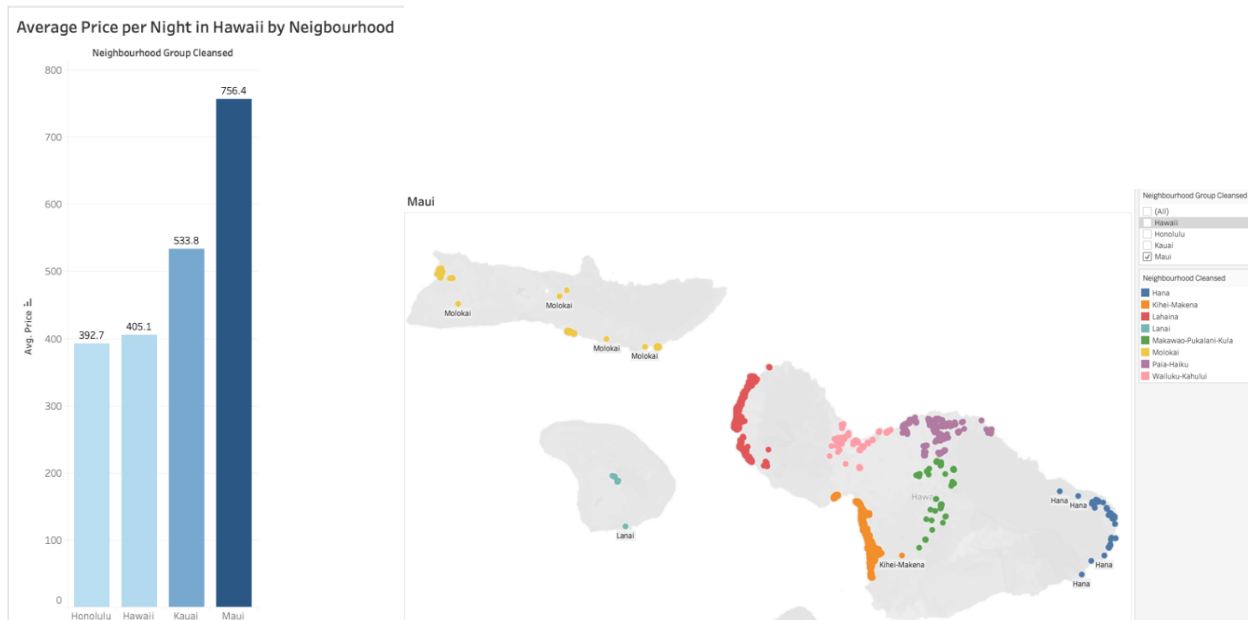


Marketing Analytics: Airbnb Midterm - Hawaii

Group 6: Aditi, Airi, Pankhuri, Pooja

Overview: *Hawaii, United States*

In our project, we picked **Maui**, which includes the most popular and expensive neighborhoods, Kihei and Lahaina. Maui also has the highest average price per night in Hawaii.



Part 1: Identify a Problem

As a **quest** of Airbnb, they would like to know the **average price per night**, which is the dependent variable (y variable). Guests may have questions, like how much is the average price per night in a certain neighborhood with certain number of people, room type, and the number of bedrooms. These questions are important in determining and predicting the budget for the stay in Maui.

