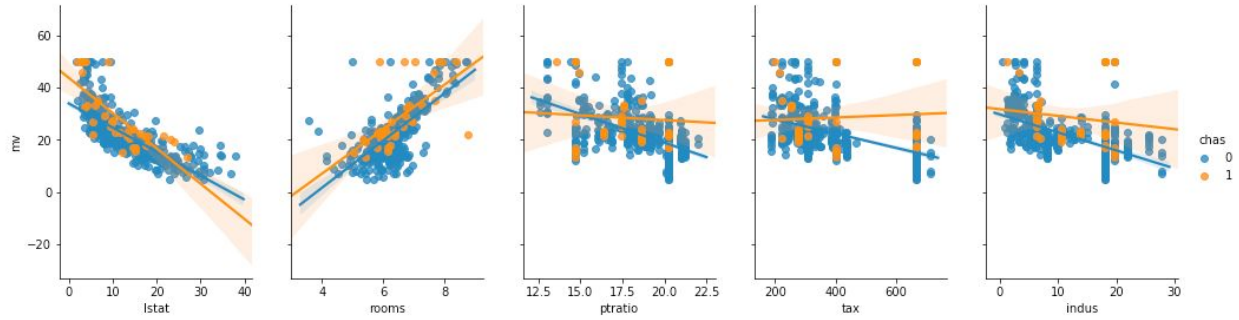
Colab

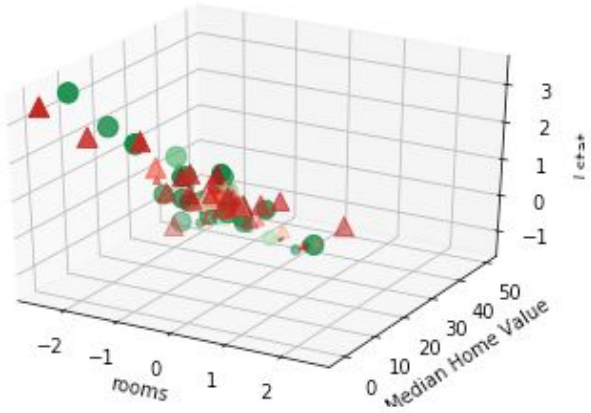
Problem Definition: The goal of this assignment was to use multiple linear regression and ElasticNet to create a model that accurately predicted the median value of homes in thousands. Model efficacy was evaluated using RMSE and R^2.

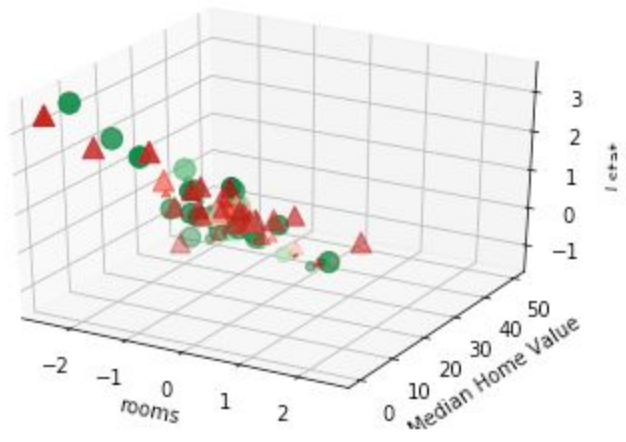
Research Design/Programming: To begin, features of interest were selected using a correlation matrix. Features with a relationship to the dependent variable ≥ 0.5 or $\leq$-0.5 were viewed closer using scatterplots. Ultimately 5 features were selected, including



"lstat","rooms","ptratio","age" and "nox." An 85/15 train/test split was used . All independent variables were normalized accordingly.

Below are two visualizations, demonstrating the difference between the predicted and expected y values for both Linear (Left) and ElasticNet (right) models. The red triangles represent the model predictions and the green circles represent the expected values.





Recommendations for Management: The ElasticNet model outperformed the Multiple Linear Regression model during cross validation and the final out of sample evaluation. The RMSE explains that on average the ElasticNet predicted values were approximately $3731.19 off the expected values. Moving forward I recommend collecting more data. A larger training set will likely improve model accuracy and performance. We will also need more test data to accurately gauge the models generalization accuracy.

```
Cross-Val Comparison:

RMSE Linear Regression: 5463.93
RMSE ElasticNet: 5456.91
R2 Linear Regresssion: 65.82
R2 ElasticNet:65.95


Final Model Comparisons:

RMSE Linear Regression: 3808.03
RMSE ElasticNet: 3731.19
R2 Linear Regresssion: 77.99
R2 ElasticNet:78.87
```