Appendix

Group 2-B: Syrena Hilgendorf, Neha Maddali, Loropu Kokoi, Marie Klapacz

Relevant Variables

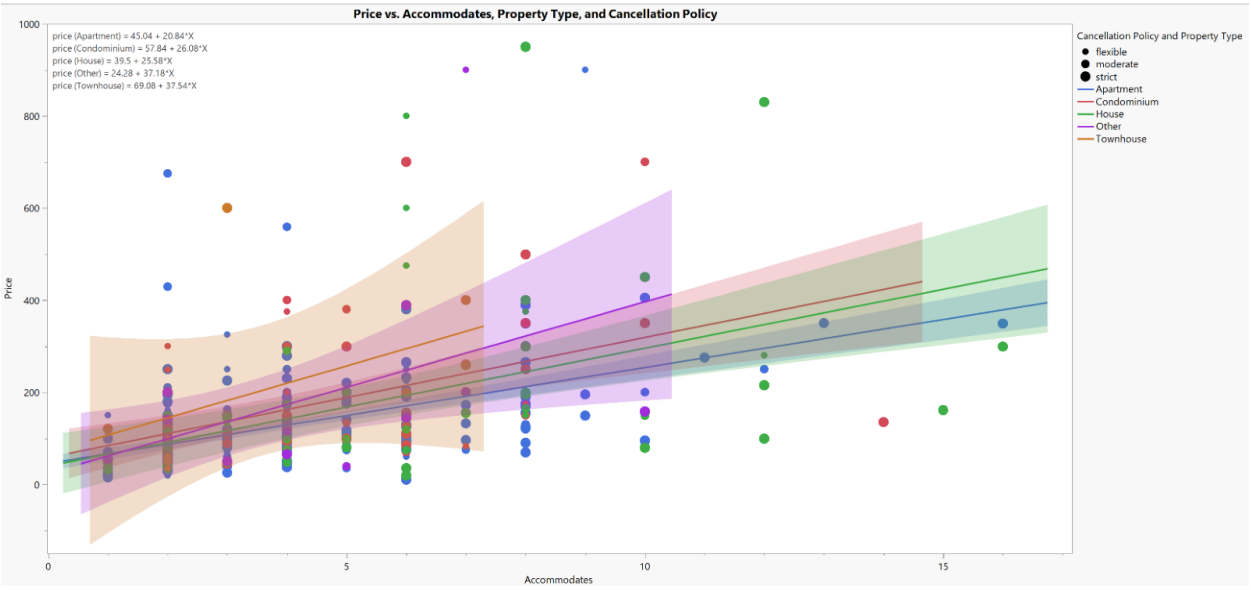

Figure 1: Graph from Graph Builder including the variables price, accommodates, property_type, and cancellation_policy. Price is the response variable. Accommodates is a quantitative explanatory variable. Property_type and cancellation_policy are qualitative explanatory variables. Accommodates is plotted along the x-axis with price on the y-axis. Cancellation policy is indicated by the size of the circle data point. The color is representative of the property type. The slopes of each of the lines are not that significant in difference. The predicted price for Apartments = 45.04 + 20.84*accommodates. The predicted price for Condominiums = 57.84 + 26.08*accommodates. The predicted price for Houses = 39.5 + 25.58*accommodates. The predicted price for Other = 24.28 + 37.18*accommodates. The predicted price for Townhouses = 69.08 + 37.54*accommodates. There is no missing data for our model, so this is not a concern for this dataset. The sample size for this dataset is 500 Airbnb's. When constructing models, independence is met because the data was collected by using a simple random sample, meaning that the Airbnb's did not influence one another.

Full Model

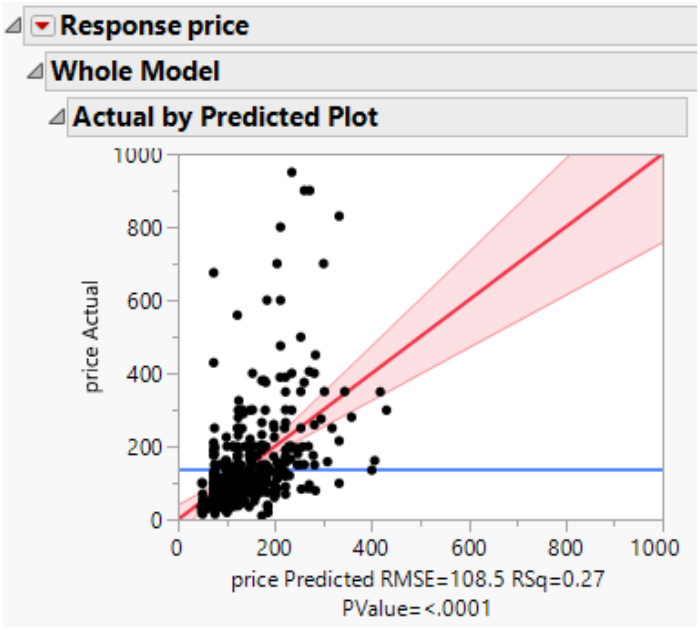

Figure 2: Actual by Predicted Plot visual assessment of model fit plotting the observed values of price (response variable) against the marginal predicted values of price. Based on the scatterplot, linearity is met because there is not a curve present.

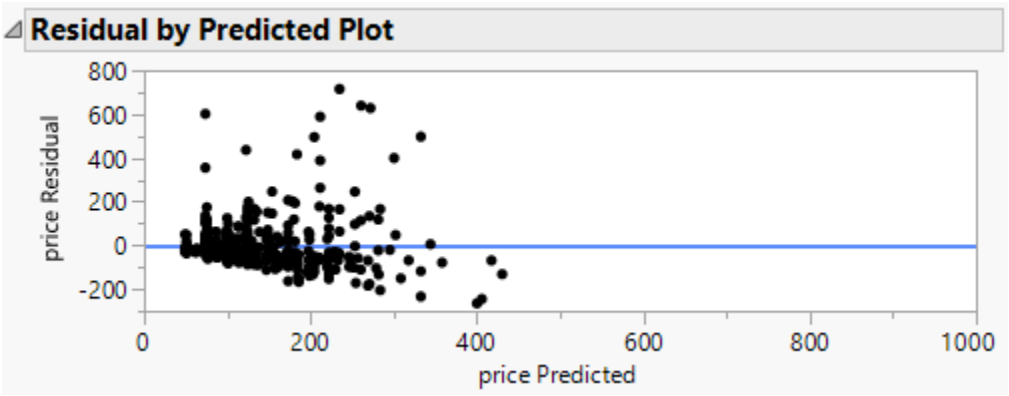

Figure 3: Residual by Predicted Plot indicating the difference between the observed price and fitted price values. The plot shows a fan and the means are not equally distributed above and below the mean of zero, indicating that equal variance is not met.