Data: We used the detailed listing data from Hawaii, Inside Airbnb

- Listings_detailed

Part 2. (Preliminary) Data Analyses

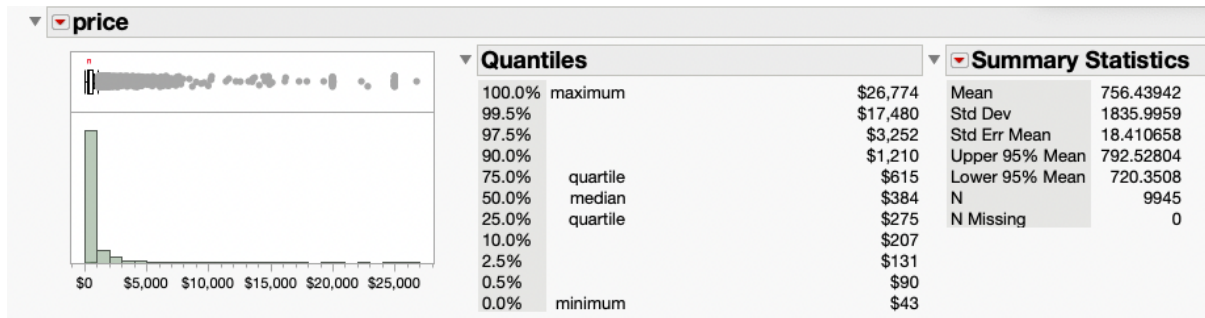
Based on intuition, the key variables/factors that affect and can predict your phenomenon of interest, (independent variables = x variables) are:

- Room Type
- Instant Bookable
- Neighborhood in Maui
- Number of Bedrooms
- Number of People (Accommodates)

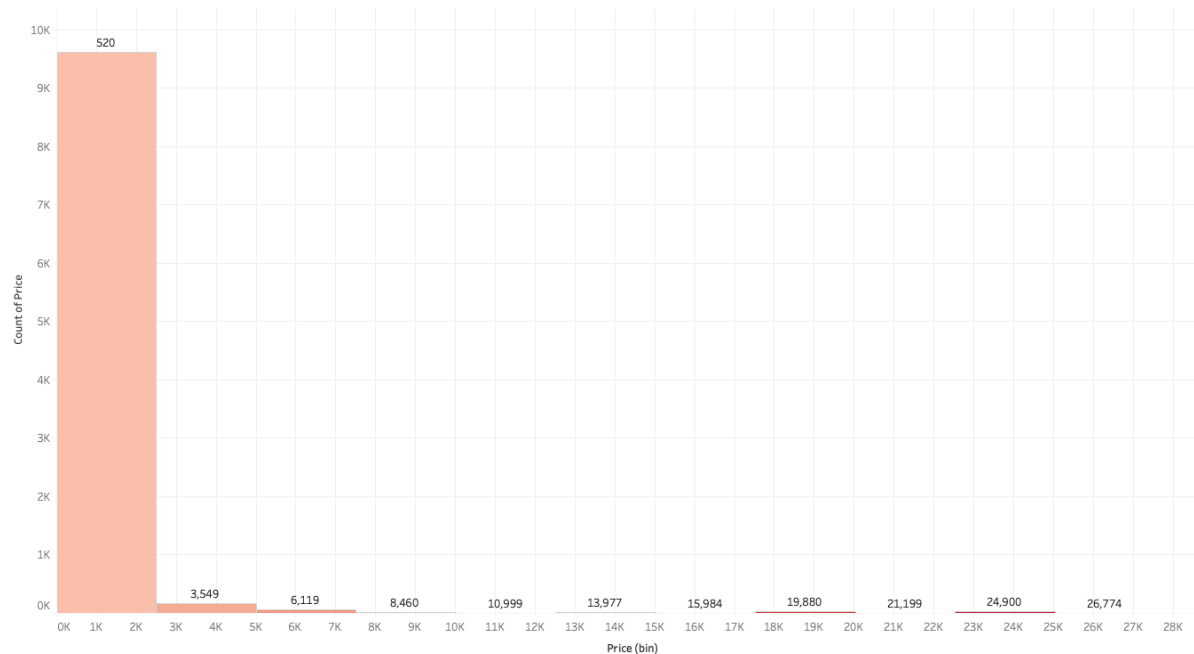
A) Summarize these variables (Tableau and JMP)

- Univariate (with the y variable)

Since the Maui listings have a variety range of prices per night, from \$43 up to \$26,774, and those 8,586 listings consists of \$43 to \$1,000-\$3,000 range, the histogram of the price is extremely right skewed.

