A Data-Driven Approach to Improving Hospital Observation Unit Operations: Optimizing Patient Transfers and Enhancing Efficiency in Montanaro Hospital's Observation Unit

CHAPTER - 1

INTRODUCTION

1.1 SUMMARY

This project focuses on improving the operational efficiency of the 23-bed Observation Unit (OU) at Montanaro Hospital. The key issue addressed is the frequent transfer of patients from observation status to inpatient care, which leads to increased length of stay and inefficient use of hospital resources. Using data-driven approaches, including predictive analytics and machine learning techniques, this study aims to identify the primary factors that contribute to these patient transfers.

By analyzing patient data from the hospital's Electronic Health Records (EHR), the study examines variables such as patient demographics, diagnoses, medications, and vital signs. These variables are used to build predictive models that can guide hospital management in optimizing bed capacity, reducing overcrowding in the Emergency Department (ED), and improving overall patient flow.

The outcomes of this project are expected to provide actionable insights for hospital administrators to streamline observation unit operations, enhance resource allocation, and improve patient care. These improvements are not only beneficial for Montanaro Hospital but can also serve as a model for other healthcare institutions looking to enhance their observation unit processes.

1.3 PROBLEM STATEMENT

The problem statement of this study is:

- What are the key factors influencing patient transfers from observation status to inpatient status in the Montanaro Hospital Observation Unit?

This problem statement, along with previous research and hospital operational data, forms the foundation for the analysis conducted in this study. By addressing this issue, we aim to improve resource utilization and patient flow efficiency within the hospital.

CHAPTER – 2

METHODS

2.1 DATA SET

The dataset for this project was provided by the professor, which includes patient records from the Observation Unit (OU) at Montanaro Hospital. This dataset comprises various patient attributes such as vital signs, demographics, and hospital stay characteristics, offering a comprehensive foundation for our analysis. The dataset plays a pivotal role in predicting patient transfers from observation to inpatient status, which is the core focus of this study.

2.1.1 DATA CLEANING & TRANSFORMATION

1. **Data Cleaning:** The dataset was cleaned by converting character variables to numeric where applicable, removing non-essential variables like ”ObservationRecordKey” and imputing missing values with mean values for numerical variables.
2. **Feature Engineering:** Categorical variables such as ”Gender” and ”PrimaryInsuranceCategory” were converted to factor variables to make them suitable for predictive modeling. Dummy variables were created for categorical variables to ensure proper representation in the models.
3. **Missing Value Imputation:** Missing data were handled using techniques such as mean imputation. For the ”BloodPressureDiff” variable, missing values were computed using the difference between systolic (upper) and diastolic (lower) blood pressure reading

2.1.2 DATA DICTIONARY

2.1.3 CATEGORICAL GROUPING

The dataset contains both quantitative and qualitative variables. Quantitative variables such as temperature, pulse, and respiration rates are used directly for analysis. Qualitative variables like gender and insurance category are converted into factors for better processing. Additionally, some variables required grouping or transformation. For instance, the variable ”Flipped,” which indicates whether a patient’s status was transferred, was recoded from numerical (0 and 1) to categorical (”No” and ”Yes”). This transformation ensures better interpretability of the results.

Column (Variable) Name	**Definition**
Age	Age of patient, in years
Gender	Patient gender (recorded as Male/Female)
PrimaryInsuranceCategory	Insurance provider for the 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
BloodPressureLower	Diastolic, or lower, blood pressure number in mm Hg
BloodPressureUpper	Systolic, or upper, blood pressure number in mm Hg
BloodPressureDiff	Difference between systolic and diastolic blood pressure
Pulse	Patient pulse
Pulse Oximetry	Measure of level of oxygen in patient’s blood
Respirations	Number of breaths patient takes per minute
Temperature	Patient’s temperature in Fahrenheit

