# Airbnb Project

## Maui, Hawaii

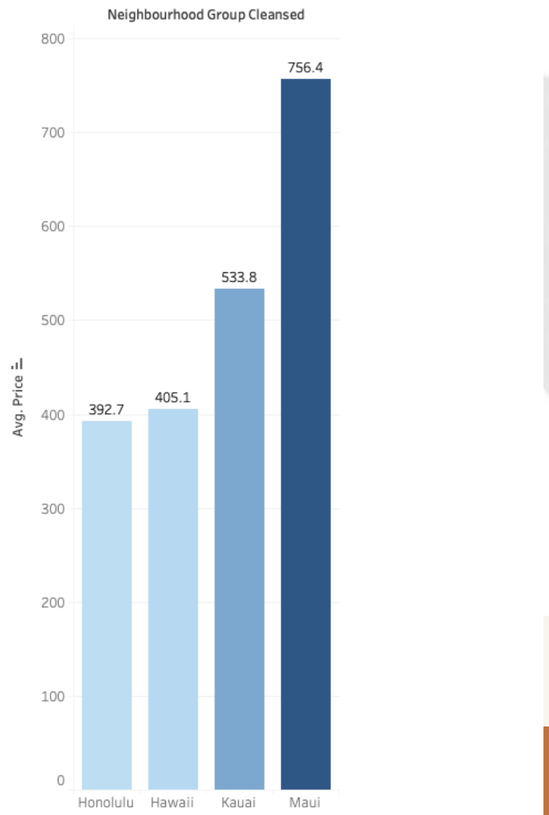
Group 6: Aditi, Airi, Pankhuri, Pooja

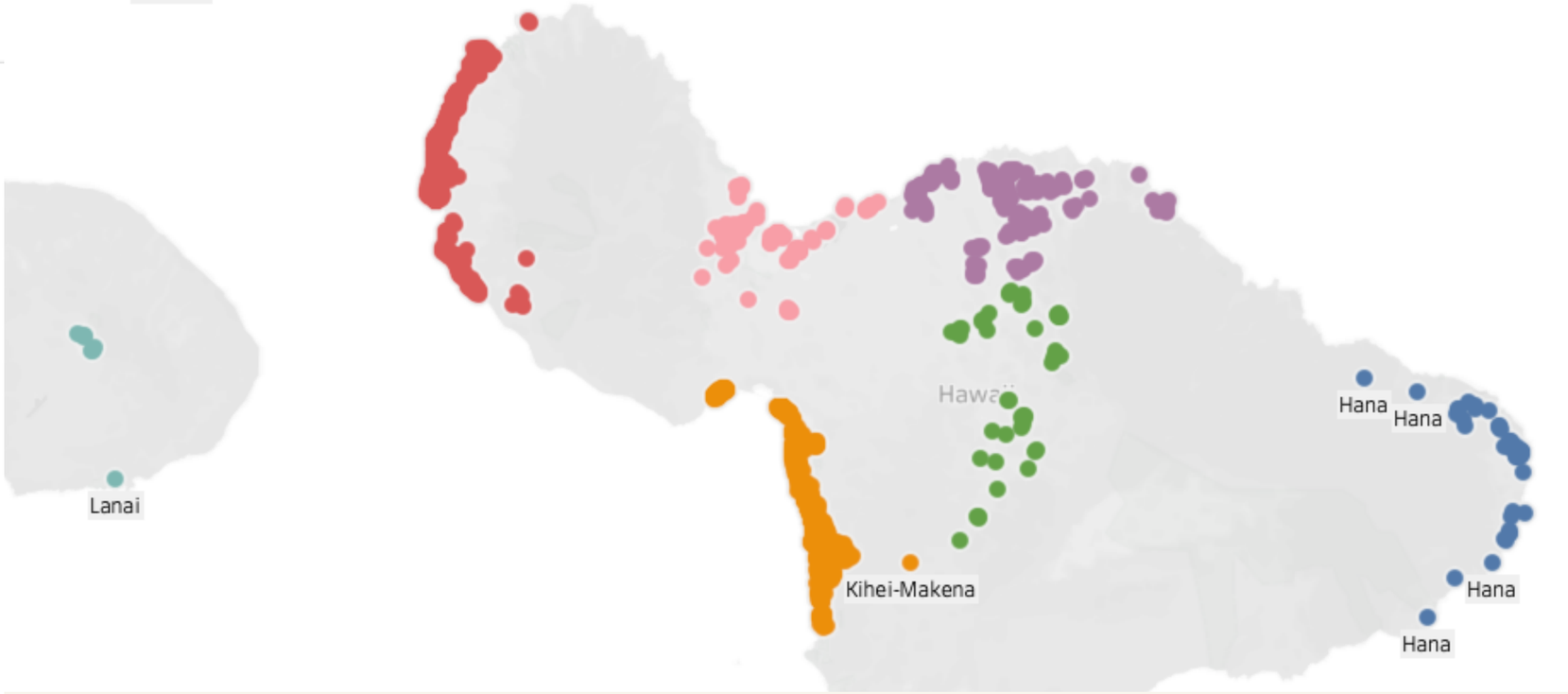
Maui

# Overview

## *Airbnb Market in Maui*


- The highest price per night in the neighborhood is Maui.
- There are total of 9,945 listings in Maui.
- In Maui, the highest average price per night is Lahaina, followed by Kihei.




