**ENHANCING HOSPITAL OBSERVATION UNIT EFFICIENCY**

**Final Project Report**

**Abhishek Gundala**

**Sajjiynie Suraweera**

**Problem Description**

In the dynamic healthcare landscape, Montanaro Hospital's Observation Unit (OU) stands as a key asset, designed to cater to patients with non-life-threatening symptoms efficiently. The OU serves a crucial role in reducing ED overcrowding, optimizing hospital capacity, and delivering streamlined patient care. However, challenges have surfaced, primarily revolving around the average length of stay for OU patients and a significant number transitioning to inpatient status.

The central challenge lies in the precise identification of patients who are best suited for observation-level care versus those requiring inpatient attention. This necessitates a refinement of the existing patient placement protocols, including the OU exclusion list. Achieving a more accurate categorization based on clinical needs is paramount for optimizing OU bed capacity and ensuring efficient resource utilization.

The successful optimization of the OU holds profound implications for Montanaro Hospital. By honing in on patients requiring observation-level care, the hospital can elevate bed turnover rates, curtail inpatient transitions, and alleviate the ED's burden. Beyond operational efficiency, this initiative aligns with broader healthcare objectives, emphasizing timely and appropriate care delivery.

**Methodology**

**1. Data Understanding and Exploration**

1.1 Exploratory Data Analysis (EDA):

Conducted thorough exploratory data analysis to comprehend the dataset's nature and characteristics.

Leveraged bar charts and histograms for in-depth insights into the distribution of insurance types across various age groups, shedding light on potential socioeconomic trends.

Explored the relationship between length of stay in the observation unit, patient age, and insurance types, identifying patterns that could impact resource planning and patient care.

**2. Data Preprocessing**

2.1 Handling Missing Data:

Identified and addressed missing values using appropriate imputation techniques, ensuring the integrity of the dataset.

2.2 Encoding Categorical Variables:

Applied one-hot encoding to represent different categories within categorical variables consistently across the dataset.

2.3 Variable Scaling:

Standardized numerical variables to mitigate issues related to variable magnitudes, crucial for models sensitive to scale differences.

**3. Model Development:**

3.1 Decision Tree:

Implemented a decision tree model as an initial exploration into hierarchical decision-making processes.

Assessed variable importance to gain insights into the factors contributing to patient transitions and length of stay.

Examined the model's interpretability to inform subsequent modeling decisions.

3.2 Logistic Regression:

Employed logistic regression as a baseline model for binary classification, aiming to predict patient transitions to inpatient care.

Included relevant predictors, such as age, length of stay, gender, and insurance types, to estimate the probability of transitioning.

Evaluated the statistical significance of coefficients, providing insights into the impact of each predictor on the outcome.

3.3 Random Forest:

Utilized a Random Forest classifier to capture non-linear relationships and interactions among predictors.

Evaluated variable importance based on MeanDecreaseGini and MeanDecreaseAccuracy, providing a comprehensive understanding of the features driving patient transitions.

**4. Model Evaluation:**

4.1 Performance Metrics:

Calculated accuracy, confusion matrices, and area under the ROC curve (AUC) to comprehensively assess the performance of each model.

Analyzed specific timeframes in the observation stay when the risk of transitioning to inpatient care is notably higher.

4.2 Simplified Logistic Regression:

Derived a simplified logistic regression model by selecting significant predictors based on coefficients and significance levels.

Assessed the performance of the simplified model in predicting patient transitions, offering a streamlined approach for practical implementation.

This methodology encapsulates the comprehensive journey from data exploration and preprocessing to the development and evaluation of predictive models. The integration of insightful visualizations from EDA enhances the understanding of patient demographics, insurance trends, and length of stay dynamics, setting the stage for informed modeling decisions. The progression from decision trees to logistic regression and Random Forest reflects a deliberate approach to progressively enhance model complexity and interpretability.

**Results**

**1. Decision Tree Model**

Findings:

The Decision Tree model provided valuable insights into patient categorization.

