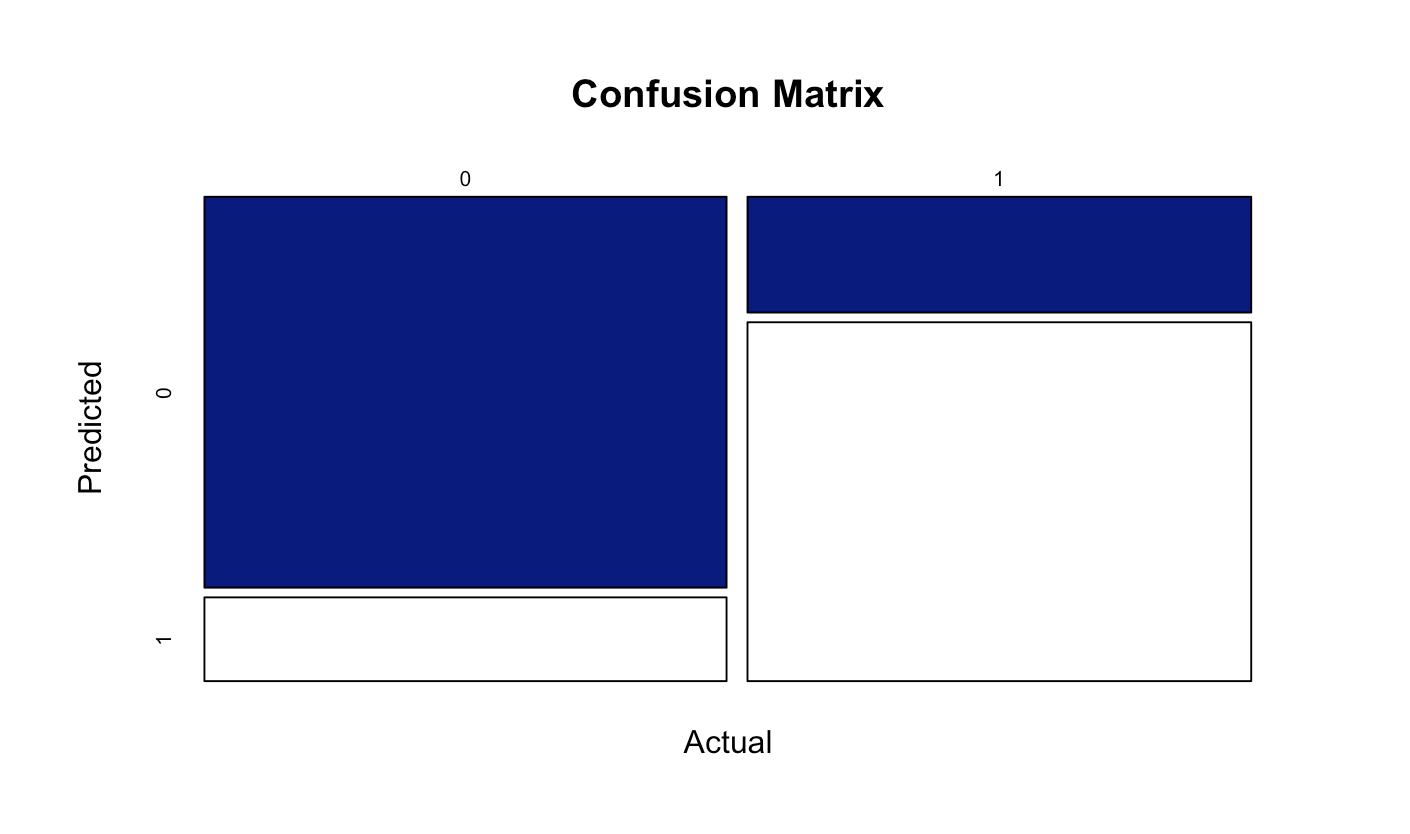
Firstly, we run a full logistic model with all potential variables and use the Stepwise Selection Method to identify the important predictors. The outcomes revealed that “Age”, “DRG01”, “PrimaryInsuranceCategory” and “OU\_LOS\_hrs” exhibited statistical significance. Therefore, we only keep these as our predictors and eliminate the insignificant variables from the model for better prediction. In this case, the dependent variable is “Flipped”, while the chosen independent variables are “Age”, “DRG01”, “PrimaryInsuranceCategory”, “OU\_LOS\_hrs”. The model is written as below:

*P(Flipped | Age, Primary Insurance Category, DRG Code, Length of Stay)*

**4. Results**

The reduced model is then applied and evaluated on the test set using all techniques. Decision Tree emerged as the most suitable choice based on its higher accuracy of 79.04% and specificity of 80.52% according to the confusion matrix.