Neighbourhood Group Cleansed

- [ ] (All)
- [ ] Hawaii
- [ ] Honolulu
- [ ] Kauai
- [x] Maui

Neighbourhood Cleansed

- Hana
- Kihei-Makena
- Lahaina
- Lanai
- Makawao-Pukalani-Kula
- Molokai
- Paia-Haiku
- Wailuku-Kahului

# Identify a Problem

- As a guest of Airbnb, we would like to know ***Price*** per night (Y variable).

  <What factors influence price?>
- Number of people?
- Neighborhood?
- Room Type?
- Number of bedrooms?
- Which neighborhood is the lowest/highest average price?
- Whether or not the listing is instantly bookable

- **These questions are important in determining and predicting the budget for their stay.**

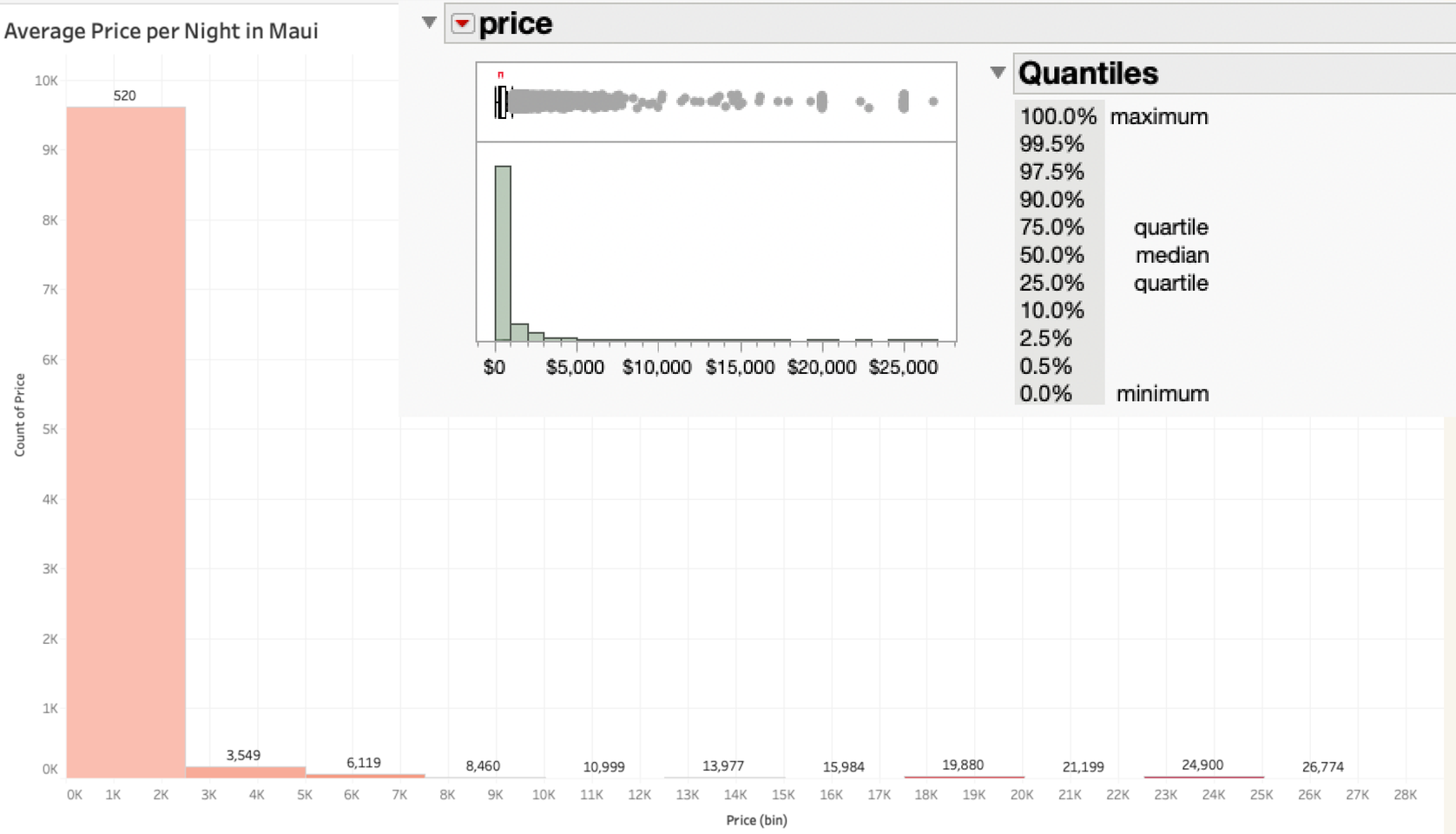
# Independent variables (x-variables)

- Room type (Shared room, Private Room, etc.)
- Instant Bookable
- Neighborhood in Maui
- Number of Bedrooms
- Number of people (Accommodation)