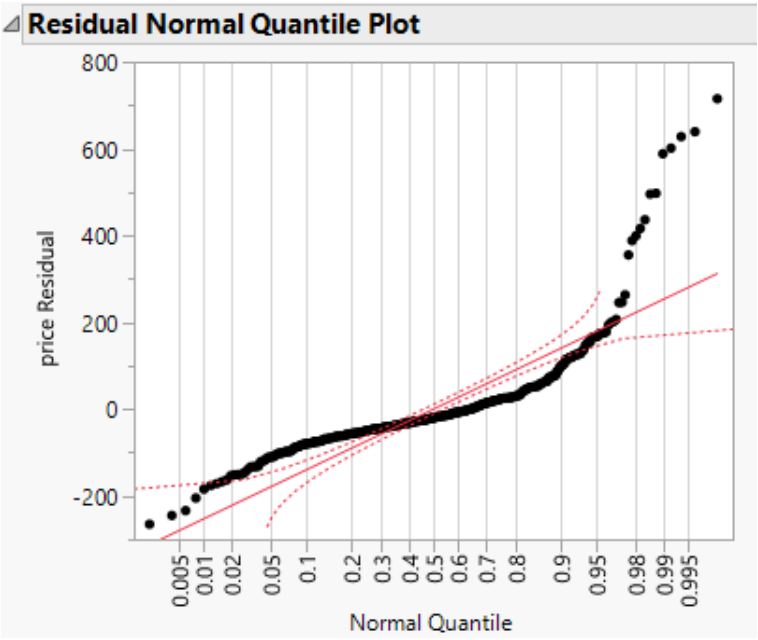

Figure 4: Based on the normal quantile plot, normality is not met because the data is not within the recommended boundaries on the normal quantile plot.

Summary of Fit

RSquare	0.267954
RSquare Adj	0.257539
Root Mean Square Error	108.4958
Mean of Response	135.416
Observations (or Sum Wgts)	500

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	2119893.6	302842	25.7270
Error	492	5791501.8	11771	Prob > F
C. Total	499	7911395.5		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	39.321291	18.42959	2.13	0.0334*
Apartment	-12.83288	14.67372	-0.87	0.3822
Condominium	18.523587	18.98226	0.98	0.3296
Other	24.821948	27.01857	0.92	0.3587
Townhouse	71.062938	38.64681	1.84	0.0666
accommodates	24.434905	1.977633	12.36	<.0001*
flexible	25.584839	12.83829	1.99	0.0468*
moderate	-1.77235	11.64699	-0.15	0.8791

Figure 5: Summary of Fit, Analysis of Variance and Parameter Estimates are pictured above. The full model includes 7 explanatory variables: Apartment, Condominium, Other, Townhouse, accommodates, flexible, and moderate. Where Apartment, Condominium, Other and Townhouse are indicator variables of the property_type categorical variable with a reference group of Houses. Flexible and moderate are indicator variables of the cancellation_policy categorical

variable with a reference group of strict. 25.75% of variability in price can explained by the linear model with the Airbnb property type, the number of accommodates, and the cancellation policy after adjusting for the complexity of the model. Accommodates has a test statistic of 12.36 and a p-value of <0.0001. There is overwhelming evidence to suggest that accommodates is a predictor of price. Flexible has a test statistic of 1.99 and a p-value of 0.0468. There is moderate evidence to suggest that flexible is a predictor of price.

Figure 6: Predicted Expression of the Full Model. y^ = 39.322 -12.833*Apartment + 18.524*Condominium + 24.822*Other + 71.063*Townhouse + 24.435*accommodates + 25.585*flexible -1.772*moderate. Where y^ is the predicted price.

Interaction Model

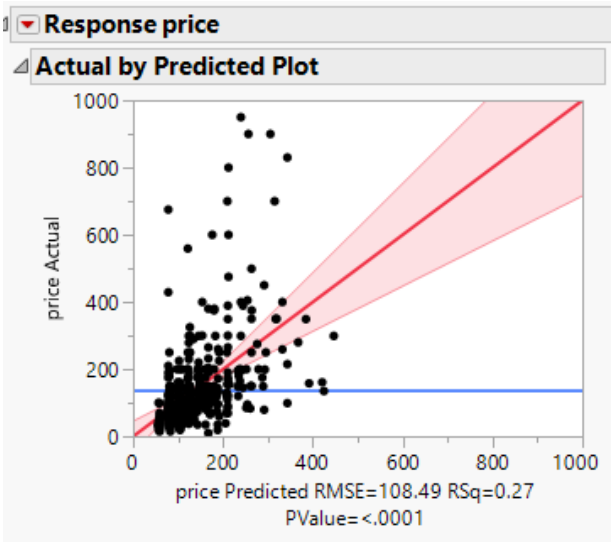

Figure 7: Actual by Predicted Plot visual assessment of model fit plotting the observed values of price (response variable) against the marginal predicted values of price. Based on the scatterplot, linearity is met because there is not a curve present.

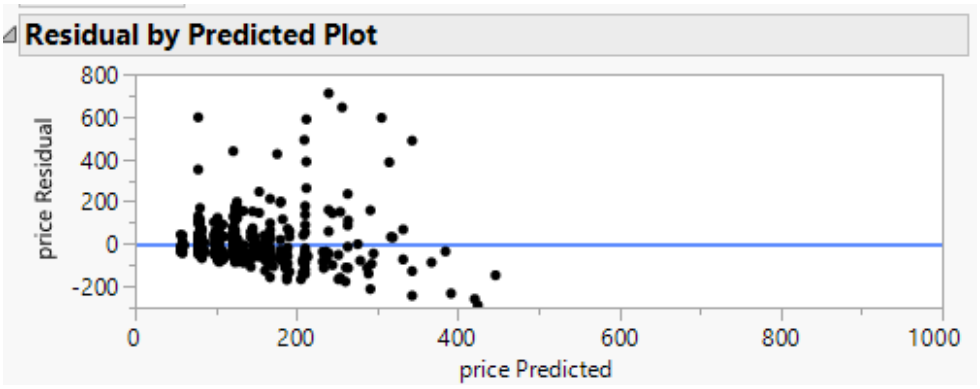

Figure 8: Residual by Predicted Plot indicating the difference between the observed price and fitted price values. The plot shows a fan and the means are not equally distributed above and below the mean of zero, indicating that equal variance is not met.