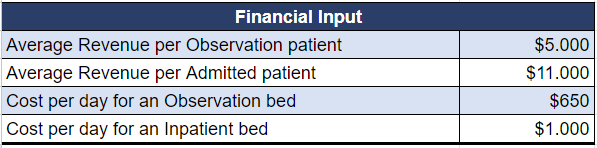
The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) at a threshold of 0.5, 80.52% of patients who were actually discharged were correctly identified as such by the model and the model correctly identified 78% of patients who were actually admitted. Moreover, the AUC value is 0.816. This indicates that, on average, the model has an 81.6% chance of ranking a randomly chosen positive instance higher than a randomly chosen negative instance.



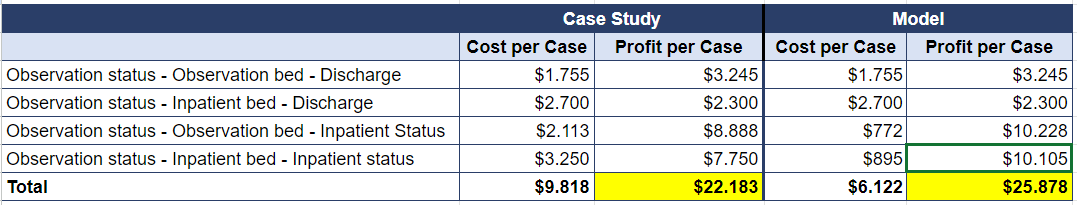
The coefficient of DRG in the logistic regression model (Appendix 7) shows that Colitis and Pancreatitis have higher association with the flipped rate. That means patients in these groups are associated with flipped rate, which aligns with the visualizations in actual data. Besides, the true positive rate by DGR (Appendix 8) from the confusion matrix of Classification Tree shows the ability of the model in predicting the percentage of difference by DRG. We can see the highest percentage derive from Urinary Tract Infection and Dehydration.

To compare the profit between the case study and model, we collected data from the case study and calculated the OU patient flow using the model and tree rules derived from the classification tree (Appendix 9). A noticeable change can be seen in the percentage of changes to inpatient status decreasing from 45% to 7%. Meanwhile, the average OU length of stay in transferring to an inpatient ward decreased to 0.35 days.

It is crucial to consider various revenue and cost components for calculating profit. We assumed the following financial input:



The result shows that the model resulted in reduced costs in the case where observation patient is placed in observation bed and then change to inpatient status, as well as where observation patient is placed in inpatient bed and then change to inpatient status. This has contributed to an overall decrease in the total cost per case to $6,122 in the model, leading to greater the profit using the model as compared to the current model of the hospital.



**5. Recommendation**

* **Adopt the Predictive Model**

Compared to a random baseline, the model was able to reduce the flipping rate from 45% to 9%, underlying substantial decline in operational disruption and improvement in patient care. The cost per case decreased from $9,818 to $6,122 (a 38% decrease), which is $18,855 per day and $6,882,352 per year, assuming that the hospital operates every year. Such reduction will lead to improved workflow, increase OU staff satisfaction, and eventually enhance patient’s experience.

On the other hand, the clinic should modify the OU exclusion list based on the most-likely-to-flip Diagnosis-Related Group using the logistic regression result. Specifically, considering adding Pancreatitis (577), Colitis (578), Urinary Tract Infection (599), Dehydration (276) and Congestive Heart Failure (428) to the exclusion list due to their positive association with the likelihood of flip rate (Appendix 7, Appendix 8)

* **Resource Allocation**

With a lower flipping rate and an increase in lower-acuity observation patients, the workload and time required for OU nurse and inpatient staff would require adjustment. The clinic can utilize the predictive model’s insights regarding patient demographics, primary insurance company and DRG Code to adjust staffing levels accordingly. Particularly, allocating extra nurse and support personnel during periods when patients with characteristics associated with higher likelihood of status flip are expected (i.e. when the patient DRG code falls in one of the specified categories, or when they have Medicare Insurance) and intensifying monitoring efforts for those patients. This can contribute to increased patient safety, as it ensures that individuals with greater likelihood of status change receive appropriate levels of care and monitoring, thus reducing the risk of adverse cases and improving patient’s outcomes.

Additionally, by aligning staffing levels with the anticipated needs, the hospital can achieve the right staff-to-patient ratio, which helps to reduce stress and burn out among OU staff. Assigning staff members to care for the patients whose characteristics correspond to their specialized field helps them to feel a greater sense of recognition. Such acknowledgement can boost morale and increase retention rate for the overall clinic.