# Univariate (y = price)

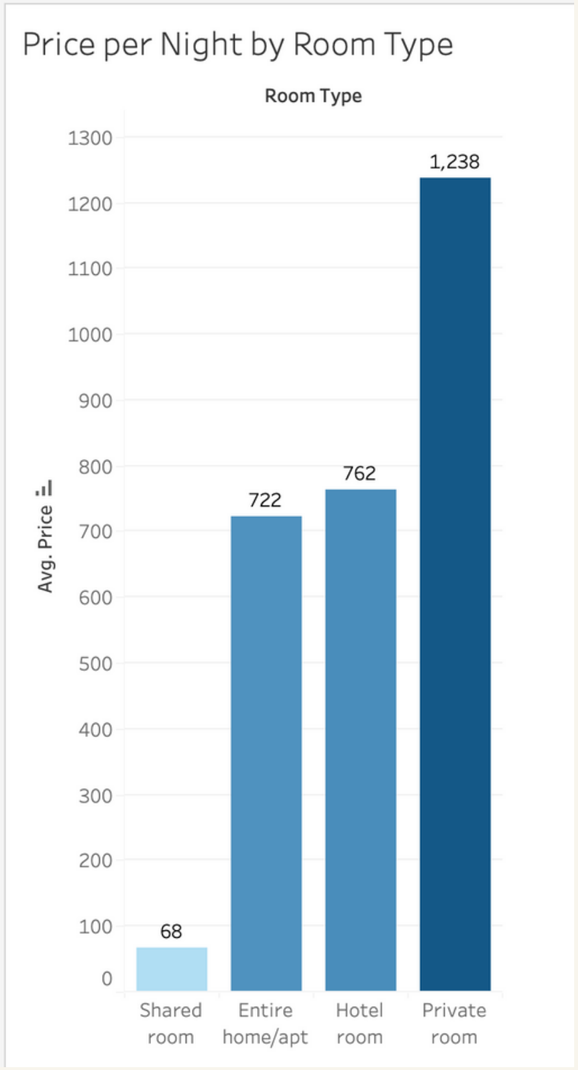
## Histogram of Price per night

- Range of prices/night from $43 up to $26,774
- 8,586 listings consists of $43 to $5,000 range, the histogram of the price is extremely right skewed.



# Bivariate (with y = price and Xs)

## Price and Room type




- 4 room types in Maui, private rooms highest at $1,238, followed by hotel room at $762, Entire home/apt at $722, and shared room at $68.
- The prob > F is smaller than 0.05, which indicates that there is a correlation between the price and the room type.
- The values of means are not equal as well.