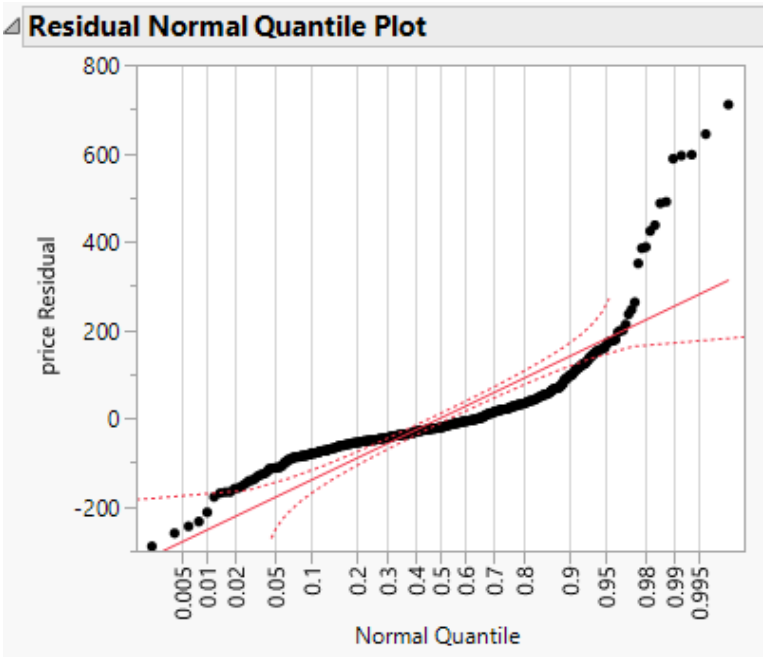

Figure 9: Based on the normal quantile plot, normality is not met because the data is not within the recommended boundaries on the normal quantile plot.

Summary of Fit

RSquare	0.273937
RSquare Adj	0.257571
Root Mean Square Error	108.4935
Mean of Response	135.416
Observations (or Sum Wgts)	500

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	2167221.5	197020	16.7380
Error	488	5744174.0	11771	Prob > F
C. Total	499	7911395.5		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	33.013552	23.71625	1.39	0.1645
Apartment	3.9739898	24.65798	0.16	0.8720
Condominium	15.560778	33.5818	0.46	0.6433
Other	-8.82858	44.57784	-0.20	0.8431
Townhouse	26.002961	70.77929	0.37	0.7135
accommodates	25.850193	3.652998	7.08	<.0001*
flexible	24.05302	12.9054	1.86	0.0630
moderate	-2.255463	11.68353	-0.19	0.8470
Apartment*accommodates	-4.129442	4.441006	-0.93	0.3529
Condominium*accommodates	1.0080468	6.395612	0.16	0.8748
Other*accommodates	10.876992	10.05572	1.08	0.2799
Townhouse*accommodates	13.157815	16.38304	0.80	0.4223

Figure 10: Summary of Fit, Analysis of Variance and Parameter Estimates are pictured above. The interaction model includes 11 explanatory variables: Apartment, Condominium, Other, Townhouse, accommodates, flexible, and moderate. The interaction variables are Apartment * accommodates, Condominium * accommodates, Other * accommodates, and Townhouse * accommodates. Where Apartment, Condominium, Other and Townhouse are indicator variables of the property_type categorical variable with a reference group of Houses. Flexible and moderate are indicator variables of the cancellation_policy categorical variable with a reference group of strict. 25.75% of variability in price can explained by the linear model with the Airbnb property type, the number of accommodates, and the cancellation policy after adjusting for the complexity of the model. Accommodates has a test statistic of 7.08 and a p-value of <0.0001. There is overwhelming evidence to suggest that accommodates is a predictor of price.

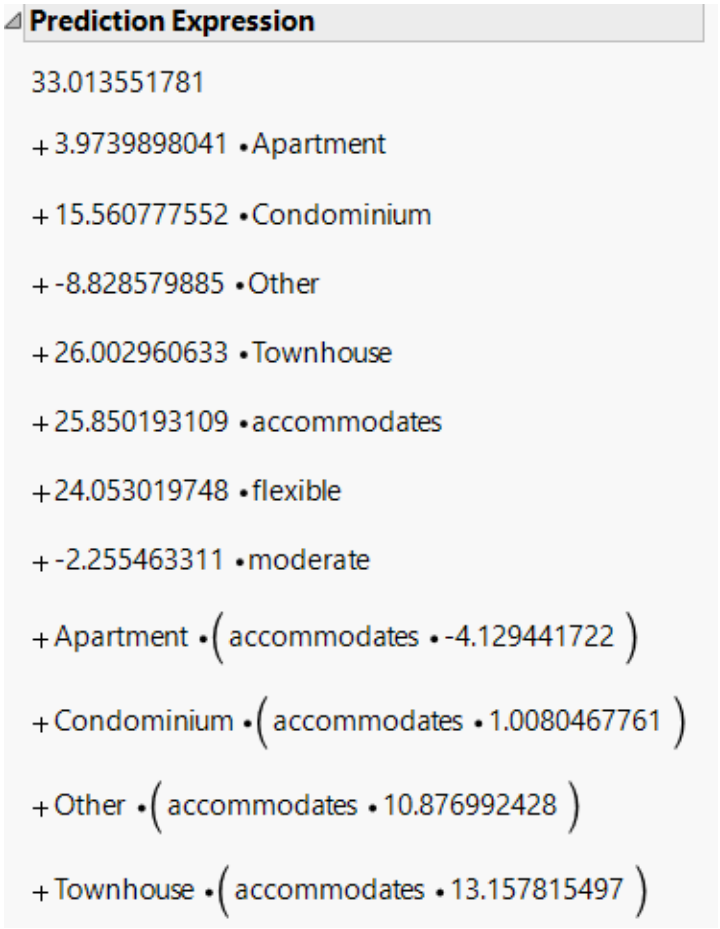

Figure 11: Predicted Expression of the Interaction Model. y^ = 33.014 +3.974*Apartment + 15.561*Condominium - 8.829*Other + 26.003*Townhouse + 25.850*accommodates + 24.053*flexible - 2.255*moderate + Apartment * (accommodates * -4.129) + Condominium * (accommodates * 1.008) + Other * (accommodates * 10.877) + Townhouse * (accommodates * 13.158). Where y^ is the predicted price.

