

# ISYE 7406 - Homework 2

Matthew Bernstein

February 1, 2023

## Introduction

## Exploratory Data Analysis

## Methodology

Eight models were created from the data in order to understand the different types of Linear Regressions

## Full Model

A least squares linear model was fitted on all features, excluding *siri*, *density*, *free* as noted above.

## 5 Best Feature Selection

Two models were created by selecting the 5 best features and performing a regression against those. The first model was created by selecting the 5 features with the lowest p-scores in the full model. The second model was selected by performing an exhaustive search over all combinations of 5 features and selecting the combination with the lowest Mallows'  $C_p$  score.

## Stepwise Regression

A model was created using a bi-directional stepwise regression. Beginning with an empty model, a single feature was added or removed and the AIC of the resulting model was calculated. This process was repeated until no changes to the model from the previous step improved upon the model.

## **Ridge Regression**

A Ridge regression model was created using the full dataset. The  $\lambda$  parameter was selected using a 5-fold cross validation selecting for the minimum MSE of the training set.

## **Results**

## **Conclusion**