Table 1: Variable Names and Definitions

2.1.5 EXPLORATORY DATA ANALYSIS

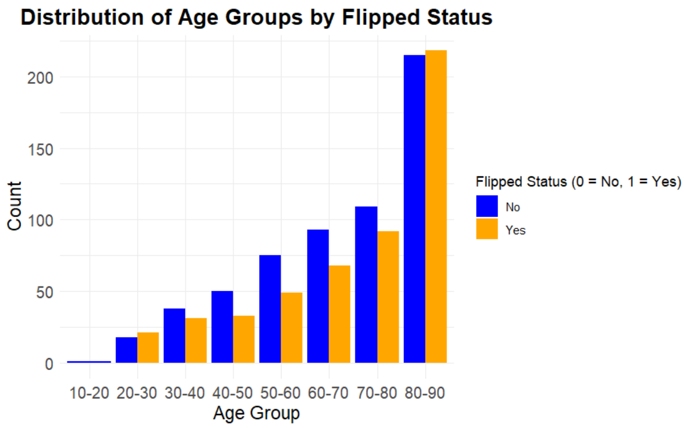

Figure 1: Distribution of Age Groups

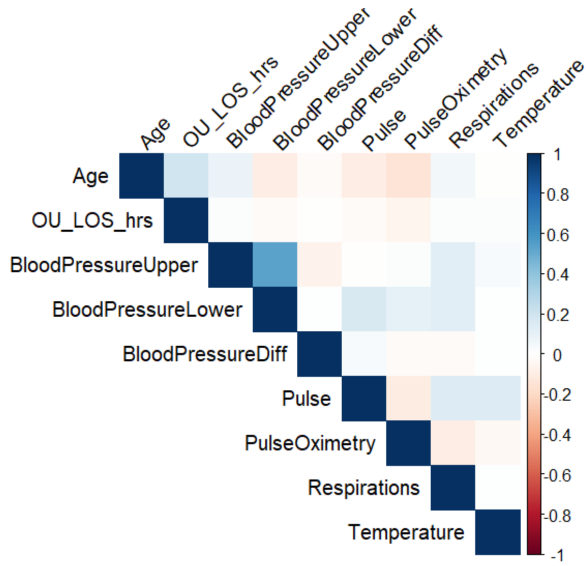

Figure 4: Correlation Matrix of Variable

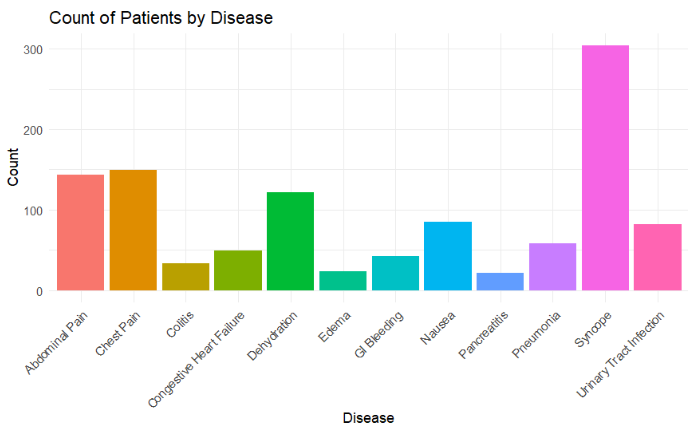

Figure 2: Count of Patients by Disease

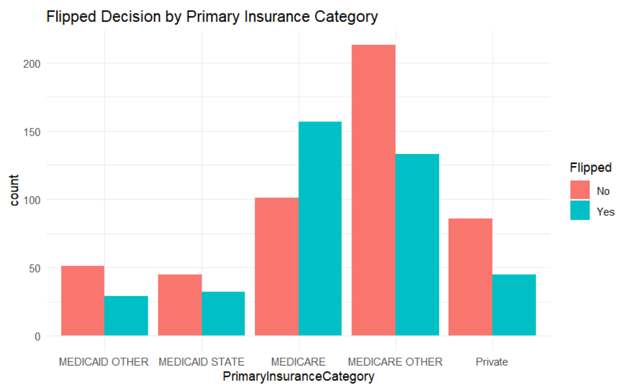

Figure 5: Flipped Decision by Primary Insurance Category

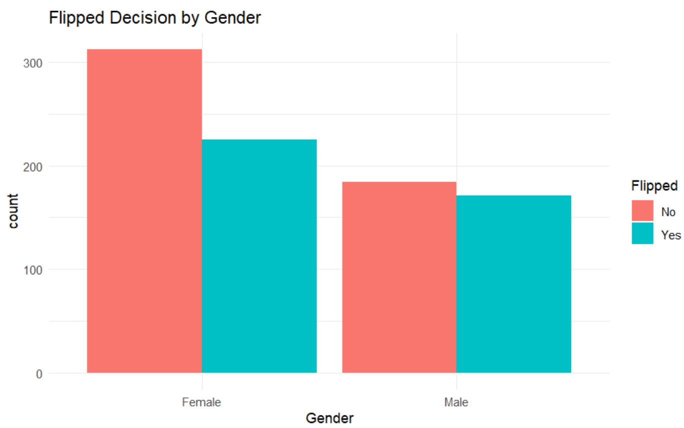

Figure 3: Flipped Decision by Gender

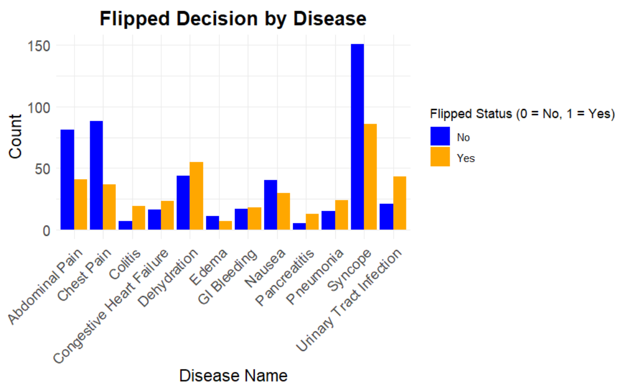

Figure 6: Flipped Decision by Disease

2.1.6 FINDINGS IN THE EDA

1. Distribution of Age Groups by Flipped Status

The chart shows that older age groups (60 and above) have higher rates of being transferred to inpatient status (Flipped = Yes) compared to younger age groups. In particular, the 80-90 age group has nearly equal numbers of transferred and non-transferred patients, while younger groups (below 50) are predominantly not flipped. Overall, the likelihood of being flipped increases with age, indicating a potential relationship between age and the need for inpatient care.

2. Correlation Matrix of Variables

This heatmap displays the correlation between various patient variables. Strong positive correlations (shown in darker blue) exist between BloodPressureUpper, BloodPressureLower, and BloodPressureDiff, indicating that these measures are closely related. Additionally, Pulse and OU_LOS_hrs show some moderate correlations with other variables, while Temperature shows a slight negative correlation with several factors, notably Respirations. Overall, the chart highlights relationships between key vitals, with blood pressure measures showing the strongest inter-variable correlations.

3. Count of Patients by Disease