Transformation Model

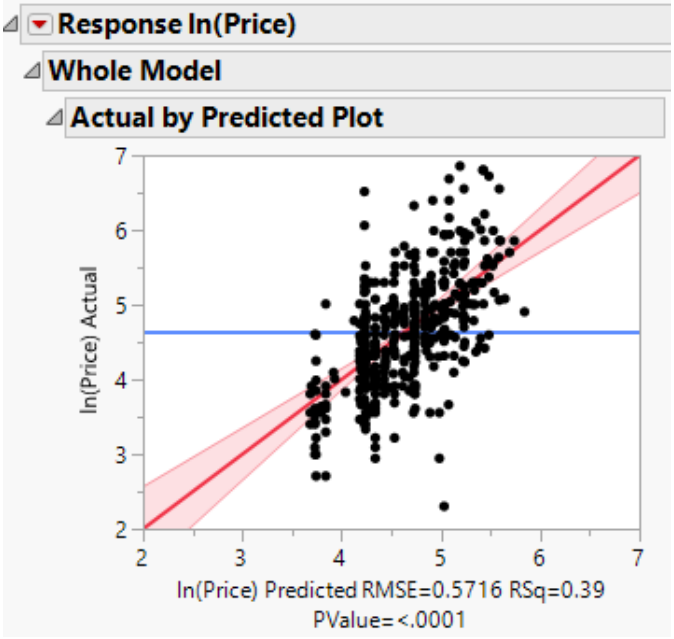

Figure 12: Actual by Predicted Plot visual assessment of model fit plotting the observed values of price (response variable) against the marginal predicted values of price. Based on the scatterplot, linearity is met because there is not a curve present.

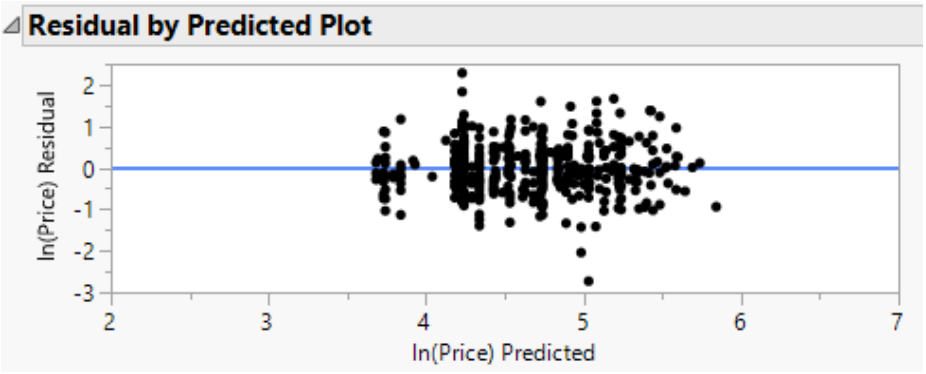

Figure 13: Residual by Predicted Plot indicating the difference between the observed price and fitted price values. Equal variance is met because the plot doesn't show a fan and the means are equally distributed above and below the mean of zero.

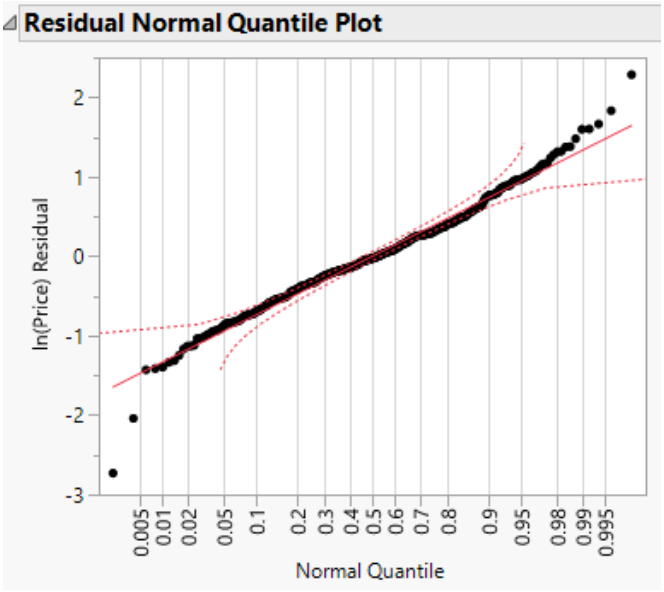

Figure 14: Based on the normal quantile plot, normality is met because the data is within the recommended boundaries on the normal quantile plot.

Summary of Fit

RSquare	0.388179
RSquare Adj	0.379474
Root Mean Square Error	0.571637
Mean of Response	4.628356
Observations (or Sum Wgts)	500

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	102.00310	14.5719	44.5939
Error	492	160.77016	0.3268	Prob > F
C. Total	499	262.77326		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.6947429	0.100427	36.79	<.0001*
Apartment	0.0472837	0.076597	0.62	0.5373
Condominium	0.2472841	0.099695	2.48	0.0135*
Other	0.235355	0.142362	1.65	0.0989
Townhouse	0.4319124	0.203383	2.12	0.0342*
flexible	0.0978347	0.067602	1.45	0.1485
moderate	-0.011432	0.061271	-0.19	0.8521
ln(Accommodates)	0.7200794	0.043114	16.70	<.0001*

Figure 15: Summary of Fit, Analysis of Variance and Parameter Estimates are pictured above. The transformation model includes 7 explanatory variables: Apartment, Condominium, Other, Townhouse, ln(accommodates), flexible, and moderate where the response variable has been transformed to be ln(price). Apartment, Condominium, Other and Townhouse are indicator variables of the property_type categorical variable with a reference group of Houses. Flexible and moderate are indicator variables of the cancellation_policy categorical variable with a

reference group of strict. 37.95% of variability in ln(price) can explained by the linear model with the Airbnb property type, the ln(accommodates), and the cancellation policy after adjusting for the complexity of the model. Condominium has a test statistic of 2.48 and a p-value of 0.0135. There is strong evidence to suggest that condominium is a predictor of ln(price). Townhouse has a test statistic of 2.12 and a p-value of 0.0342. There is moderate evidence to suggest that townhouse is a predictor of ln(price). ln(accommodates) has a test statistic of 16.70 and a p-value of <0.0001. There is overwhelming evidence to suggest that ln(accommodates) is a predictor of ln(price).

