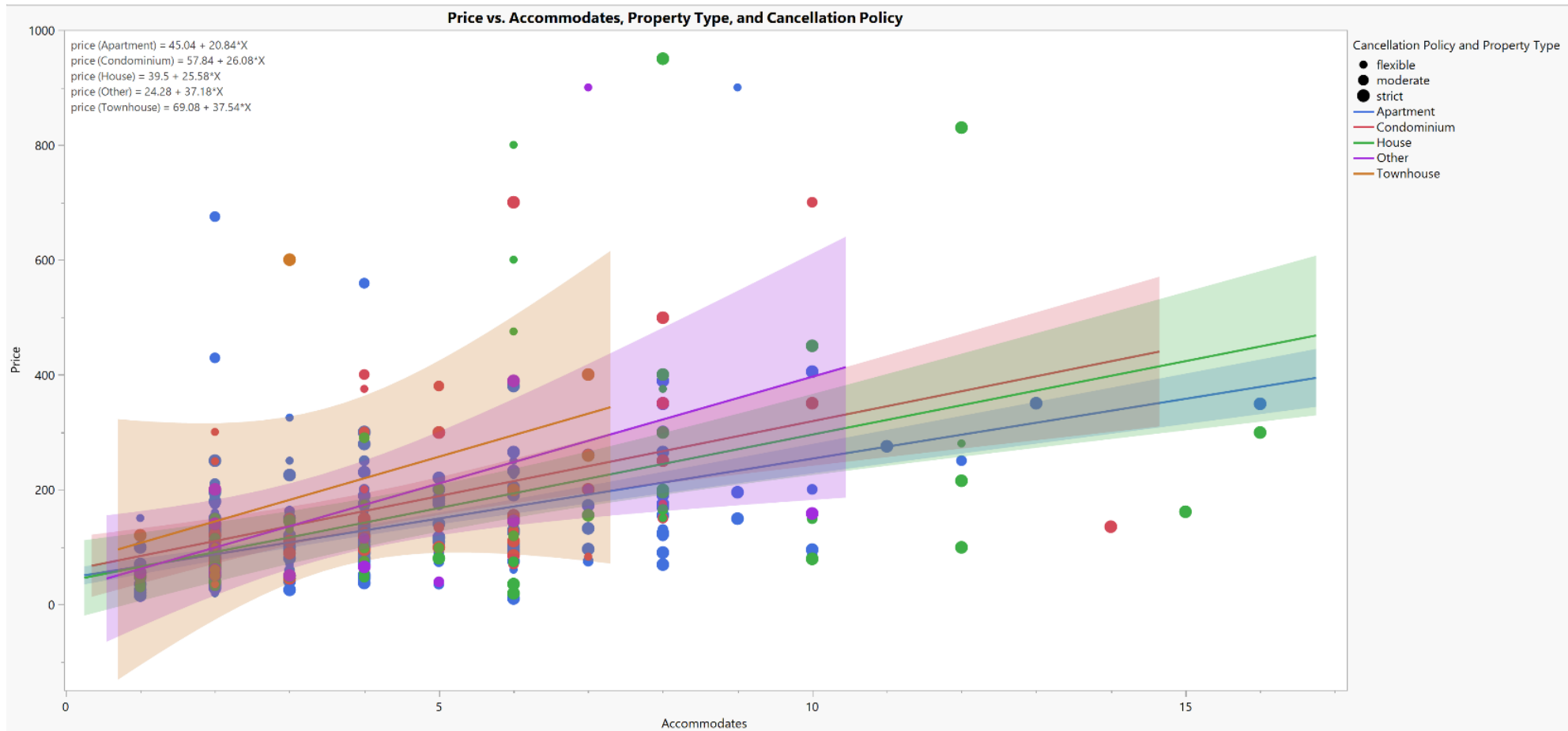# Airbnb Predicted Pricing

Airbnb Dataset

Syrena Hilgendorf, Neha Maddali, Lorpu Kokoi, Marie Klapacz

Group 2-B

Project Background and Data Description

- Dataset: Airbnb listings subset
- Random sample from Airbnb hosts
- Listings in Chicago, IL from Aug 2008-May 2017
- Response variable: price
- Quantitative explanatory variable: accommodates
- Qualitative explanatory variables: cancellation_policy (strict as reference group), property_type (house as reference group)