Accuracy: Achieving an accuracy of approximately 76.64% on the test set, the decision tree model demonstrates a reasonable ability to classify patients into observation and inpatient categories.

Variable Importance:

**OU\_LOS\_hrs**: With a significant Mean Decrease in Accuracy of 62.12, the length of stay in the Observation Unit emerges as a crucial predictor. Longer stays contribute substantially to the accuracy of patient classification.

**PrimaryInsuranceCategory**: This variable, with a Mean Decrease in Accuracy of 14.04, emphasizes the importance of insurance type in determining the level of care. Different insurance categories impact patient pathways within the healthcare system.

**Age**: While contributing to accuracy (Mean Decrease in Accuracy - 3.85), age seems less impactful compared to the length of stay and insurance category.

Interpretation:

The decision tree captures non-linear relationships, providing insights into the complex interplay of factors influencing patient categorization.

Longer stays and specific insurance categories significantly influence the decision to transition a patient to inpatient care.

**2. Logistic Regression Model**

Findings:

The Logistic Regression model yielded insights into the probability of inpatient transitions.

Accuracy: The model achieved an accuracy of approximately 77.37% on the test set, demonstrating improved performance compared to the decision tree.

Coefficients:

OU\_LOS\_hrs: A positive coefficient of 0.086 with a p-value < 2e-16 indicates that as the length of stay increases, the odds of transitioning to inpatient care also increase significantly.

PrimaryInsuranceCategoryMEDICARE OTHER: This category shows a negative coefficient of -0.610 with a p-value of 0.0371, suggesting that having this type of insurance decreases the odds of transitioning to inpatient status.

Age: While statistically significant (p-value - 0.000949), the coefficient is relatively small (0.032), indicating a less pronounced impact compared to other variables.

Interpretation:

The logistic regression model provides a quantifiable understanding of how each predictor contributes to the likelihood of inpatient transitions.

Longer stays and specific insurance types remain critical factors influencing patient outcomes.

**3. Random Forest Model**

Findings:

The Random Forest model offered a comprehensive ensemble-based perspective:

Accuracy: Accuracy: The model achieved an accuracy of approximately 79.19% on the test set, reflecting the model's ability to generalize to unseen data.

Variable Importance:

**OU\_LOS\_hrs**: With a Mean Decrease in Accuracy of 76.71, the length of stay remains a dominant predictor, emphasizing its importance in patient classification.

**Age**: Despite a relatively lower Mean Decrease in Accuracy (18.45), age contributes significantly to the model's understanding of patient outcomes.

**PrimaryInsuranceCategory:** This variable, though less influential, still plays a role in determining the optimal level of care (Mean Decrease in Accuracy - 6.28).

Interpretation:

The Random Forest model reaffirms the importance of length of stay in predicting patient outcomes, maintaining consistency with the decision tree and logistic regression findings.

The ensemble nature mitigates overfitting, enhancing the model's generalizability.

**4. Simplified Logistic Regression Model**

Findings:

The Simplified Logistic Regression model provides a more interpretable model:

Accuracy: Achieving an accuracy of approximately 75.91%, this simplified model retains predictive power while enhancing interpretability.

Coefficients:

**OU\_LOS\_hrs**: With a coefficient of 0.0287 (p-value < 2e-16), the length of stay remains a significant predictor, aligning with findings from other models.

**PrimaryInsuranceCategoryMEDICARE OTHER:** The negative coefficient of -0.610 (p-value - 0.0371) emphasizes the role of insurance type in patient categorization.

**GenderMale:** While not statistically significant (p-value - 0.4743), the coefficient suggests a potential influence of gender on outcomes.

Interpretation:

The simplified model retains predictive accuracy while focusing on the most influential predictors, facilitating easier implementation in a real-world setting.

Insurance type and length of stay continue to be critical factors.

These detailed results provide a nuanced understanding of each model's performance and the specific contributions of predictor variables, offering actionable insights for optimizing patient care in the Montanaro Hospital's Observation Unit.

Business Problem Relevance:

The results from the models directly address the identified business problem:

**Reduced capacity in the OU**: The models highlight the significance of the length of stay in determining the level of care, aiding in the appropriate allocation of beds and resources.