Prediction Expression
3.6947429006
+ 0.0472836546 • Apartment
+ 0.2472841186 • Condominium
+ 0.2353550284 • Other
+ 0.4319124037 • Townhouse
+ 0.0978346746 • flexible
+ -0.011431748 • moderate
+ 0.7200794257 • ln(Accommodates)

Figure 16: Predicted Expression of the Transformation Model. y^ = 3.695 + 0.047*Apartment + 0.247*Condominium + 0.235*Other + 0.432*Townhouse + 0.098*flexible - 0.011*moderate + 0.720*ln(accommodates). Where y^ is the predicted ln(price). When accommodates increases by 10%, we predict the median price to change by a factor of 1.07 which comes from (1.1)^0.720.

Quadratic Model

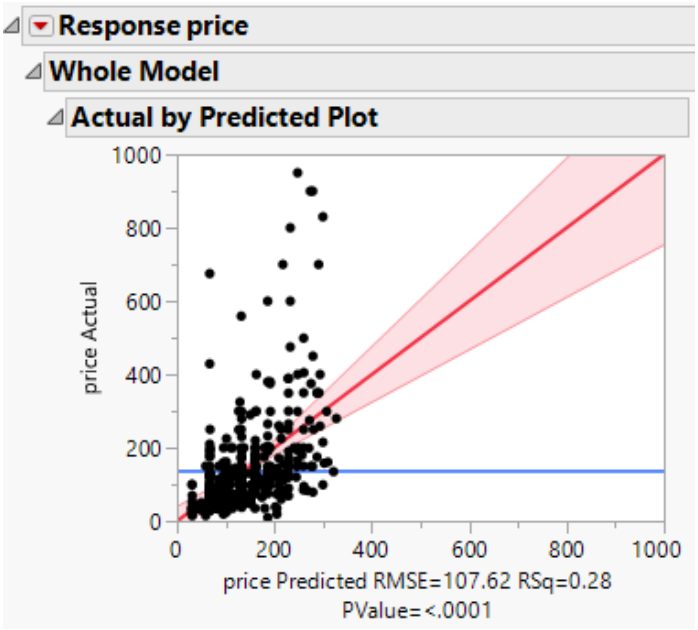

Figure 17: Actual by Predicted Plot visual assessment of model fit plotting the observed values of price (response variable) against the marginal predicted values of price. Based on the scatterplot, linearity might be violated because there seems to be a slight curve present.

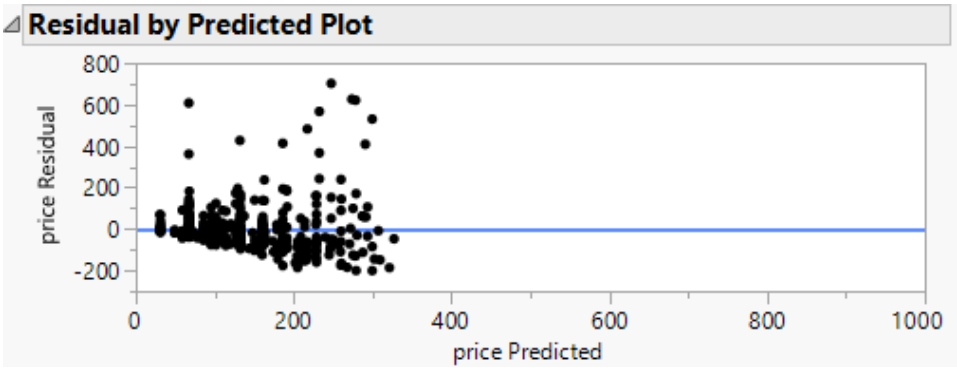

Figure 18: Residual by Predicted Plot indicating the difference between the observed price and fitted price values. Equal variance is not met because the plot shows a fan and the means are not equally distributed above and below the mean of zero.

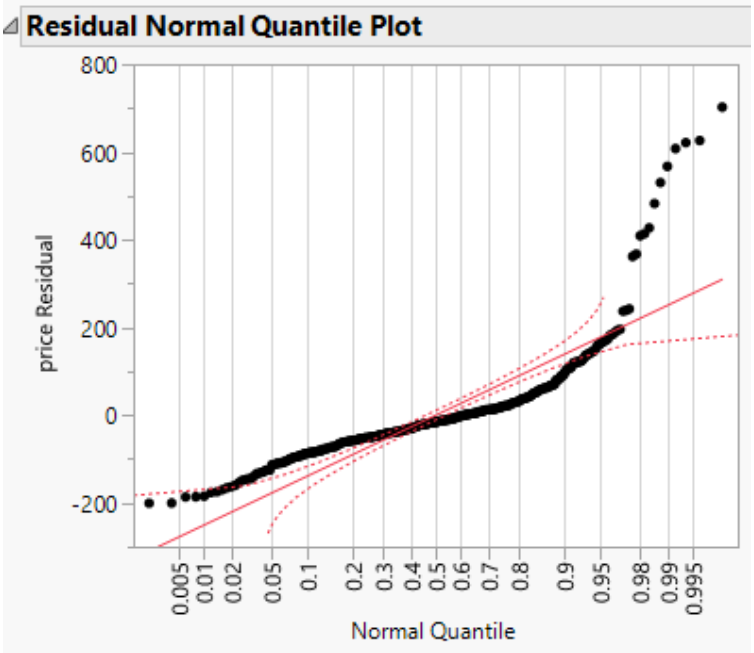

Figure 19: Based on the normal quantile plot, normality is not met because the data is not within the recommended boundaries on the normal quantile plot.