Scatterplot

Price vs. Accommodates, Property Type, and Cancellation Policy

price (Apartment) = 45.04 + 20.84*X
price (Condominium) = 57.84 + 26.08*X
price (House) = 39.5 + 25.58*X
price (Other) = 24.28 + 37.18*X
price (Townhouse) = 69.08 + 37.54*X

Cancellation Policy and Property Type
● flexible
● moderate
● strict
— Apartment
— Condominium
— House
— Other
— Townhouse

Price

Accommodates

Results from Multiple Regression Analysis

	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: 25.7270	<0.0001 F-test: 16.7380	<0.0001 F-test: 44.5939	<0.0001 F-test: 24.0159	<0.0001 F-test: 28.3686
Largest p-value for all individual coefficient t-tests: $\beta_i = 0$ vs $\beta_i \neq 0$	0.8791 (moderate)	0.8748 (condo*accommodates)	0.8521 (moderate)	0.9493 (moderate)	0.9185 (moderate)

Models

Full: property_ type, cancellation_policy, accommodates, (response price)

Interaction: property_type*accommodates + Full

Transformation: property_type, cancellation_policy, ln(accommodates), (response ln(price))

Quadratic: accommodates*accommodates + Full

Transformation + Interaction: property_type*ln(accommodates), property_type, cancellation_policy, ln(accommodates), (response ln(price))

Chosen Model

- Best model: transformation model
- Highest adjusted R^2 of 0.38
- Lowest RMSE of 107.62
- Lowest Mallow's Cp of 8
- Lowest F-test value of 44.59, with very small p-value of <0.0001
- Fewest variables

Chosen Model Results

- predicted ln(price) = 3.695 + 0.047*Apartment + 0.720*ln(Accommodates) - 0.011*moderate + 0.098*flexible + 0.432*Townhouse + 0.235*Other + 0.247*Condominium
- Sample size: 500
- 37.95% of variability in ln(price) can be explained by a linear model with property_type, ln(accommodates), and cancellation_policy after adjusting for the complexity of the model

