Streamlining Observation Care: Leveraging Data for Improved Patient Management

By:

Jairo Onate

Jun Ming Li

Mazen Alhaffar

TABLE OF CONTENTS
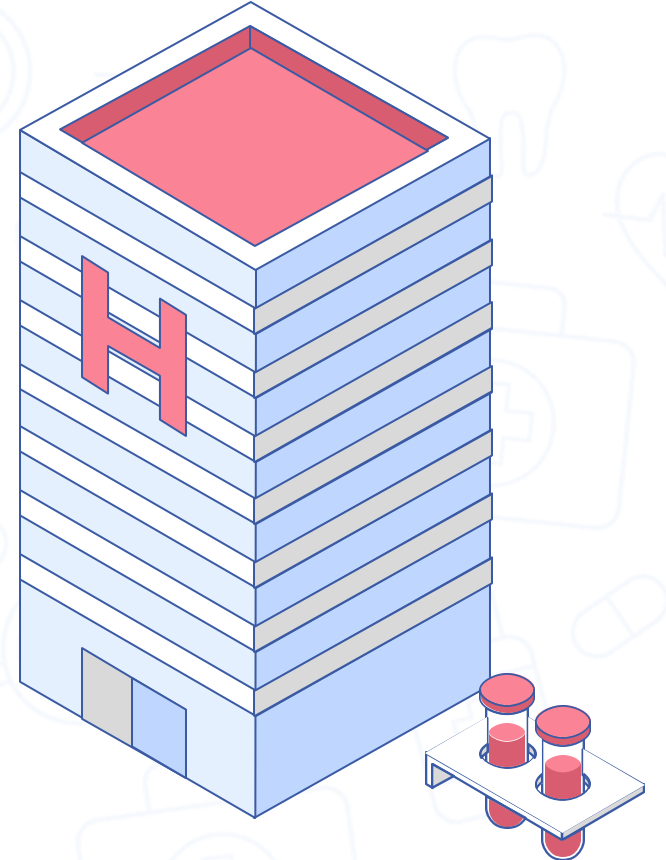
INTRODUCTION

The hospital's Observation Unit (OU) at Montanaro is struggling with inefficient patient flow and resource utilization.

The goal is to **develop a data-driven, predictive model** to optimize the exclusion list, improve patient placement decisions, reduce the rate of patients "flipping" to inpatient care, and enhance overall efficiency, ultimately improving patient care and the hospital's financial performance.

VARIABLES DICTIONARY

Variable	Description
ObservationRecordKey	Unique identifier for each patient visit
Age	Age of the patient
Gender	Gender of the patient (Male/Female)
PrimaryInsuranceCategory	The patient's primary insurance category (e.g., MEDICAID, MEDICARE, etc.)
InitPatientClassAndFirstPostOUClass	Classification of the patient's status upon admission and after observation
Flipped	Boolean indicator (1 if the patient status was changed to inpatient, 0 if not)
OU_LOS_hrs	Length of stay in the Observation Unit in hours
DRG01	The primary diagnosis-related group (DRG) code for the patient
BloodPressureUpper	Upper value of the patient's blood pressure
BloodPressureLower	Lower value of the patient's blood pressure
BloodPressureDiff	Difference between upper and lower blood pressure values
Pulse	The patient's pulse rate (beats per minute)
PulseOximetry	The patient's oxygen saturation level (percentage)
Respirations	The patient's respiration rate (breaths per minute)
Temperature	The patient's body temperature in degrees Fahrenheit

DRG Code	Description
428	Heart failure and shock
599	Cardiac arrhythmia
786	Disorders of the nervous system
780	Acute renal failure and related conditions
428	Respiratory infections and inflammations
599	Injuries and poisoning of the musculoskeletal system
111	Major small and large bowel procedures
820	Orthopedic conditions

Summary Statistics

Metric	Mean	Standard Deviation	Min	25th Percentile	50th Percentile (Median)	75th Percentile	Max
Age	69.03	18.43	19	57	73	86	89
OU_LOS_hrs	62.25	51.04	1.2	24.4	45.8	84.4	405.5
BloodPressureUpper	139.25	25.39	80	120	136	155	229
BloodPressureLower	75.27	15.19	0	65	74	85	146
BloodPressureDiff	63.35	20.16	3	49	61	76	136
Pulse	79.95	16.13	37	68	79	90	148
PulseOximetry	96.63	2.66	75	95	97	99	100
Respirations	17.63	3.92	8	16	18	19	73
Temperature	97.98	1.02	94.7	97.3	97.9	98.4	103.7

Summary Statistics

➢ **The Length of Stay:** has a mean of 62.25 hours but with a high standard deviation of 51.04 hours, indicating a large variation in how long patients stay in the observation unit, spanning from a minimum of 1.2 hours to a maximum of 405.5 hours.