This bar chart illustrates the distribution of patients across different diseases. Syncope has the highest patient count, with over 300 cases, followed by Abdominal Pain and Chest Pain, each with approximately 100-150 patients. Other diseases, such as Dehydration and Nausea, also have significant patient counts, while conditions like Congestive Heart Failure, Pancreatitis, and Pneumonia have relatively fewer cases. Overall, Syncope stands out as the most common reason for observation unit admissions in this dataset.

4. Flipped Decision by Gender

This bar chart shows the distribution of patient transfer decisions (Flipped = Yes or No) by gender. Female patients have a higher overall count compared to males, with a notable difference in the number of patients who were not transferred (Flipped = No). Among male patients, the number of those transferred (Flipped = Yes) and not transferred is nearly equal. In contrast, for female patients, a significantly larger portion were not flipped to inpatient status. This indicates that the flipping decision varies by gender, with female patients less likely to be transferred compared to male patients.

5. Flipped Decision by Primary Insurance Category

This chart shows the relationship between primary insurance categories and the decision to transfer patients (Flipped = Yes or No). Medicare Other has the highest number of patients who were not transferred (Flipped = No), while Medicare has a more balanced distribution of transfers and non-transfers. Medicaid State and Medicaid Other show similar patterns, with a smaller patient count but more patients not being flipped. Private insurance has a relatively low count of patients, with a slight preference for non-transfer decisions. Overall, patients covered by Medicare Other and Medicare represent the largest groups, and they are more likely to remain in observation rather than being transferred.