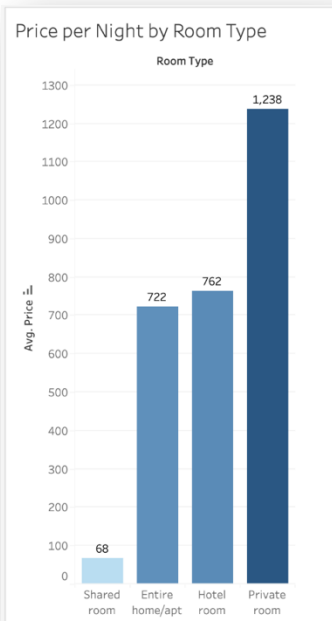


Average Price per Night in Maui



- Bivariate (with Y and X variables)

Price and Room Type: There are 4 room types in Maui, and private rooms have the highest price per night at \$1,238, followed by hotel room at \$762, Entire home/apt at \$722, and shared room at \$68. All three means are not equal, and the probability > F is smaller than 0.05, which indicates that there is a correlation between the price and the room type.

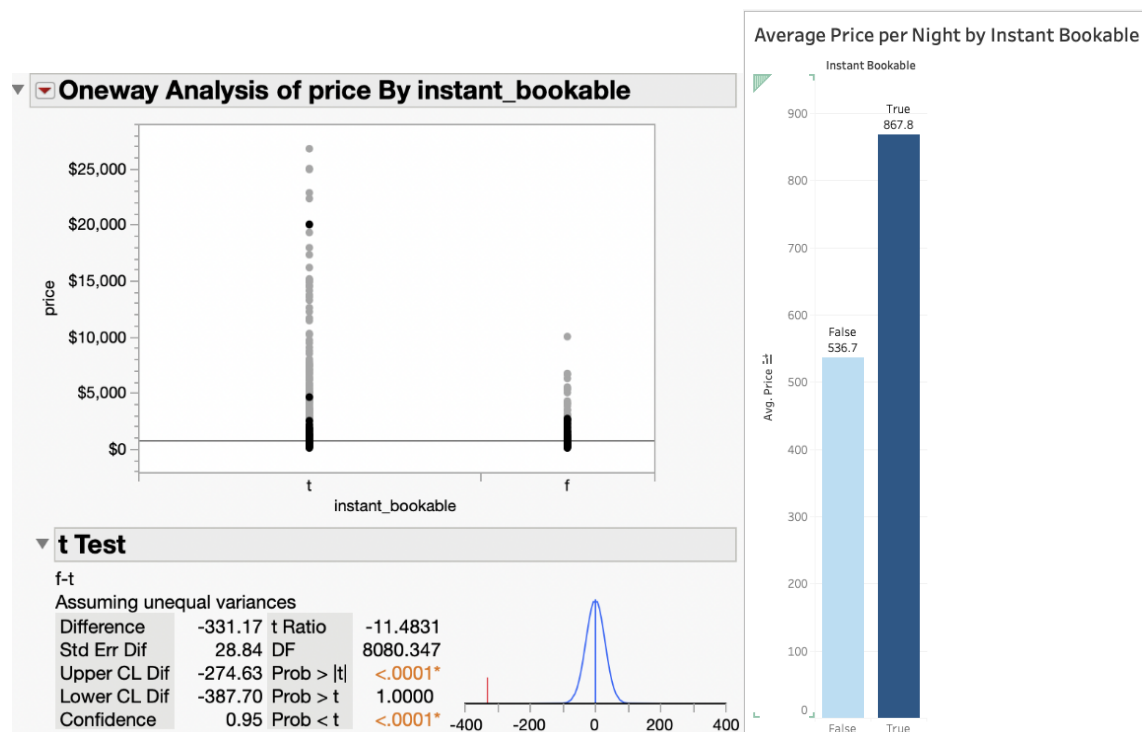


Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
room_type	3	176749185	58916395	17.5654	<.0001*
Error	9941	3.3343e+10	3354118.4		
C. Total	9944	3.352e+10			