* **Technology Integration**

Montanaro hospital is recommended to implement a robust electronic health record (EHR) system with real-time data analytics capabilities to capture patient’s ongoing data within the OU. Next, ensuring that the predictive model is fully incorporated into the EHR, and integrate the updated EHR system with the existing databases to facilitate data exchange and interoperability across departments and care settings. Upon the completion of the new system, it is essential to provide hands-on training sessions for staff members to familiarize them with the utilization of EHR system in decision-making, while also acknowledging them on data integrity, privacy, and security protocols to maintain compliance with regulatory requirements (i.e. protect patient information). The integration of a comprehensive EHR system with predictive algorithm and real-time analytics is expected to leverage risk assessment, treatment planning and resource allocation within the OU.

* **Performance tracking**

To track the model’s impact on flipping rates, patient outcomes, and operational costs, Montanaro can use several key performance indexes (KPIs) such as the percentage reduction in flipping rate, length of stay, readmission rate or mortality rate. By assessing these KPIs, the hospital can effectively measure the impact of the predictive model on flipping rate and patient outcomes, enabling data-driven decision making and continuous improvement efforts to optimize care delivery within the OU.

By optimizing resource utilization, reducing unnecessary admissions, and improving patient throughput, the integrated EHR system contributes to cost savings for the hospital. Additionally, enhanced patient care and outcomes may attract more patients to the hospital, leading to increased revenue generation over time.