**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 | | | |

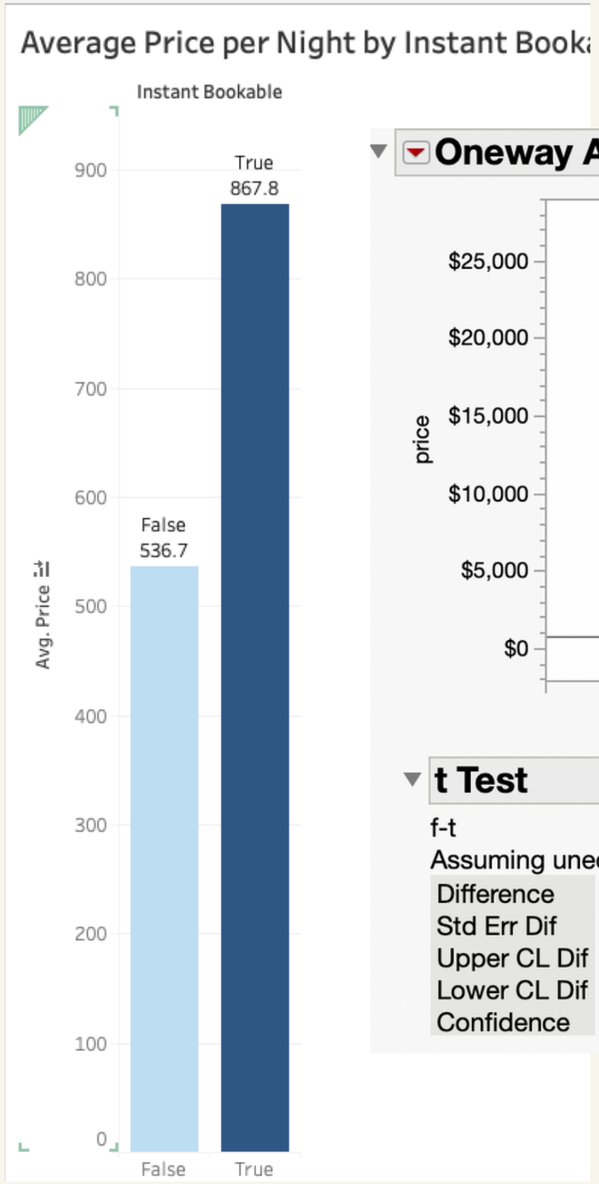
**Means for Oneway Anova**

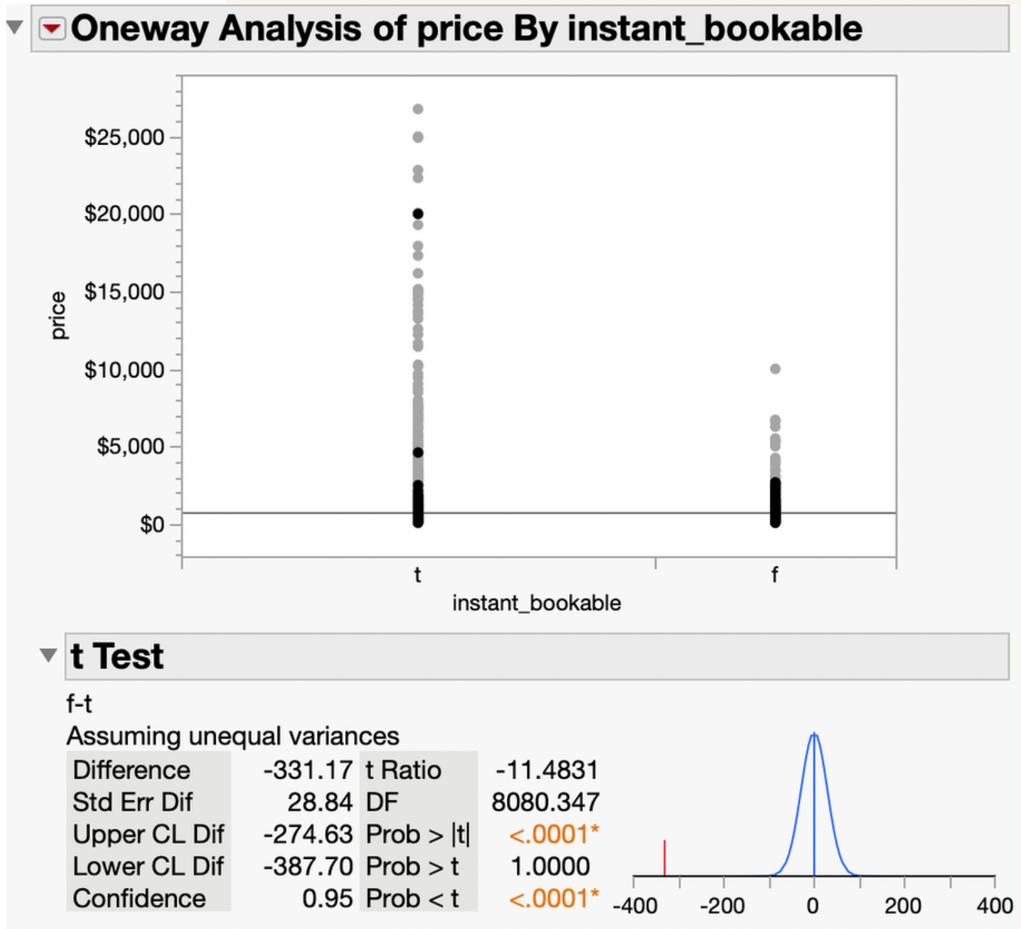
| Level | Number | Mean | Std Error | Lower 95% | Upper 95% |
|---|---|---|---|---|---|
| Entire home/apt | 9231 | 722.16 | 19.06 | 685 | 759.5 |
| Hotel room | 15 | 762.33 | 472.87 | -165 | 1689.3 |
| Private room | 682 | 1237.51 | 70.13 | 1100 | 1375.0 |
| Shared room | 17 | 67.65 | 444.19 | -803 |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