Summary of Fit

RSquare	0.281247
RSquare Adj	0.269536
Root Mean Square Error	107.6157
Mean of Response	135.416
Observations (or Sum Wgts)	500

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	2225054.9	278132	24.0159
Error	491	5686340.5	11581	Prob > F
C. Total	499	7911395.5		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	10.507742	20.62987	0.51	0.6107
accommodates	40.705434	5.744726	7.09	<.0001*
accommodates*accommodates	-1.385018	0.459625	-3.01	0.0027*
Apartment	-18.65371	14.68231	-1.27	0.2045
Condominium	12.380187	18.93833	0.65	0.5136
Other	23.210049	26.80473	0.87	0.3870
Townhouse	65.960649	38.37069	1.72	0.0862
flexible	27.641604	12.75242	2.17	0.0307*
moderate	-0.73486	11.55764	-0.06	0.9493

Figure 20: Summary of Fit, Analysis of Variance and Parameter Estimates are pictured above. The quadratic model includes 8 explanatory variables: Apartment, Condominium, Other, townhouse, accommodates, accommodates*accommodates, flexible, and moderate. Apartment,

Condominium, Other and Townhouse are indicator variables of the property_type categorical variable with a reference group of Houses. Flexible and moderate are indicator variables of the cancellation_policy categorical variable with a reference group of strict. 26.95% of variability in price can explained by the quadratic model with the Airbnb property type, the number of accommodates, and the cancellation policy after adjusting for the complexity of the model. Accommodates has a test statistic of 7.09 and a p-value of <0.0001. There is overwhelming evidence to suggest that accommodates is a predictor of price. Accommodates*Accommodates has a test statistic of -3.01 and a p-value of 0.0027. There is strong evidence to suggest that accommodates*accommodates is a predictor of price. Flexible has a test statistic of 2.17 and a p-value of 0.0307. There is moderate evidence to suggest that flexible is a predictor of price.

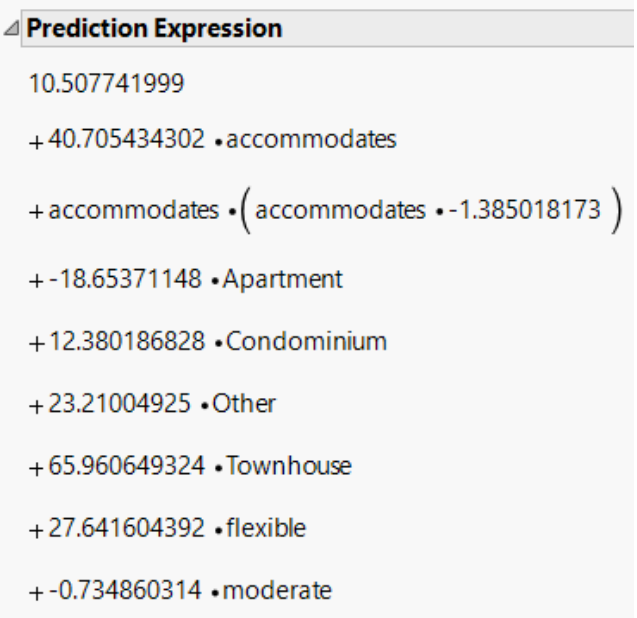

Figure 21: Predicted Expression of the Quadratic Model. y^ = 10.508 + 40.705*accommodates + accommodates*(accommodates * -1.385) - 18.654 * Apartment + 12.380*Condominium + 23.210*Other + 65.961*Townhouse + 27.642*flexible - 0.735*moderate. Where y^ is the predicted price.

Transformation and Interaction Model

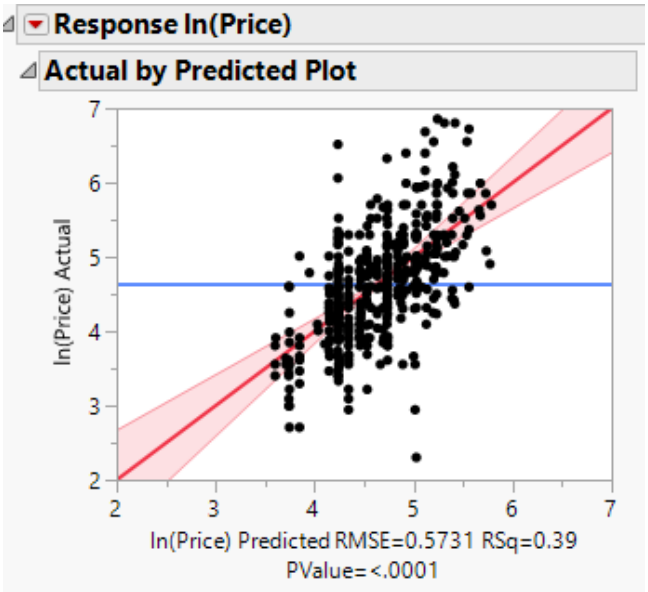

Figure 22: Actual by Predicted Plot visual assessment of model fit plotting the observed values of price (response variable) against the marginal predicted values of price. Based on the scatterplot, linearity is met because there is not a curve present.

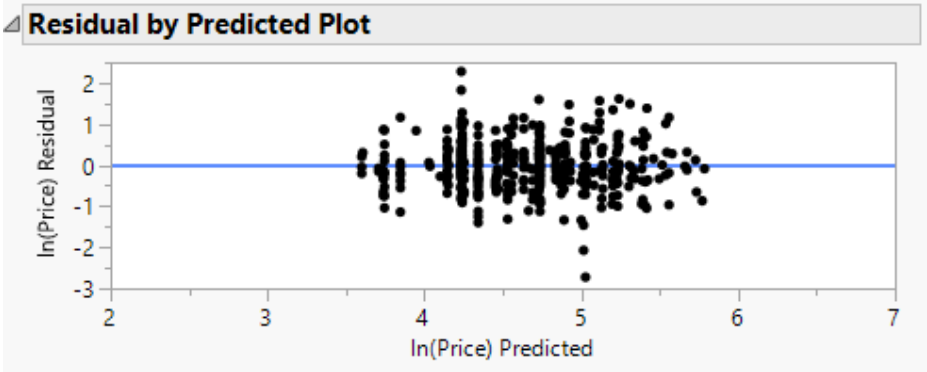

Figure 23: Residual by Predicted Plot indicating the difference between the observed price and fitted price values. Equal variance is met because the plot doesn't show a fan and the means are equally distributed above and below the mean of zero.

