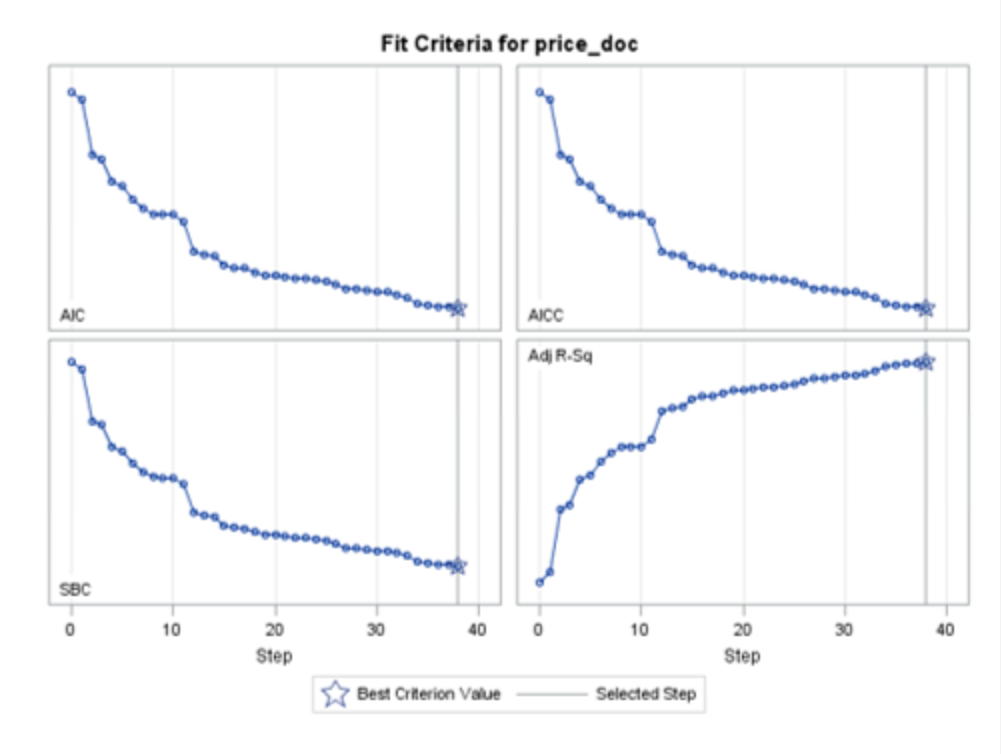
**TODO**

1. **plot box plot comparisons of high leverage points vs. other points to see if there is difference**
2. **residual plots for models from caret + histograms for residuals**
3. **we may also want to add some plots similar to SAS like this for the train process**



1. **Need to document AIC, BIC, etc.**
2. **Do we need standardization in LASSO,**
3. **try model with transformation on y3**

**REASONS for not such a good fit**

1. **Most inputs were not directly correlated to a large extent with the output. Hence, we may need to apply domain expertise and do manual feature engineering**
2. **Or apply other advanced methods such as tree based methods, Neural Networks to get a better fitting model.**

Without Cross validation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Algorithm** | **DF (num)** | **Adj R2** | **RMSE (Train)** | **MSE (Test)** |
| Full Model |  |  |  |  |
| Forward Selection (full data) | **55** | 0.229 | 9.401 | 95.67 |
| Forward Selection (filtered data) | **64** | 0.3329 | 7.243 | 98.21 |
|  |  |  |  |  |
|  |  |  |  |  |
| Stepwise Selection (full data) | **47** | 0.2293 | 9.412 | 95.46 |
| Stepwise Selection (filtered data) | **72** | 0.3268 | 7.265 | 96.95 |
| LASSO (full data) |  |  |  | 94.36 |
| LASSO (filtered data) |  |  |  | 97.31 |

