Healthcare Multivariable Linear Regression Assignment

Scenario:

A hospital wants to optimize patient care by predicting the length of hospital stay (LOS) based on demographic and clinical data.

Your Task:

Use multivariable linear regression to predict the length of hospital stay using the provided dataset.

Predictor Variables:

1. Age (years)
2. Sex (0 = Female, 1 = Male)
3. BMI (Body Mass Index, kg/m2)
4. Severity Score (scale 1 to 10)
5. Comorbidities Count (number of chronic illnesses)
6. Previous Admissions (hospitalizations in last year)
7. Admission Type (0 = elective, 1 = emergency)
8. Blood Pressure (systolic, mmHg)
9. Heart Rate (beats/min)
10. WBC Count (white blood cells per microliter)

Dependent Variable:

- Length of Stay (LOS, in days)

Instructions:

1. Data Preparation:

- Perform exploratory data analysis (EDA).
- Handle missing values, if any.
- Encode categorical variables appropriately.

2. Model Building:

- Fit a multivariable linear regression model.
- Validate model assumptions: linearity, homoscedasticity, normality of residuals, and multicollinearity.

3. Evaluation:

- Use R-squared, Mean Squared Error (MSE), and Mean Absolute Error (MAE) for model evaluation.
- Discuss significant predictors and their impact on LOS.

4. Reporting:

- Clearly document your findings, interpretations, and implications for hospital management.