Group 3: James Min, Anh Nguyen, Cam Nguyen, Andy Nguyen, Eddie Pham

City: Austin, Texas as an Airbnb Host

Part 1:

Who are you? Host: As a host, we want to create the best experience for our guests while still being able to offer optimal and competitive pricing. The ultimate goal of this research is to find key variables that are important and significant to price and see how we can predict the best average price a host should charge for their place. Therefore, we take this project on with the price in mind as the dependent variable and hopefully we are able to find independent variables that have a significant impact on price and how the hosts can utilize the information to make smarter business decisions.

Phenomenon of interest: As mentioned earlier, our phenomenon of interest is price charged per night at an available listing. We believe price is most important because to be competitive in a dynamic environment, you as the host would want to price competitively against other listings

Part 2:

2A Data Analyses:

- **Accommodates** would be significant in predicting price because accommodates represents the number of people that are able to stay at an Airbnb. If more people were able to stay at Airbnb it would make sense that the rental price would be more expensive. Airbnb likely has more rooms or space and more people staying at an Airbnb gives a host more leverage to raise the rental price.
- **Hosting Listing Count:** I believe we can use the host listing count variable to predict the price of rent. If a host has a high listing count, it may indicate that they are a successful and experienced host who has a good reputation on the Airbnb platform. This can result in higher demand for their properties and potentially higher rental rates. Additionally, a host with multiple listings may be able to offer discounts or bundles of deals, which can also impact the rental price.

- **Review Score Ratings,** I believe review score ratings would have a significant impact in predicting the price charged per night of stay. That is because I believe that having a higher review score rating would convince guests to pay a little premium for a peace of mind. While review score ratings are not everything, it is the first thing we look at when looking up a restaurant. Similarly, we believe that by having a high review score rating, we are able to charge a little higher price than our competitors who may not have a high review score rating. If we are able to predict that review score ratings do have an impact on price, we are able to focus on some of its attributes to help boost our rate of charge.
- **Number of reviews,** Low review counts, or unfavorable comments may have the opposite impact, diminishing customer trust in the item or service and possibly resulting in reduced costs. So, the link between the quantity of reviews and the price paid will rely on a range of circumstances, thus it is crucial to take numerous elements into account when setting prices.
- **Room Types:** Room Type is one of the factors that can affect the price. There are so many types of rooms such as Entire Homes, Apartments, Private Rooms, or Hotel Rooms. The price is high or low depending on the type of room. For example, the Entire Home or Apartment might be more expensive because their acreage is bigger. Also, the bigger house is more convenient for a big family, it might save a lot of money for them instead of renting many hotel rooms.
- **Host Identity Verified:** Because it is essential to foster trust between hosts and guests, host identification verification on Airbnb can have a major influence on price. Possible guests may be certain that a host is a true, reputable person who has been approved by the site if they see the Airbnb logo on their listing. As a result of increased demand from customers ready to pay more for a listing that they believe to be safer and more secure, this may have a favorable impact on the listing price. Additionally, a certified host is more likely to get more favorable ratings from visitors, which can also help to increase demand and raise pricing. However, if a host is not verified, it can result in less interest in the listing since customers might be less willing to take the

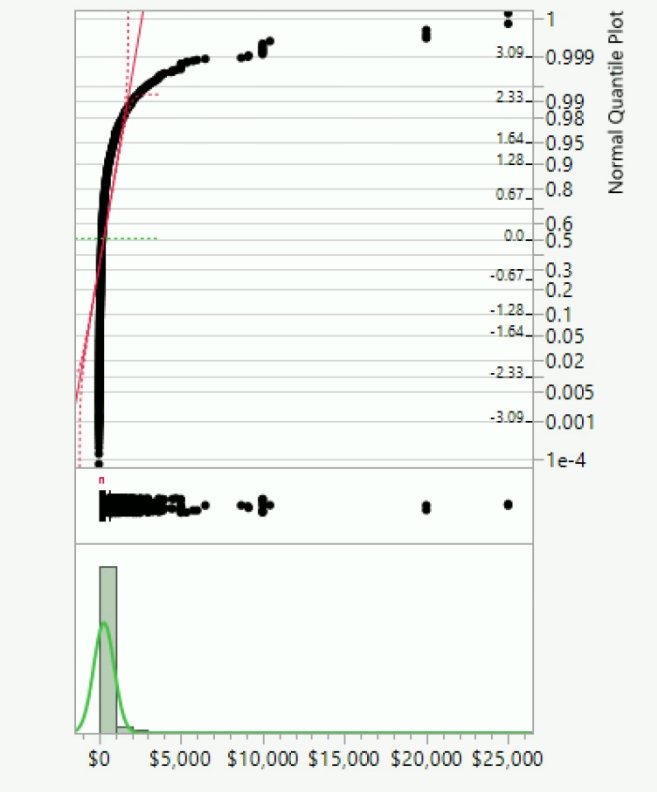
chance of making a reservation with an unverified host. As a result, the listing's price may be reduced to entice reservations.