**Inefficient resource utilization**: By emphasizing the importance of insurance types, the models guide resource allocation, ensuring that limited resources are directed towards patients with the highest likelihood of requiring inpatient care.

**Ineffective patient flow**: The predictive power of the models facilitates the identification of critical timeframes for potential transitions, enabling interventions to improve patient flow and reduce unnecessary medical interventions.

**RECOMMENDATION**

Montanaro Hospital's commitment to excellent patient care and operational efficiency is furthered by implementing data-driven recommendations within the Observation Unit (OU). Using predictive modeling, a Length of Stay (LOS) threshold of 70.25 hours has been identified as the optimal point for transitioning patients to inpatient care, which will be a pivotal component of our patient management strategy.

**Background and Analytical Insights**:

Through the utilization of various predictive models, our data analysis has determined that an average LOS of 70.25 hours is the critical juncture at which patients in the OU should be evaluated for inpatient care transition ("flipped" status). This recommendation stems from an in-depth analysis using Logistic Regression, Random Forest, Simplified Logistic Regression, and Decision Tree models. The ensemble of model predictions provides us with a data-supported LOS threshold that aligns with efficient patient management and resource utilization.

**Strategic Recommendations with LOS Threshold Implementation**:

**Early Intervention and LOS Threshold Protocol**:

Introduce a protocol where patients nearing the 70.25-hour threshold in the OU receive a comprehensive review to determine the necessity of inpatient care.

This threshold will act as an automatic trigger for a decision-making process involving both OU and Emergency Department physicians.

**Age-Related Criteria within LOS Threshold**:

Apply age-related criteria in conjunction with the LOS threshold to streamline the transition process, acknowledging that older patients may require earlier intervention.

**Dynamic Adjustment of OU Operations:**

Incorporate the 70.25-hour LOS threshold into the OU exclusion criteria, making it a dynamic measure that is regularly evaluated and adjusted according to patient outcomes and operational data.

**LOS Threshold in Collaborative Decision-Making:**

Use the 70.25-hour threshold as a critical point in the collaborative decision-making process, enabling timely and effective transitions from the OU to inpatient care.

**Operationalization Plan:**

Update OU protocols to include the 70.25-hour threshold, ensuring that all staff are trained and aware of this new operational benchmark.

Implement a system that alerts physicians as patients approach the LOS threshold, facilitating timely evaluations.

Modify the patient management system to include the LOS threshold as a factor in the patient care pathway.

**Outcomes and Benefits:**

**By adopting the 70.25-hour LOS threshold**:

Patient care is expedited, with the decision to transition to inpatient care being made at the data-supported optimal time.

