

**BDA 640 Data Driven Decision Making and Optimization**

**Project Title: Streamlining Observation Care - Leveraging Data for Improved  
Patient Management**



**Prepared By:**

**Jun Ming Li – MUID: 11022435**

**Jairo Onate – MUID: 11062465**

**Mazen Alhaffar – MUID: 11054744**

**Submitted To:**

**Reza Maihami, Ph.D.**

**Stetson-Hatcher School of Business**

**Mercer University**

## 1. Executive Summary

This report presents an in-depth analysis of operations within Montanaro hospital observation unit, applying a range of advanced data analysis techniques to improve patient management and overall operational efficiency. The analysis integrates Clustering, Random Forest, and Logistic Regression models to uncover patterns in patient data and forecast health outcomes, focusing on vital metrics like age, blood pressure, pulse, and respiration. By clustering patients based on these characteristics, the study identifies distinct groups with specific needs, enabling targeted interventions. Random Forest models were employed to determine the key variables influencing patient outcomes, while Logistic Regression provided insights into binary predictions, such as the likelihood of patients requiring higher levels of care. Additionally, a cost analysis simulation was carried out to assess the financial impact of various strategies, offering recommendations for optimizing resource use, reducing costs, and improving hospital operations. Overall, this analysis emphasizes the role of data-driven decision-making in enhancing patient care, streamlining resources, and achieving operational cost savings.

## 2. Problem Description:

The main problem that is facing the hospital is the inefficient categorization and placement of their patients that arrive at their ER. Many patients are placed in the OU when they should have been passed along to inpatient care from the onset. This leads to a compounding effect whereby the OU capacity is reached and inpatient rooms become pseudo-OU units. This also leads to inefficient use of hospital resources and staff when “flips” occur, and patients must be moved in and out of the OU and inpatient care units. More than just the immediate cost of wasted resources on mismanaged patient placements, there are also potential losses in potential revenue from LWBS patients and the difference in OU and inpatient pricing. Patients that are mismanaged not only receive ineffective care but are also costly to the hospital.

## 3. Methodology:

First, the dataset was reviewed for missing and NA values. These values were places with the mean for numerical variables and the mode for categorical variables. Upon review of the main purpose of the case and the data dictionary, the following variables were removed: ObservationRecordKey, InitPatientClassAndFirstPostOUClass, and OU\_LOS\_hrs. The Record Key was just for identification. The Inpatient/OU class variable was redundant with the Flipped target variable. The OU\_LOS\_hrs was considered irrelevant to our analysis because the focus is sorting the patient correctly from the onset, and not when to optimally move from a patient from OU to Inpatient care. Outliers were then removed via the IQR method, and the data was partition into 70% training and 30% testing. Finally, the data was scaled using min-max standardization before performing logistic, random forest, and clustering.

### 3.1. Logistic Model:

Three models were created for logistic regression. First, a Full model with all 11 variables was created as a baseline and the accuracy of this initial model was run against the testing data set using a confusion matrix. To improve the predictability of the model, data reduction was conducted using forward stepwise regression, which eliminated 6 variables: BloodPressureUpper, BloodPressureDiff, Pulse, PulseOximetry, Temperature, and Respirations. Finally, a third model of self-selected variables was created using the significant information found from the Full model and the Forward stepwise model and from reviewing the primary needs of the hospital. This third model used 4 variables: Gender, DRG01, PrimaryInsuranceCategory, and BloodPressureLower. A summary of the Logistic model statistics is summarized in Table 1.

### 3.2. Random Forest:

Four Random Forest models were built, adjusting the mtry parameter from 3 to 6 to evaluate their impact on accuracy, sensitivity, and specificity. Each model utilized the following variables: Age, DRG01, BloodPressureUpper, BloodPressureLower, BloodPressureDiff, Pulse, PulseOximetry, Respirations, and Temperature. Performance was assessed using a confusion matrix across all mtry configurations, allowing for a comprehensive comparison of model effectiveness in predicting patient transitions from Observation Unit (OU) to Inpatient (IP). A summary of the Random Forest model statistics is summarized in [Table 2](#).