Std Error uses a pooled estimate of error variance

# Bivariate (with y = price and Xs)

## Price and Instant Bookable

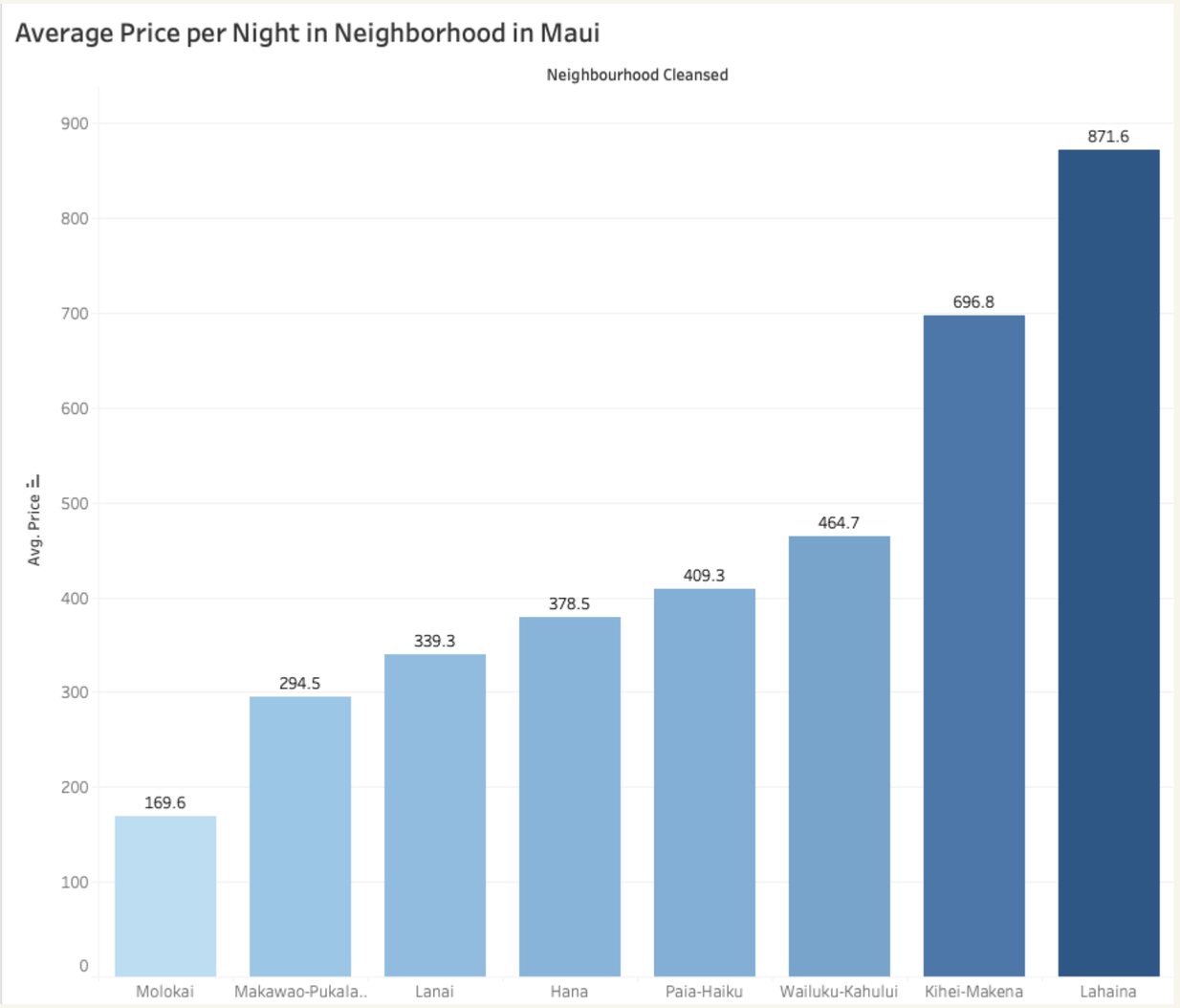



- Instantly bookable = the average price is $868 and $537 when it's not.
- Instant bookable listings have higher average price per night, and the difference is at $331.
- t-test = smaller than 0.05, there is a correlation.

# Bivariate (with y = price and Xs)

## Price and Neighborhood




- The Prob > F = smaller than 0.05, which indicates there is a correlation.
- Lahaina seems to have the highest average price at $872, followed by Kihei at $697 and others.

### 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 | | | |

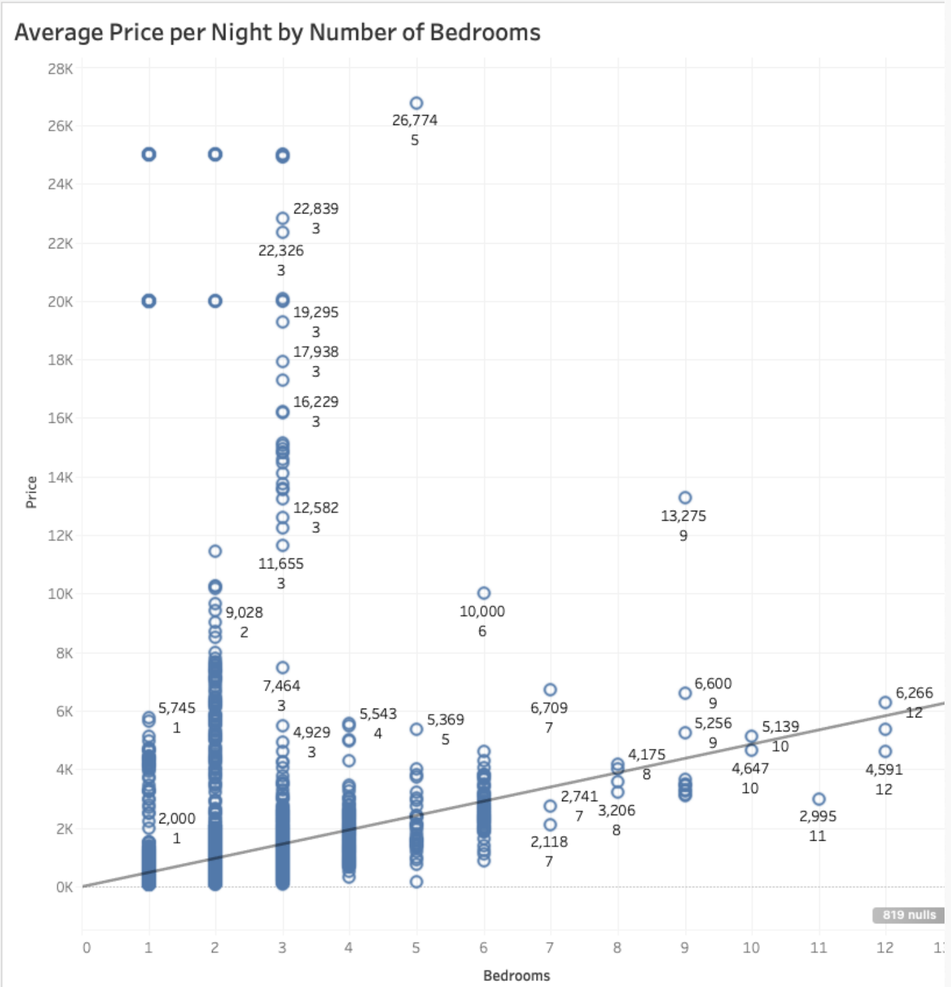
### Means for Oneway Anova

| 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 |
| Wailuku-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

