Capstone_Two_Presentation

Analysis of Hospital Data for Cost and Length of Stay Predictions

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. The key questions include:
- 1. Predicting hospital costs and length of stay based on patient data.
- 2. Identifying patterns in demographics and medical history that impact these variables.
- 3. Offering actionable insights for resource allocation and financial planning.

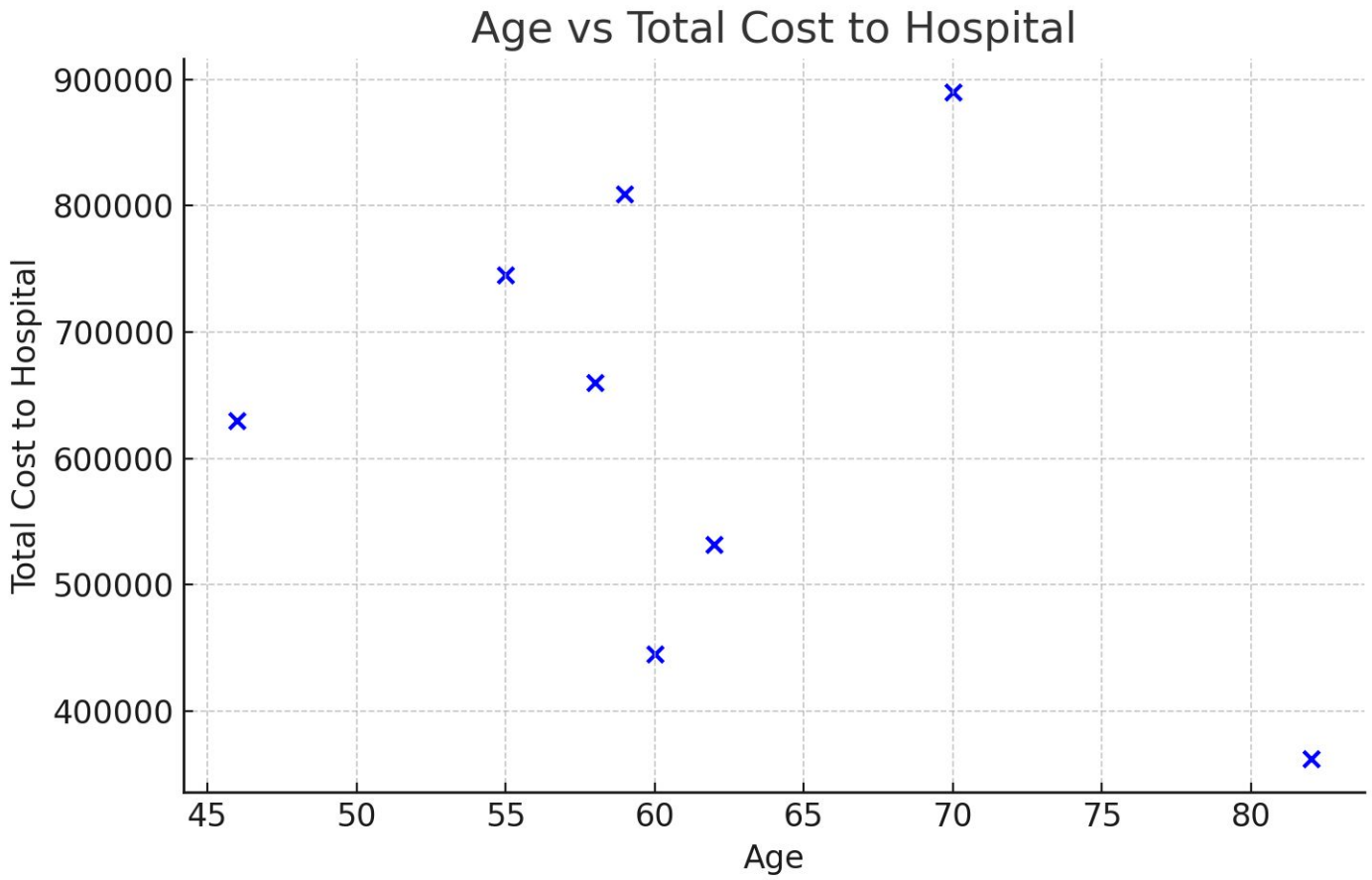
Approach

- The project was divided into the following phases:
- 1. Data Wrangling: Cleaned and prepared the dataset by dropping irrelevant columns.
- 2. Exploratory Data Analysis (EDA): Explored relationships between demographics, medical history, and hospital costs.
- 3. Preprocessing and Model Training: Prepared the dataset for machine learning by encoding and scaling features.
- 4. Modeling: Trained models (e.g., Random Forest) to predict hospital costs and length of stay.