6. Flipped Decision by Disease

This chart displays the relationship between specific diseases and the decision to transfer patients (Flipped = Yes or No). Syncope has the highest number of patients, with most not being transferred (Flipped = No). Abdominal Pain and Chest Pain also have significant patient counts, with a majority not flipped to inpatient status. Dehydration and Urinary Tract Infection show a more balanced distribution between flipped and non-flipped cases, while Congestive Heart Failure and GI Bleeding have more patients being transferred to inpatient care. Overall, Syncope, Abdominal Pain, and Chest Pain are the most common diagnoses, but they have lower transfer rates compared to more severe conditions like Congestive Heart Failure and GI Bleeding.

2.2 MODELING

The following models were applied to predict the "Flipped" status (whether a patient was transferred from observation to inpatient):

Logistic Regression

Logistic regression was used to model the binary outcome variable "Flipped" (Yes/No). This model estimates the probability of a patient being transferred based on input features such as age, blood pressure, and pulse. The logistic function was applied to map the predictions to a range between 0 and 1, providing the likelihood of a patient's transfer.

Decision Tree

A decision tree model was implemented to classify the patient's likelihood of transfer based on the decision rules derived from the independent variables. This model was intuitive and easy to interpret, especially when visualized, providing clear decision paths that highlighted the most influential variables, such as the patient's length of stay and primary insurance category.

Random Forest

The random forest model was employed to improve prediction accuracy by building multiple decision trees and averaging their predictions. Random forest works by creating multiple subsets of data and performing decision tree classification on each subset. It provided better accuracy compared to individual decision trees by reducing the risk of overfitting.

2.3 PERFORMANCE CRITERIA

To evaluate the performance of the models, several metrics were used:

Accuracy

The percentage of correct predictions made by the model. This metric was crucial for assessing the overall model performance.

Precision and Recall

Precision measured the accuracy of the positive predictions (patients predicted to transfer), while recall measured the ability of the model to identify all patients who were actually transferred.

Confusion Matrix

A confusion matrix was used to visualize the performance of the classification models, detailing true positives, false positives, true negatives, and false negatives.

CHAPTER – 3
RESULTS

3.1 Logistic Regression Model

Model 1: Full Logistic Regression Model

The first logistic regression model included several variables to predict the likelihood of a patient being "Flipped." Below are the key results and interpretations:

Variable	Coefficient	z-value	p-value	Significance
Intercept	-14.40	-1.314	0.18882	Not significant
Age	0.000932	0.149	0.88135	Not significant
Gender (Male)	0.2804	1.657	0.09748	Borderline
Primary Insurance (Medicaid State)	0.07501	0.187	0.85191	Not significant
Primary Insurance (Medicare)	1.129	3.093	0.00198	Significant **
DRG01558	1.339	2.340	0.01929	Significant *
DRG01780	-0.8803	-3.032	0.00243	Significant **
DRG01786	-0.8090	-2.397	0.01651	Significant *

Table 2: Key Results for Full Logistic Regression Model

Model Performance:

- Null Deviance: 982.12
- Residual Deviance: 885.49
- AIC: 935.49
- Accuracy on Test Set: 65.36%
- Sensitivity: 70.75%
- Specificity: 57.53%

Interpretation: This logistic regression model suggests that certain variables, such as Medicare Insurance, DRG01558, DRG01780, and DRG01786, are statistically significant in predicting whether a patient will be flipped. The gender variable shows borderline significance. Overall, the model provides reasonable predictive performance, with an accuracy of 65.36%.

Model 2: Reduced Logistic Regression Model

The second model simplifies the predictor set by including only Gender, Primary Insurance Category, and DRG01.