Means for Oneway Anova					
Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Entire home/apt	9231	722.16	19.06	685	759.5
Hotel room	15	762.33	472.87	-165	1689.3
Private room	682	1237.51	70.13	1100	1375.0
Shared room	17	67.65	444.19	-803	938.3

Std Error uses a pooled estimate of error variance

Instant Bookable: This is the data of price and instant bookable. When the listing is instantly bookable, the average price is \$868 and \$537 when it is not. Since the value of t-test is smaller than 0.05, there is a correlation between the price and whether or not the listing is instant bookable. Instant bookable listings have higher average price per night.



Price and Neighborhood in Maui: The average price for each neighborhood is different. The Prob > F smaller than 0.05, which indicates there is a correlation between the price and the neighborhood group. Lahaina seems to have the highest average price at \$872, followed by Kihei at \$697 and other neighborhoods.

▼ **Analysis of Variance**

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
neighbourhood_cleansed	7	215367247	30766750	9.1798	<.0001*
Error	9937	3.3305e+10	3351582.3		
C. Total	9944	3.352e+10			

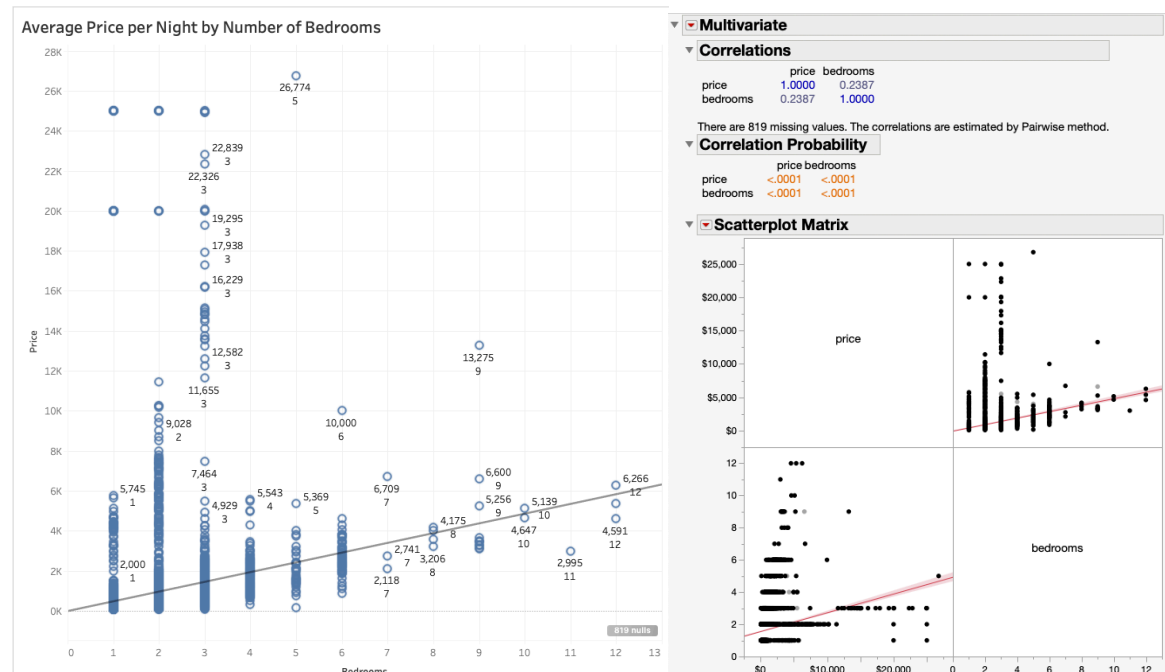
▼ **Means for Oneway Anova**

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Hana	73	378.452	214.27	-41.6	798.5
Kihei-Makena	4309	696.770	27.89	642.1	751.4
Lanai	8	339.250	647.26	-929.5	1608.0
Makawao-Pukalani-Kula	48	294.458	264.24	-223.5	812.4
Paia-Haiku	224	409.272	122.32	169.5	649.0
Waikuku-Kahului	191	464.702	132.47	205.0	724.4
Lahaina	4889	871.563	26.18	820.2	922.9
Molokai	203	169.601	128.49	-82.3	421.5