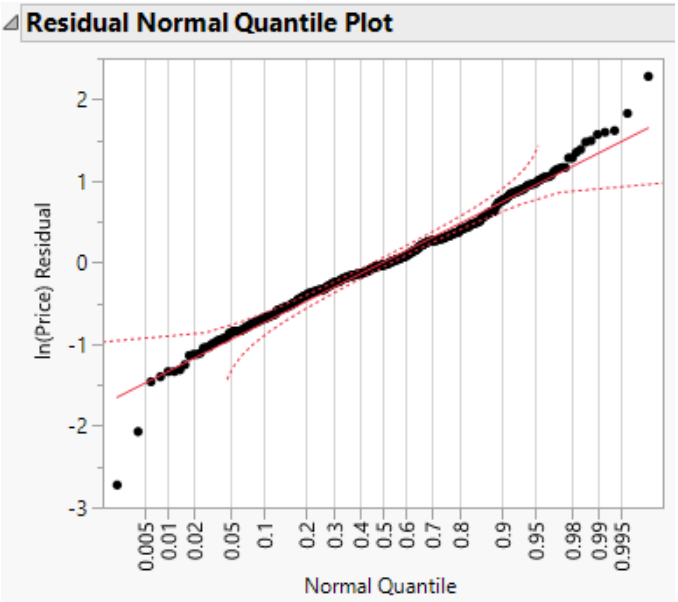

Figure 24: Based on the normal quantile plot, normality is met because the data is within the recommended boundaries on the normal quantile plot.

Summary of Fit

RSquare	0.390042
RSquare Adj	0.376293
Root Mean Square Error	0.573101
Mean of Response	4.628356
Observations (or Sum Wgts)	500

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	102.49250	9.31750	28.3686
Error	488	160.28076	0.32844	Prob > F
C. Total	499	262.77326		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.6067863	0.146358	24.64	<.0001*
flexible	0.1015061	0.068105	1.49	0.1368
moderate	-0.006308	0.061587	-0.10	0.9185
ln(Accommodates)	0.7856418	0.092214	8.52	<.0001*
Apartment	0.1386542	0.153705	0.90	0.3675
Condominium	0.3927513	0.217823	1.80	0.0720
Other	0.4286054	0.240566	1.78	0.0754
Townhouse	0.3405758	0.392036	0.87	0.3854
Apartment*ln(Accommodates)	-0.070687	0.105928	-0.67	0.5049
Condominium*ln(Accommodates)	-0.11448	0.154268	-0.74	0.4584
Other*ln(Accommodates)	-0.182642	0.189995	-0.96	0.3369
Townhouse*ln(Accommodates)	0.1014794	0.309147	0.33	0.7429

Figure 25: Summary of Fit, Analysis of Variance and Parameter Estimates are pictured above. The transformation and interaction model includes 11 explanatory variables: Apartment,

Condominium, Other, Townhouse, ln(accommodates), flexible, and moderate where the response variable has been transformed to be ln(price). The interaction variables are Apartment * ln(accommodates), Condominium * ln(accommodates), Other * ln(accommodates), and Townhouse * ln(accommodates). Where Apartment, Condominium, Other and Townhouse are indicator variables of the property_type categorical variable with a reference group of Houses. Flexible and moderate are indicator variables of the cancellation_policy categorical variable with a reference group of strict. 37.63% of variability in ln(price) can explained by the linear model with the Airbnb property type, ln(accommodates), and the cancellation policy after adjusting for the complexity of the model. ln(Accommodates) has a test statistic of 8.52 and a p-value of <0.0001. There is overwhelming evidence to suggest that ln(accommodates) is a predictor of ln(price).

Figure 26: Predicted Expression of the Transformation and Interaction Model. y^ = 3.607 + 0.102*flexible - 0.006*moderate + 0.786*ln(Accommodates) + 0.139*Apartment + 0.393*Condominium + 0.429*Other + 0.341*Townhouse + Apartment * (ln(Accommodates) * -0.071) + Condominium * (ln(Accommodates) * -0.114) + Other * (ln(Accommodates) * -0.183) + Townhouse * (ln(Accommodates) * 0.101). Where y^ is the predicted ln(price). When accommodates increases by 10%, we predict the median price to change by a factor of 1.078 which comes from (1.1)^0.786.

Summary of Models

	Full Model	Interaction Model	Transformation Model	Quadratic Model	Transformation +Interaction Model
Number of Explanatory Variables	7	11	7	8	11
Adjusted R^2	0.257539	0.257571	0.379474	0.269536	0.376293
Mallow's Cp	8	12	8	9	12
RMSE	108.49583	108.49354	0.5716368	107.61571	0.5731005
F-test p-value	<0.0001 F-test: 25.727	<0.0001 F-test: 16.738	<0.0001 F-test: 44.594	<0.0001 F-test: 24.016	<0.0001 F-test: 28.369
Largest p-value for all individual coefficient t-tests: $\beta_i = 0$ vs $\beta_i \neq 0$	0.8791 (moderate)	0.8748 (condominium * accommodates)	0.8521 (moderate)	0.9493 (moderate)	0.9185 (moderate)

Table 1: Summary of all the 5 models constructed. The best model is the Transformation Model because it has the highest Adjusted R^2, the lowest RMSE, lowest Mallow's Cp, lowest F-test p-value and has the lowest number of variables.