➢ **Blood Pressure Upper:** has a mean of 139.25 mmHg, with a standard deviation of 25.39 mmHg, showing a significant variation in the upper blood pressure values.

➢ **Blood Pressure Lower:** has a mean of 75.27 mmHg, with a standard deviation of 15.19 mmHg, indicating moderate variation.

➢ **Blood Pressure Difference:** has a mean of 63.35 mmHg and a standard deviation of 20.16 mmHg, showing that there is moderate variability in the gap between the upper and lower blood pressure readings.

➢ **Age:** The interquartile range (25th and 75th percentiles) shows that most patients fall between 57 and 86 years of age, and the mean age of patients is 69 years.

➢ **Respiration Rate:** the mean is 17.63 breaths per minute, which is low compared to the other variables, ranging from 8 breaths per minute to 73 breaths per minute.

➢ **Temperature:** The mean body temperature of patients is 97.98°F, with a minimum of 94.7°F and a maximum of 103.7°F.

DATA CLEANING & PREPROCESSING

Find and replace missing values.
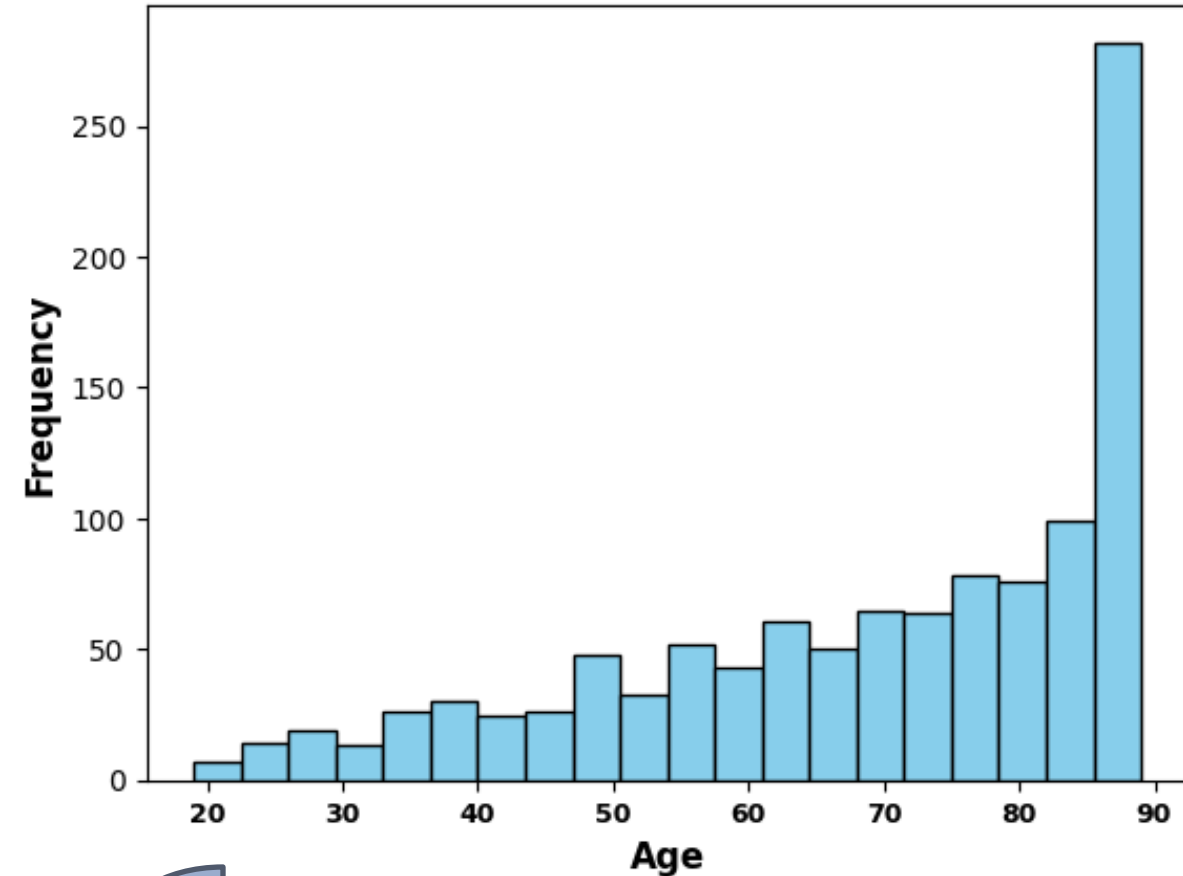
Observe the distribution of data.