Graphs for Checking Conditions
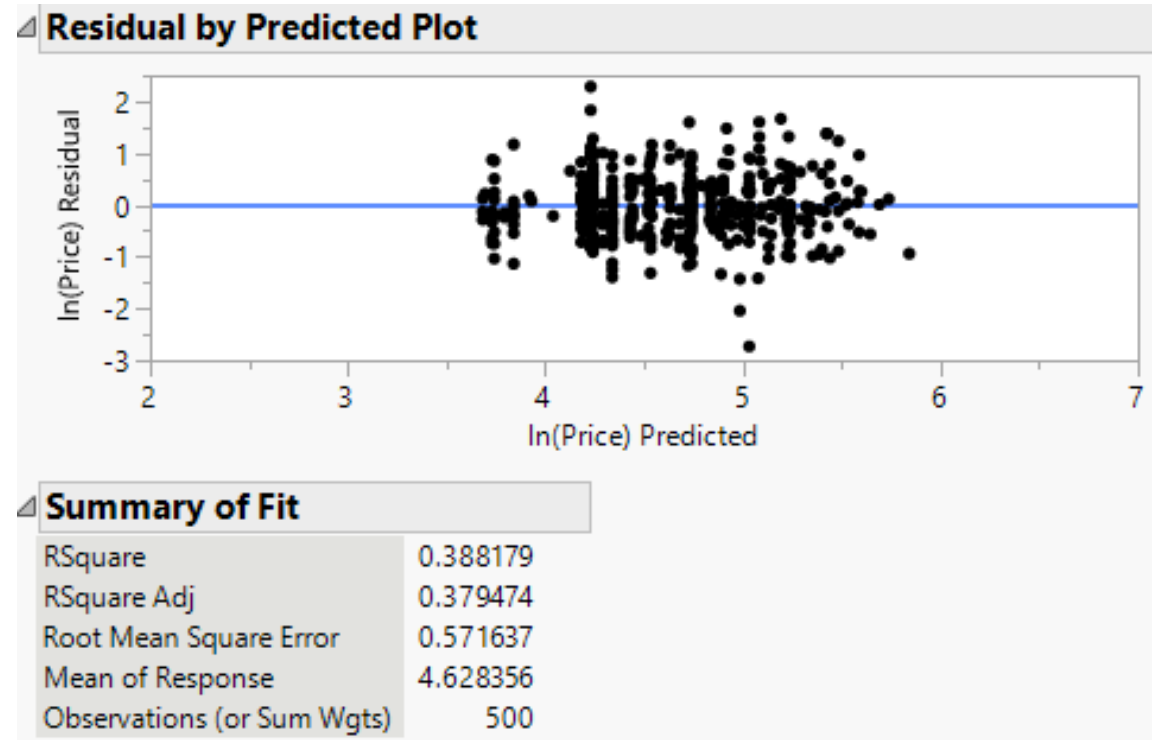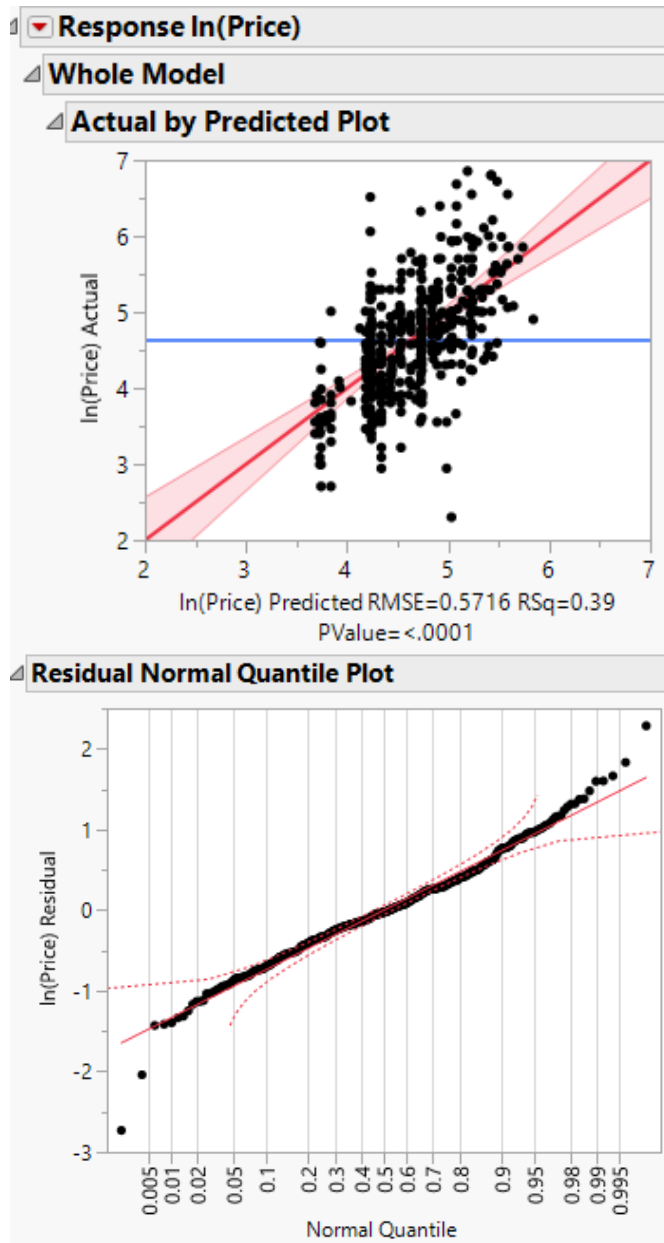
Checking Conditions

- Independence met: random sample
- Linearity met: no curve in scatterplot
- Normality met: data within recommended boundaries of normal quantile plot
- Equal Variance met: no fan in residual by predicted plot, means equally distributed above and below mean of zero