# Bivariate (with y = price and Xs)

## Price and Number of Bedrooms




**Multivariate**

**Correlations**

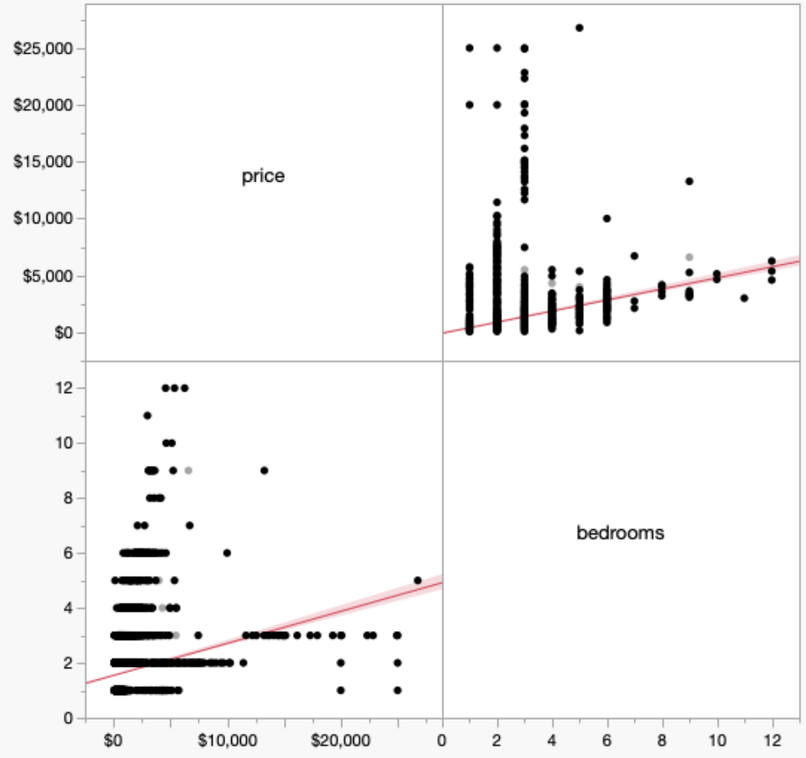
| | price | bedrooms |
|---|---|---|
| price | 1.0000 | 0.2387 |
| bedrooms | 0.2387 | 1.0000 |

There are 819 missing values. The correlations are estimated by Pairwise method.

**Correlation Probability**

| | price | bedrooms |
|---|---|---|
| price | <.0001 | <.0001 |
| bedrooms | <.0001 | <.0001 |

**Scatterplot Matrix**




- Correlation is >0, the variable increases as the other variable increases.
- The correlation probability is smaller than 0.05, thus there is a correlation.
- The scatter plot also indicates that price increases as the number of bedrooms increases.

# Bivariate (with y = price and Xs)

## Price and Accommodation

**Multivariate**

**Correlations**

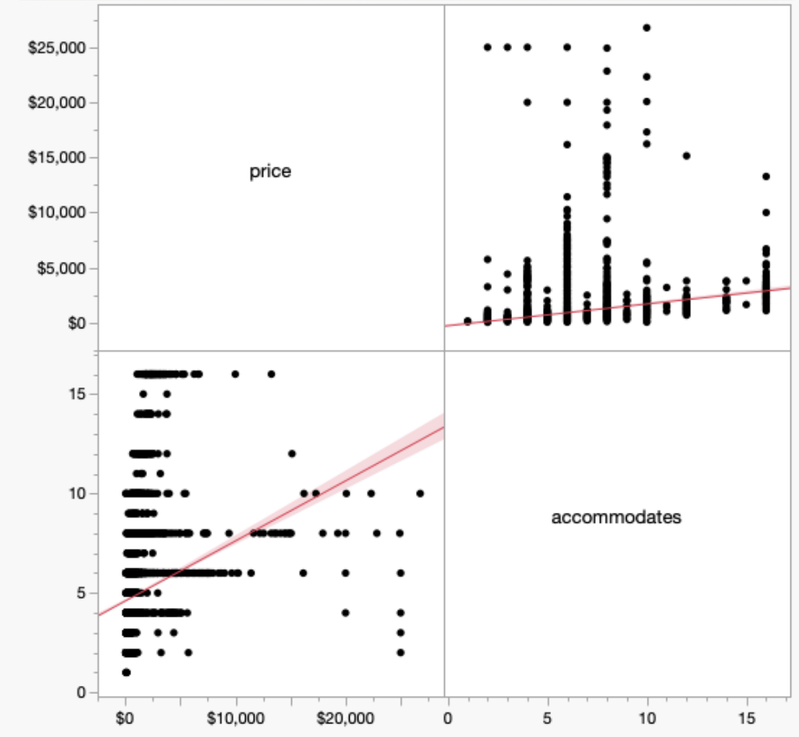
| | price | accommodates |
|---|---|---|
| price | 1.0000 | 0.2446 |
| accommodates | 0.2446 | 1.0000 |

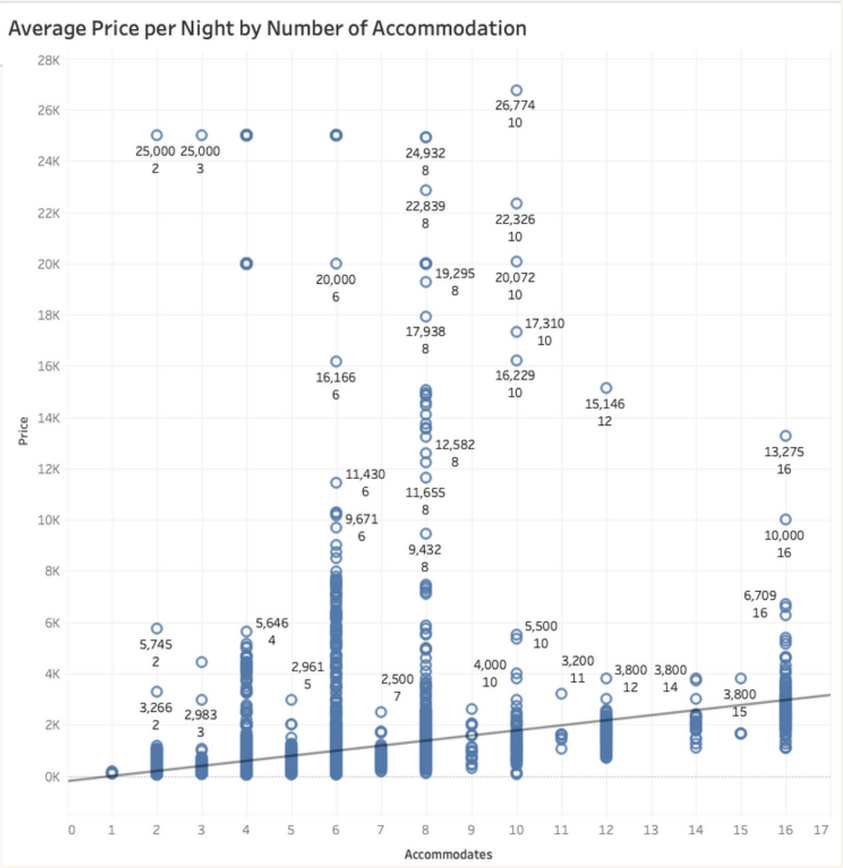
The correlations are estimated by Row-wise method.