Std Error uses a pooled estimate of error variance

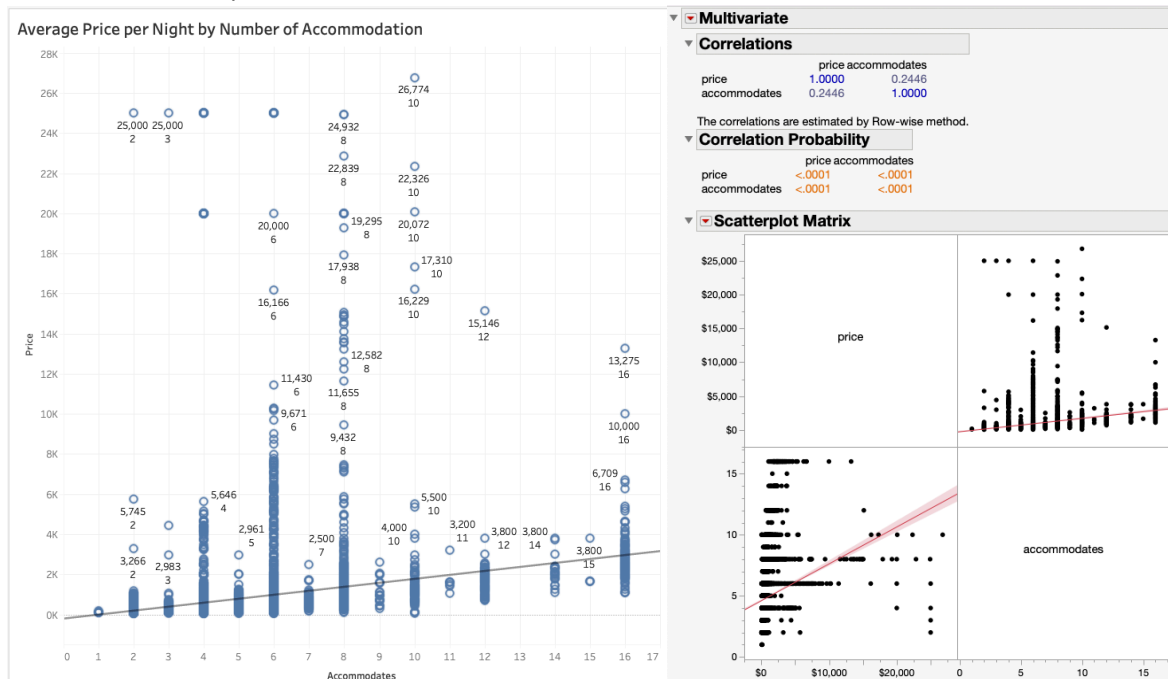


Price and Number of Bedrooms: Since the correlations are >0, the x- variable increases as the other y- variable also increases. The correlation probability is smaller than 0.05, thus there is a correlation between price and the number of bedrooms. The scatter plot also indicates that price increases as the number of bedrooms increases.



Price and number of People (accommodates): Since the correlation is >0, the x- variable increases as the other y- variable also increases. The correlation probability is smaller than 0.05, thus there is a correlation between price and the number of people that the listing can accommodate. The scatter plot

also indicates that price increases as the number of accommodates increases.



