

Marks

Student ID:

2	5	1	4	0	0	7	F	
---	---	---	---	---	---	---	---	--

Name Stefan Faulkner

Final Model Chosen Housing.step – This was our final model chosen

Were there any outliers present. If yes what methods did you use to find them.

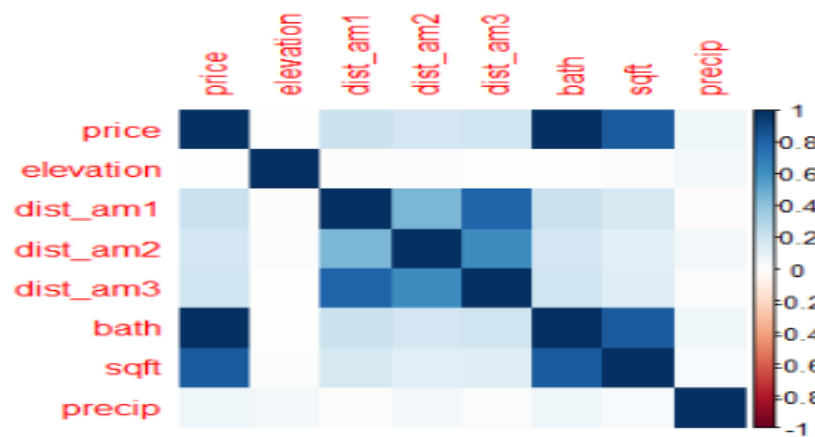
Yes, methods used to find them was when I was performing the EDA and using the outlierTest function from the R Library package car of our final model.

What methods did you use to choose the final model

Our final Model was chosen based on examining our model diagnostics and by the model R^2 (adj) score we got.

Exploratory analysis:

Graph of exploratory analysis



Comments on exploratory analysis

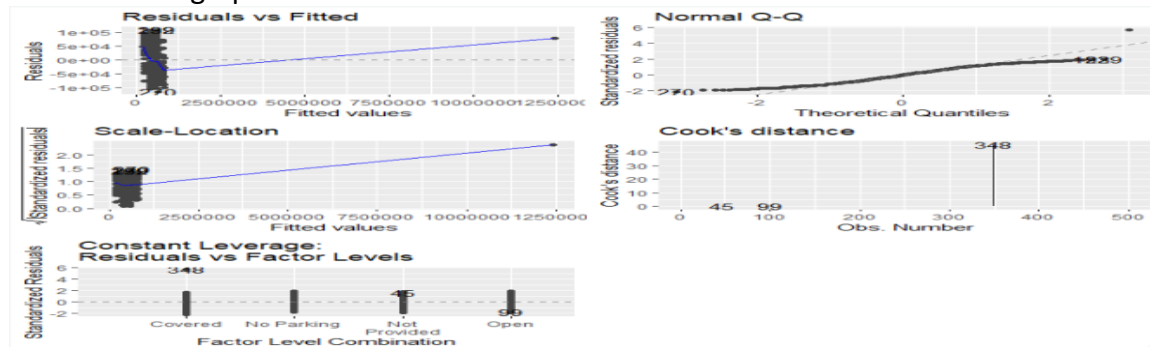
Looking on our correlation plot we can see that price is positively correlated with dist_amt1, dist_am2, dist_am3, bath and sqft. (NOTE: We can see that It doesn't have a strong relation with dist_amt1 and dist_amt2 and dist_amt 3. In fact, we can see there is 0 positive relation with elevation as it relates to our response variable price)

Also, we see that there exists a higher positive correlation with bath and sqft as it is the closest to 1.

For our correlation plot we had to remove our categorical variable Parking.

Model Diagnostics and outlier detection

Methods and graphs



Comments on Model selection

We used Stepwise Regression for our Model Selection. We can see that for the Normal Q Q plot it has lighter tails meaning that the residuals deviate from the diagonal line in both the upper and lower tail. In other words it indicates that it has smaller values than we would expect under the standard modelling assumptions.

Model selection

Methods and steps

For our model we have used the Stepwise Regression as a way of building out our model. It uses the AIC criterion. At each step it drops a predictor and evaluates the AIC of the new model without that predictor. In essence we start with the full model that had an AIC value of 10929.1 and let's dropped a variable in our case the first variable it dropped was parking then the AIC now would be 10927.04. This is the lowest possible AIC value by removing one variable from the full model. Continuing this same procedure, we end up with our final model with two explanatory variables which are bath and sqft. So our final model is $\text{price} \sim \text{bath} + \text{sqft}$. However, note that there are a lot of other model selections which could also have been chosen which could give us a different model.

Final Model selection

Our Final Model Selection was the Stepwise Regression after carefully compare with two other model selections Forward and backward where one uses the p value to remove variables starting from the full model while the other start from the null model respectively.

Concluding remarks

Remarks

In building out our final model, I believe the Stepwise Regression which was chosen as the method can be very effective since it has an advantage since we had a lot of potential explanatory variables in our dataset.

When we examine our R^2 (adj) model we see that it's really good (I decided to use this value since we had more than one explanatory variable although the value was the same with the R^2). The score was approximately 99% of the variability for our response variable price, which can be predicted by the covariates. This indicates a really good fit, hence a good model.