**Correlation Probability**

| | price | accommodates |
|---|---|---|
| price | <.0001 | <.0001 |
| accommodates | <.0001 | <.0001 |

**Scatterplot Matrix**




- The correlation is >0 = the variable increases as the other variable increases.
- The correlation probability is smaller than 0.05, thus there is a correlation between price and the accommodation.
- The scatter plot also indicates that price increases as the number of accommodates increases.

# Regression Models

## Linear

**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_cleansed | 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 | Prob > F |
| Total Error | 9112 | 3.008e+10 | | <.0001* |
| | | | | Max RSq |
| | | | | 0.2462 |

**Summary of Fit**

| | |
|---|---|
| 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 | Prob > F |
| C. Total | 9125 | 3.2729e+10 | | <.0001* |

- ***Price = Bo + B1*Room type + B2*Instant_Bookable + B3*Neighborhood in Maui + B4*# of bedrooms + B5*accommodates***

- Adjusted R-square is 7.9%

# Regression Models

## Log-Log

**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 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* |

- ***Log (Price) = Bo + B1*Room type + B2*Instant_Bookable + B3*Neighborhood in Maui + B4*Log(# of bedrooms) + B5*Log(accommodates)***

- Adjusted R squared is 38%

# Regression Models

## Semi-Log

### 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 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 | 0.39008 |
| 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* |

### 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* |

- Best Model: Semi-Log with highest Adj. RSquare at 39%

# Regression Models

## Semi-Log Model Equation

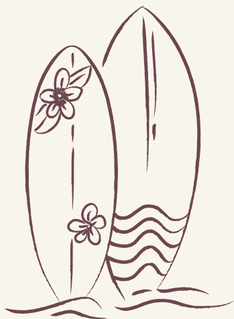
- ***Log (Price) = $15.47 + 352.67*Room type(Entire Home) + 222.20*Room type(Hotel Room) + 542.37*Room type(Private Room) + 13.88*Instant_Bookable + 95.42*Neighborhood(Hana)+ 129.33*Neighborhood(Kihei-Makena) + 60*Neighborhood(Lahai) + 55.27*Neighborhood(Makawao) +89.65*Neighborhood(Paia) + 75.07*Neighborhood(Wailuku)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

# Regression Models

## 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* |

On an average with each additional bedroom, the price increases by 28.40%

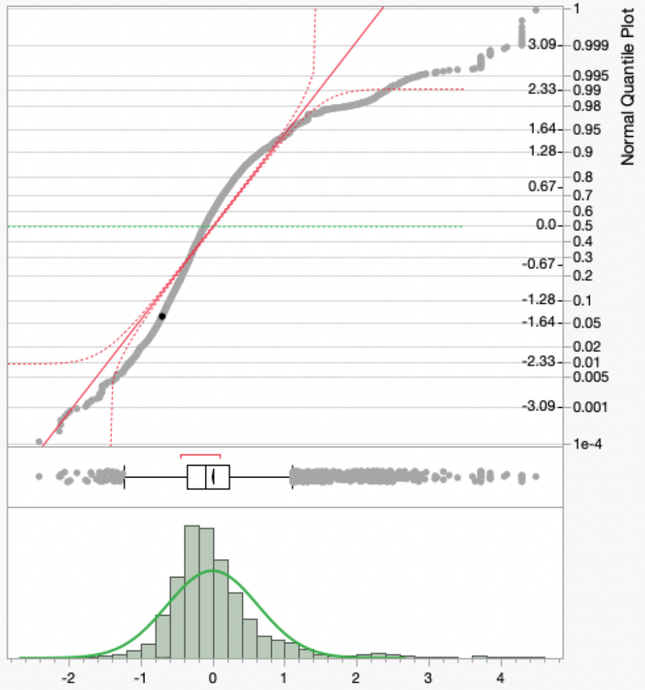


# Regression Models

## Residual Diagnostic for Semi Log Model

**Distributions**

**Residual Log[price]**




**Compare Distributions**

| Show | Distribution | AICc | BIC | -2*LogLikelihood |
|---|---|---|---|---|
| ☑ | Normal | 17413.438 | 17427.675 | 17409.437 |

**Quantiles**

| | | |
|---|---|---|
| 100.0% | maximum | 4.4833206 |
| 99.5% | | 2.8634114 |
| 97.5% | | 1.518275 |
| 90.0% | | 0.6887864 |
| 75.0% | quartile | 0.2298454 |
| 50.0% | median | -0.107484 |
| 25.0% | quartile | -0.356945 |
| 10.0% | | -0.594815 |
| 2.5% | | -0.897718 |
| 0.5% | | -1.326917 |
| 0.0% | minimum | -2.403366 |