- **Number of Minimum Nights:** We believe this variable might affect the price. In seasonal markets, the most apparent theme is the uptick in the number of three-night listings during the summer months. From April to August, over 1/3 of the market requires three-night minimum stays. One- and two-night stays run neck and neck until the end-of-summer push when one-nighters experience a significant boost. Additionally, there are over twice as many listings with 7–27-night minimums as those with 4–6-night minimums.

Part 2B: Summary of Variables

2B: Univariate with Price:

From the summary statistics shown below, there are 14368 Airbnb listings listed in the Austin dataset. From using this data, we can see that the average rate an Airbnb host charges for a night in Austin Texas is approximately $307.24. The maximum rate a host charges is $25,000 and the minimum one charges is $0. We believe that the maximum is an outlier since the average is well below $500. Additionally, we can assume that approximately 50% of the hosts/listings charge more than $180 per night, but also 50% of the hosts/listings charge a lower rate than $180. Additionally, we can see that approximately 75% of the listings are under $320 per night while 25% of the listings are under $108 per night.



Quantiles

100.0%	maximum	$25,000
99.5%		$2,900
97.5%		$1,290
90.0%		$600
75.0%	quartile	$320
50.0%	median	$180
25.0%	quartile	$108
10.0%		$68
2.5%		$40
0.5%		$24
0.0%	minimum	$0

Summary Statistics

Mean	307.23518
Std Dev	622.20457
Std Err Mean	5.1908089
Upper 95% Mean	317.40983
Lower 95% Mean	297.06052
N	14368

2B: Bivariate with Y and Xs

- **Price and Review Score Ratings:** By looking at the p-value, we are able to see that the review scores rating does have a significant impact in price charged per night. Since the p-value is .0001, which is less than .05, we can assume that these two variables do have an impact on each other.

This aligns with our prediction as we predicted that review scores rating do have a significant impact on price. And by looking at the parameter estimates, we can assume that for every 1 unit increase in review scores rating, the price charged per night can increase by $35.28. On the other hand, for every 1 unit decrease in review scores rating, the price charged per night may decrease by $35.28.