Outliers, Influential Points, and High Leverage Values

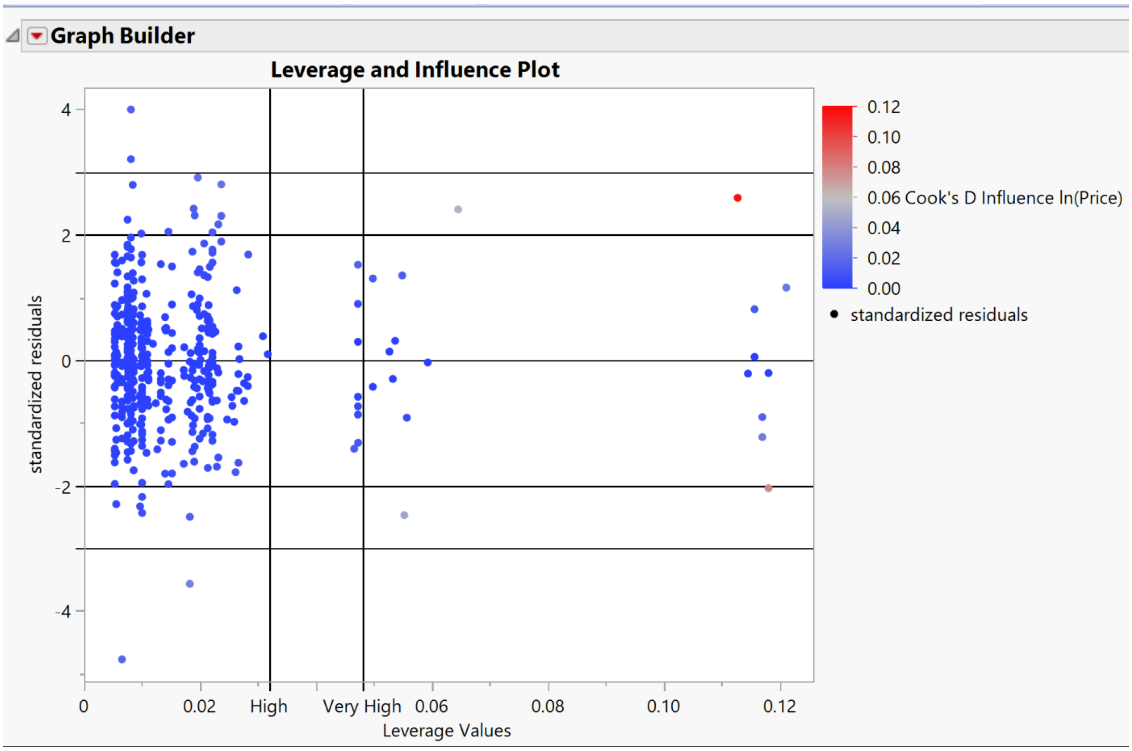

Figure 27: Leverage and Influence Plot for the Transformation Model. High Leverage was calculated from: $h > 2((1+7/500) = h > 2(8/500) = h > 0.032$. Very High Leverage was calculated from: $h > 3((1+7)/500) = h > 3(8/500) = h > 0.048$. There are approximately twenty outliers that are high leverage. There are approximately ten outliers that are above a standardized residual of 2 and approximately five outliers below a standardized residual of -2. There are no outliers with a Cook's D value above 0.5 or 1, so there is not any high influence.

Partial F-tests

Custom Test

Parameter				
Intercept	0	0	0	0
Apartment	1	0	0	0
Condominium	0	1	0	0
Other	0	0	1	0
Townhouse	0	0	0	1
flexible	0	0	0	0
moderate	0	0	0	0
ln(Accommodates)	0	0	0	0
=	0	0	0	0
Value	0.0472836546	0.2472841186	0.2353550284	0.4319124037
Std Error	0.0765972582	0.099694898	0.1423622248	0.2033830822
t Ratio	0.6173021821	2.4804089619	1.653212632	2.123639779
Prob>|t|	0.5373210066	0.0134565662	0.0989254931	0.0341986394
SS	0.1245190982	2.0104206021	0.8930952355	1.4736761195

Sum of Squares	4.2171745315
Numerator DF	4
F Ratio	3.2264225269
Prob > F	0.0124822497

Figure 28: Partial F-test for the explanatory variable property_type for the Transformation Model. The conducted test has a Ho: $\beta 2 = \beta 3 = \beta 4 = \beta 5 = 0$ where $\beta 2$= slope for apartment property type, $\beta 3$= slope for condominium property type, $\beta 4$ = slope for other property type, $\beta 5$ = slope for townhouse property type. Ha: At least one $\beta i \neq 0$. The test statistic is 3.226 with a p-value of 0.012. It can be concluded that there is strong evidence to suggest that the Property Type is a predictor of ln(price).

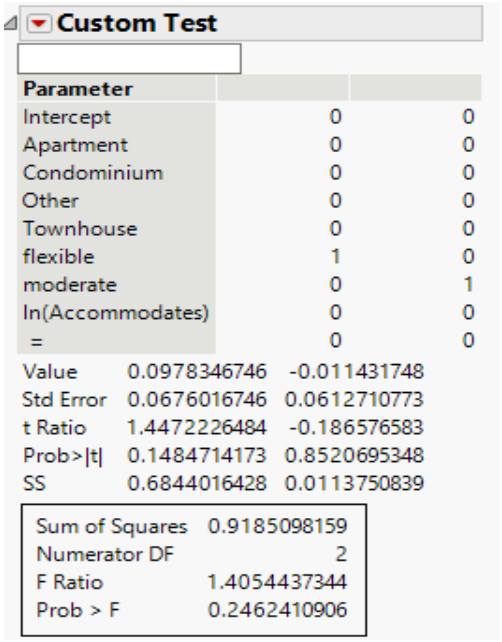

Figure 29: Partial F-test for the explanatory variable cancellation_policy for the Transformation Model. The conducted test has a Ho: $\beta 6 = \beta 7 = 0$ where $\beta 6$= slope for flexible cancellation policy, and $\beta 7$= slope for moderate cancellation policy. Ha: At least one $\beta i \neq 0$. The test statistic is 1.405 with a p-value of 0.246. It can be concluded that there is little to no evidence to suggest that the Cancellation Policy is a predictor of ln(price).

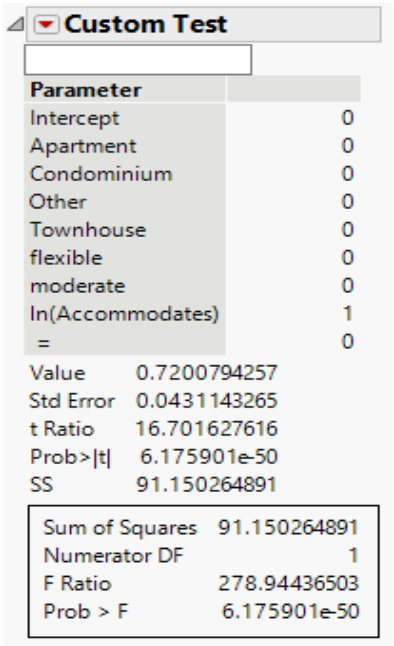

Figure 30: Partial F-test for the explanatory variable ln(accommodates) for the Transformation Model. The conducted test has a Ho: $\beta 8 = 0$ where $\beta 8$= slope for ln(Accommodates). Ha: $\beta 8 \neq 0$. The test statistic is 278.944 with a p-value of 6.18e-50. It can be concluded that there is overwhelming evidence to suggest that ln(accommodates) is a predictor of ln(price).