### **3.3. Cluster Analysis:**

Cluster analysis was conducted on numerical variables only, and we used Euclidean distance to figure out the clusters. The elbow method was used to determine the optimal number of clusters (Fig. 1), and the clusters profiling was figured out using the clusters plot (Fig. 2)

## **4. Empirical Results**

### **4.1. Results for Logistic Model:**

The Full model has the highest accuracy among all 3 models and boasts the highest sensitivity and second-highest specificity. The Final model has only slightly lower accuracy, with a difference of only .0032 in favor of the Full model. However, we ultimately chose the Final model as the best-performing model to predict patient flip status. While the Full model does have higher accuracy, it is not a parsimonious model, needing 11 variables for a negligible increase compared to the only 4 variables needed in the Final model. Moreover, the study is interested in the conditions that lead a patient to “flip”, so the Final model’s high specificity rate means it is better suited to predicting these conditions.

### **4.2. Results for Random Forest:**

The Random Forest model was evaluated to determine the best mtry value for predicting whether a patient flips status from Observation Unit (OU) to Inpatient (IP) based on key physiological features such as age, blood pressure, pulse, pulse oximetry, respirations, and temperature. Among the tested values, mtry = 4 performed the best, achieving the highest accuracy (0.6188), AUC (0.6192), and sensitivity (0.6084) while maintaining a relatively high specificity (0.6299). This suggests that mtry = 4 provides the optimal balance between correctly identifying patients who will transition to inpatient care and minimizing false classifications.

AUC (Area Under the Curve) measures the model’s ability to distinguish between patients who remain in the Observation Unit (OU) and those who flip to Inpatient (IP). Since the highest observed AUC in this experiment is 0.6192 (for mtry = 4), the model is better than random guessing but still not exceptionally strong. This indicates that while the model provides useful predictive power, there is room for improvement, potentially by incorporating additional clinical or demographic features or tuning hyperparameters further.

Sensitivity represents the proportion of patients who actually transitioned from Observation Unit (OU) to Inpatient (IP) and were correctly identified by the model. A higher sensitivity means fewer at-risk patients are missed. The highest sensitivity (0.6084 at mtry = 4) indicates that this configuration is the most effective at detecting patients who require inpatient care, making it valuable for early intervention in clinical settings. On the other hand, specificity measures the proportion of patients who remained in the Observation Unit (OU) and were correctly classified. A higher specificity reduces false positives, ensuring that fewer patients are mistakenly predicted to need inpatient care. The highest specificity (0.6299 at mtry = 4 and mtry = 5) suggests the model is effective at preventing unnecessary admissions, aiding in resource allocation and hospital bed management.

Adjusting mtry involves a trade-off between sensitivity and specificity. Increasing mtry tends to improve specificity by reducing false positives but may lower sensitivity, potentially missing at-risk

patients. Conversely, decreasing mtry enhances sensitivity by capturing more true positives but can increase false positives, leading to unnecessary inpatient admissions. In this case, mtry = 4 provides the optimal balance, effectively identifying at-risk patients while keeping unnecessary admissions to a minimum.

### **4.3. Results for Clustering Analysis**

The following is the profiling results from conducting the cluster analysis:

**Cluster 1:** It has relatively low values across most variables, with the lowest values in Age and hours of stay in OU. This cluster possibly represents patients that are creating the bottleneck at the OU since most of their readings are normal.

**Cluster 2:** This cluster shows high values in Age and longer stays in OU, and overall normal readings in all other aspects. This suggests that the individuals within this Cluster are those that may be part of the true subset of individuals that belong in the OU and do not flip. Moreover, this group may contribute to the congestion in the OU unit, thus the demographic of this cluster is significant to review.

**Cluster 3:** It is characterized by higher values in Blood Pressure readings, the highest pulse and temperature of all patients, and the lowest stays in OU. This cluster likely represents individuals who need to be flipped to inpatient status since they have critical readings.

**Cluster 4:** This cluster has more varied characteristics but has the highest Blood Pressure readings. The higher pulse and respiration levels might indicate a cluster of individuals with increased activity or stress, as well as potential hypertension, which usually stays at the OU to be observed.

