Real Estate Price Prediction with Elastic-net Regression

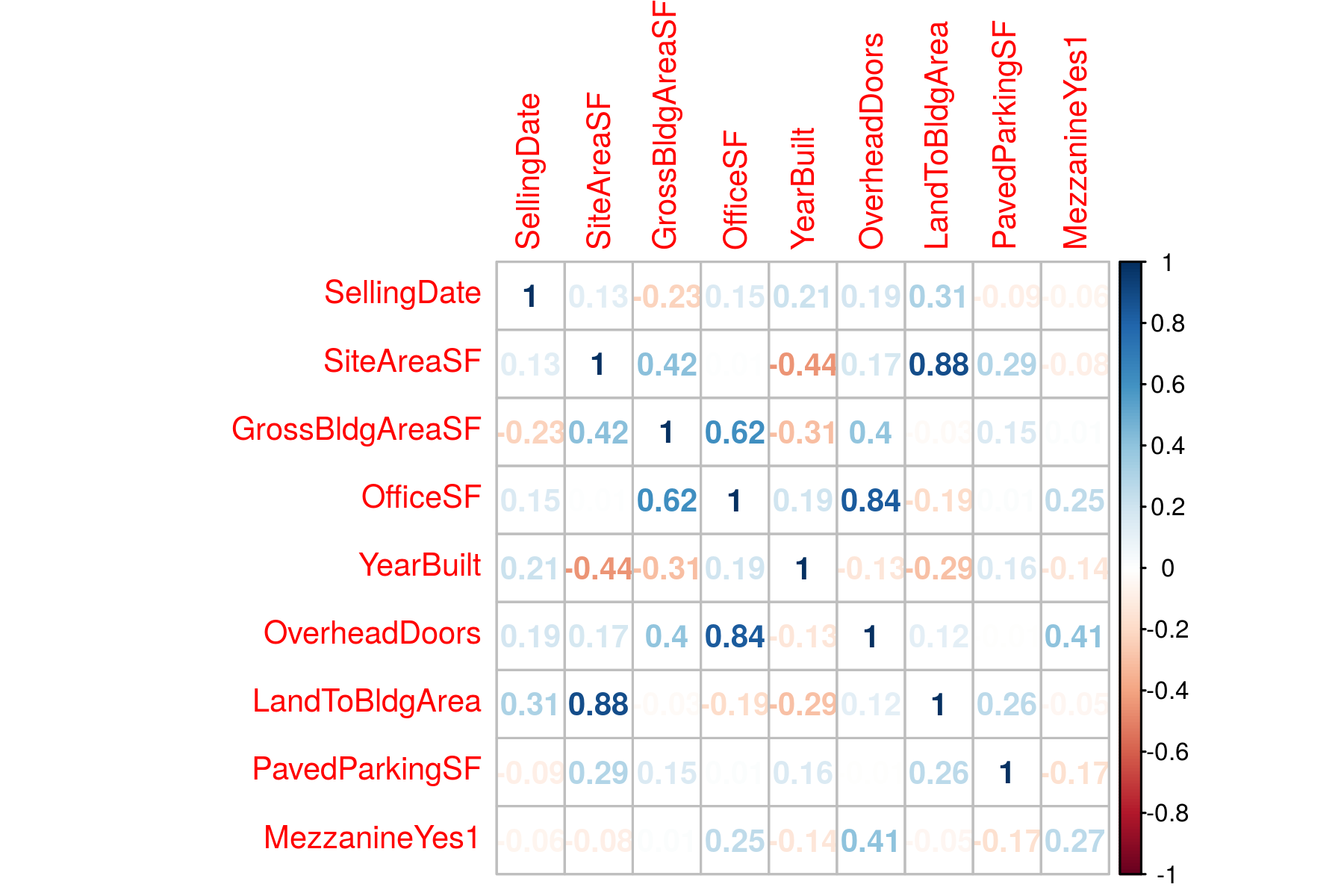
Input Data file: data/Dataset7Obs.csv

## Basic summary statistics

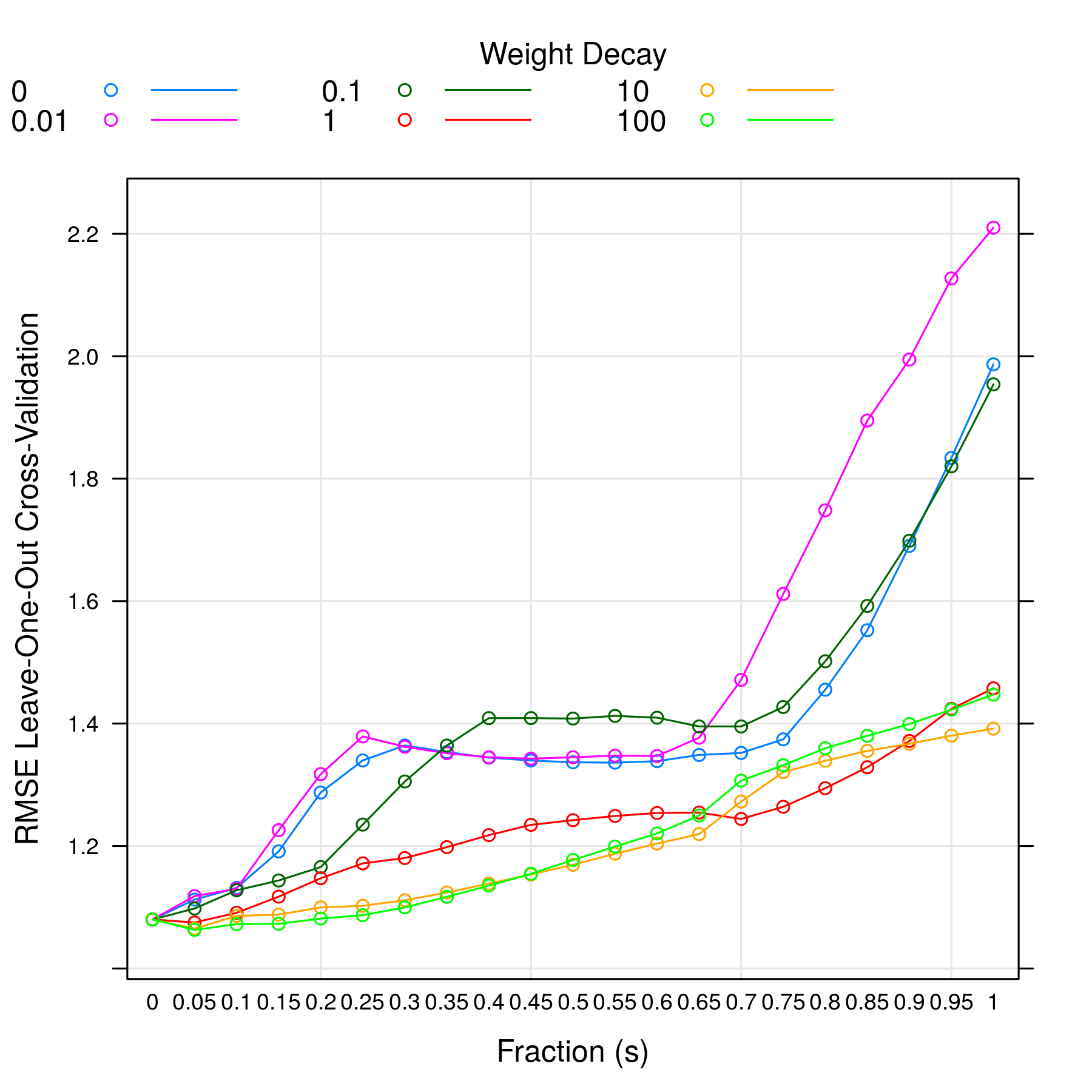
|  | **N** | **Mean** | **Std Dev** | **Min** | **Max** |
| --- | --- | --- | --- | --- | --- |
| SellingPrice | 007 | 364285 | 101916 | 185000 | 450000 |
| SellingDate | 007 | 40597 | 1109 | 39006 | 42308 |
| SiteAreaSF | 007 | 30408 | 13137 | 14015 | 50094 |
| GrossBldgAreaSF | 007 | 6408 | 1123 | 5400 | 8400 |
| OfficeSF | 007 | 1189 | 555 | 768 | 2400 |
| MezzanineYes1 | 007 | 000 | 000 | 000 | 001 |
| YearBuilt | 007 | 2003 | 003 | 1999 | 2010 |
| OverheadDoors | 007 | 002 | 001 | 001 | 006 |
| LandToBldgArea | 007 | 004 | 001 | 002 | 007 |
| PavedParkingSF | 007 | 7890 | 5653 | 1800 | 17088 |

NOTE - No summary statistics are provided for categorical variables.

## Correlations Between Predictors



## Tuning Parameter Selection Using LOOCV



From above plot, lambda =100.00 and s =0.05 gives the minimum RMSE model.

## Variable Importance



## Standardized Model Coefficients

| **Variable** | **Estimate** |
| --- | --- |
| SellingDate | 0.000 |
| SiteAreaSF | 0.115 |
| GrossBldgAreaSF | 0.058 |
| OfficeSF | 0.000 |
| YearBuilt | 0.000 |
| OverheadDoors | 0.000 |
| LandToBldgArea | 0.020 |
| PavedParkingSF | 0.000 |
| MezzanineYes1 | 0.000 |

NOTE std. errors are calculated using bootstrapping which is the only way to determine coef. errors for a penalized regression. But the errors should be only used for reference. It is yet unclear how meaningful the std. errors are in penalized regression.

## Model Prediction

| **Predicted Value** | **Prediction Error** | **R2** |
| --- | --- | --- |
| 366869.386 | 1.063 | 0.266 |