**Healthcare Analytics Report**

**Introduction:** Healthcare institutions constantly strive to improve patient care and resource allocation. Predicting patient length of stay (LOS) upon admission is crucial for efficient resource management and planning. This report outlines a healthcare analytics project aimed at predicting LOS using machine learning techniques.

**Data Generation:** Synthetic patient data was generated, including features such as age, gender, weight, height, smoker status, and LOS. The dataset was saved as "patient\_data\_large.csv" and served as the foundation for our analysis.

**Data Preprocessing:** Categorical variables were encoded using one-hot encoding to convert them into numerical values. The dataset was then split into training and testing sets in an 80-20 ratio. Standard scaling was applied to ensure feature consistency and to bring all features to the same scale.

**Model Training:** Two models were trained: a Multi-layer Perceptron (MLP) Regressor and an XGBoost Regressor. The MLP Regressor consisted of hidden layers with sizes (100, 50), while the XGBoost Regressor utilized the default hyperparameters.

**Model Evaluation:** The trained models were evaluated on the testing set to assess their predictive performance. The MLP Regressor achieved a marginally negative accuracy of approximately -0.0016, while the XGBoost Regressor performed slightly better with a negative accuracy of approximately -0.1448.

**Actionable Insights:** Feature importance analysis was conducted on the XGBoost model to identify the significant predictors influencing LOS. Age, gender, weight, and height were identified as the most important factors, with smoker status showing no importance in LOS prediction according to the model.

**Sample Patient Data and Predictions:** For a specific patient scenario represented by sample data, predictions were made using the neural network model. The predicted length of stay (LOS) for this patient was approximately 14.63 days. This prediction provides valuable insights for healthcare practitioners in managing patient care and resource allocation.

**Conclusion:** While the predictive accuracy of the models may be suboptimal, the feature importance analysis and individual predictions offer actionable insights for healthcare management. Further refinement and optimization of the models may be necessary to improve their predictive performance and enhance their utility in real-world healthcare settings.