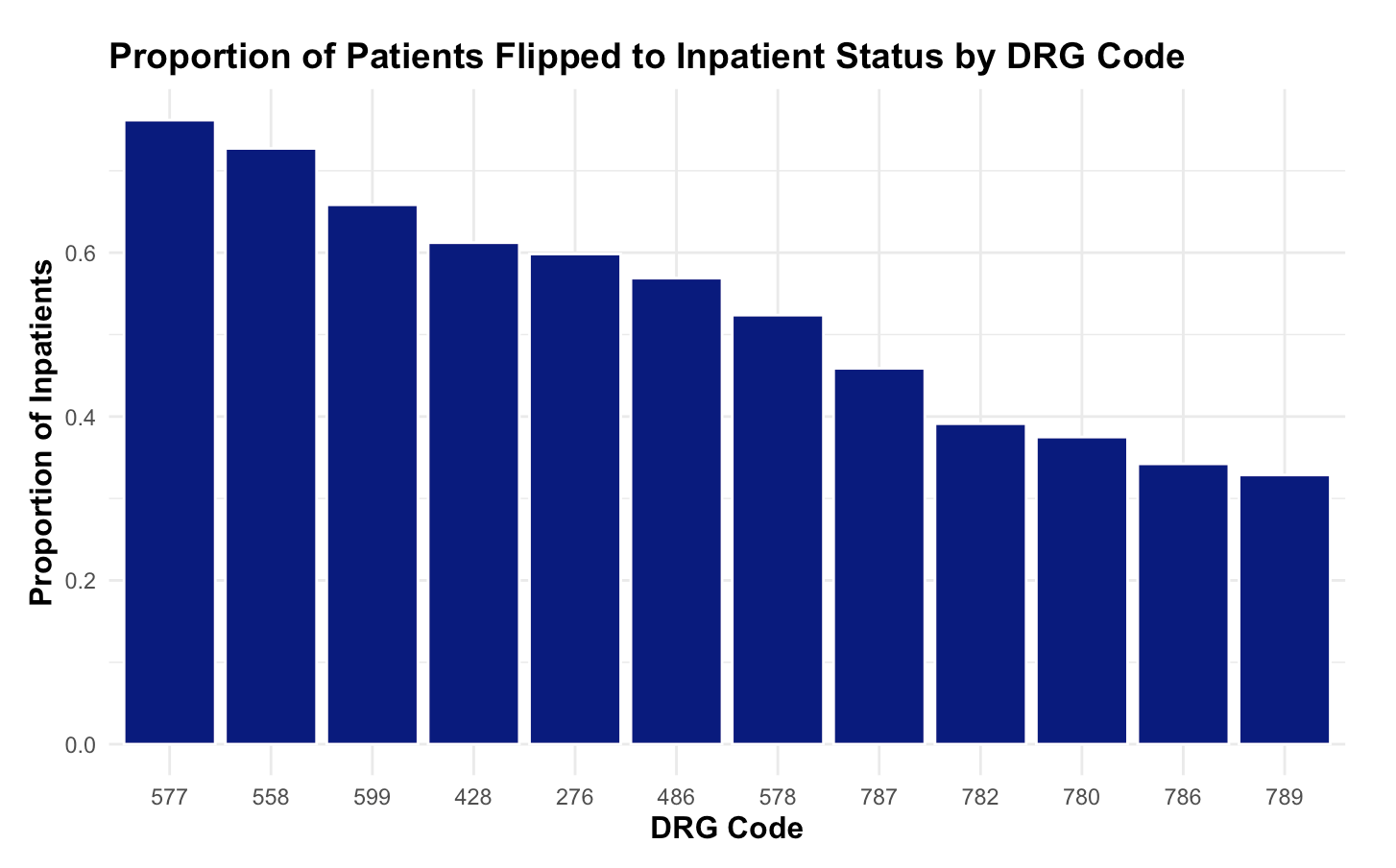
**Appendix**

**Appendix 1 - Correlation Heatmap**

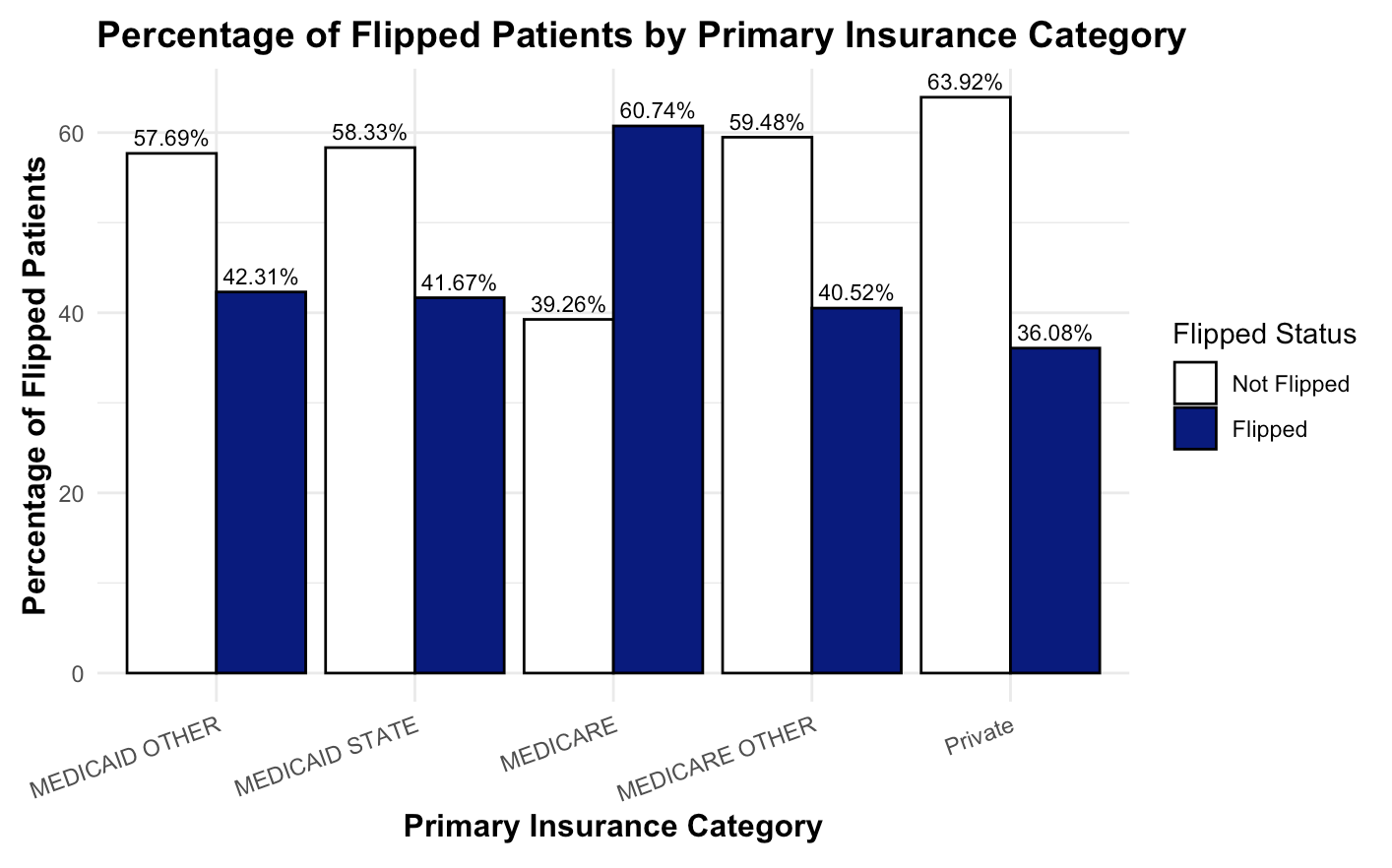
A graph showing the results of a blood pressure test

