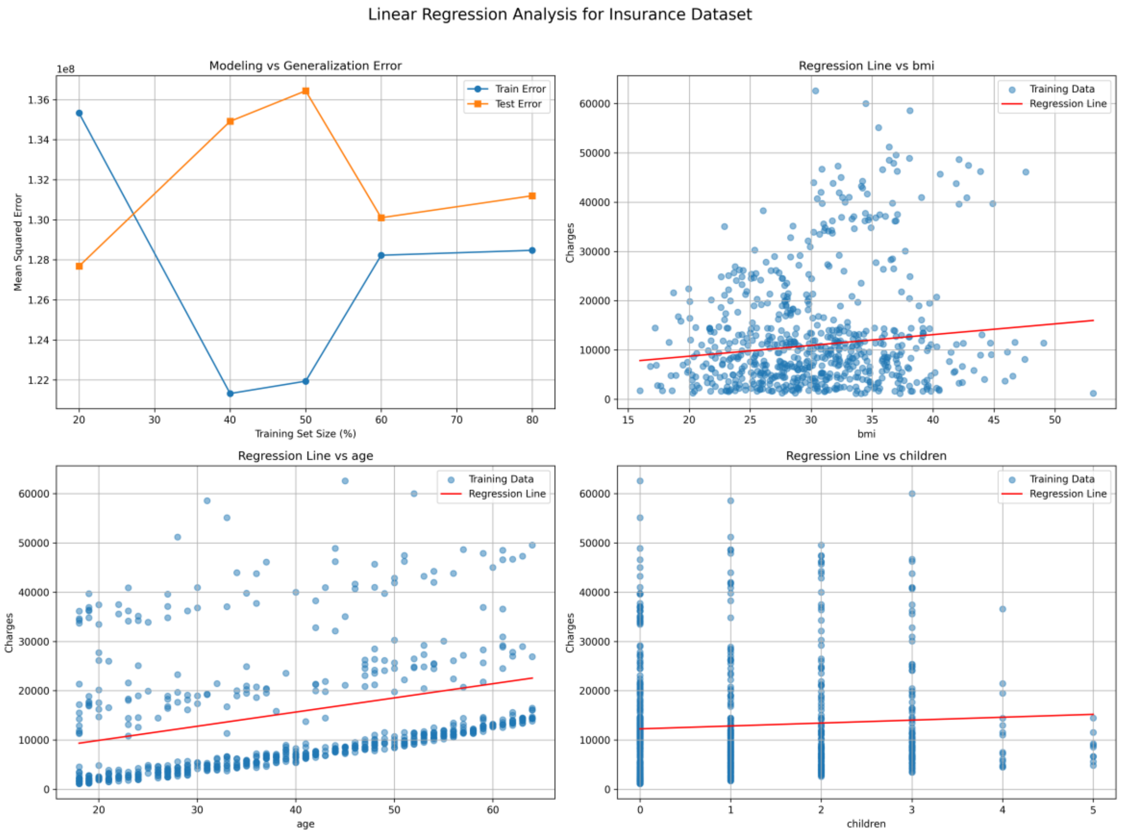
Group Members:

Derek Corniello, Nathan Grilliot, Ryan Sippy

All members contributed equally



The linear regression model shows only moderate predictive ability when restricted to numerical features (age, BMI, and children count). The modeling vs. generalization error reveals that increasing the training data from 20% to 80% consistently lowers training error as the model better fits observed data. However, test error shows inconsistent behavior, with spikes in the 40-50% range, suggesting poor generalization due to insufficient features. At the 50% split point, test and training errors reach similar values, which is the optimal balance between under and overfitting.

Examining individual feature relationships, age demonstrates the strongest linear correlation with insurance charges, showing a clear positive trend. BMI exhibits only a weak relationship with substantial variance, while number of children shows almost no meaningful pattern, resulting in a nearly horizontal regression line. The model's effectiveness is limited by its basic structure and the omission of categorical variables, which may contain significant predictive information. Future improvements should incorporate categorical features and expand to non-linear models.