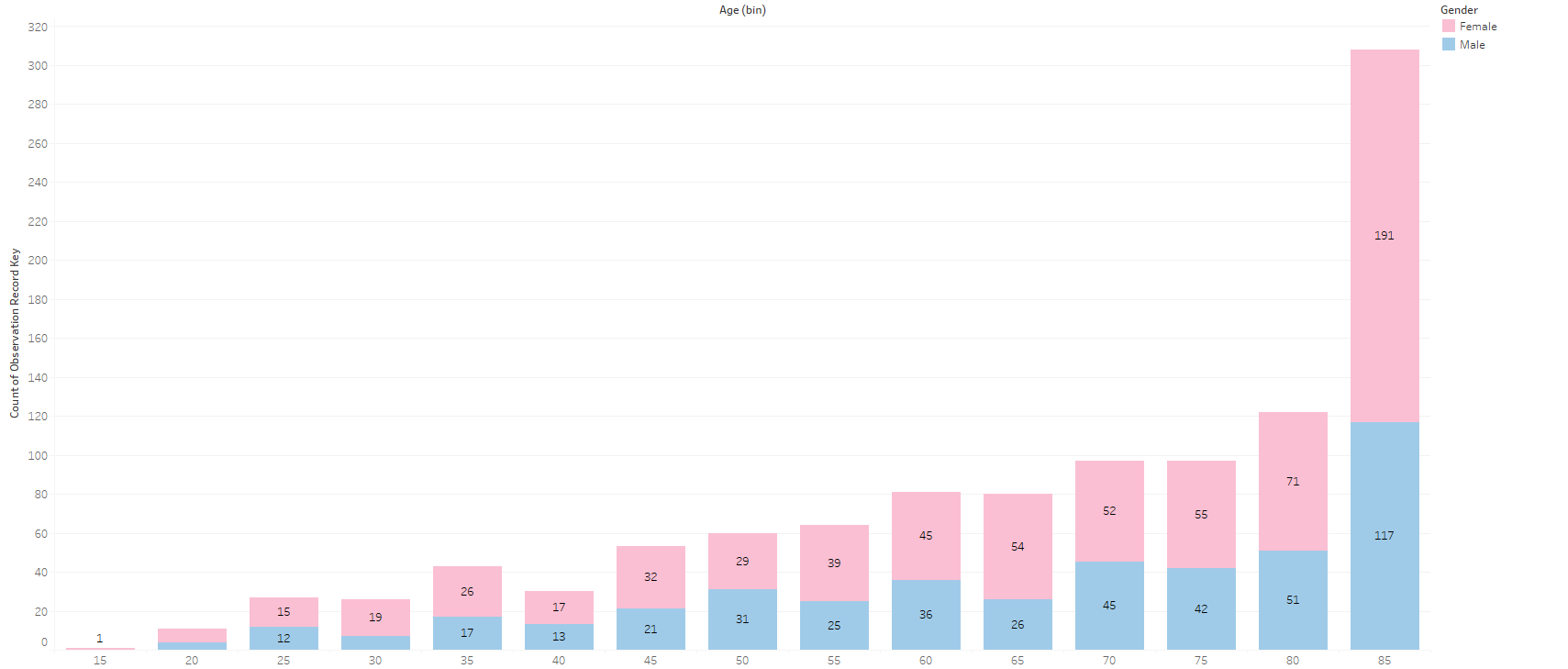
The OU can manage bed availability more effectively, potentially reducing wait times and improving patient flow.

The hospital can optimize its operations, ensuring that patients receive the right level of care at the right time.

**Conclusion:**

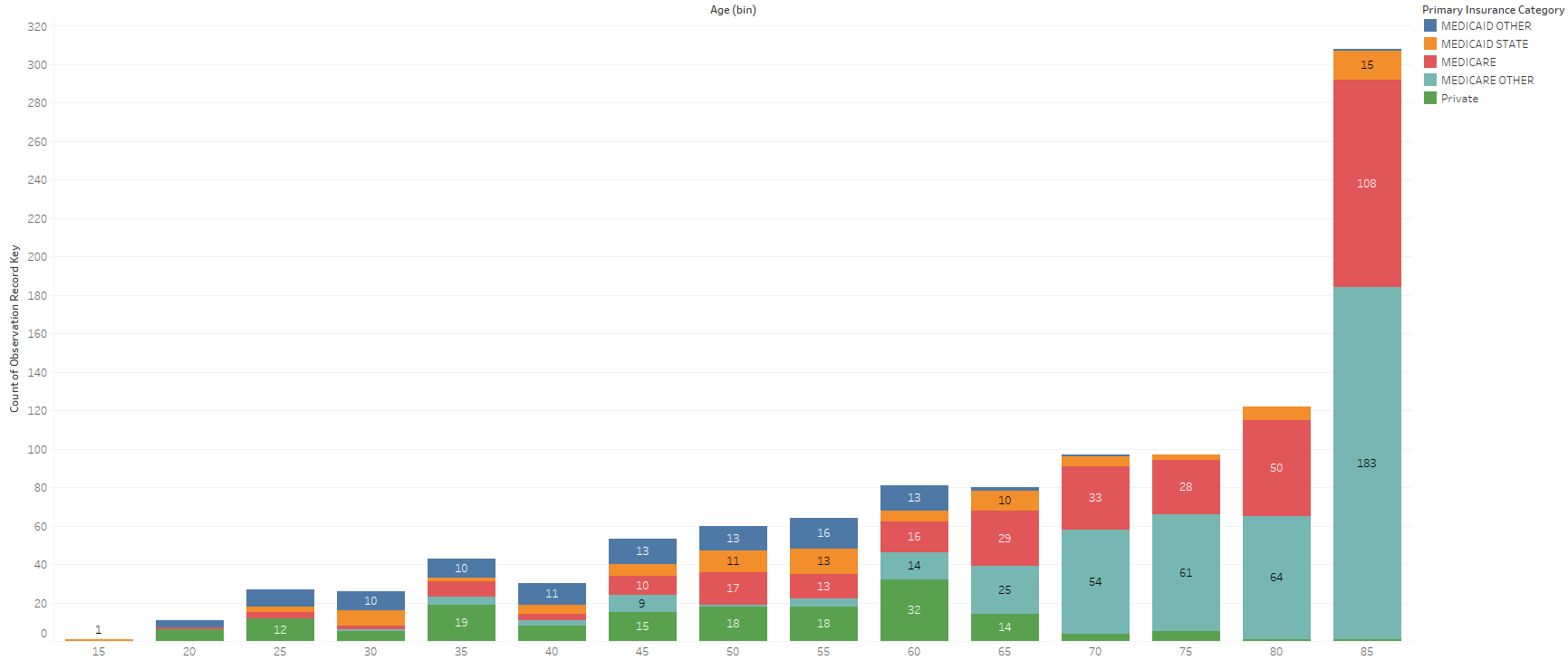
The integration of a 70.25-hour LOS threshold signifies a major step forward in optimizing patient management within Montanaro Hospital's OU. This evidence-based approach aligns with our strategic goals of enhancing patient care, maximizing resource utilization, and improving the overall efficiency of hospital operations.