Remove outlier.

Scale the data.
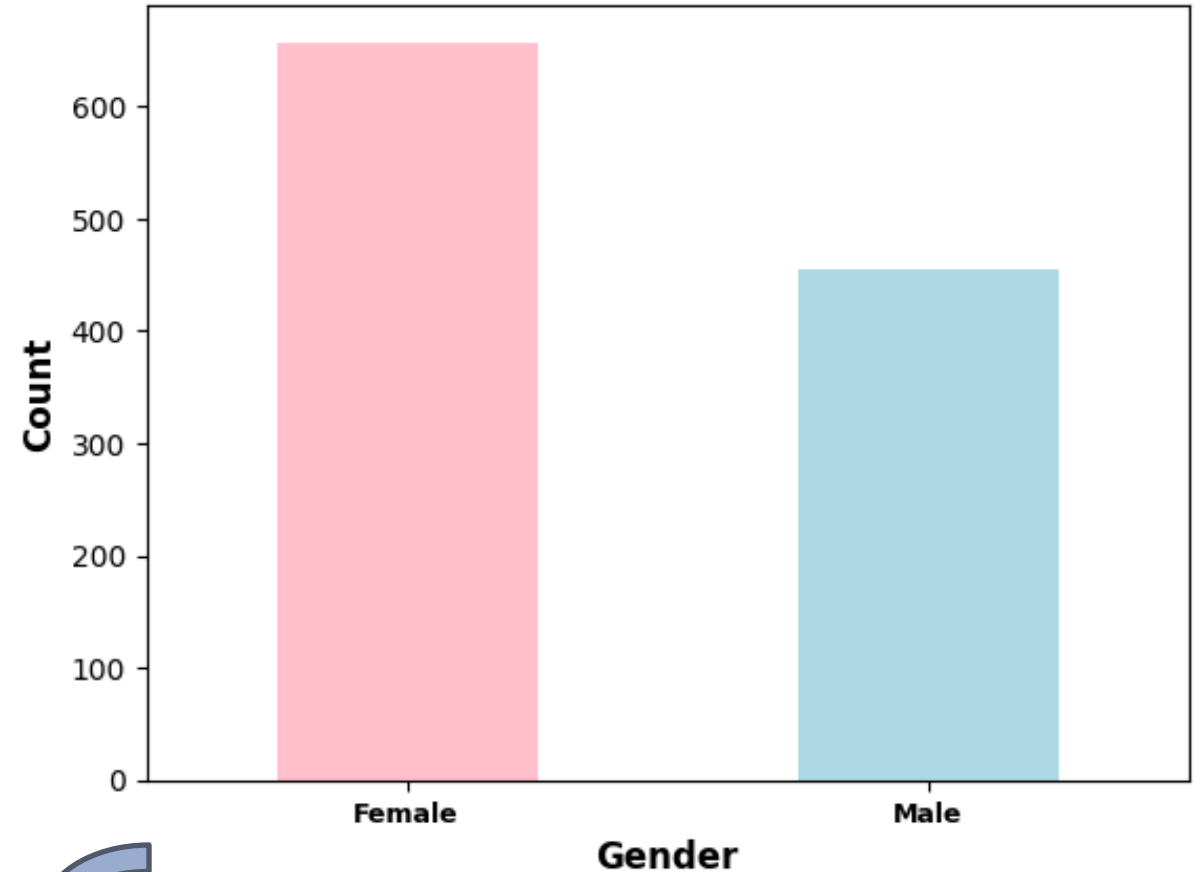
Partition and Balance data sets.