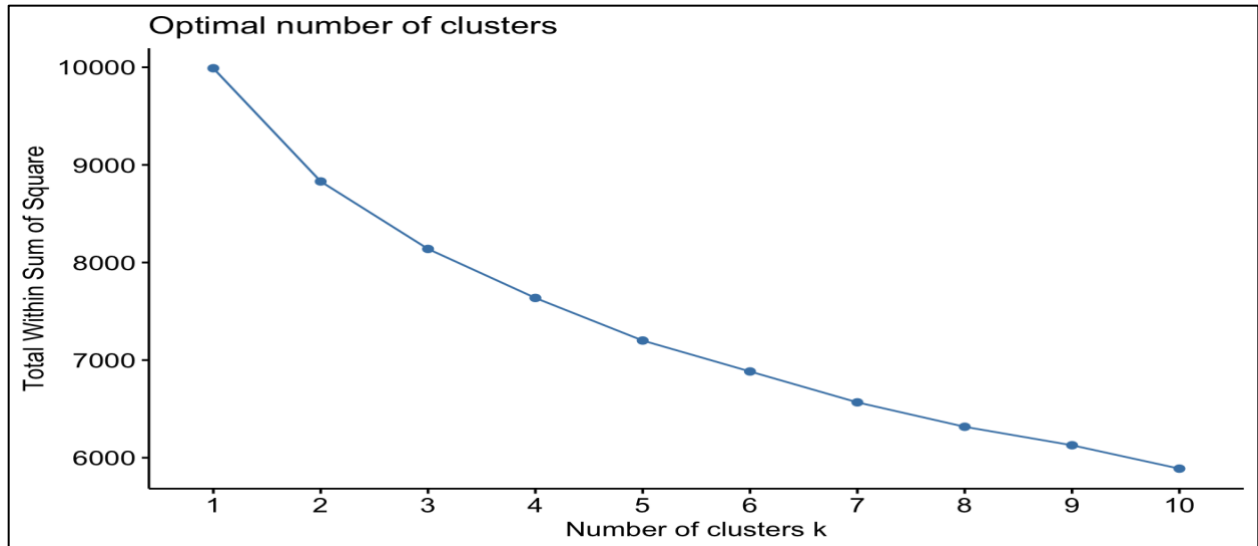
## **5. Conclusions & Recommendations**

We recommend that the hospital follow the Final logistic model to optimize how they sort their ER patients. Particularly, we recommend that the physician and nurses focus on the variables Gender, DRG01, PrimaryInsuranceCategory, and BloodPressureLower. By collecting more information on these variables for the patients that enter the inpatient ward or flip, the hospital can also define the range by which patients are more likely to flip. Utilizing the assumptions in Table 3, it was determined that a more sensitive cutoff point of .3 slightly exceeds the revenue gain from the expected return of the baseline. From a cutoff of .2, our model yields revenue gains of at least 7.2% more than the baseline (Table 4).