**Appendix**

****

Graph 01: Age and Gender Distribution of the Observation unit patients

The age groups are divided into 5-year bins from 15 years onwards. The highest number of observations recorded are in the last age bin, which is labeled as '85'. This indicates that the oldest age group had the most visits to the observation unit. Across almost all age groups, female patients have visited the observation unit more often than male patients. otably, there is a significant gender disparity in the '85' age bin, with female patients' visits more than doubling those of male patients. This could indicate either a higher survival rate for women into older age, a higher incidence of conditions requiring observation in females of this age, or both. n the younger age groups, especially from '15' to '50', the number of visits is relatively low for both genders, with a slight increase in the '30' age bin for females.



Graph 02: Insurance Distribution by Age

The chart highlights differences in the distribution of insurance types across various age groups. Here's what we can observe.Younger Ages (10-25): Medicaid and private insurance seem most prevalent.

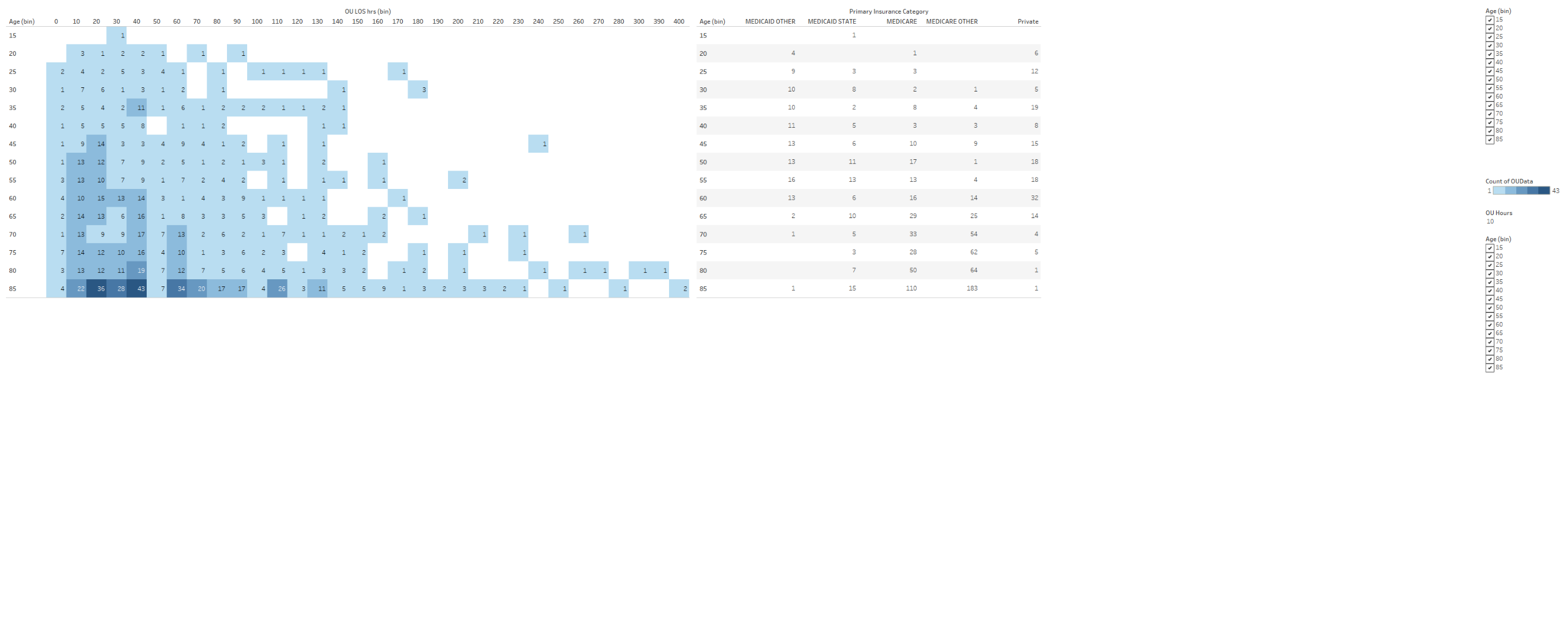
Middle Ages (25-64): Other (likely private) insurance becomes the most common category, while Medicare starts gaining prominence.

Older Ages (64+): Medicare is the dominant insurance type.

Socioeconomic Factors: The graph could hint at socioeconomic trends where younger individuals rely more on government-aided insurance or the working class rely more on private insurance.

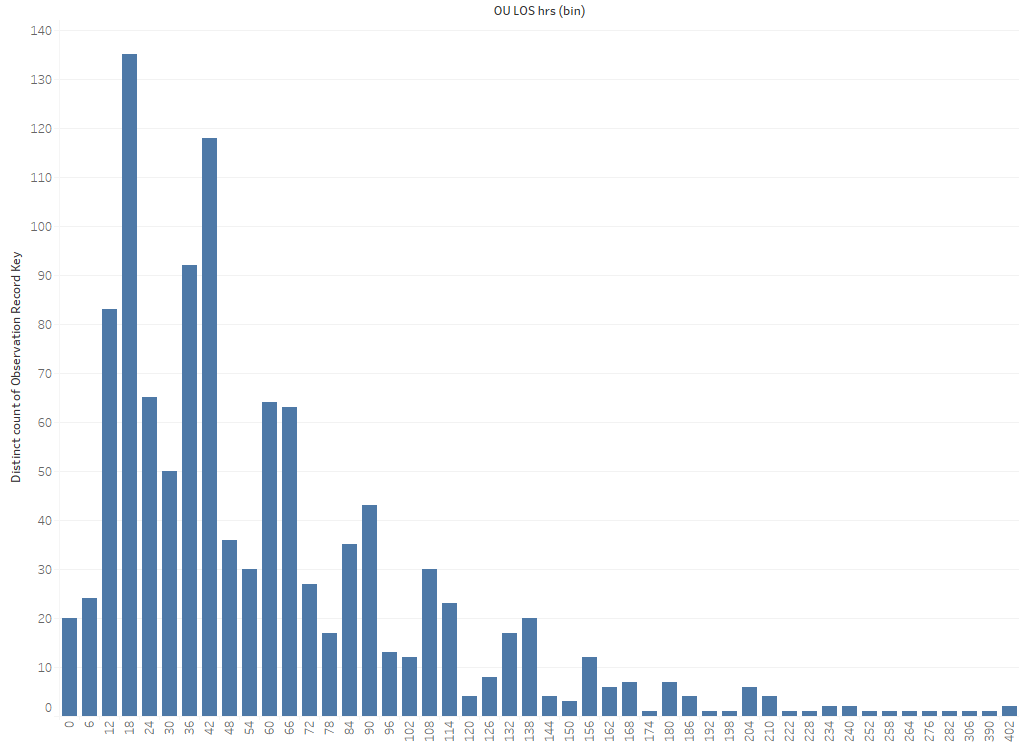
This graph reflects how insurance needs change throughout a person's lifetime.