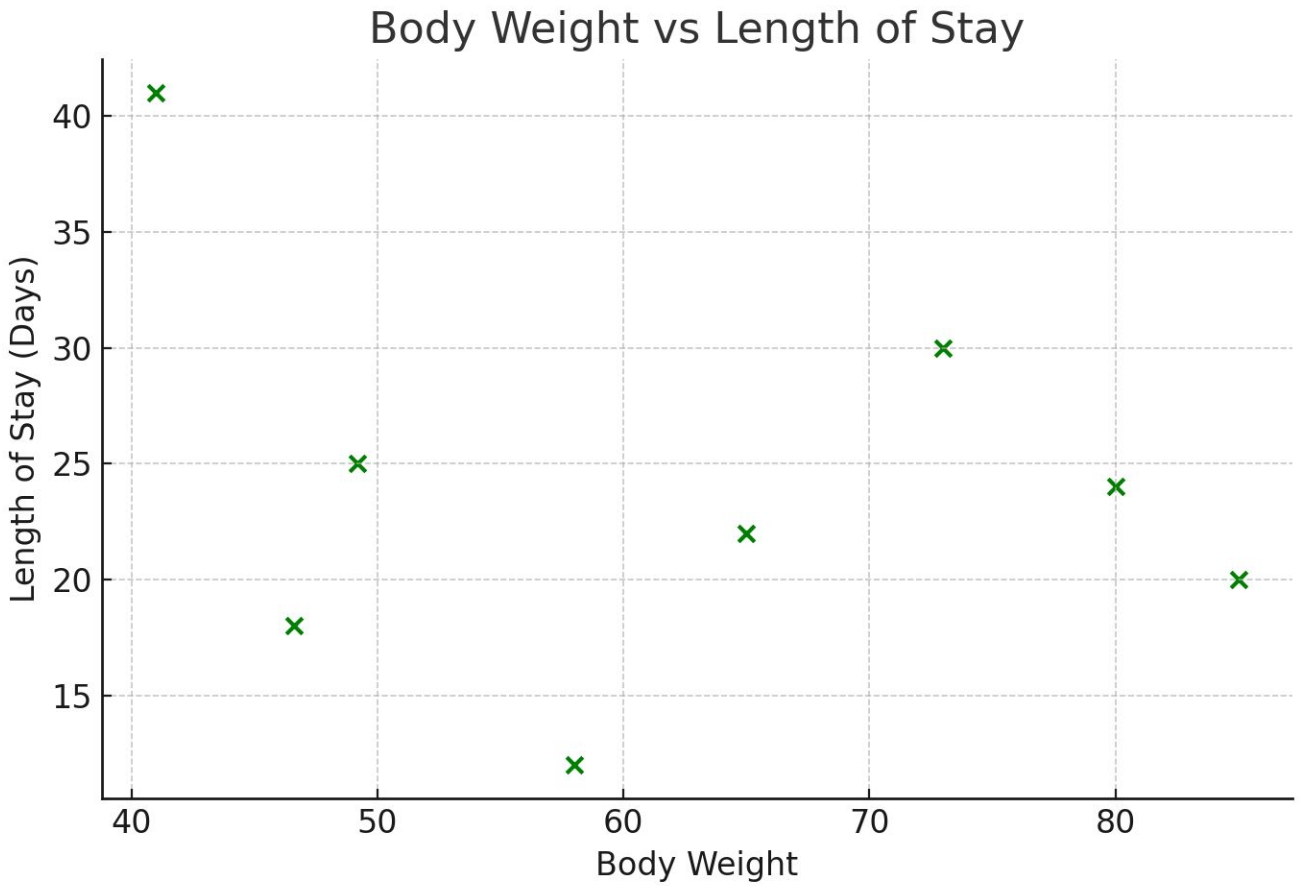
Exploratory Data Analysis (EDA)

- Key findings from EDA include:
- 1. Older patients tend to have higher hospital costs.
- 2. Patients with higher body weight tend to have longer stays.
- 3. High systolic blood pressure readings are associated with increased hospital costs.