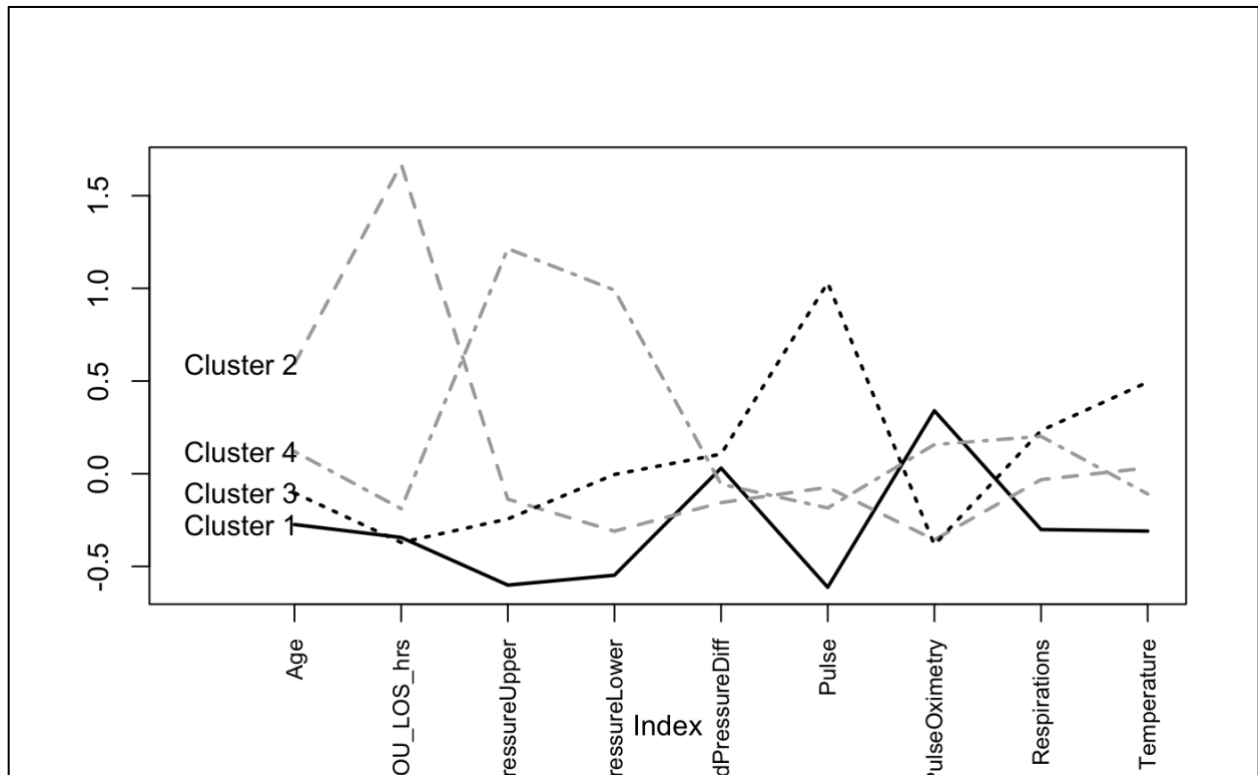
An optimization of utilization was also performed for Montanaro hospital. The simulation highlights a strategy to optimize Observation Unit (OU) utilization, which increased patient throughput from 13.5 to 22 patients per week, generating \$44,000 in OU revenue and \$330,000 in inpatient revenue. With a baseline revenue of \$798,000 per 100 ER patients, maximizing OU efficiency reduces inpatient strain and boosts profitability.

Moreover, by using clustering analysis, it was determined that cluster 1 and 2 are significant to review, thus evaluating the demographic of the individuals within these clusters should illuminate more key variables that would cause patients to flip. The hospital can utilize this information to find not only the combination of variables, but also a confidence interval for those variables to add to their OU exclusion list. Utilizing these simulations Montanaro hospital can better anticipate demand, allocate resources, and prevent bottlenecks, ensuring better patient care and financial sustainability.

## Appendix



**Figure (1): Elbow Method**



**Figure (2): Cluster Analysis Results**

#Patients	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	
Avg OU Profit	\$ 56,045.55	\$ 57,370.60	\$ 58,198.76	\$ 59,523.81	\$ 60,351.97	\$ 60,848.86	\$ 62,691.51	\$ 64,016.56	\$ 64,844.72	\$ 66,169.77	\$ 66,997.93	\$ 68,322.98	\$ 70,165.63	\$ 71,490.68	\$ 72,318.84	
\$	140,000.00	70.06%	71.71%	72.75%	74.40%	75.44%	76.06%	78.36%	80.02%	81.06%	82.71%	83.75%	85.40%	87.71%	89.36%	90.40%
28	\$ 39,232	\$ 40,159	\$ 40,739	\$ 41,667	\$ 42,246	\$ 42,594	\$ 43,884	\$ 44,812	\$ 45,391	\$ 46,319	\$ 46,899	\$ 47,826	\$ 49,116	\$ 50,043	\$ 50,623	
29	\$ 40,633	\$ 41,594	\$ 42,194	\$ 43,155	\$ 43,755	\$ 44,115	\$ 45,451	\$ 46,412	\$ 47,012	\$ 47,973	\$ 48,573	\$ 49,534	\$ 50,870	\$ 51,831	\$ 52,431	
30	\$ 42,034	\$ 43,028	\$ 43,649	\$ 44,643	\$ 45,264	\$ 45,637	\$ 47,019	\$ 48,012	\$ 48,634	\$ 49,627	\$ 50,248	\$ 51,242	\$ 52,624	\$ 53,618	\$ 54,239	
31	\$ 43,435	\$ 44,462	\$ 45,104	\$ 46,131	\$ 46,773	\$ 47,158	\$ 48,586	\$ 49,613	\$ 50,255	\$ 51,282	\$ 51,923	\$ 52,950	\$ 54,378	\$ 55,405	\$ 56,047	
32	\$ 44,836	\$ 45,896	\$ 46,559	\$ 47,619	\$ 48,282	\$ 48,679	\$ 50,153	\$ 51,213	\$ 51,876	\$ 52,936	\$ 53,598	\$ 54,658	\$ 56,133	\$ 57,193	\$ 57,855	
33	\$ 46,238	\$ 47,331	\$ 48,014	\$ 49,107	\$ 49,790	\$ 50,200	\$ 51,720	\$ 52,814	\$ 53,497	\$ 54,590	\$ 55,273	\$ 56,366	\$ 57,887	\$ 58,980	\$ 59,663	
34	\$ 47,639	\$ 48,765	\$ 49,469	\$ 50,595	\$ 51,299	\$ 51,722	\$ 53,288	\$ 54,414	\$ 55,118	\$ 56,244	\$ 56,948	\$ 58,075	\$ 59,641	\$ 60,767	\$ 61,471	
35	\$ 49,040	\$ 50,199	\$ 50,924	\$ 52,083	\$ 52,808	\$ 53,243	\$ 54,855	\$ 56,014	\$ 56,739	\$ 57,899	\$ 58,623	\$ 59,783	\$ 61,395	\$ 62,554	\$ 63,279	
36	\$ 50,441	\$ 51,634	\$ 52,379	\$ 53,571	\$ 54,317	\$ 54,764	\$ 56,422	\$ 57,615	\$ 58,360	\$ 59,553	\$ 60,298	\$ 61,491	\$ 63,149	\$ 64,342	\$ 65,087	
37	\$ 51,842	\$ 53,068	\$ 53,834	\$ 55,060	\$ 55,826	\$ 56,285	\$ 57,990	\$ 59,215	\$ 59,981	\$ 61,207	\$ 61,973	\$ 63,199	\$ 64,903	\$ 66,129	\$ 66,895	
38	\$ 53,243	\$ 54,502	\$ 55,289	\$ 56,548	\$ 57,334	\$ 57,806	\$ 59,557	\$ 60,816	\$ 61,602	\$ 62,861	\$ 63,648	\$ 64,907	\$ 66,657	\$ 67,916	\$ 68,703	
39	\$ 54,644	\$ 55,936	\$ 56,744	\$ 58,036	\$ 58,843	\$ 59,328	\$ 61,124	\$ 62,416	\$ 63,224	\$ 64,516	\$ 65,323	\$ 66,615	\$ 68,411	\$ 69,703	\$ 70,511	
40	\$ 56,046	\$ 57,371	\$ 58,199	\$ 59,524	\$ 60,352	\$ 60,849	\$ 62,692	\$ 64,017	\$ 64,845	\$ 66,170	\$ 66,998	\$ 68,323	\$ 70,166	\$ 71,491	\$ 72,319	
41	\$ 57,447	\$ 58,805	\$ 59,654	\$ 61,012	\$ 61,861	\$ 62,370	\$ 64,259	\$ 65,617	\$ 66,466	\$ 67,824	\$ 68,673	\$ 70,031	\$ 71,920	\$ 73,278	\$ 74,127	