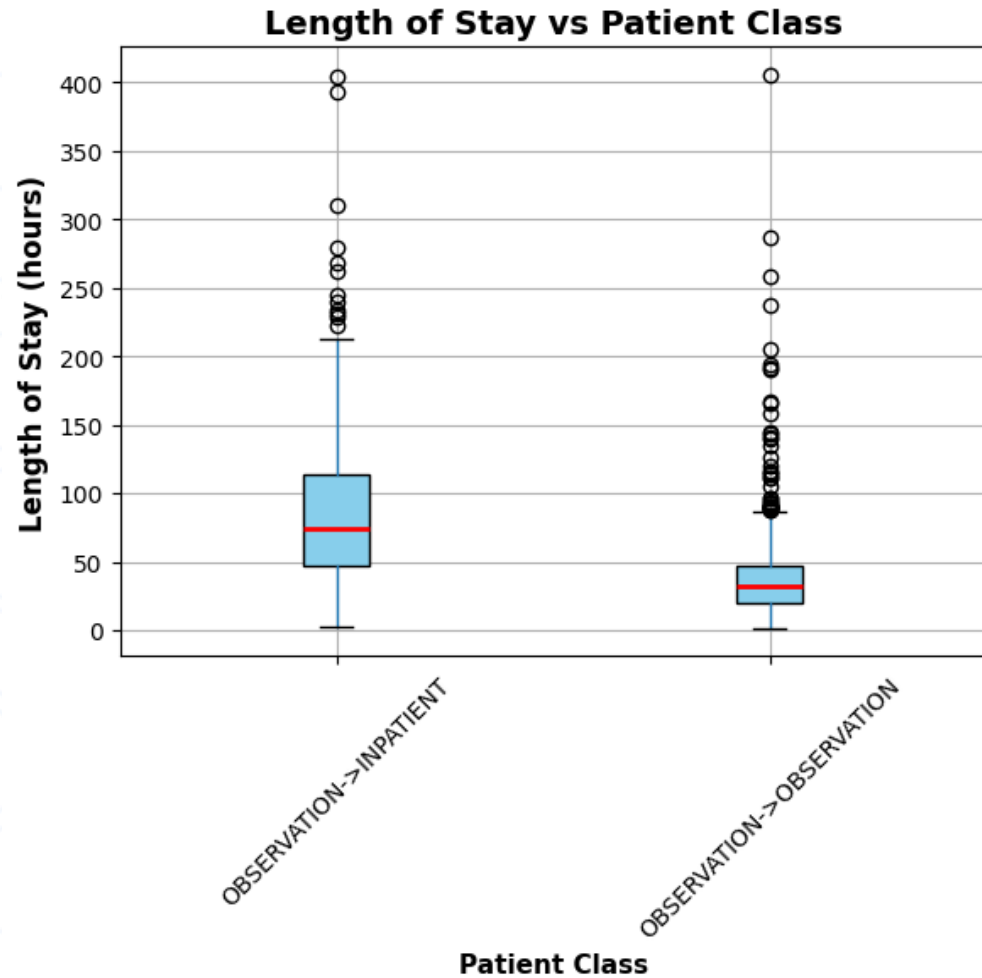# EXPLORATORY DATA ANALYSIS (EDA)

Age Distribution

Gender Distribution

Age distribution is skewed to the right, with most patients being elderly.

Higher count of female patients compared to male patients in the dataset.

EXPLORATORY DATA ANALYSIS (EDA)

Length of Stay vs Patient Class

Pulse vs Pulse Oximetry (Oxygen Levels)

Patients who flipped to inpatient status have a significantly longer length of stay (with many outliers) compared to the patients who stayed in the observation unit, who generally have a shorter and more consistent length of stay.
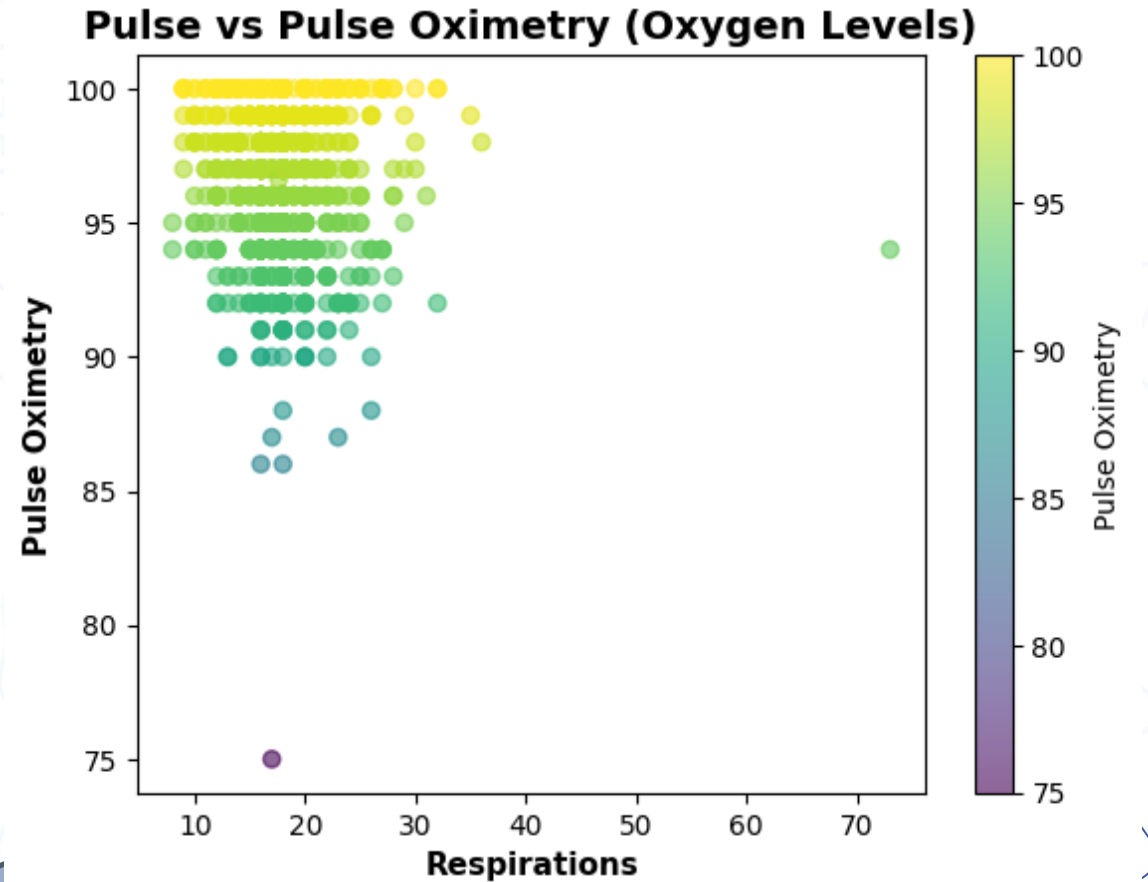
Most patients tend to have high oxygen saturation levels (close to 95-100%), regardless of their respiratory rate. However, there is a slight tendency for pulse oximetry to decrease as respirations increase, particularly for patients with lower oxygen levels.