Parameter Estimates

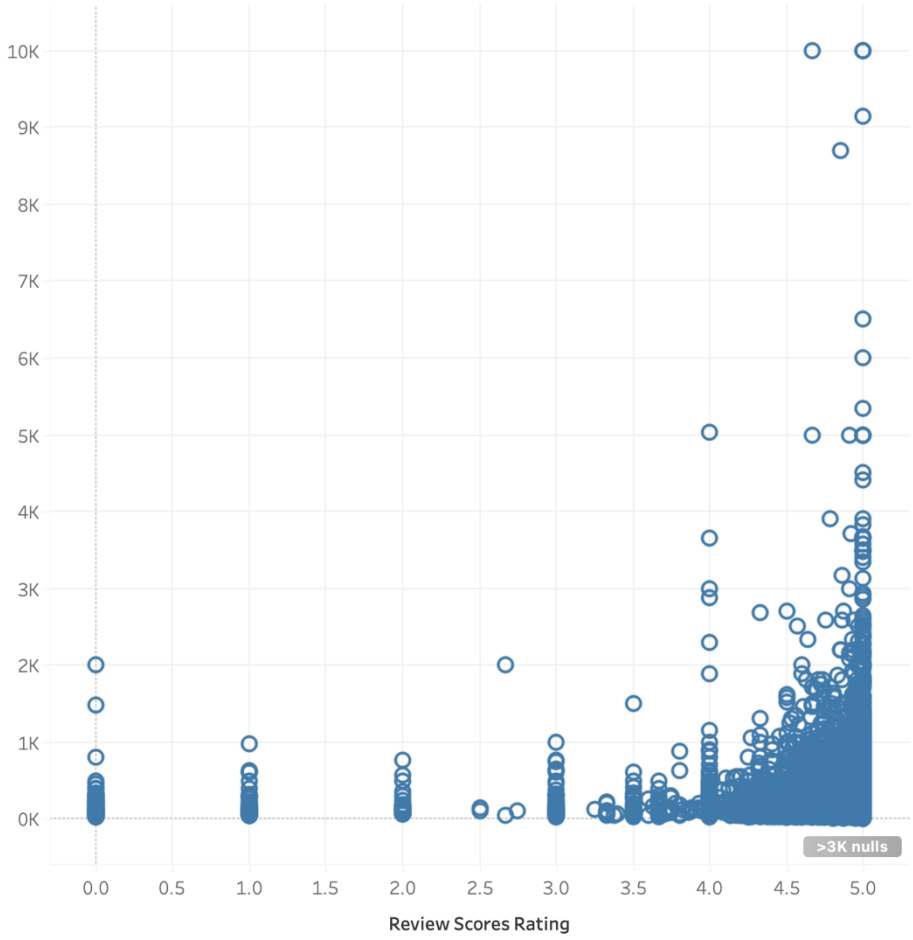
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	107.69737	34.85236	3.09	0.0020*
review_scores_rating	35.283813	7.238005	4.87	<.0001*

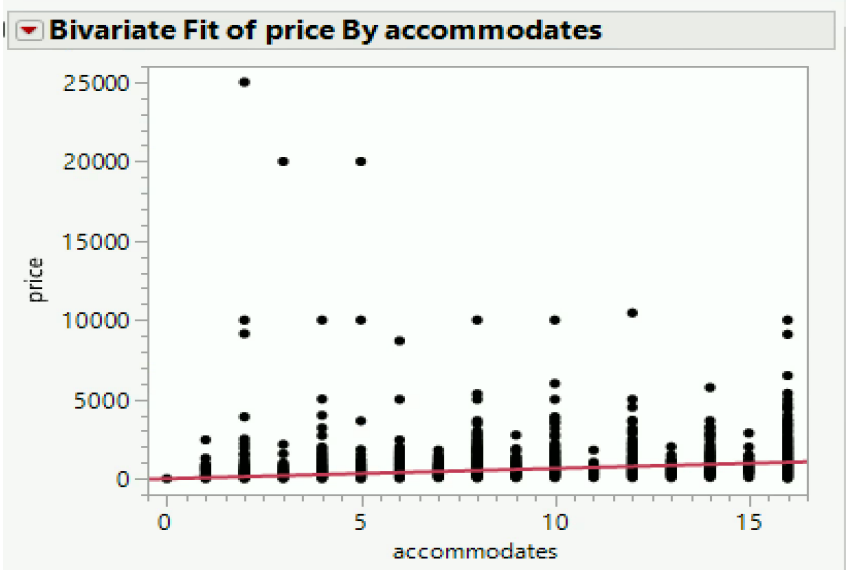
Linear Fit

price = 107.69737 + 35.283813*review_scores_rating

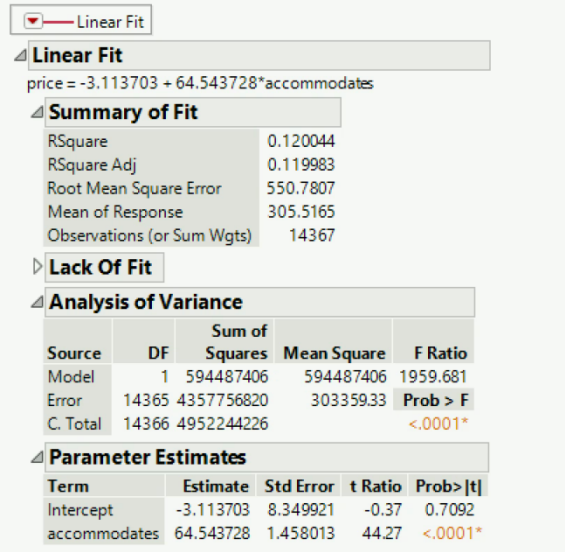
Summary of Fit

RSquare	0.002092
RSquare Adj	0.002004
Root Mean Square Error	401.7963
Mean of Response	276.5967
Observations (or Sum Wgts)	11336


10K
9K
8K
7K
6K
5K
4K
3K
2K
1K
0K
>3K nulls
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Review Scores Rating



- **Accommodates** explain 12.04% of the variance in price. It is significant with a p-value of less than .05. When accommodates goes up by 1 unit then the price goes up by $64.54.



- **Hosting Listing Count (Insignificant):** Based on the p-value we can see that host listing count by itself is insignificant in predicting the price per night. Initially we assumed that hosts with more listings would be able to charge a higher price but, in the dataset, there was no relationship in-between the two variables.

Linear Fit

Linear Fit

price = 307.08995 - 0.013296*host_listings_count

Summary of Fit

RSquare	0.000182
RSquare Adj	0.000112
Root Mean Square Error	587.1286
Mean of Response	305.5519
Observations (or Sum Wgts)	14365

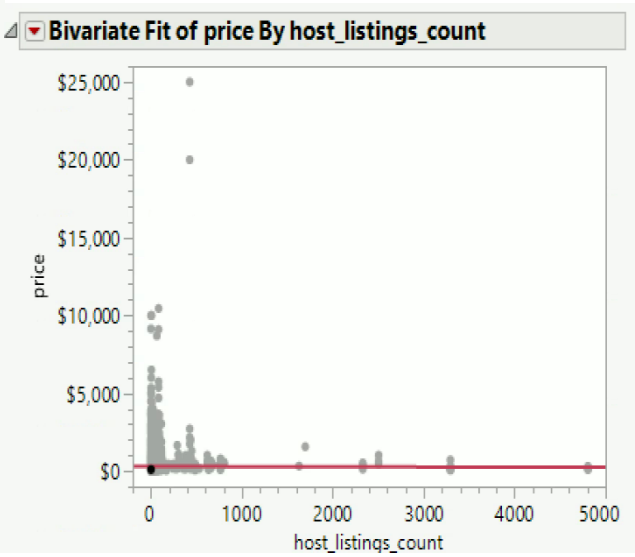
Lack Of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	900612.9	900613	2.6126
Error	14363	4951212886	344720	Prob > F
C. Total	14364	4952113499		0.1060

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	307.08995	4.990258	61.54	<.0001*
host_listings_count	-0.013296	0.008226	-1.62	0.1060

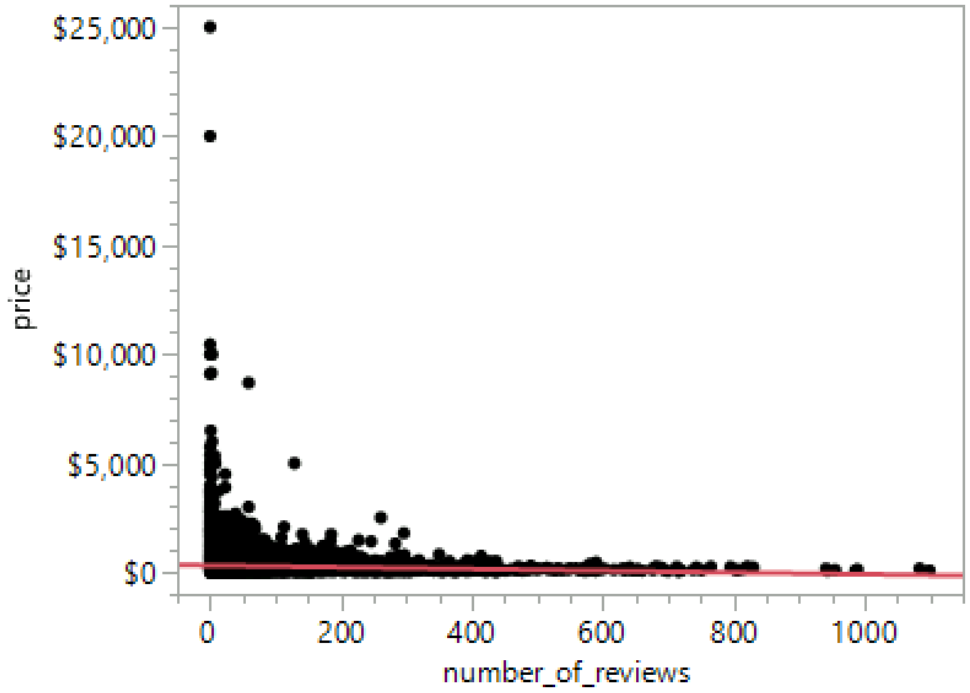


- **Number of reviews:** Based on the p-value from the model .0001, we can see that the number of reviews and the price charged per night are significant to each other. Interestingly we can see that with every increase in the number of reviews, the average price tends to go down by $.41. This may be because, with a higher number of reviews, the host may feel the need to be more competitive/responsible and therefore, the host might have strategically lowered the price.

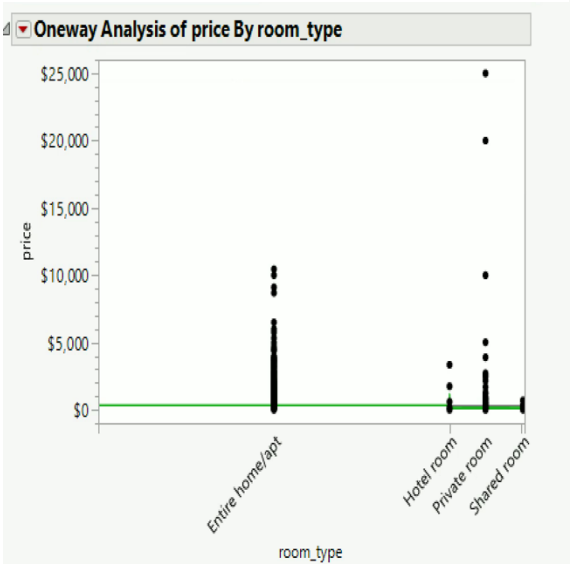
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	320.06333	5.388817	59.39	<.0001*
number_of_reviews	-0.407449	0.063333	-6.43	<.0001*

Bivariate Fit of price By number_of_reviews



- **Room Type (insignificant):** After analyzing the p-value, we can realize that the Room Type does not really have a significant impact on the price charged per night. There are 3 variables: Entire home/Apt, Hotel Room, and Private Room. In these 3 variables, only the Entire Home/Apt, and Hotel Room have a p-value less than 0.05, so these two have an impact on the price. However, the Private Room p-value is more than 0.5, so it does not have a significant impact on the price charged per night.



Response price

Effect Summary

Summary of Fit

RSquare	0.006678
RSquare Adj	0.006471
Root Mean Square Error	795.781
Mean of Response	310.0299
Observations (or Sum Wgts)	14368

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	61153048.7	20384350	32.1892
Error	14364	9096253996	63326747	Prob > F
C. Total	14367	9157407045		<.0001*

Parameter Estimates

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
room_type	3	3	61153049	32.1892	<.0001*

Indicator Function Parameterization

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	71.055046	76.222	0.93	0.3512
room_type[Entire home/apt]	268.11523	76.57223	3.50	0.0005*
room_type[Hotel room]	715.69495	291.4931	2.46	0.0141*
room_type[Private room]	105.49853	77.92175	1.35	0.1758

- **Number of Minimum Nights:** Based on the Parameter Estimates, the minimum nights variable does not have a significant impact on the price, because the p-value is more than 0.05. However, after adding it to the final set with review score rating, number of reviews, accommodations, bedrooms, and minimum nights, it worked. After running the profiler, we can see that the price increased when the minimum nights increased. It can be explained because if the customers stay there longer, Airbnb cannot get more profit from booking, which will affect the revenue. Therefore, they need to increase the price a bit to cover that part.

Linear Fit

Linear Fit

price = 307.57348 - 0.2744289*minimum_nights

Summary of Fit

RSquare	0.000133
RSquare Adj	6.345e-5
Root Mean Square Error	587.1098
Mean of Response	305.5165
Observations (or Sum Wgts)	14367

Lack Of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	658904.609	658905	1.9115
Error	14365	4951585321	344698	Prob > F
C. Total	14366	4952244226		0.1668

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	307.57348	5.119171	60.08	<.0001*
minimum_nights	-0.274429	0.19849	-1.38	0.1668

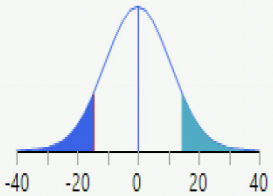
- **Host Identity Verified:** we assumed that having the host identity verified will have a significant impact on the price charged per night, but it turns out that host identity verification has little to no significance to price. The p-value was .9109, which is significantly higher than .05.

t Test

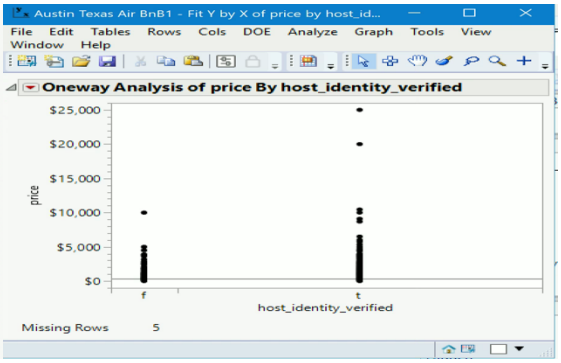
t-f

Assuming unequal variances

Difference	-14.495	t Ratio	-1.34669
Std Err Dif	10.763	DF	3975.252
Upper CL Dif	6.607	Prob > |t|	0.1782
Lower CL Dif	-35.597	Prob > t	0.9109
Confidence	0.95	Prob < t	0.0891



Missing Rows 5



2c. Explain why each certain variable did not work. Explain the introduction of new variables and why. Explain the final set of chosen variables and the results of the regression.

What did not work: Room type and Host identity verification, as well as host listing count all did not have any significance in predicting the price per night. We decided to leave these variables out since they do not display any significance in models or in relation to price.

The 5 variables we found to be significant: number of reviews, review scores rating, accommodations, bedrooms, minimum nights.

Linear Regression Model:

Price= -138.90 +17.14(review scores rating) -.10(number ofreviews)+24.25(accomodates)+105.75(bedrooms)+.693(minimum nights)

Summary of Fit

RSquare Adj	0.258046
Root Mean Square Error	354.4726
Mean of Response	284.1543
Observations (or Sum Wgts)	10715

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	468834232	93766846	746.2494
Error	10709	1345594654	125650.82	Prob > F
C. Total	10714	1814428886		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-138.8959	31.91872	-4.35	<.0001*
review_scores_rating	17.14145	6.596448	2.60	0.0094*
number_of_reviews	-0.103282	0.042721	-2.42	0.0156*
accommodates	24.25254	2.089731	11.61	<.0001*
bedrooms	105.74989	5.213919	20.28	<.0001*
minimum_nights	0.6929058	0.137447	5.04	<.0001*

Effect Summary

Source	LogWorth	PValue
bedrooms	89.053	0.00000
accommodates	30.228	0.00000
minimum_nights	6.328	0.00000
review_scores_rating	2.028	0.00937
number_of_reviews	1.806	0.01564

Semi Log Model:

ln(Price)= 3.93 + 9.5472e^-7(Number of Reviews) + 0.08(Review Scores Rating) + 0.098(Accommodates) + 0.20(Bedrooms) - 0.001(Minimum nights)

Summary of Fit

RSquare	0.449518
RSquare Adj	0.449261
Root Mean Square Error	0.627881
Mean of Response	5.239068
Observations (or Sum Wgts)	10715

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	3447.5231	689.505	1748.972
Error	10709	4221.8535	0.394	Prob > F
C. Total	10714	7669.3766		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.9292433	0.056538	69.50	<.0001*
number_of_reviews	9.5472e-7	7.567e-5	0.01	0.9899
review_scores_rating	0.0839323	0.011684	7.18	<.0001*
accommodates	0.0983034	0.003702	26.56	<.0001*
bedrooms	0.204033	0.009235	22.09	<.0001*
minimum_nights	-0.001255	0.000243	-5.15	<.0001*

Response Log[price]

Effect Summary

Source	LogWorth	PValue
accommodates	149.828	0.00000
bedrooms	105.073	0.00000
review_scores_rating	12.139	0.00000
minimum_nights	6.585	0.00000
number_of_reviews	0.004	0.98993

Log - Log Model:

Ln(price) = 3.84 - .05ln(Minimum nights) - .02ln(number of reviews) + .29ln(review scores rating) + .48ln(bedrooms) + .53ln(accommodates)

Summary of Fit

RSquare	0.487488
RSquare Adj	0.487248
Root Mean Square Error	0.605359
Mean of Response	5.240669
Observations (or Sum Wgts)	10655

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	3711.8905	742.378	2025.812
Error	10649	3902.4268	0.366	Prob > F
C. Total	10654	7614.3174		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.8402407	0.078406	48.98	<.0001*
Log[minimum_nights]	-0.046771	0.005715	-8.18	<.0001*
Log[number_of_reviews]	-0.01777	0.003761	-4.72	<.0001*
Log[review_scores_rating]	0.2934027	0.048938	6.00	<.0001*
Log[bedrooms]	0.4798168	0.020318	23.62	<.0001*
Log[accommodates]	0.5343866	0.018013	29.67	<.0001*

Response Log[price]

Effect Summary

Source	LogWorth	PValue
Log[accommodates]	185.172	0.00000
Log[bedrooms]	119.501	0.00000
Log[minimum_nights]	15.516	0.00000
Log[review_scores_rating]	8.678	0.00000
Log[number_of_reviews]	5.631	0.00000

Remove Add Edit FDR

Lack Of Fit

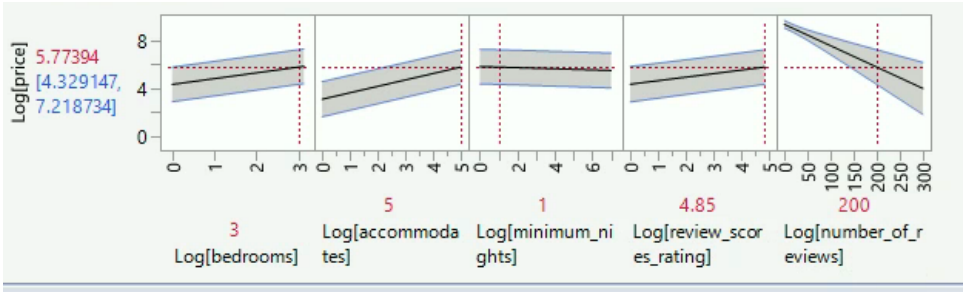
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	7567	2731.2627	0.360944	0.9498
Pure Error	3082	1171.1641	0.380001	Prob > F
Total Error	10649	3902.4268		0.9566
				Max RSq
				0.8462

2C Continued:

From the 3 different types of models we ran, we found that the log-log model had the highest adjusted R-Square which means that it can predict that variability in price the best out of these models. For example, the log-log model could explain 48.8% of the variance found in price charged per night by using just 5 variables such as: review scores rating, number of reviews, accommodates, bedrooms, and minimum nights.

Interpretation:

- When accommodation increases by 1%, the price charged per night will increase by .534%.
- When bedrooms are increased by 1%, there is a .48% increase in average price charged per night
- When review scores rating goes up by 1%, there is a .29% increase in average price charged per night
- When the number of reviews goes up by 1%, the average price charged per night goes down by .012%.
- When the number of minimum nights increases by 1%, there is a decrease in the average price charged per night of .05%.
- When all other variables = 0, 0 minimum nights, 0 reviews, 0 people accommodation, 0 bedrooms, and number of reviews is 0, the average price per night is $46.53 to stay a night in Austin Texas using Airbnb.
- When the host has 3 bedrooms, can accommodate 5 people, has a 1-night minimum stay, review score of 4.85 and has 200 reviews, the price the host should charge is $321.80
-


Log[price]
5.77394
[4.329147, 7.218734]
8
4
0
3
Log[bedrooms]
5
Log[accommodates]
1
Log[minimum_nights]
4.85
Log[review_scores_rating]
200
Log[number_of_reviews]

-

Additional Research for the Review Scores Rating Variable:

Summary of Fit

RSquare	0.839895
RSquare Adj	0.83981
Root Mean Square Error	0.152199
Mean of Response	4.813647
Observations (or Sum Wgts)	11266

Analysis of Variance

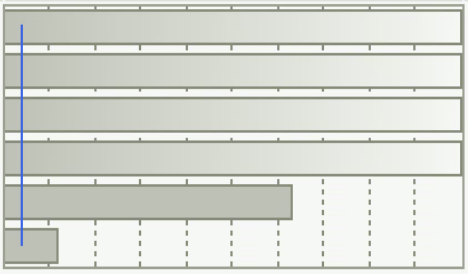
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	1368.1797	228.030	9843.938
Error	11259	260.8091	0.023	Prob > F
C. Total	11265	1628.9888		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.419013	0.025348	-16.53	<.0001*
review_scores_accuracy	0.3144756	0.007562	41.59	<.0001*
review_scores_cleanliness	0.2682645	0.005507	48.71	<.0001*
review_scores_checkin	0.0861489	0.007265	11.86	<.0001*
review_scores_communication	0.1843619	0.006965	26.47	<.0001*
review_scores_location	0.0275412	0.005628	4.89	<.0001*
review_scores_value	0.2039743	0.006252	32.62	<.0001*

Effect Tests

Source	LogWorth	PValue
review_scores_cleanliness	469.341	0.00000
review_scores_accuracy	351.066	0.00000
review_scores_value	222.426	0.00000
review_scores_communication	149.120	0.00000
review_scores_checkin	31.518	0.00000
review_scores_location	5.998	0.00000



Part 3:

Log Log regression equation:

Ln(price) = 3.84 - .05ln(minimum_nights) - .02ln(number_of_reviews) + .29ln(review_scores_rating) + .48ln(bedrooms) + .53ln(accomodates)

The regression model gives the average price for an Airbnb listing based on our chosen variables. To make a listing competitive to other Airbnb according to price, a host can use the relationship between each variable and price.

- Minimum nights and number of reviews have a negative relationship with price. To increase prices, the host can decrease the number of minimum nights required
- Review score rating and price have a positive relationship and so to help increase our review score rating, we found all 6 variables (review score of: accuracy, cleanliness, check-in, communication, location, and value) to be significant in predicting the variance in overall review score rating. Therefore, we recommend that hosts focus on cleanliness, accuracy, and communication as these 3 variables have a higher impact on review scores rating and the host has the ability to make changes if needed. Therefore, try utilizing different cleaning services or updating your listing to have a more accurate description. We also recommend always communicating with the guests and seeing if they have any additional needs that can be fulfilled.
- A high number of reviews has a negative effect on price, but the cause is inconclusive. A possibility could be that there are more negative reviews included in the high number of reviews, which are decreasing the overall review score, therefore the price might decrease. Hosts should focus on increasing the review score rating rather than the number of reviews.
- Accommodates has the highest impact on price, so a host that can accommodate more guests will be able to competitively charge a higher rate. Hosts that can accommodate more people can lower

their price compared to the average price to attract more renters. If a host wants to increase their listing price, they can find ways to accommodate more guests by adding more sleeping arrangements, such as futons.

- The number of bedrooms has a high impact on price, so if a host has a rental property with only one or two bedrooms, they may be at a disadvantage compared to larger properties that can accommodate more guests. However, the host can still find ways to increase their listing price by highlighting the quality and comfort of their bedrooms and offering additional amenities. Additionally, expanding the number of bedrooms in a rental property can be an effective way to increase potential rental income, but it can also be costly and may not be possible for all hosts. To overcome this challenge, hosts can consider utilizing existing spaces within the property, such as an office, common area, or basement, and convert them into additional sleeping areas. For example, a host can convert a basement into a bedroom, if it meets building codes and safety regulations. In that way, they can accommodate more guests and increase their listing price.