**Summary Statistics**

| | |
|---|---|
| Mean | -7.98e-16 |
| Std Dev | 0.6281059 |
| Std Err Mean | 0.006575 |
| Upper 95% Mean | 0.0128884 |
| Lower 95% Mean | -0.012888 |
| N | 9126 |
| N Missing | 819 |

**Fitted Normal Distribution**

| Parameter | | Estimate | Std Error | Lower 95% | Upper 95% |
|---|---|---|---|---|---|
| Location | μ | -7.98e-16 | 0.006575 | -0.012888 | 0.0128884 |
| Dispersion | σ | 0.6281059 | 0.0046496 | 0.6191245 | 0.6373536 |

| Measures | |
|---|---|
| -2*LogLikelihood | 17409.437 |
| AICc | 17413.438 |
| BIC | 17427.675 |

**Goodness-of-Fit Test**

| | $A^2$ | Simulated p-Value |
|---|---|---|
| Anderson-Darling | 251.58767 | <.0001* |

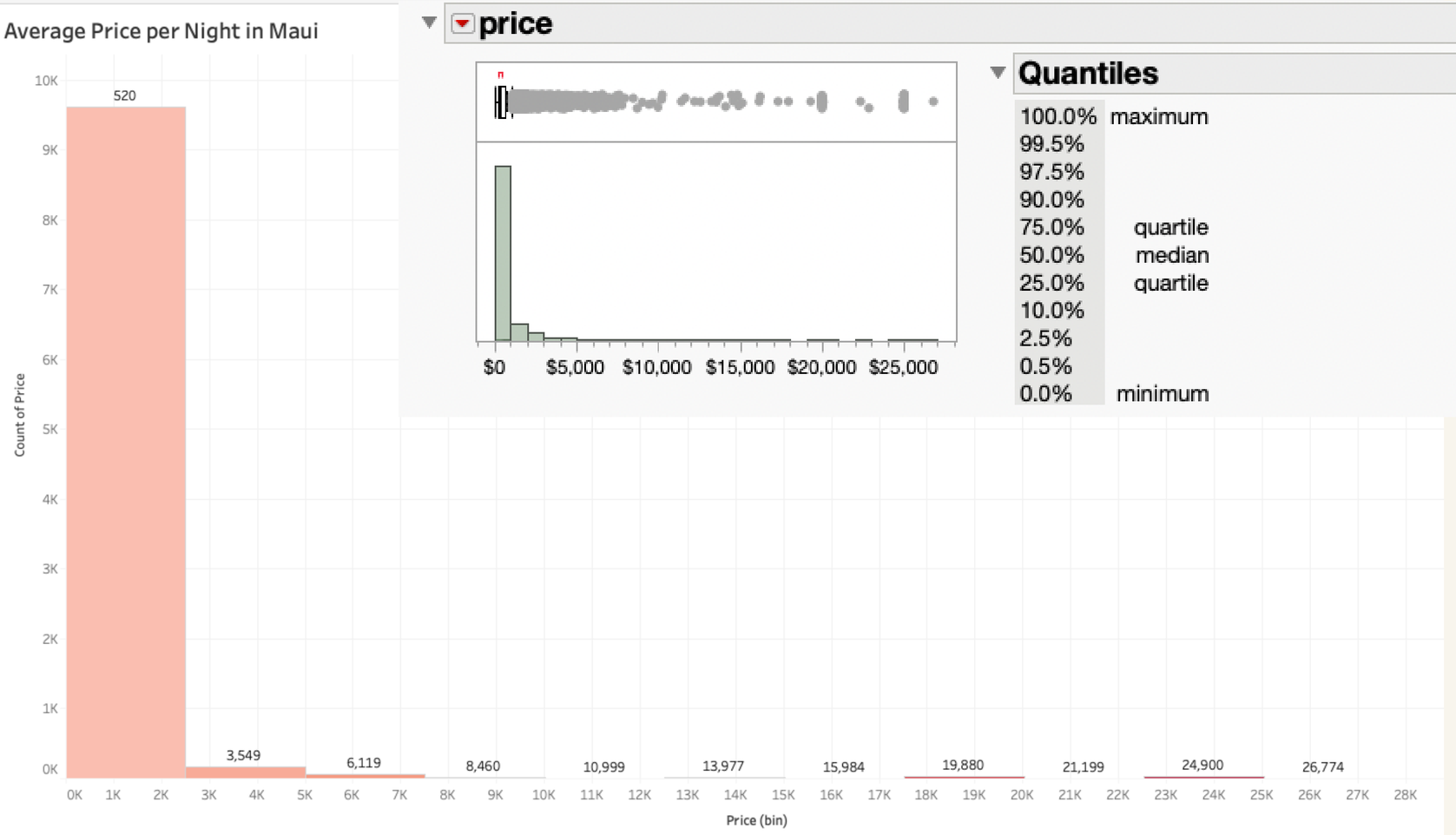
Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

- This semi-log model is not normally distributed

# Univariate (y = price)

## Histogram of Price per night

- Range of prices/night from $43 up to $26,774
- 8,586 listings consists of $43 to $5,000 range, the histogram of the price is extremely right skewed.



# Recommendation

- 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 the neighborhood 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.

# Recommendation

Price Prediction using Profiler for Semi-Log Model

- **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 bedroom - 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

# Questions


Mahalo!
(thank you!)