|  |  |
| --- | --- |
| **Algorithm** | **Variables** |
| Full Model |  |
| Forward Selection (full data) | sqrt.x18 + x7 + stat98 + stat110 + x9 + x4 + x16 + x17 + x10 + x8 + stat14 + stat41 + x21 + x11 + stat4 + stat26 + stat54 + stat187 + stat48 + stat92 + stat151 + stat172 + stat100 + stat23 + stat43 + stat214 + stat94 + stat46 + stat37 + stat13 + stat87 + x14 + stat19 + stat35 + stat45 + stat158 + stat156 + stat199 + stat33 + stat70 + stat207 + stat204 + stat25 + stat22 + stat144 + stat38 + stat146 + stat3 + stat60 + x12 + stat89 + stat10 + stat175 + stat141 + stat161 |
| Forward Selection (filtered data) | sqrt.x18 + x7 + stat98 + stat110 + x9 + x4 + x10 + x16 + x17 + x8 + stat14 + x11 + stat172 + x21 + stat128 + stat26 + stat41 + stat4 + stat146 + stat23 + stat54 + stat127 + stat87 + stat100 + stat99 + stat141 + stat15 + stat38 + stat50 + stat70 + stat13 + stat62 + stat144 + x23 + stat156 + stat3 + stat39 + x13 + stat35 + x5 + stat91 + stat167 + stat207 + stat194 + stat92 + stat34 + stat216 + stat73 + stat10 + stat33 + x6 + stat149 + stat214 + stat80 + stat60 + stat61 + stat48 + stat5 + stat183 + stat43 + stat46 + stat175 + stat199 + stat95 |
|  |  |
|  |  |
| Stepwise Selection (full data) | sqrt.x18 + x7 + stat98 + stat110 + x9 + x17 + x16 + x4 + stat14 + x10 + x11 + x8 + x21 + x14 + stat41 + stat4 + stat23 + stat3 + stat60 + stat100 + stat144 + stat33 + stat200 + stat149 + x13 + stat26 + stat24 + stat151 + stat175 + stat126 + stat146 + stat92 + stat22 + stat205 + stat135 +  stat25 + stat73 + stat55 + stat46 + x19 + stat84 + stat19 + stat217 + stat76 + stat13 + stat156 + stat104 |
| Stepwise Selection (filtered data) | sqrt.x18 + x7 + stat98 + stat110 + x9 + x4 + x16 + x17 + x10 + stat14 + x11 + x8 + stat60 + x21 + stat23 + stat41 + stat3 + stat146 + x13 + stat4 + stat149 + stat128 + stat100 + stat144 + stat172 + stat24 + stat85 + stat38 + stat13 + stat33 + stat35 + stat15 + stat175 + stat11 + stat135 + stat61 + stat26 + stat126 + stat84 + stat92 + stat90 + stat207 + x19 + stat59 + stat99 + stat105 + stat87 + stat73 + stat50 + stat182 + stat54 + stat156 + stat32 + stat76 + x5 + stat153 + stat200 + stat177 + x14 + stat210 + stat34 + stat21 + stat174 + stat39 + stat162 + stat55 + stat216 + stat115 + stat9 + stat71 + stat133 + stat217 |

**UPDATED**

**With Cross validation (using caret)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Algorithm** | **DF (num)** | **R2** | **Adj R2** | **RMSE** | **MAE** | **MSE (Test)** |
| Full Model | 240 | 0.2532 | 0.2221 | 9.527 |  |  |
| Filtered Model | 240 | 0.3562 | 0.328 | 7.351 |  |  |
| Forward Selection (full data) | 16 | 0.215 |  | 9.565 | 7.16 | 89.73 |
| Forward Selection (filtered data) | 25 | 0.314 |  | 7.427 | 5.96 | 91.22 |
| Backward Elimination (full data) | 19 | 0.216 |  | 9.562 | 7.16 | 90.30 |
| Backward Elimination (filtered data) | 25 | 0.314 |  | 7.427 | 5.96 | 91.22 |
| Stepwise Selection (full data) | 19 | 0.216 |  | 9.562 | 7.16 | 90.30 |
| Stepwise Selection (filtered data) |  |  |  |  |  |  |
| LASSO (full data) |  | 0.216 |  | 9.564 | 7.18 | 89.96 |
| LASSO (filtered data) |  | 0.316 |  | 7.421 | 5.97 | 90.96 |
| LARS (full data) |  | 0.216 |  | 9.564 | 7.18 | 89.97 |
| LARS (filtered data) |  | 0.316 |  | 7.421 | 5.97 | 90.97 |

|  |  |
| --- | --- |
| **Algorithm** | **Variables** |
| Full Model |  |
| Forward Selection (full data) | (Intercept) x4 x7 x8 x9  x10 x11 x14 x16 x17  x21 stat14 stat60 stat98 stat110  stat149 sqrt.x18 |
| Backward Elimination (full data) | (Intercept) x4 x7 x8 x9  x10 x11 x14 x16 x17  x21 stat13 stat14 stat24 stat60  stat98 stat110 stat144 stat149 sqrt.x18 |
| Stepwise Selection (full data) | (Intercept) x4 x7 x8 x9  x10 x11 x14 x16 x17  x21 stat13 stat14 stat24 stat60  stat98 stat110 stat144 stat149 sqrt.x18 |
| LASSO |  |
| LARS |  |