Variable Significance	**Coefficient**	**z-value**	**p-value**
Gender (Male) Borderline	0.27784	1.666	0.09570
Primary Insurance (Medicare) Significant ***	1.10397	3.358	0.00078
DRG01558 Significant *	1.42489	2.514	0.01194
DRG01780 Significant **	-0.81689	-2.861	0.00422
DRG01786 Significant *	-0.85288	-2.566	0.01027

Table 3: Key Results for Reduced Logistic Regression Model

Model Performance:

- Residual Deviance: 891.50
- AIC: 925.5
- Accuracy on Test Set: 61.45%
- Sensitivity: 69.81%
- Specificity: 49.32%

Interpretation: In this reduced model, Medicare Insurance, DRG01558, DRG01780, and DRG01786 remain significant predictors. The model sacrifices some accuracy (61.45%) compared to the full model but provides a more concise and interpretable set of variables.

3.2 Decision Tree Model

The decision tree model was trained using a similar set of predictors, and the tree visualization helps identify key decision points for predicting flipping.

Model Performance:

- Accuracy on Test Set: 64.8%
- Sensitivity: 68.87%
- Specificity: 58.90%
- Balanced Accuracy: 63.89%

Interpretation: The decision tree model performs similarly to the logistic regression model, with an accuracy of 64.8%. It provides a clear, interpretable set of decision rules based on DRG codes and Primary Insurance Category.

3.3 Random Forest Model

A random forest model was built using 100 trees to aggregate predictions from multiple decision trees.

Model Performance:

- Out-of-Bag Error Rate: 38.85%
- Accuracy on Test Set: 61.45%
- Sensitivity: 70.75%
- Specificity: 47.95%
- Balanced Accuracy: 59.35%

Interpretation: The random forest model shows comparable performance to the logistic regression and decision tree models, with an accuracy of 61.45%. It reduces variance and overfitting, making it more robust, though at the cost of slightly lower interpretability.

3.4 MODEL EVALUATION

When evaluating the performance of different models using the logarithm of the price, we consider several accuracy metrics: ME (Mean Error), RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), MPE (Mean Percentage Error), and MAPE (Mean Absolute Percentage Error).

Model Performance: Logarithmic Price Prediction

Model	ME	RMSE	MAE	MPE	MAPE
Initial Model	-0.00497	0.417	0.325	-0.904%	7.09%
Stepwise Model	-0.00495	0.417	0.325	-0.903%	7.09%
Regression Tree	-0.00326	0.428	0.334	-0.922%	7.29%
Random Forest	-0.00666	0.403	0.314	-0.914%	6.84%

Table 4: Model Evaluation for Logarithmic Price Prediction

Among these models, the Random Forest model stands out as the best performer. It exhibits the lowest values for RMSE, MAE, MPE, and MAPE, indicating superior predictive accuracy compared to the other models. The Random Forest model's lower MAPE of 6.84% suggests that, on average, its predictions deviate by approximately 6.84% from the actual log prices, making it the most accurate model for predicting the logarithm of the price.

Model Accuracy for Flipped Status Prediction

Model	Accuracy	Sensitivity	Specificity
Logistic Regression 1	0.6313	0.7553	0.4941
Logistic Regression 2	0.6201	0.7447	0.4824
Decision Tree	0.6536	0.7447	0.5529
Random Forest	0.6480	0.7766	0.5059

Table 5: Model Accuracy for Flipped Status Prediction

3.5 FINAL MODEL SELECTION

Across all models, Medicare Insurance, specific DRG codes, and Gender (to a lesser extent) are consistently identified as important predictors of whether a patient will be flipped. The logistic regression model offers the best balance between interpretability and performance, making it the recommended choice for predicting flipping decisions. However, the decision tree and random forest models offer alternative, more flexible approaches that could be further explored in future work.

3.6 RESULTS AND CONCLUSION

After selecting the final model to predict the probability of a patient flipping (from observation to inpatient) based on other variables, we predict the probability for each observation in the dataset. The results were exported to an Excel file, adding a new column called *PredictedProbability.* Based on a chosen cutoff point (e.g., 0.5), patients are predicted as ”No” (if the predicted probability is less than the cutoff point) or ”Yes” (if it is greater than or equal to the cutoff point) in a new *PredictedFlipped* column.

We created a confusion matrix to evaluate the accuracy of our predictions. We then assumed that all patients predicted as ”Yes” (likely to flip) were moved to inpatient care. The new flipped rate was calculated based on those patients predicted as ”No” but with an actual flipped status of ”Yes” (misclassified by the model).