EXPLORATORY DATA ANALYSIS (EDA)

Correlation Matrix
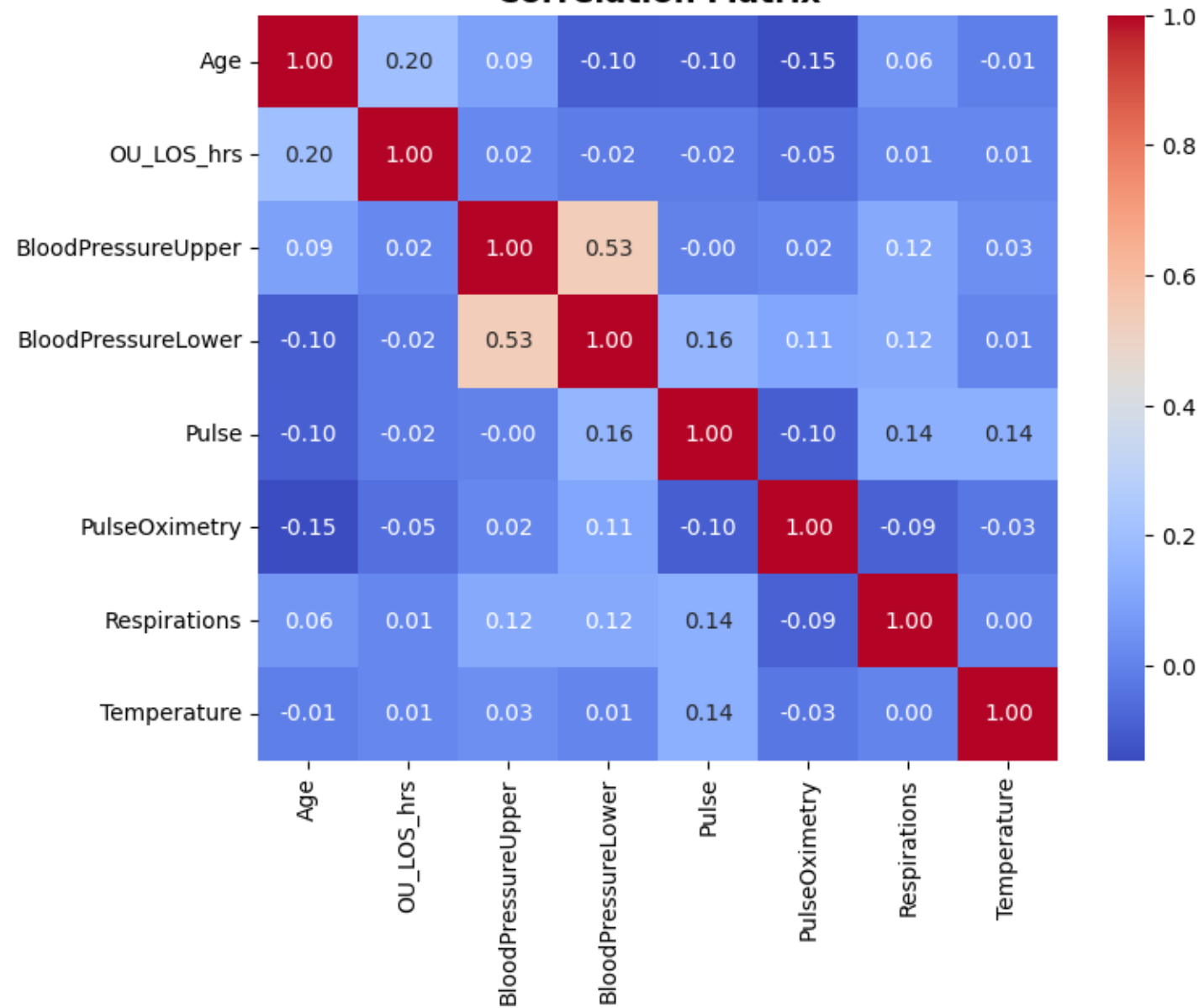
Blood Pressure Lower moderately correlates with Blood Pressure Upper (0.53).