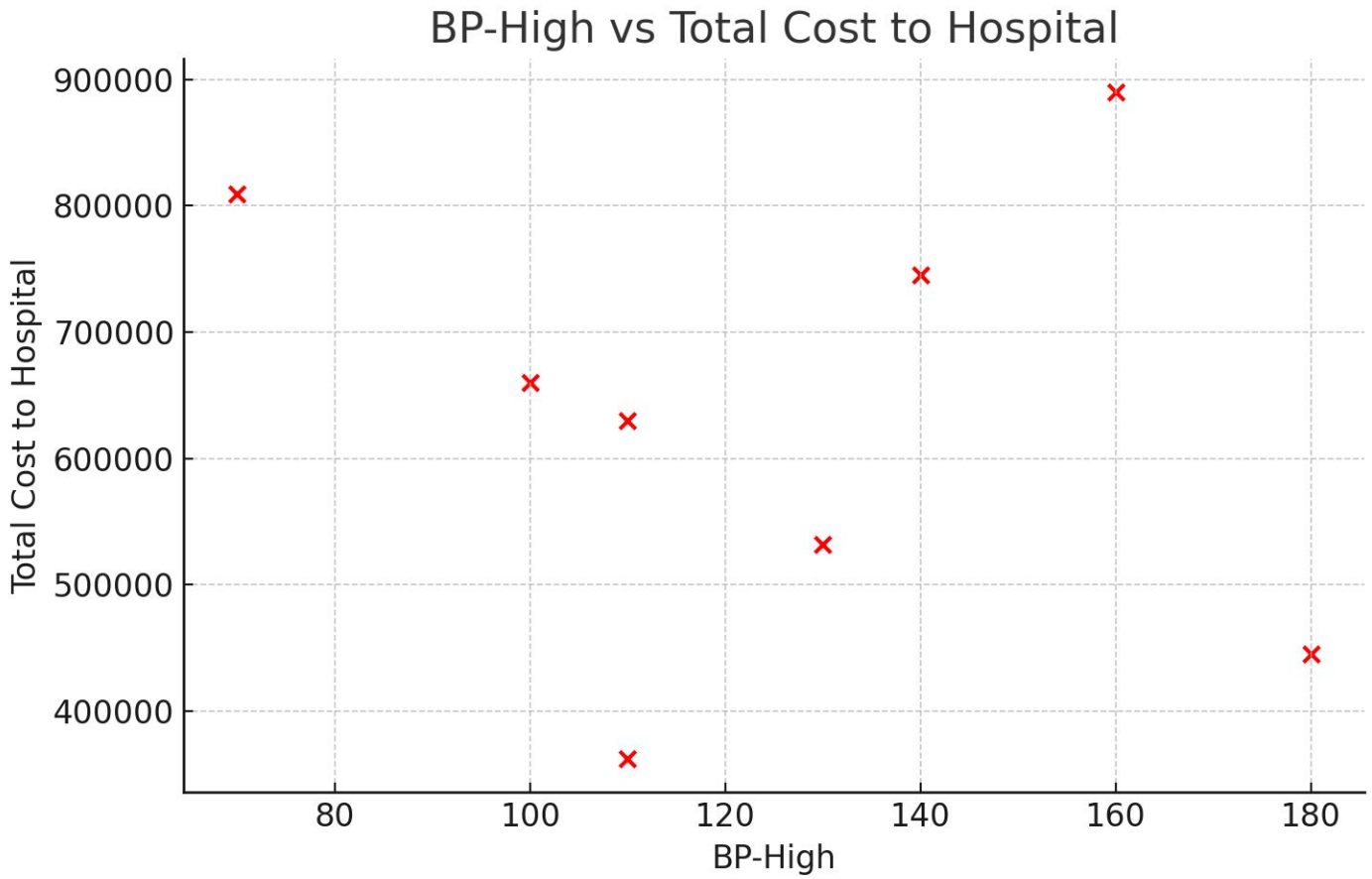
Age vs Total Cost to Hospital



Body Weight vs Length of Stay



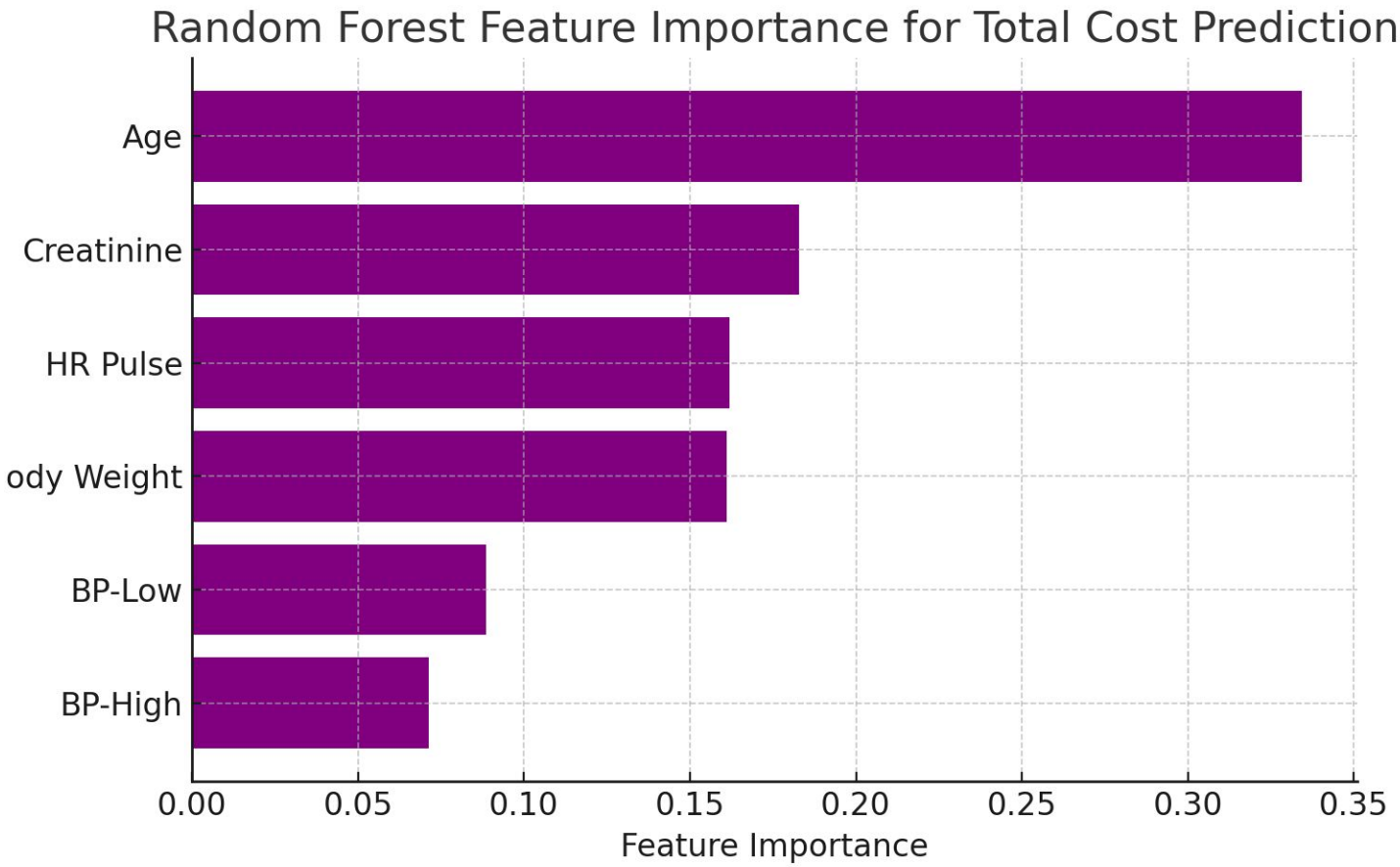
BP-High vs Total Cost to Hospital



Modeling Results

- Random forest and gradient boosting models were tested for predicting hospital costs and length of stay. The Random Forest model performed the best, capturing complex relationships between patient features and outcomes.

Feature Importance from Random Forest



Findings

- 1. Key variables influencing hospital costs include age, body weight, and creatinine levels.
- 2. Length of stay was most significantly impacted by mode of arrival (e.g., ambulance) and medical conditions like heart diseases.
- 3. Random forest models outperformed simpler models like linear regression.

Recommendations

- 1. Optimize Patient Flow: Focus on early intervention for high-risk patients to reduce length of stay.
- 2. Use Cost Prediction Models: Integrate predictive models into financial systems for better resource allocation.
- 3. Enhance Resource Allocation: Preemptively allocate ICU resources based on patient profiles.

Further Research

- 1. Predicting Patient Outcomes: Extend the model to predict patient outcomes, such as readmission rates.
- 2. Incorporate External Data: Use socioeconomic or insurance data to improve prediction accuracy.
- 3. Real-Time Decision Support: Build a system that assists hospital staff with real-time decision-making.