Description automatically generated

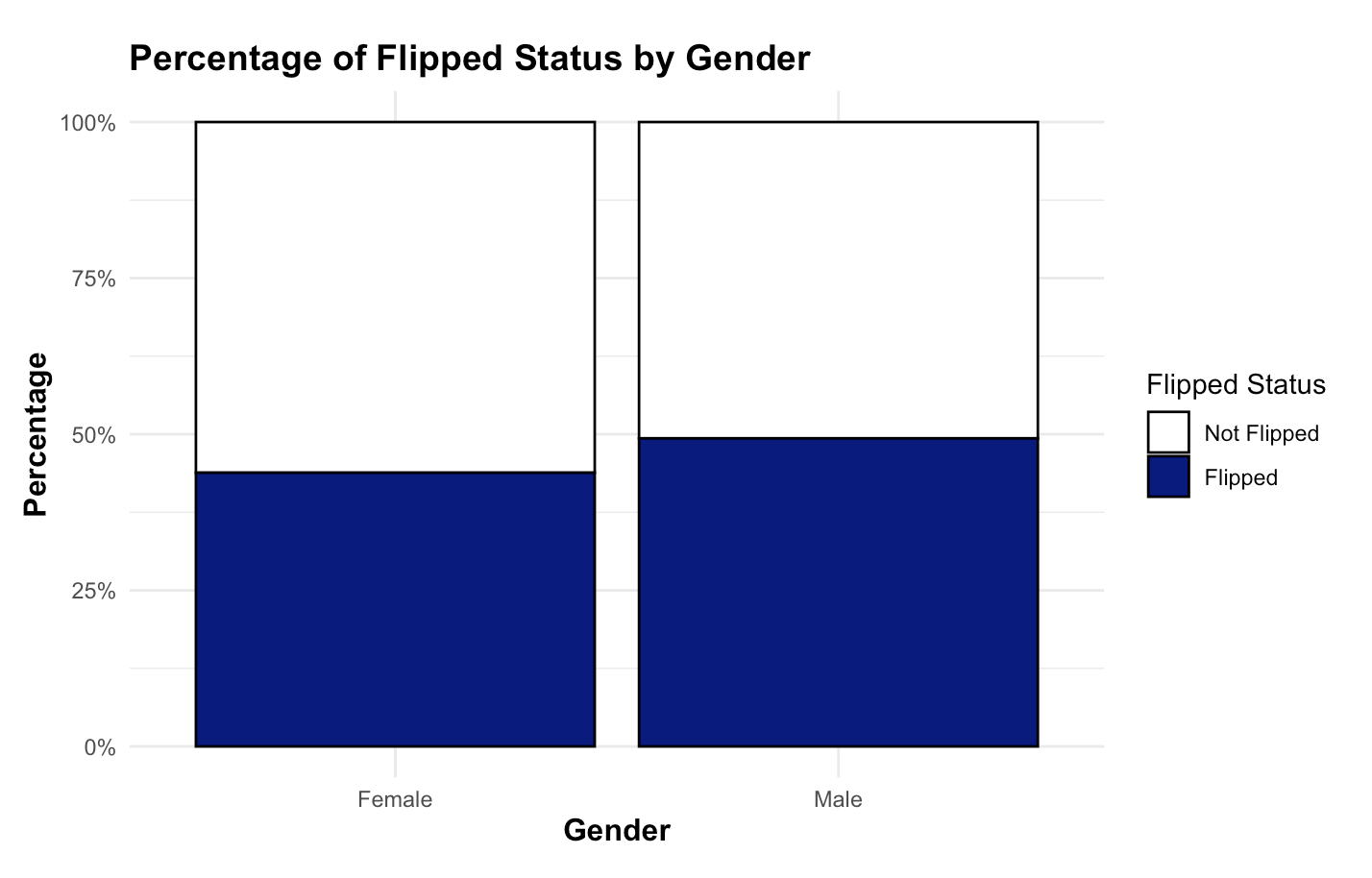
**Appendix 2 - Bar Chart of the proportion of flipped patient by DRG code**