Gender, DRG01, and Primary Insurance Category show minimal correlation with other variables, suggesting independent predictive value.

DATA MODELING — RANDOM FOREST

Four Models Created

- mtry = 3 used as a baseline
- Sensitivity helps find at-risk patients
- Specificity reduces false positives for better hospital planning

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

DATA MODELING — LOGISTIC REGRESSION

Three Models Created

Variable Selection

Our Choice

Compared: Accuracy, AUC, Sensitivity, & Specificity
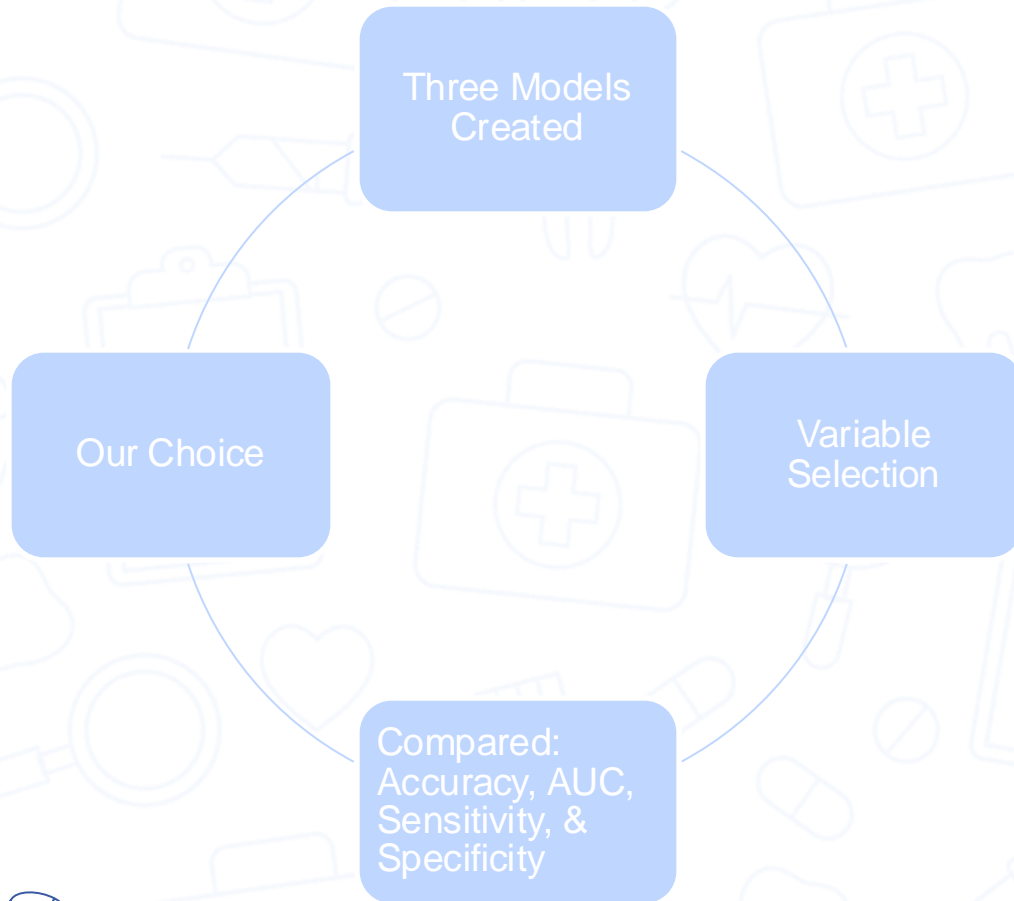
Model	AUC	Accuracy	Sensitivity	Specificity
Logistics (11 vars.)	0.6371	0.6188	0.6145	0.6234
Forward (5 vars.)	0.6433	0.6	0.6024	0.5974
Final (4 vars.)	0.6547	0.6156	0.5904	0.6429