Influence and Leverage Plot

- ≈ 20 outliers indicating high leverage
- ≈ 10 outliers with standardized residual > 2
- ≈ 5 outliers with standardized residual < -2
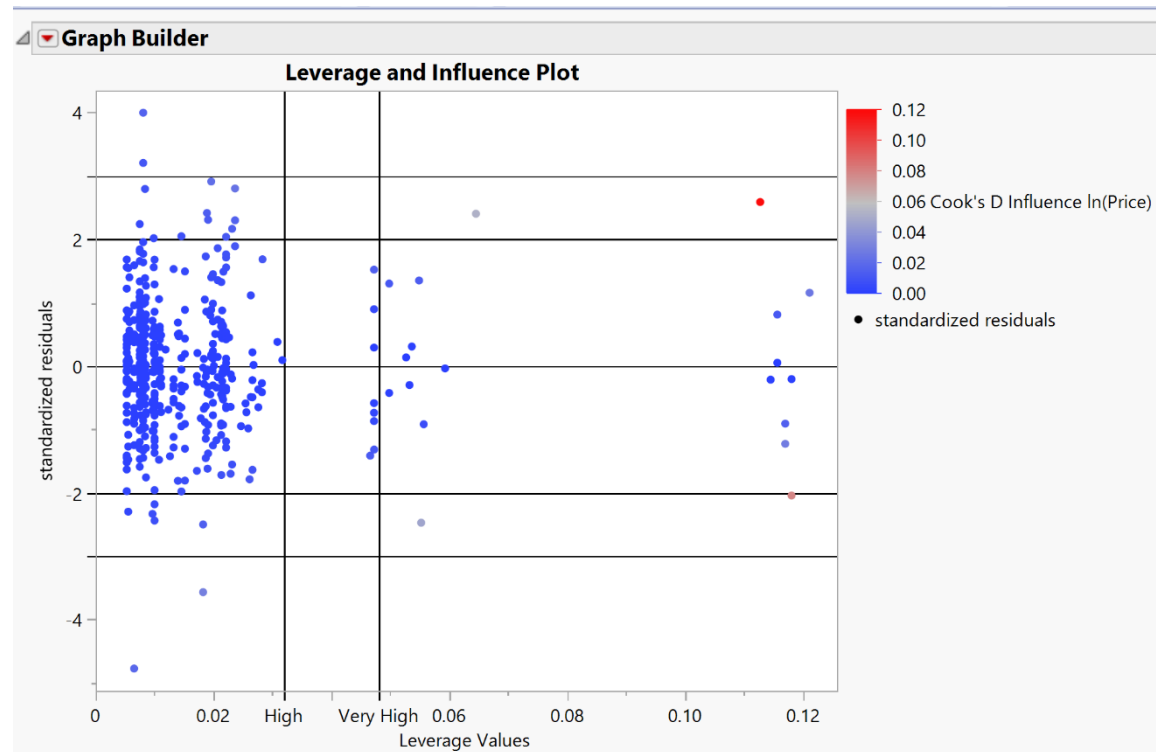
Hypothesis Tests

H0: $\beta2= \beta3= \beta4= \beta5=0$ where $\beta2=$ apartment slope, $\beta3=$ condominium slope, $\beta4 =$ other slope, $\beta5 =$ townhouse slope.

Ha: At least one $\beta i \neq 0$

Test statistic = 3.2264

p-value = 0.01248

Conclusion: There is strong evidence to suggest that the Property Type is a predictor of ln(Price).

▲ ▼ **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
In(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	0.034198694
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

H0: β6= β7=0 where β6= flexible cancellation policy slope, β7= moderate cancellation policy slope

Ha: At least one βi ≠ 0

Test statistic = 1.405

p-value = 0.246

Conclusion: There is little to no evidence to suggest that the Cancellation Policy is a predictor of ln(Price).

◢ ▼ Custom Test

Parameter		
Intercept	0	0
Apartment	0	0
Condominium	0	0
Other	0	0
Townhouse	0	0
flexible	1	0
moderate	0	1
ln(Accommodates)	0	0
=	0	0
Value	0.0978346746	-0.011431748
Std Error	0.0676016746	0.0612710773
t Ratio	1.4472226484	-0.186576583
Prob>|t|	0.1484714173	0.8520695348
SS	0.6844016428	0.0113750839

Sum of Squares	0.9185098159
Numerator DF	2
F Ratio	1.4054437344
Prob > F	0.2462410906

H0: β8=0 where β8= slope for ln(Accommodates)

Ha: β8 ≠ 0

Test statistic = 278.94

p-value = 6.18e-50

Conclusion: There is overwhelming evidence to suggest that the ln(Accommodates) is a predictor of ln(Price).

Custom Test

Parameter	
Intercept	0
Apartment	0
Condominium	0
Other	0
Townhouse	0
flexible	0
moderate	0
ln(Accommodates)	1
=	0

Value	0.7200794257
Std Error	0.0431143265
t Ratio	16.701627616
Prob>|t|	6.175901e-50
SS	91.150264891

Sum of Squares	91.150264891
Numerator DF	1
F Ratio	278.94436503
Prob > F	6.175901e-50

Overall Conclusions

The number of accommodates that an Airbnb can support, as well as the type of property of the Airbnb influence the price of the Airbnb.

Cancellation price does not influence the price of the Airbnb.