





Reflection:

In this project, we used **Linear Regression** to predict healthcare costs based on patient demographics and lifestyle factors from the "Medical Cost Personal Dataset." This dataset was chosen because it includes critical features like **age**, **BMI**, **smoking status**, and **region**, which are common factors in determining medical expenses. It provides a realistic scenario for modeling healthcare costs and has a good balance of categorical and continuous variables, making it ideal for regression analysis.

Key predictors like **smoking status** and **BMI** were strongly associated with higher medical costs, while **age** had a moderate correlation. **Region** and **gender** had less impact on costs. The model was evaluated using **Mean Squared Error (MSE)**, showing reasonable performance, though more complex models could potentially capture non-linear relationships better.

Overall, the dataset was a good fit for understanding healthcare cost drivers, and the model offers useful insights while highlighting opportunities for further improvement.