DATA MODELING – CLUSTER ANALYSIS

❑ Numerical variables only

❑ Euclidean distance

❑ Elbow method suggests 4 clusters as optimal number of clusters

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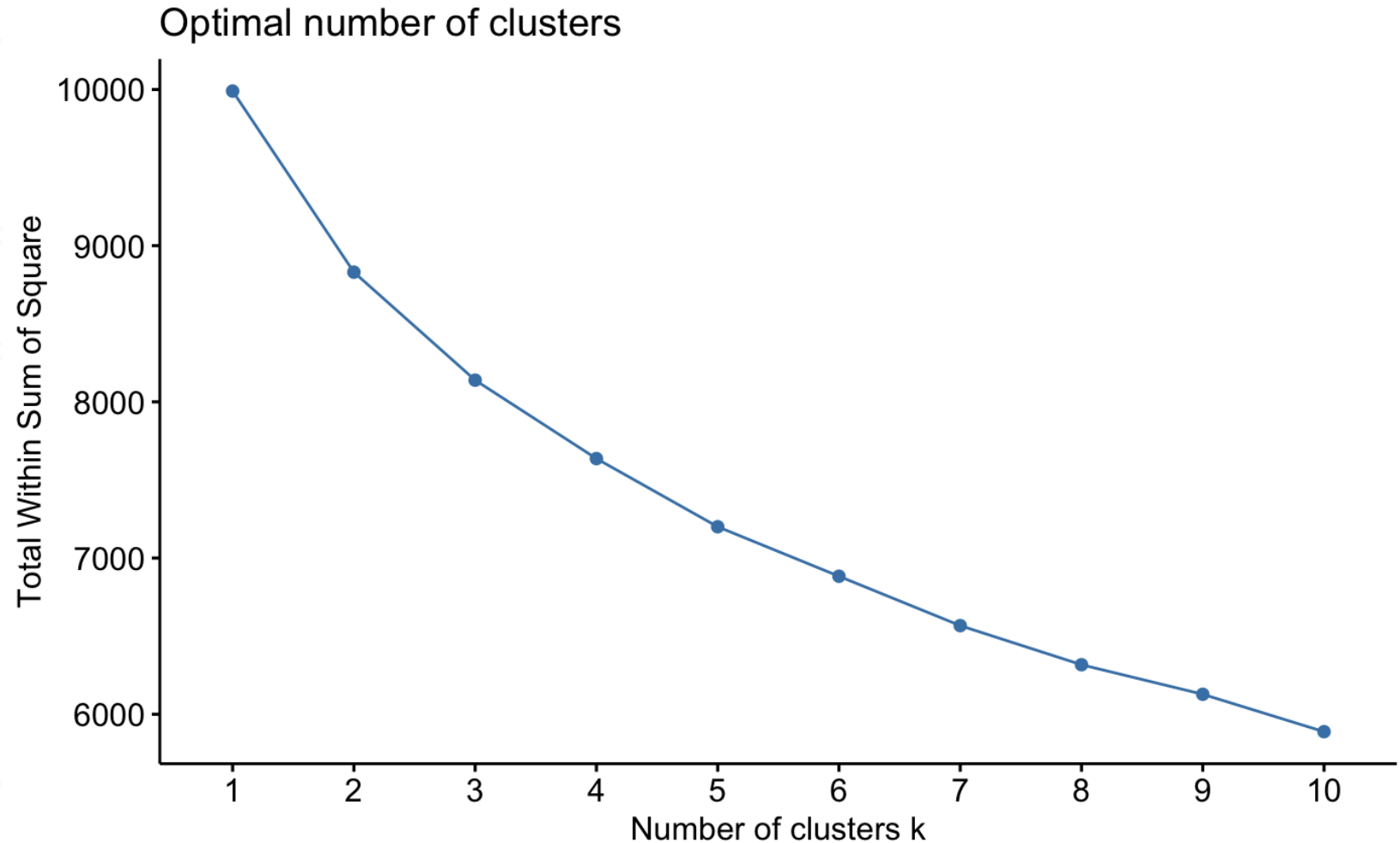
CLUSTER ANALYSIS EVALUATION
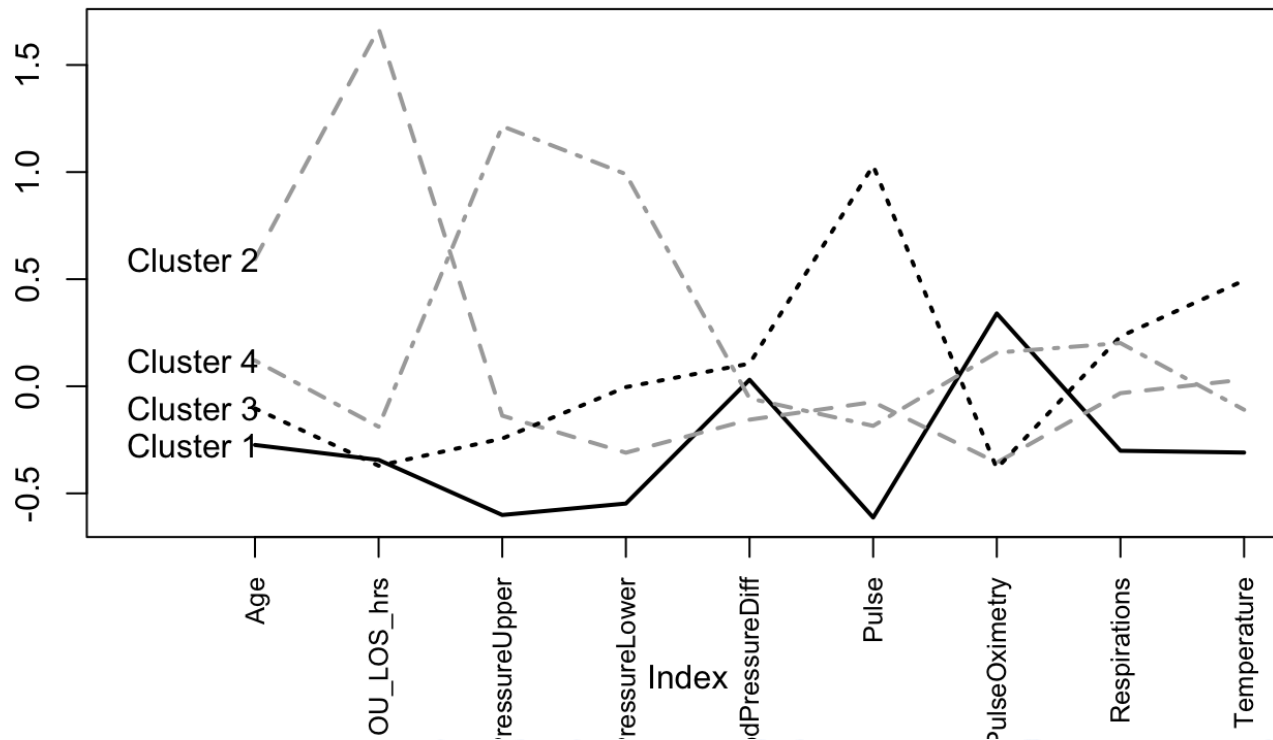
Cluster 1: It has relatively low values across most variables, with the lowest values in Age and hours of stay in OU. This cluster possibly represents patients that are creating the bottleneck at the OU since most of their readings are normal.