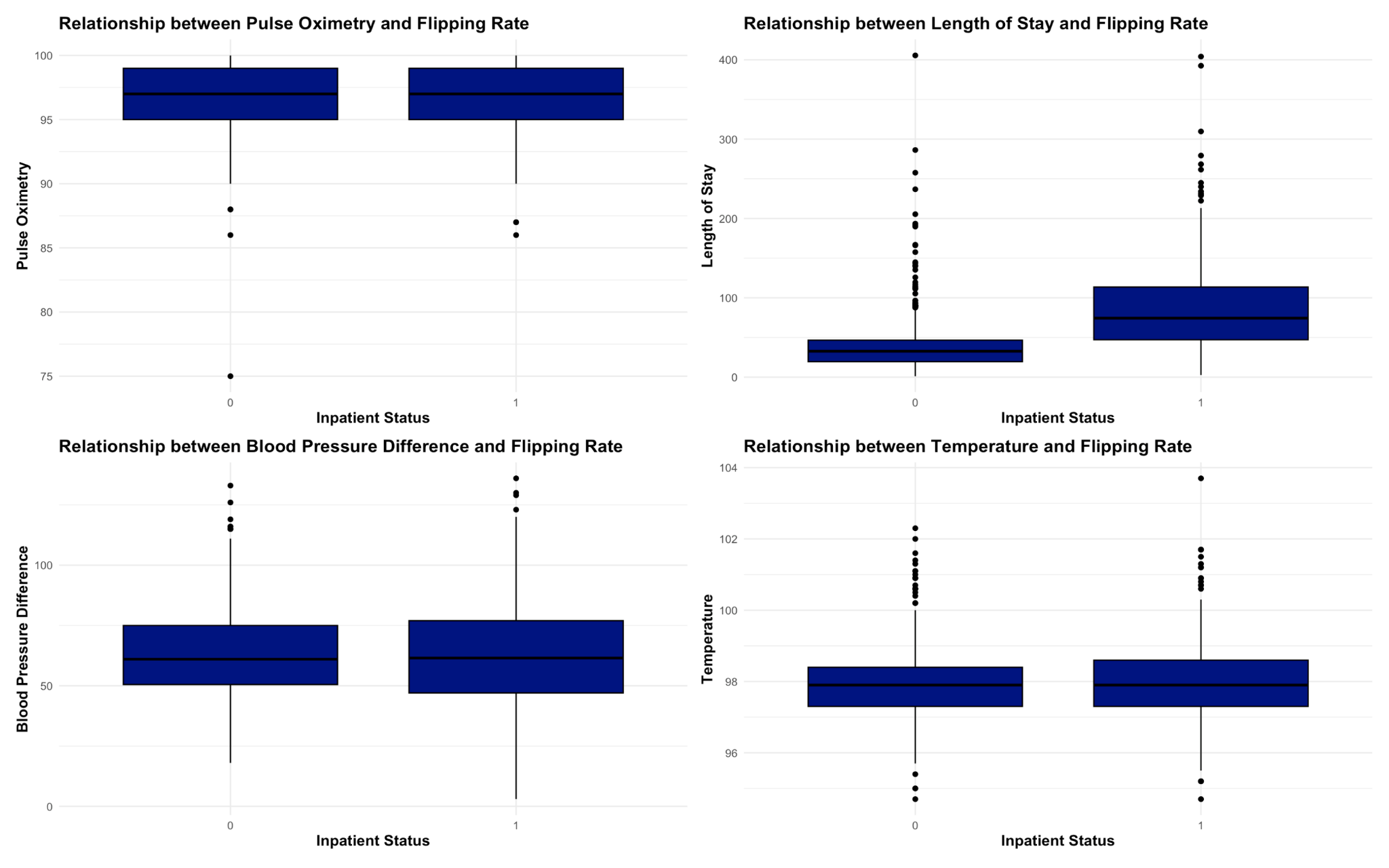
**Appendix 3 - Bar Chart of flipping proportion by primary insurance company**



