Capstone_Two_Final_Report

Problem Overview:

The objective of this data science project is to analyze patient data to gain insights that could improve hospital operations, patient care, and cost management. Specifically, the analysis aims to address the following questions:

1. How can patient information be used to predict hospital costs and length of stay?
2. What patterns can be identified in the patients' demographic and medical history that impact these variables?
3. What actionable insights can help in resource allocation, patient management, and financial planning?

Approach:

The project was divided into several phases, corresponding to the Jupyter notebooks provided:

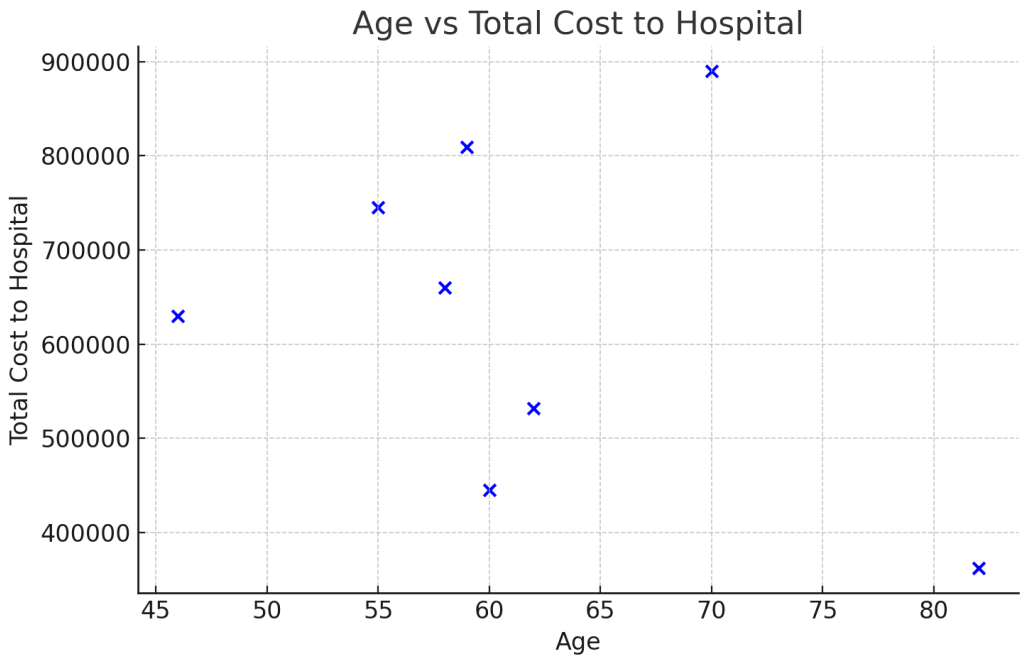
1. Data Wrangling:

This phase involved cleaning and preparing the dataset by dropping irrelevant columns like 'Total Length of Stay-ICU', 'Ward Stay', and 'Implant Used'. Important features such as patients' medical history, vital signs, and cost data were retained for analysis.