Part 2. (Preliminary) Data Analysis

B) Run Regression Models (JMP)

Based on the bivariate analysis, we chose the following variables that could have effect on y variables (price).

- Linear Regression: Adj. RSquare: 8% (7.9%)

$$\text{Price} = B_0 + B_1 * \text{Room type} + B_2 * \text{Instant Bookable} + B_3 * \text{Neighborhood in Maui} + B_4 * \text{\# of bedrooms} + B_5 * \text{accommodates}$$

The Effect summary that shows all x variables have passed the blue line indicates that all the variables are significantly important in effecting and predicting our y variable, price.

Response price

Effect Summary

Source	Logworth	PValue
room_type	21.183	0.00000
instant_bookable	16.604	0.00000
bedrooms	16.117	0.00000
accommodates	5.595	0.00000
neighbourhood_cleaned	4.029	0.00009

[Remove](#) [Add](#) [Edit](#) ☐ FDR

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	275	5410001475	19672733	7.0470
Pure Error	8837	2.467e+10	2791627.4	
Total Error	9112	3.008e+10		

Prob > F <.0001*
Max RSq 0.2462

Summary of Fit

Statistic	Value
RSquare	0.080937
RSquare Adj	0.079626
Root Mean Square Error	1816.893
Mean of Response	792.1564
Observations (or Sum Wgts)	9126

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	13	2648943641	203764895	61.7264
Error	9112	3.008e+10	3301098.8	
C. Total	9125	3.2729e+10		

Prob > F <.0001*

- Log-Log: Adj. RSquare: 38.23%

$$\text{Log}(\text{Price}) = B_0 + B_1 * \text{Room type} + B_2 * \text{Instant Bookable} + B_3 * \text{Neighborhood in Maui} + B_4 * \text{Log(\# of bedrooms)} + B_5 * \text{Log(accommodates)}$$

The Effect summary that shows all x variables have passed the blue line indicates that all the variables are significantly important in effecting and predicting our y variable, price.

Response Log[price]				
Effect Summary				
Source		Logworth		PValue
Log[bedrooms]		142.166		0.00000
neighbourhood_cleansed		131.361		0.00000
room_type		58.588		0.00000
Log[accommodates]		45.743		0.00000
instant_bookable		21.253		0.00000
Remove Add Edit <input type="checkbox"/> FDR				
Lack Of Fit				
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	275	507.0754	1.84391	5.1910
Pure Error	8837	3139.0138	0.35521	Prob > F
Total Error	9112	3646.0891		<.0001*
				Max RSq
				0.4689
Summary of Fit				
RSquare		0.383146		
RSquare Adj		0.382266		
Root Mean Square Error		0.632567		
Mean of Response		6.150165		
Observations (or Sum Wgts)		9126		
Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	13	2264.6952	174.207	435.3643
Error	9112	3646.0891	0.400	Prob > F
C. Total	9125	5910.7843		<.0001*

- Semi-Log Regressions: Adj. RSquare: 39%

$$\text{Log}(\text{Price}) = B_0 + B_1 * \text{Room type} + B_2 * \text{Instant Bookable} + B_3 * \text{Neighborhood in Maui} + B_4 * \# \text{ of bedrooms} + B_5 * \text{accommodates}$$

The Effect summary that shows all x variables have passed the blue line indicates that all the variables are significantly important in effecting and predicting our y variable, price.

Response Log[price]				
Effect Summary				
Source		Logworth		PValue
neighbourhood_cleansed		137.984		0.00000
bedrooms		69.418		0.00000
accommodates		64.076		0.00000
room_type		55.627		0.00000
instant_bookable		19.341		0.00000
Remove Add Edit <input type="checkbox"/> FDR				
Lack Of Fit				
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	275	460.9538	1.67620	4.7189
Pure Error	8837	3139.0138	0.35521	Prob > F
Total Error	9112	3599.9676		<.0001*
				Max RSq
				0.4689
Summary of Fit				
RSquare		0.390949		
RSquare Adj		0.390008		
Root Mean Square Error		0.628554		
Mean of Response		6.150165		
Observations (or Sum Wgts)		9126		
Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	13	2310.8167	177.755	449.9220
Error	9112	3599.9676	0.395	Prob > F
C. Total	9125	5910.7843		<.0001*

- The best model that predicts the best (the highest model fit based on R-Square) is the semi-log model. The low RSquare is a result of the raw data with over 9,000 listings that have the outliers, and that the data is not normally distributed.

