**Linear Regression Analysis**

This repository contains an example of a regression analysis on the [used car dataset](https://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekho). This should be used as a guide as to expectations on format, best practices, and what should be included in your analyses.

**Dataset Description**

The data being explored comes from [scikit-learn](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_california_housing.html#sklearn.datasets.fetch_california_housing) and [statsmodel](https://www.statsmodels.org/stable/index.html). It is a dataset containing attributes that relate to used car prices.

**Business Question/Problem Statement**

Being able to understand car information and predict used car values can be used in policy and investment decisions. From a usage, car model perspective, understanding what drives car prices can be used to normalize variances in car prices and create greater general affordability. From an seller perspective, being able to predict values could be used to identify under-valued and most-valued cars.

**Regression Results**

We captured 97% of the variation in car values with a simple regression model. We found engine, max power and seats were the two strongest drivers of car values.

**Predictions Using this Dataset**

We were able to predict median home values with a r-squared of 73% on a test dataset. We attempted regularization via Ridge but found no benefit to adding regularization to our models.

**Potential Next Steps and Follow-ups**