2. Exploratory Data Analysis (EDA):

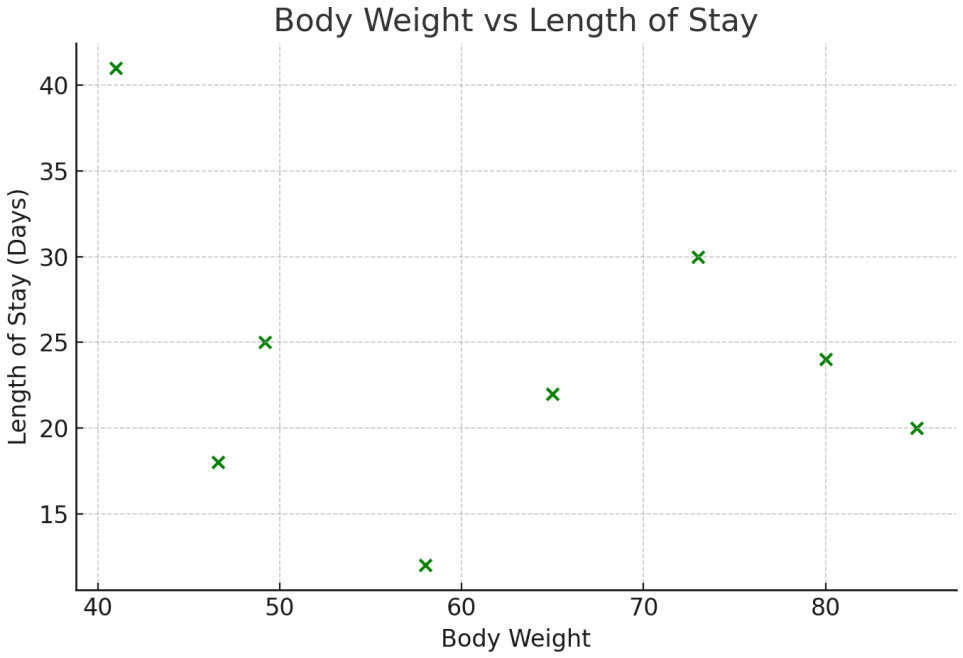
Relationships between patient demographics, medical history, and hospital costs were explored. Key variables like age, body weight, and medical history were found to influence the total hospital cost and length of stay.

Chart 1: Age vs Total Cost to Hospital



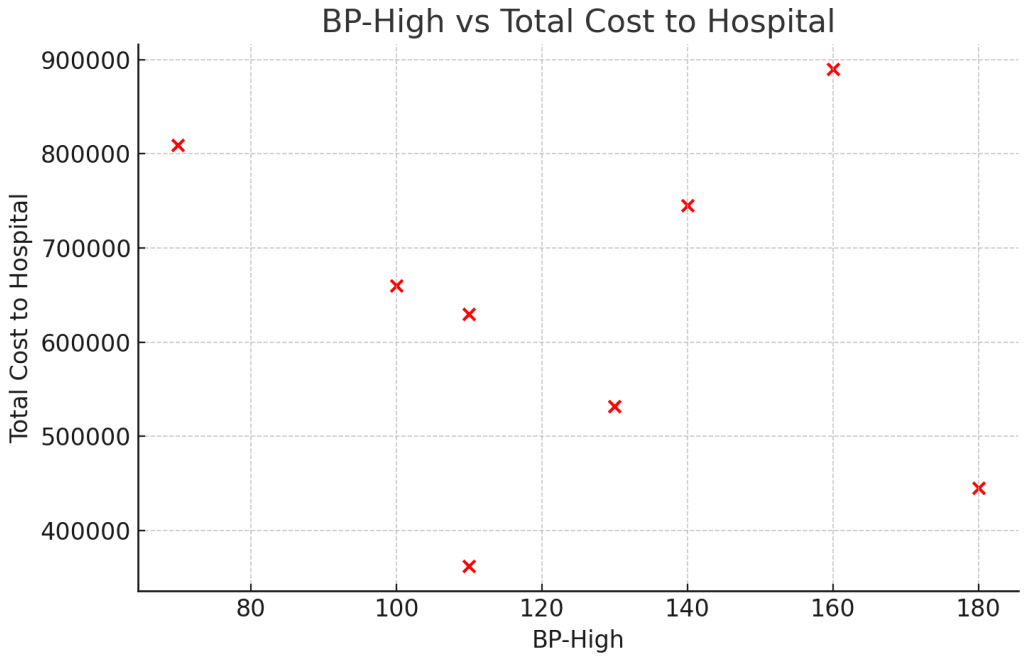
This chart shows that older patients generally incur higher hospital costs, likely due to the increased complexity of care and extended treatments required for elderly patients.

Chart 2: Body Weight vs Length of Stay



This chart highlights that patients with higher body weight tend to have longer hospital stays, likely due to obesity-related complications and longer recovery times.