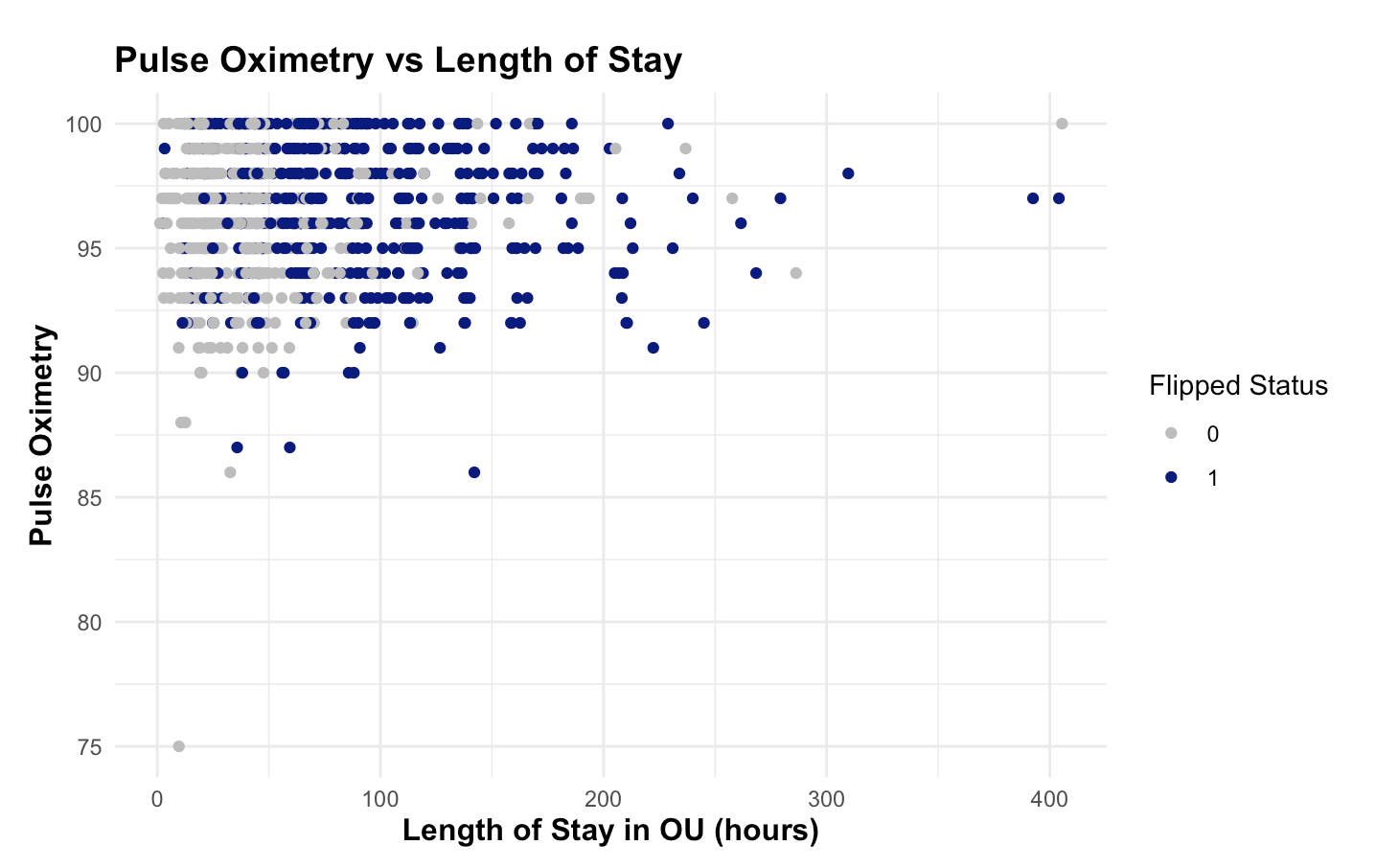
**Appendix 4 - Proportion of Flipped and Non-Flipped patient by Gender**



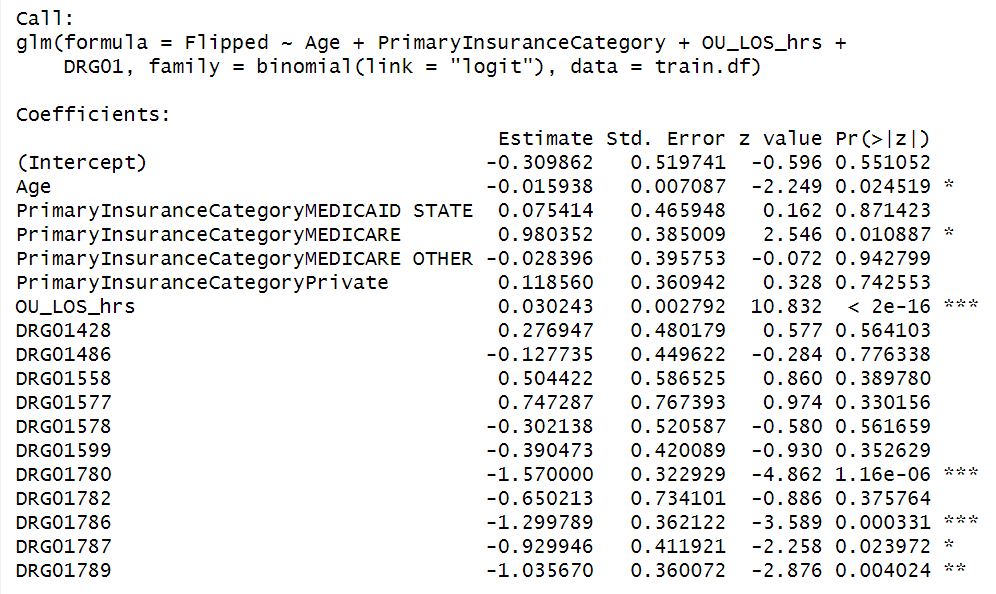
**Appendix 5 - Box plots of Pulse Oximetry, Length of Stay, Blood Pressure Difference and Temperature for “Flipped” and “Non-Flipped” groups**