42	\$ 58,848	\$ 60,239	\$ 61,109	\$ 62,500	\$ 63,370	\$ 63,891	\$ 65,826	\$ 67,217	\$ 68,087	\$ 69,478	\$ 70,348	\$ 71,739	\$ 73,674	\$ 75,065	\$ 75,935	
43	\$ 60,249	\$ 61,673	\$ 62,564	\$ 63,988	\$ 64,878	\$ 65,413	\$ 67,393	\$ 68,818	\$ 69,708	\$ 71,133	\$ 72,023	\$ 73,447	\$ 75,428	\$ 76,852	\$ 77,743	
44	\$ 61,650	\$ 63,108	\$ 64,019	\$ 65,476	\$ 66,387	\$ 66,934	\$ 68,961	\$ 70,418	\$ 71,329	\$ 72,787	\$ 73,698	\$ 75,155	\$ 77,182	\$ 78,640	\$ 79,551	
45	\$ 63,051	\$ 64,542	\$ 65,474	\$ 66,964	\$ 67,896	\$ 68,455	\$ 70,528	\$ 72,019	\$ 72,950	\$ 74,441	\$ 75,373	\$ 76,863	\$ 78,936	\$ 80,427	\$ 81,359	
46	\$ 64,452	\$ 65,976	\$ 66,920	\$ 68,452	\$ 69,405	\$ 69,976	\$ 72,095	\$ 73,619	\$ 74,571	\$ 76,095	\$ 77,048	\$ 78,571	\$ 80,690	\$ 82,214	\$ 83,167	
47	\$ 65,854	\$ 67,410	\$ 68,384	\$ 69,940	\$ 70,914	\$ 71,497	\$ 73,663	\$ 75,219	\$ 76,193	\$ 77,749	\$ 78,723	\$ 80,280	\$ 82,445	\$ 84,002	\$ 84,975	
48	\$ 67,255	\$ 68,845	\$ 69,839	\$ 71,429	\$ 72,422	\$ 73,019	\$ 75,230	\$ 76,820	\$ 77,814	\$ 79,404	\$ 80,398	\$ 81,988	\$ 84,199	\$ 85,789	\$ 86,783	
49	\$ 68,656	\$ 70,279	\$ 71,293	\$ 72,917	\$ 73,931	\$ 74,540	\$ 76,797	\$ 78,420	\$ 79,435	\$ 81,058	\$ 82,072	\$ 83,696	\$ 85,953	\$ 87,576	\$ 88,591	
50	\$ 70,057	\$ 71,713	\$ 72,748	\$ 74,405	\$ 75,440	\$ 76,061	\$ 78,364	\$ 80,021	\$ 81,056	\$ 82,712	\$ 83,747	\$ 85,404	\$ 87,707	\$ 89,363	\$ 90,399	
51	\$ 71,458	\$ 73,148	\$ 74,203	\$ 75,893	\$ 76,949	\$ 77,582	\$ 79,932	\$ 81,621	\$ 82,677	\$ 84,366	\$ 85,422	\$ 87,112	\$ 89,461	\$ 91,151	\$ 92,207	
52	\$ 72,859	\$ 74,582	\$ 75,658	\$ 77,381	\$ 78,458	\$ 79,104	\$ 81,499	\$ 83,222	\$ 84,298	\$ 86,021	\$ 87,097	\$ 88,820	\$ 91,215	\$ 92,938	\$ 94,014	

Figure (3): Simulation Matrix

Logistics	Accuracy	AUC	Sensitivity	Specificity	Variables
Full Logistics	0.6188	0.6371	0.6145	0.6234	11
Forward	0.6	0.6433	0.6024	0.5974	5
Final	0.6156	0.6547	0.5904	0.6429	4

**Table (1): Logistic Analysis Results**

Random Forest	Accuracy	AUC	Sensitivity	Specificity
mtry = 3	0.6031	0.6034	0.5964	0.6104
mtry = 4	0.6188	0.6192	0.6084	0.6299
mtry = 5	0.6094	0.6101	0.5904	0.6299
mtry = 6	0.6031	0.6032	0.6024	0.6039

**Table (2): Random Forest Analysis Results**

Assumptions	
Patients come to ER	100
Avg Inpatient Revenue	\$ 15,000.00
Avg OU Revenue	\$ 2,000.00
Baseline Avg Revenue per 100	\$ 798,000.00

**Table (3): Assumptions and Baselines**

Profit	
Cutoff: .4	\$ 655,187.97
Cutoff: .3	\$ 803,609.02
Cutoff: .2	\$ 856,240.60

**Table (4): Profit by Cutoff**