Chart 3: BP-High vs Total Cost to Hospital

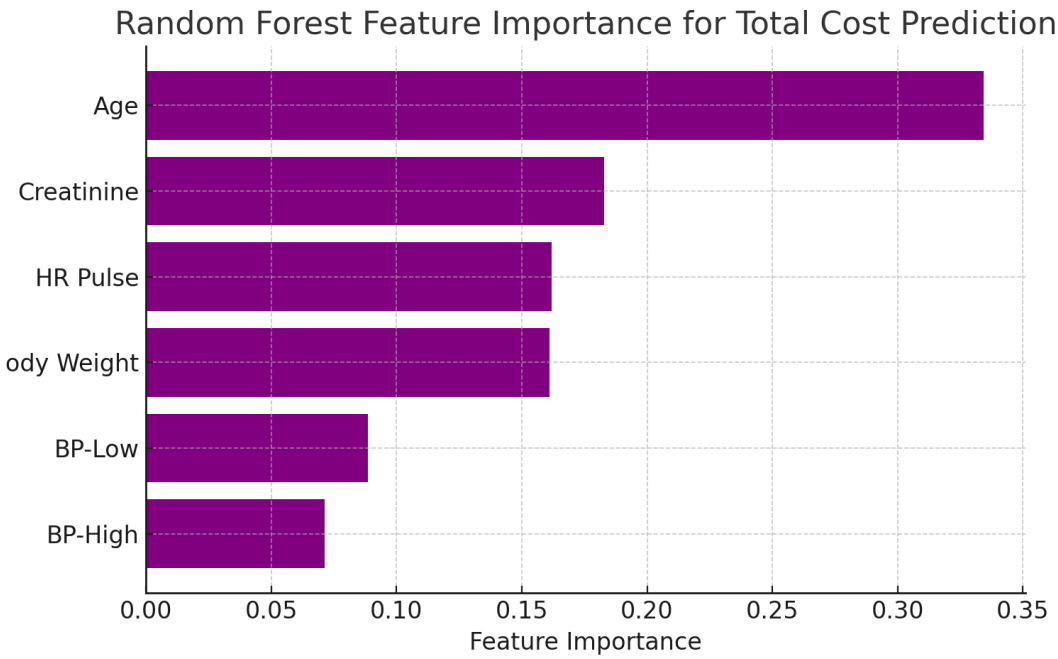


This chart suggests that higher systolic blood pressure readings at admission are associated with increased hospital costs, indicating a need for more intensive monitoring and treatment.

4. Modeling:

Several machine learning models were tested, including linear regression, decision trees, and random forest models, to predict hospital costs and length of stay. Random forest performed the best in capturing non-linear relationships and predicting outcomes.

Feature Importance from Random Forest Model:



The Random Forest model's feature importance chart shows that age and body weight were the most critical features in predicting hospital costs, followed by blood pressure levels and creatinine levels.

Findings:

1. Key Variables Influencing Hospital Costs:

Age, body weight, and creatinine levels were found to be strong predictors of hospital costs. Older patients and those with higher body weight tended to incur more significant expenses, possibly due to increased complications and longer stays.

2. Length of Stay Predictors:

The length of stay was significantly affected by the mode of arrival (e.g., ambulance) and patient health condition at admission. Patients with heart conditions or chronic diseases like diabetes and hypertension had longer hospital stays.

Recommendations:

1. Optimize Patient Flow:

Hospitals should focus on early intervention for high-risk patients (elderly, obese, or those with chronic conditions) to reduce the length of stay. This might involve faster diagnostics and preventive care measures.

2. Cost Prediction for Budgeting:

Integrating the cost prediction model into hospital financial systems will help in resource allocation and budget planning. Using admission data to predict total costs will allow hospitals to plan ahead and allocate resources efficiently.

3. Enhanced Resource Allocation:

Identifying patients likely to have longer stays can help hospitals better allocate ICU resources, staff, and beds. Prioritizing high-risk patients for early intervention will reduce bottlenecks in care.