The data could help a hospital understand trends in insurance types among their patient population and adjust resource planning accordingly.



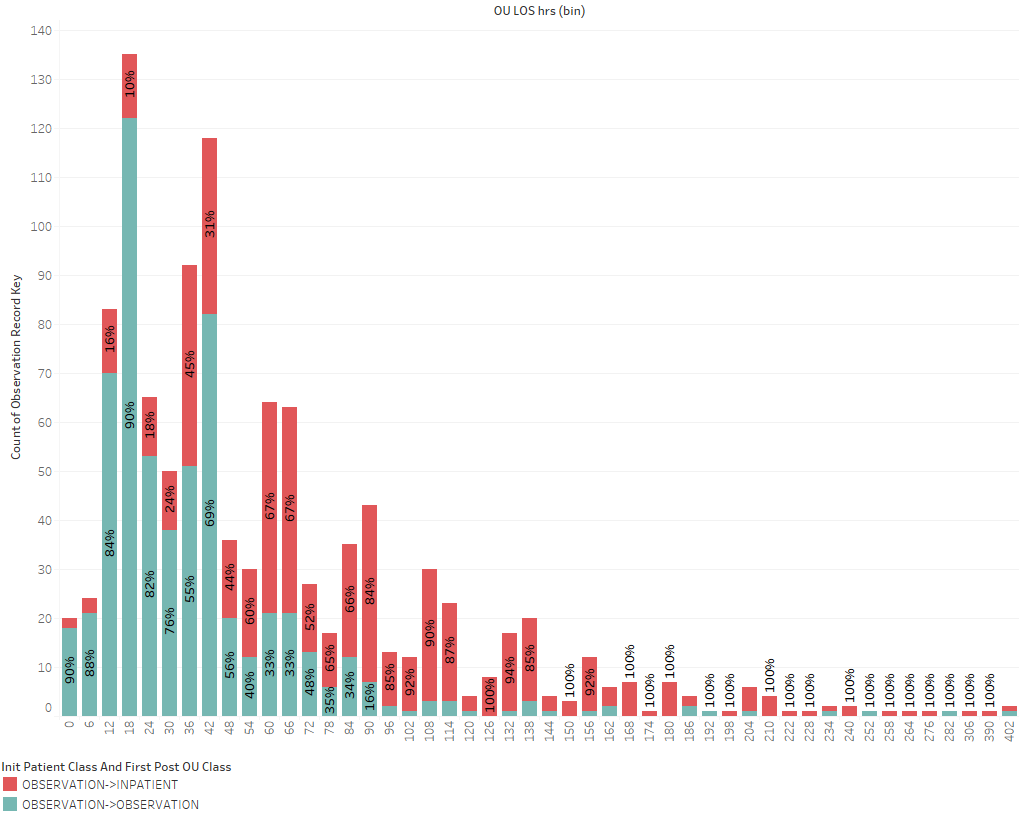
Graph 03:No of Hours in the OU Vs type of Insurance with age

The graph allows you to visualize how the length of stay in the observation unit might be related to a patient's age and insurance type.