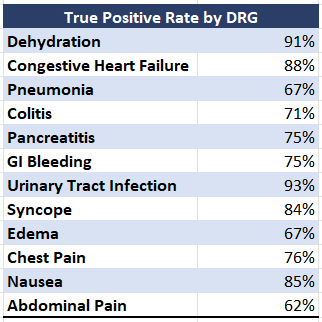
**Appendix 6 - Scatter Plot of Pulse Oximetry and Length of Stay, colored by Flipping Status**



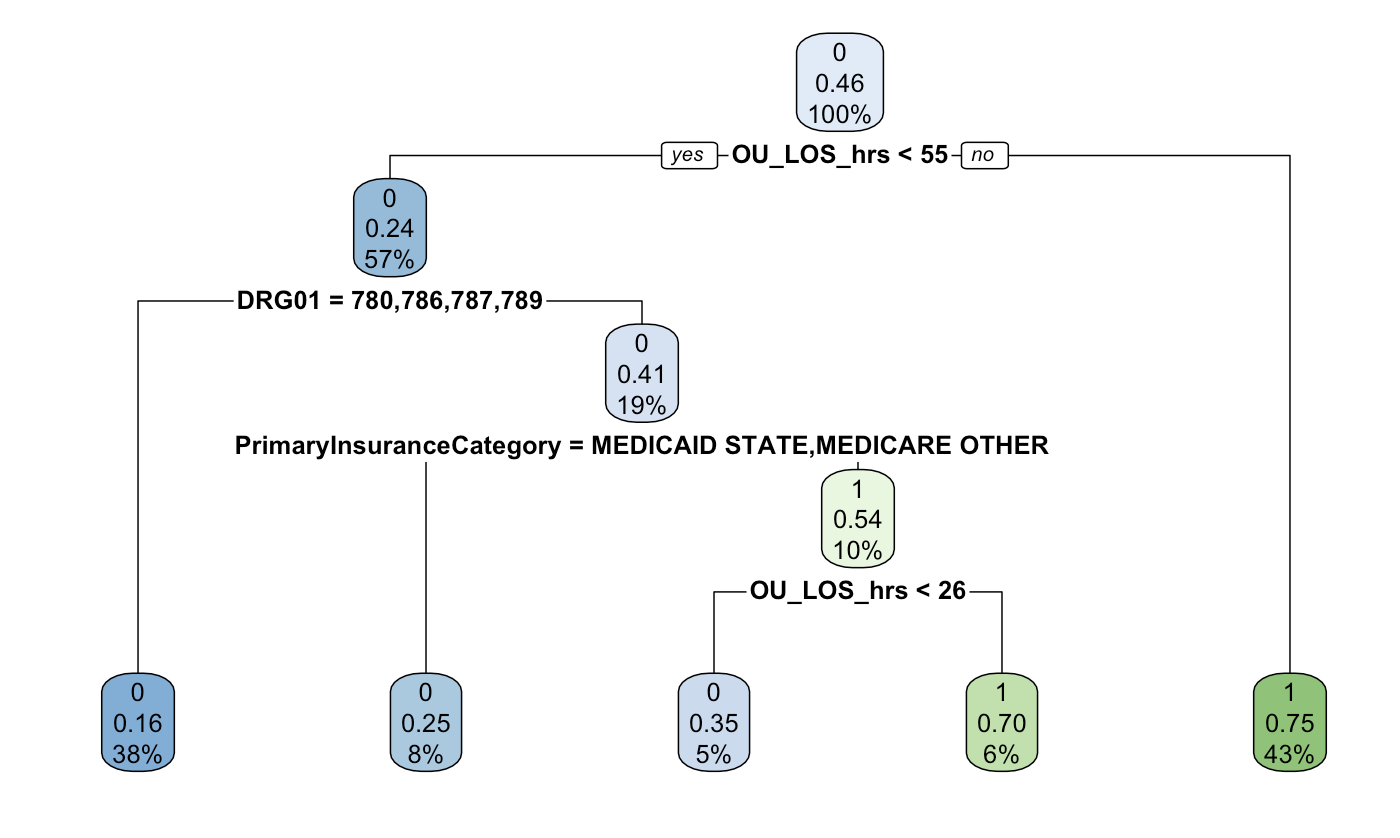
**Appendix 7 - Logistic Regression Result Using Stepwise Selection Method**



**Appendix 8 - True Positive Rate by DGR**



**Appendix 9 - Classification Tree**

****