We also adjusted the cutoff point between 0.2 and 0.9 to observe how it impacted the flipped rate:

Cutoff Point	New Flipped Rate
0.2	11%
0.3	22%
0.4	29%
0.5	32%
0.6	37%
0.7	40%
0.8	44%
0.9	44%

Table 6: New Flipped Rate for Different Cutoff Points

We chose cutoff points of 0.3 and 0.4 for simulation, aligning with Dr. Kelly's estimate. By reducing the percentage of flipped patients from 45% to 20% (achieving 22% with a cutoff point of 0.3), the unit can treat an additional 570 patients annually. Even with a reduction to one-third (29%), the OU can treat 260 additional patients annually.

Category	Patients			Hours			
	Before	0.3	0.4	Before	0.3	0.4	
Total OU patients	66	83	74	24	1584	1980	1764
Postsurgery patients	22	28	25		0	0	0
Medicine service patients	44	55	49	16	704	880	784
Remain in OU (No Flipped)	24	43	35		0	0	0
Impatient (Flipped)	20	12	14	38	742	450	535
Ended up in a hospital ward	5	3	4		0	0	0
Not transfer before discharge	15	9	11	11	161	98	116
Total hours of 23 Beds					**3864**		
Total Hours				3191	3408	3199	
Utilization				83%	88%	83%	

Table 7: OU Patient Count and Hours Served

The impact on OU utilization, including bed occupancy, is summarized below:

	Before	0.3 Cutoff	0.4 Cutoff
Total Hours of 23 Beds	3864	3191	3408
Utilization	83%	88%	83%

Table 8: OU Bed Utilization

Revenue Increase Simulation

By choosing a cutoff point of 0.3, we estimate a reduction in LWBS (Left Without Being Seen) cases by 17 patients weekly. This leads to an annual reduction of 884 LWBS cases, resulting in a projected revenue increase of $618,800.

	Before	0.3 Cutoff	0.4 Cutoff
Increase in OU Patients	0	17	8
Revenue Increase	$0	$11,900	$5,600

Table 9: Revenue Impact Based on Cutoff Points

Monte Carlo Simulation

A Monte Carlo simulation was conducted to assess how increasing the number of patients affects OU utilization. We simulated 100 runs with the number of patients increasing from 44 to 55, confirming that the OU could accommodate the additional load without compromising quality of care.

3.7 CONCLUSION

This OU operation analysis aimed to improve patient flow and efficiency by reducing the percentage of patients who "flip" from observation to inpatient status. This project involved developing predictive models and analyzing the operational and financial impact of changing the flipped rate.

After running Logistic Regression, Decision Tree, and Random Forest models, we identified the best model to predict whether a patient would flip based on various factors like age, gender, diagnosis, and vital signs. By adjusting the cutoff point for predicting flipping, we evaluated how different scenarios would affect the OU's capacity.

Based on Dr. Kelly's estimation, reducing the flipped rate from 45% to 20% would enable the OU to treat 55 patients per week instead of 44, resulting in approximately 570 additional patients per year. We analyzed the significant impact of the project on improving OU operations.

The key findings included:

- Lowering the flipped rate improves patient flow, reduces ED crowding, and decreases the number of patients leaving without being seen (LWBS). This could lead to reducing 46% LWBS by 884 patients annually, which translates into a potential revenue increase of $618,800 (approximately $11,900 monthly).
- Operationally, reducing the flipped rate improves the utilization of the OU's 23 beds without overburdening the unit. For example, with a flipped rate of 22% (cutoff 0.3), the unit's bed utilization increased from 83% to 88%, serving more patients and providing a more efficient healthcare service.
- Using a Monte Carlo simulation, we modeled how increasing the number of medicine service patients from 44 to 55 would impact bed utilization and efficiency, confirming that the OU could accommodate the additional patient load without sacrificing quality of care.

Overall, the project demonstrated that reducing the flipped rate not only improves operational efficiency by reducing unnecessary inpatient admissions but also increases patient throughput and revenue. The strategic reduction of flipped patients allows the OU to serve more patients, reduce crowding, and enhance financial performance, providing a clear roadmap for improving the hospital's operations.