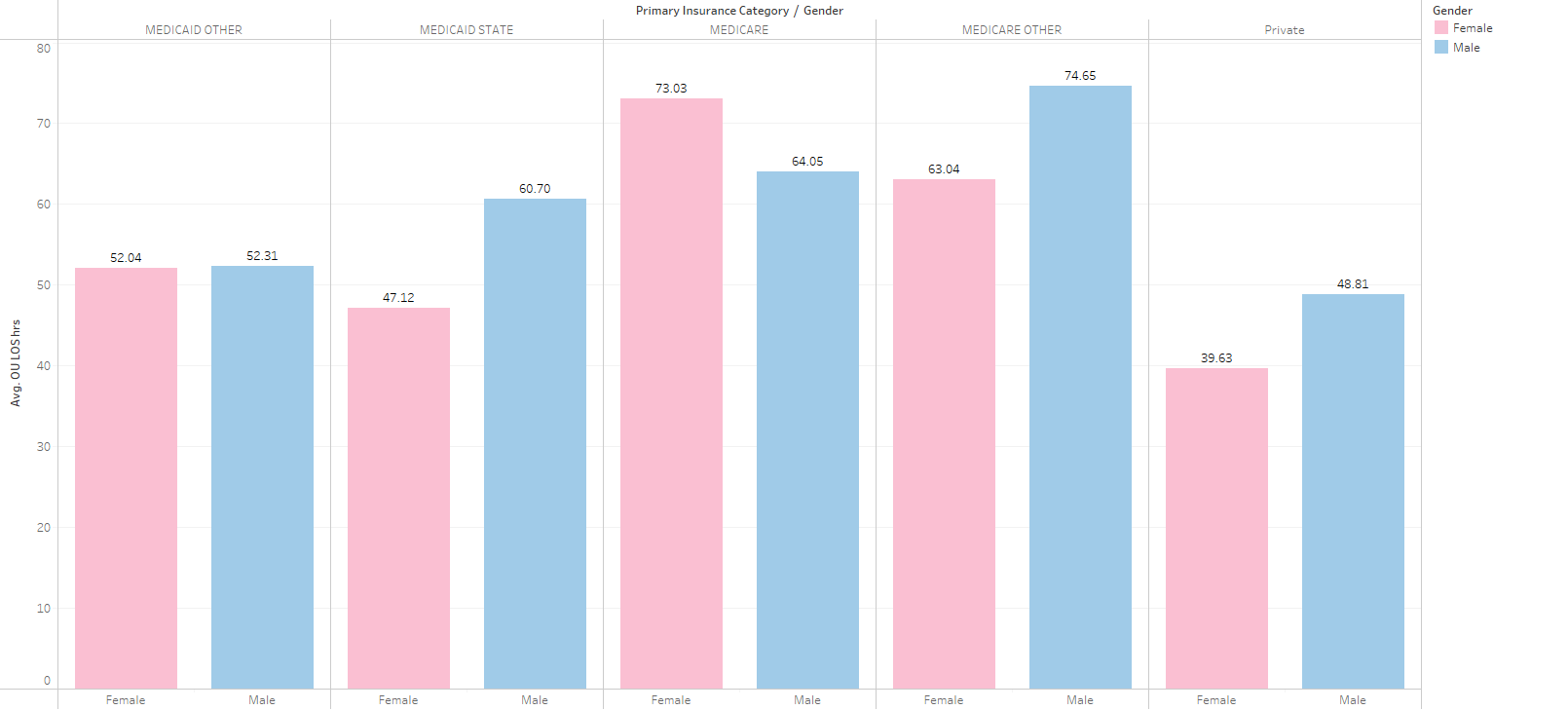
Graph 04: Number of patients in the Observation Unit and their length of stay

The graph aligns with how observation units are intended to work – most patients stay for a short period for evaluation, and only a few require more extended observation periods.



Graph 05: length of stay in the OU against if they were flipped or not

Higher percentage of patients transition to inpatient care the longer they stay in observation. This might indicate a need for closer monitoring or earlier intervention for patients who are in the observation unit for longer durations.



Graph 06: Type of Insurance and Gender

The graph reveals potential differences in the male-to-female ratio across different insurance groups. This might be due to various factors like specific health needs, insurance coverage options, or sociodemographic factors affecting healthcare access.