<Interpret the parameters (coefficients)>

<Semi-Log Model>

Indicator Function Parameterization				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	2.7389355	0.16485	16.61	<.0001*
room_type[Entire home/apt]	1.5105947	0.157239	9.61	<.0001*
room_type[Hotel room]	1.1727695	0.226169	5.19	<.0001*
room_type[Private room]	1.8567379	0.157966	11.75	<.0001*
instant_bookable[t]	0.1294526	0.014078	9.20	<.0001*
neighbourhood_cleansed[Hana]	0.6717352	0.092705	7.25	<.0001*
neighbourhood_cleansed[Kihei-Makena]	0.8250106	0.049381	16.71	<.0001*
neighbourhood_cleansed[Lanai]	0.4650648	0.242537	1.92	0.0552
neighbourhood_cleansed[Makawao-Pukalani-Kula]	0.438851	0.106069	4.14	<.0001*
neighbourhood_cleansed[Paia-Haiku]	0.6374606	0.066919	9.53	<.0001*
neighbourhood_cleansed[Wailuku-Kahului]	0.5581198	0.072225	7.73	<.0001*
neighbourhood_cleansed[Lahaina]	1.0328426	0.049512	20.86	<.0001*
bedrooms	0.2450435	0.013721	17.86	<.0001*
accommodates	0.0971367	0.005669	17.13	<.0001*

Log (Price) = \$15.47 + 352.67*Room type (Entire Home) + 222.20*Room type (Hotel Room) + 542.37*Room type (Private Room) + 13.88*Instant_Bookable + 95.42*Neighborhood (Hana)+ 129.33*Neighborhood (Kihei-Makena) + 60*Neighborhood (Lahai) + 55.27*Neighborhood (Makawao) +89.65*Neighborhood (Paia) + 75.07*Neighborhood (Wailuku) + 180.11*Neighborhood (Lahaina) + 28.40*Number of Bedrooms + 10.52*Accommodates

Intercept: \$15.47

Intercept: Shared Room, Instant Bookable = False, and Neighborhood = Molokai

Room Type:

Entire home/apt: On average, when it is the entire home/apt, then the price would increase by 352.67(%) than the shared room, keeping all other variables in the model constant.

Hotel Room: When it is a hotel room, the price would increase by 222.20% than the shared room, keeping all the other variables in the model constant.

Private Room: When it is a private room, the price would increase by 542.37% than the shared room keeping all other variables in the model constant.

Instant Booking: If the listing is instantly bookable, then the price increases by 13.88% in comparison to when the listing is not instantly bookable, keeping all other variables in the model constant.

Neighbourhood cleansed:

Hana: When the neighborhood is Hana, then the price of the Airbnb listing would increase by 95.42% compared to Molokai keeping all the other variables in the model constant.

Kihei: When the neighborhood is Kihei, then the price of the Airbnb listing would increase by 129.33% compared to Molokai, keeping all other variables in the model constant.

Lanai: When the neighborhood is Lanai, then the price of the Airbnb listing would increase by 60% compared to when the neighborhood is Molokai, keeping all other variables in the model constant.

Makawao: When the neighborhood is Makawao, then the price of the Airbnb listing would increase by 55.27% compared to when the neighborhood is Molokai, keeping all other variables in the model constant.

Paia: When the neighborhood is Paia, then the price of the Airbnb listing would increase by 89.65% compared to when the neighborhood is Molokai, keeping all other variables in the model constant.