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.

MODEL EVALUATION

Model	AUC	Accuracy	Sensitivity	Specificity
Baseline		0.4608		
Logistic Regression	0.6547	0.6156	0.5904	0.6429
Random Forest	0.6192	0.6188	0.6084	0.6299

Compared to our baseline, our improved logistic and random forest models have significantly improved performance.

COST ANALYSIS

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

Standard Cutoff	0.5
Proposed Cutoff	0.3

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

- Using a cutoff point of 0.5, our model does worse than the baseline.

- We must use at least a cutoff of at least 0.3 and below to generate more revenue.

- Using 0.1 would mean that all patients are admitted to Inpatient.

SIMULATION

#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	$ 66,929	$ 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

- Increasing OU utilization from 13.5 to 22 patients per week aligns with cost trends in the matrix, keeping expenses manageable while maximizing revenue potential.

- Higher patient throughput leads to economies of scale, reducing cost per patient and significantly improving profit margins as seen in the gradient of the cost matrix.

CONCLUSIONS & RECOMMENDATIONS

- Random Forest helped to determine the optimal number of variables to choose for analysis.

- We recommend using logistic model to improve hospital patient management. For this model, we propose using a cutoff of at least .3 to improve hospital revenue.

  - Key Variables: Gender, DRG01, PrimaryInsuranceCategory, and BloodPressureLower

- Patients in cluster 1 & 2 are the demographics we need to focus on to figure out the possible bottlenecks in the Observation Unit.

- Increasing OU utilization to 22 patients per week optimizes revenue while keeping costs within a sustainable range, maximizing profitability.

Thank you!