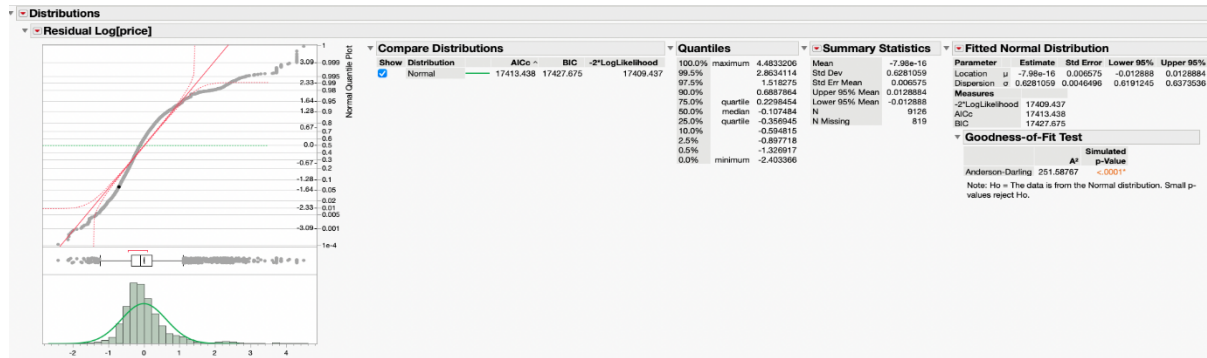
Wailuku: When the neighborhood is Wailuku, then the price of the Airbnb listing would increase by 75.07% compared to when the neighborhood is Molokai, keeping all other variables in the model constant.

Lahaina: When the neighborhood is Lahaina, then the price of the Airbnb listing would increase by 180.11% compared to when the neighborhood is Molokai, keeping all other variables in the model constant.

Bedroom: On average, when the bedroom increases by 1 unit, then the price of the Airbnb listing would increase by 28.40%, keeping all other variables in the model constant.

Accommodates: On average, when the accommodation (number of accommodating people) increases by 1 person, then the price of the Airbnb listing would increase by 10.52%, keeping all other variables in the model constant.

Residual Diagnostics: Although it seems better than the distributions in linear model, it is still definitely not normally distributed as the grey curve is way over the red curve. Also, the p-value is smaller than 0.05, which rejects the hypothesis and follows the alternatives, thus the model is not a good fit nor normally distributed.



Part 3. Recommend Decisions

- All of the variables influence the price, but the neighborhood influences the price the most in our model.
- Kihei and Lahaina are the most visited neighborhoods in Maui and that's where we can find the most Airbnb listings. As a guest, you have to make adjustments to this variable in considering the budget.
- If you want to stay in such places, like Lahaina and Kihei, then you need to compromise the price being high. You can make compromises to other variables, such as choosing a shared room instead of a private room, if you want to minimize the cost being too high.
- If you want to save money, then stay in Molokai, the cheapest average price per night, but maybe able to afford a private room there.

Price Prediction using Profiler for Semi Log model:

We used the prediction profiler for this model to be more specific.

- Students vacation trip: Private room, Instantly Bookable True, 4 people, Molokai, 2 bedroom - on average, **\$270** price per night
- Family Vacation: Entire Home, Instantly Bookable True, 5 people, Kihei-Makena, 3 bedrooms - on average, **\$614** price per night
- Couple Vacation: Private room, Instantly Bookable True, 2 people, Lahaina, 1 bedroom - on average, **\$493** price per night
- Family get-together: Entire Home, Instantly Bookable True, 20 people, Lahaina, 10 bedroom - on average, **\$18,033** price per night

Presentation:

https://www.canva.com/design/DAFhOyOMvWM/HKg8PmxXaNXBa3tSolIO5w/edit?utm_content=DAFhOyOMvWM&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton