

airbnb

Tai-Hua Chung, Olivia
Natasha, Randy Shi,
Jared Sinsheimer, Dan To

Airbnb Price Recommendation



How much should
you charge?

Objective:

- Provide **Price Recommendation Model** for new listing
 - Find out price factors

Benefits:

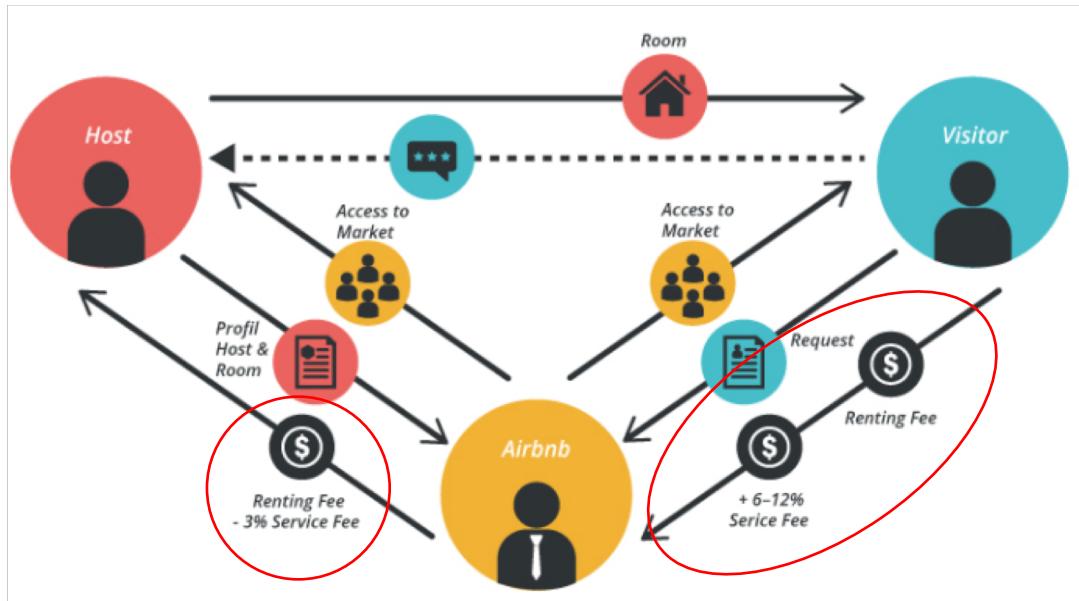
- **For Host:**
 - Set competitive pricing
 - Increase booking potential
- **For Traveler:**
 - Find reasonable pricing accommodation
- **For Business (see next page)**

Business Importance

Revenue Source:

- From Traveler:
 - 6 - 12% Service fee
(based on trip duration)
 - Renting fee
 - From Host:
 - 3% Booking fee
- Incentive to optimize number of bookings and trip price

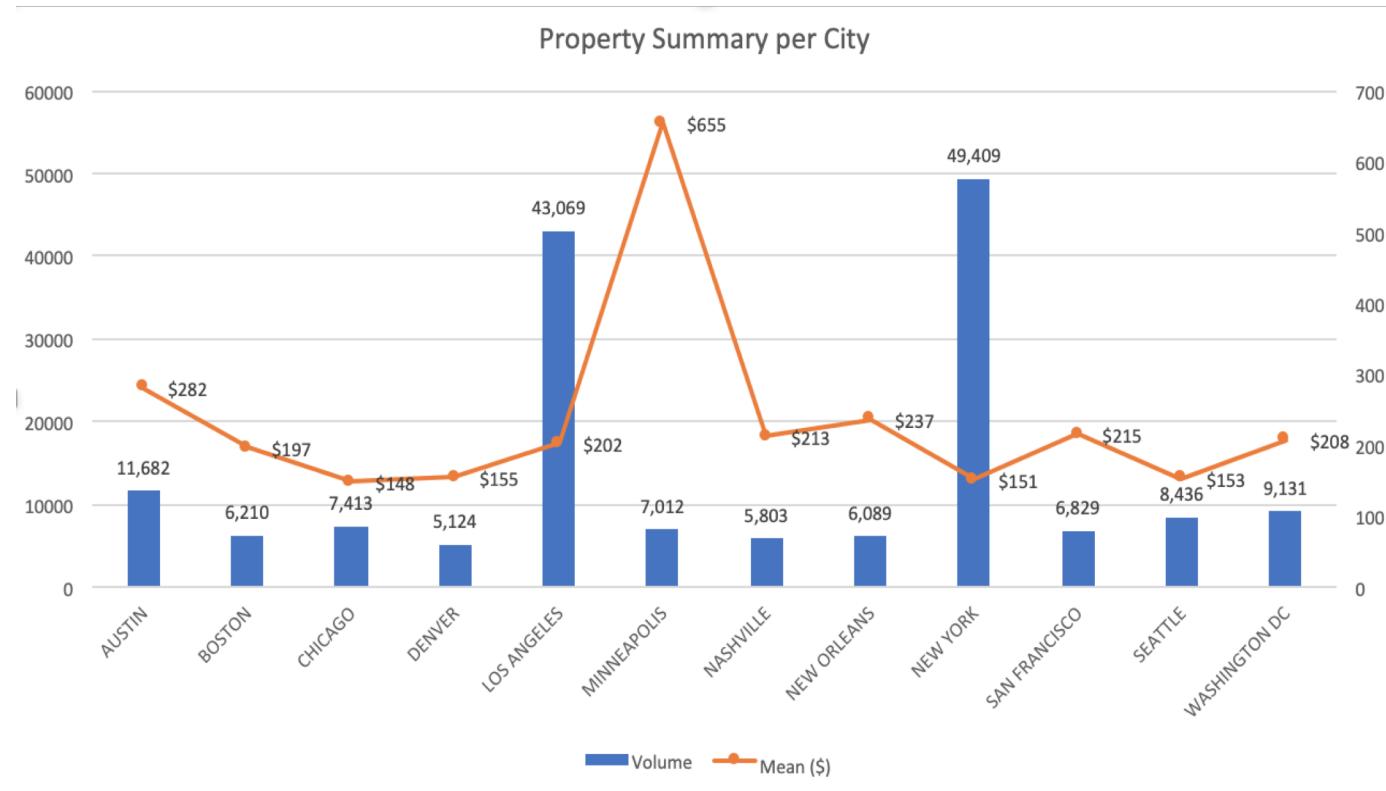
Airbnb Business Model



Data Summary

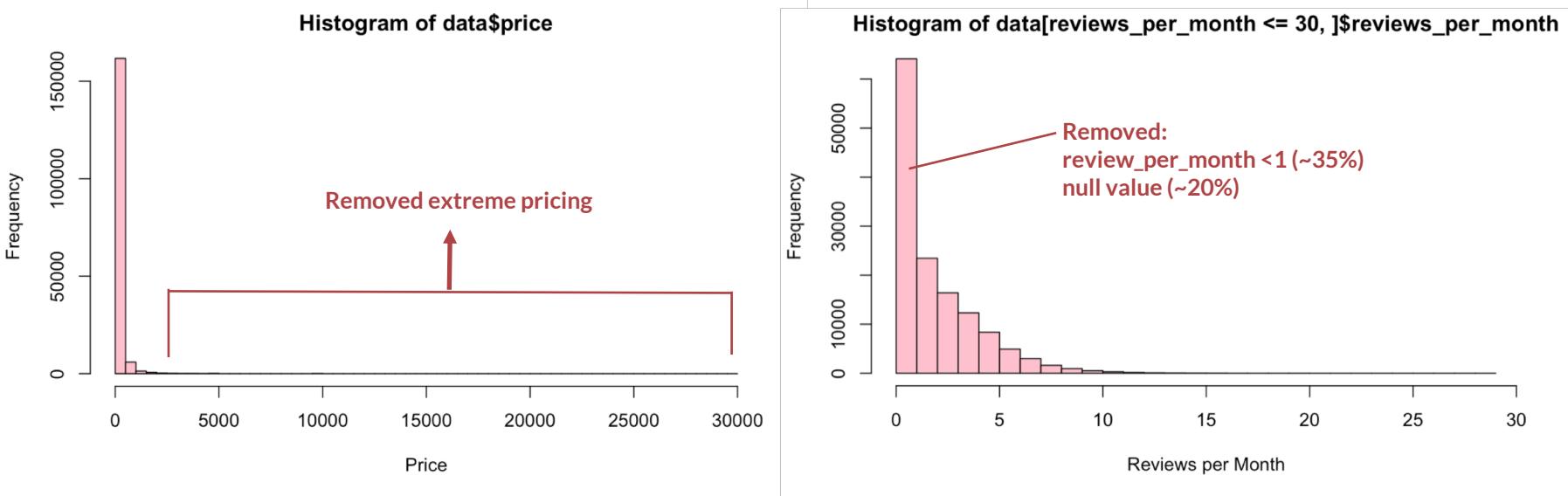
- Overview
 - Statistics
-

Summary Statistics Before Cleaning

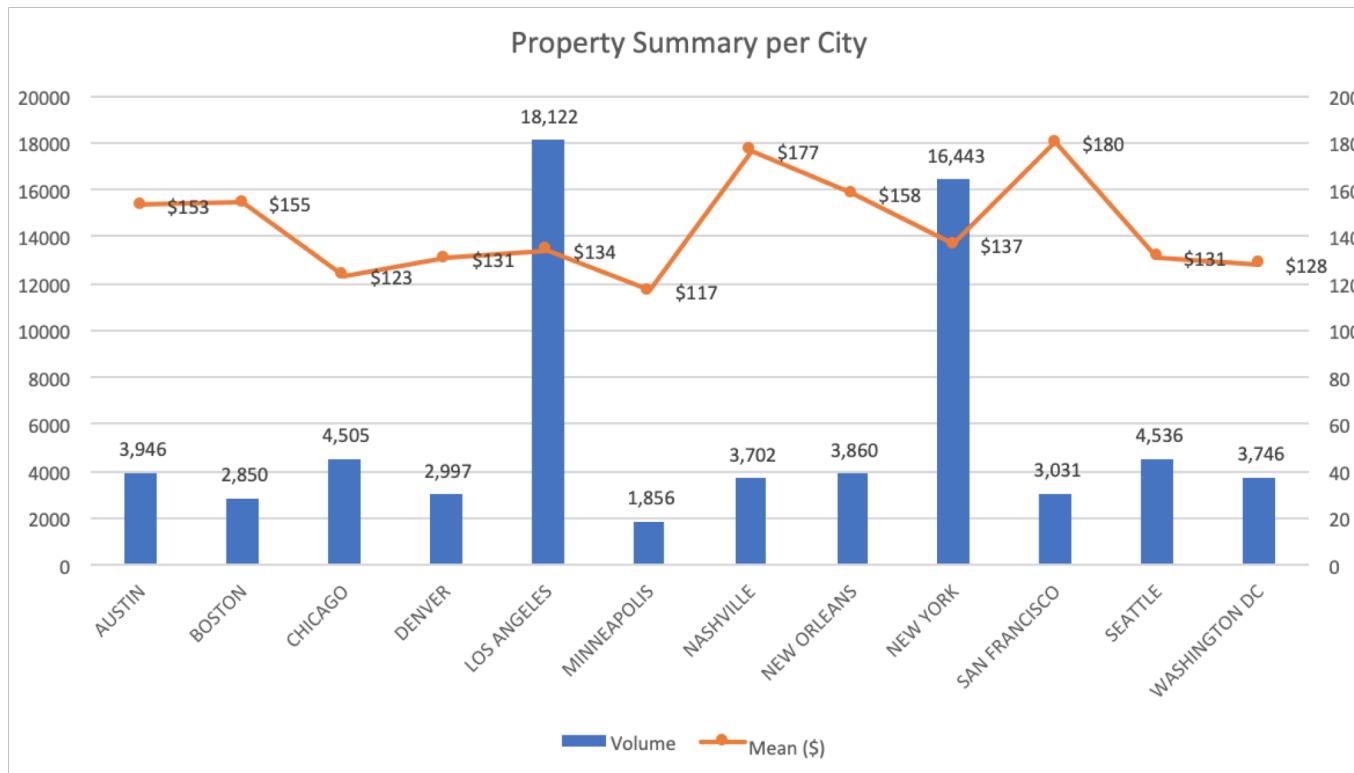


- Data Source: Insideairbnb.com which performs monthly webscrapes of Airbnb data
- Listing information is gathered from 12 major cities across all four census regions in the United States in November 2018
- Total 166,207 observations dominated by Los Angeles (25.91%) and New York (29.73%).
- Overall average of \$234.60/ night and median of \$133.29/night

Modeling Sample Selection

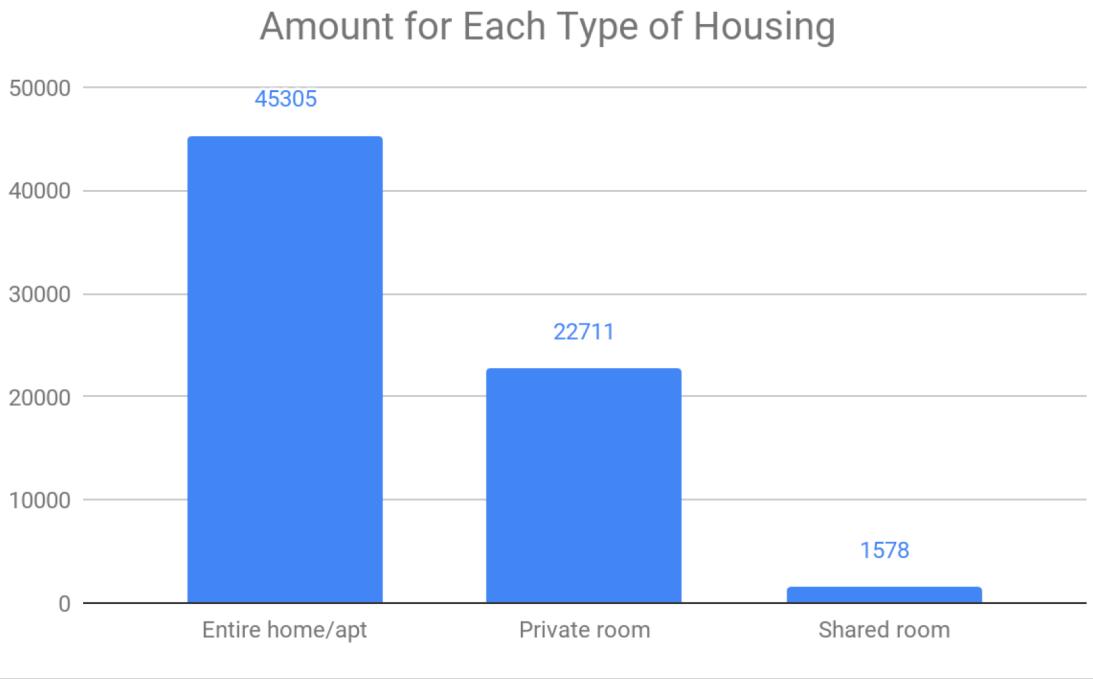


Summary Statistics After Cleaning



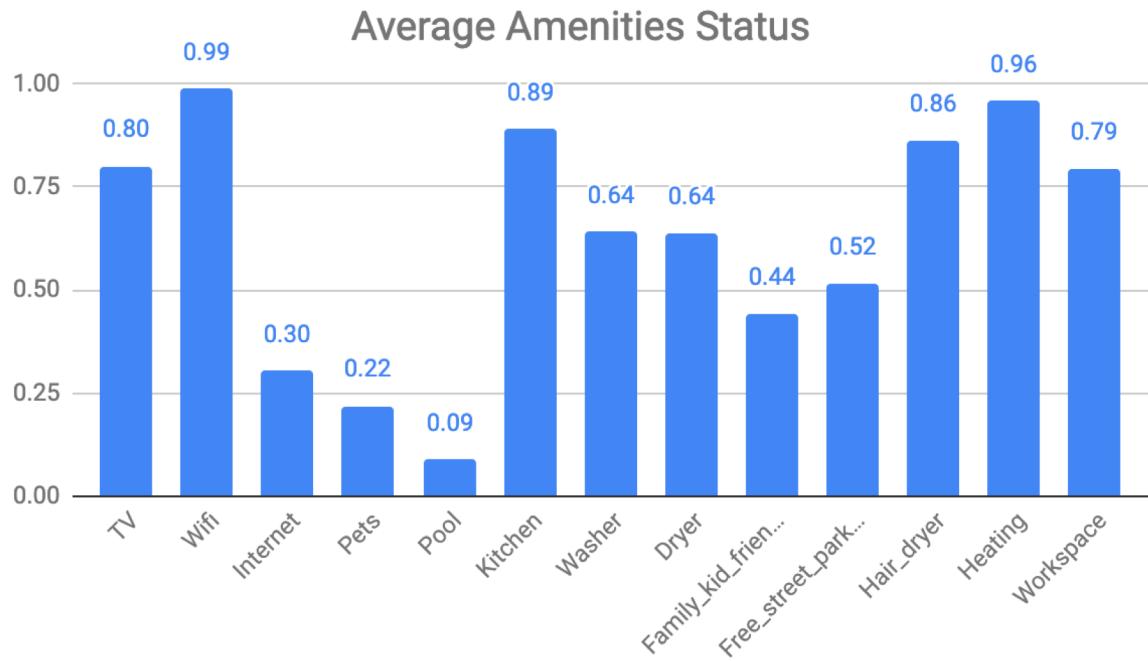
- Removed observation that had less than 1 reviews per month
- Total 69,594 cleaned observations dominated by Los Angeles (26.04%) and New York (23.63%).
- Overall average of \$143.67/ night and median of \$108.83/night

Summary Statistics



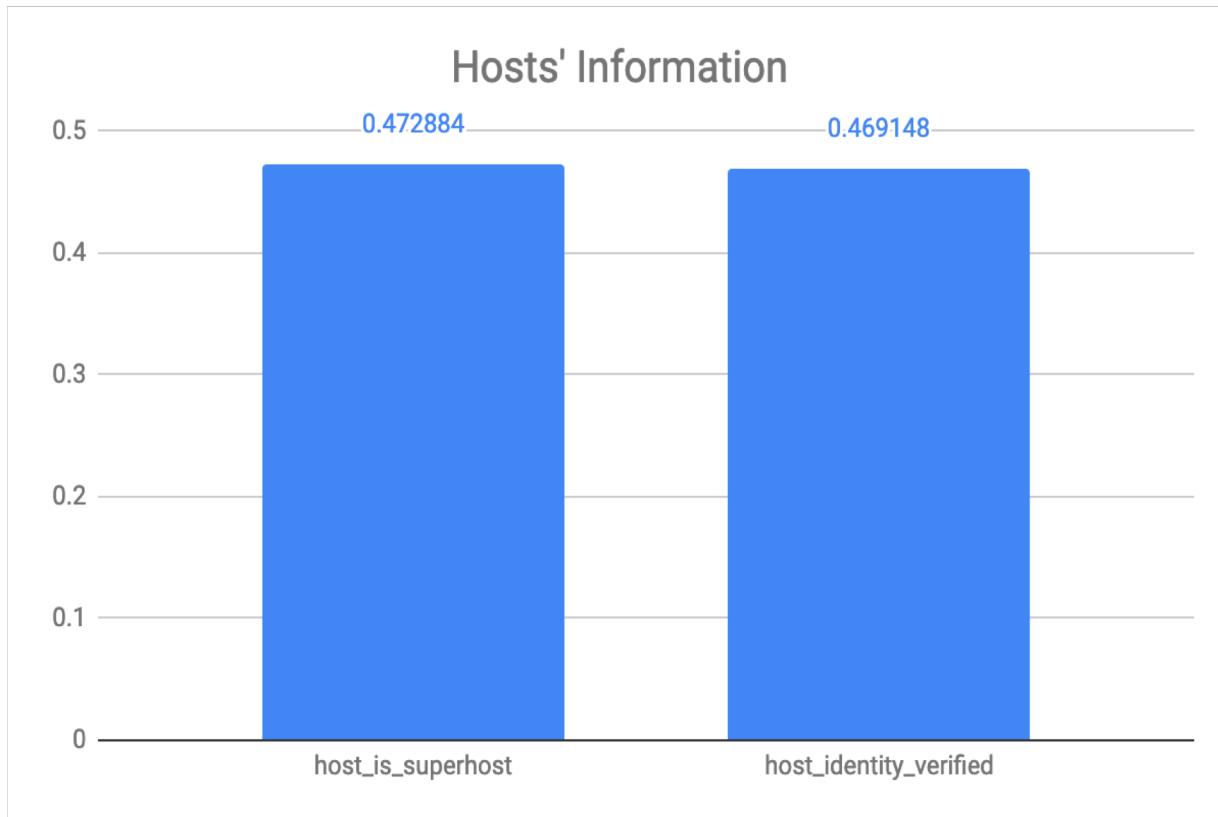
- Most hosts rent out their entire home or apartment (65.1%).

Summary Statistics



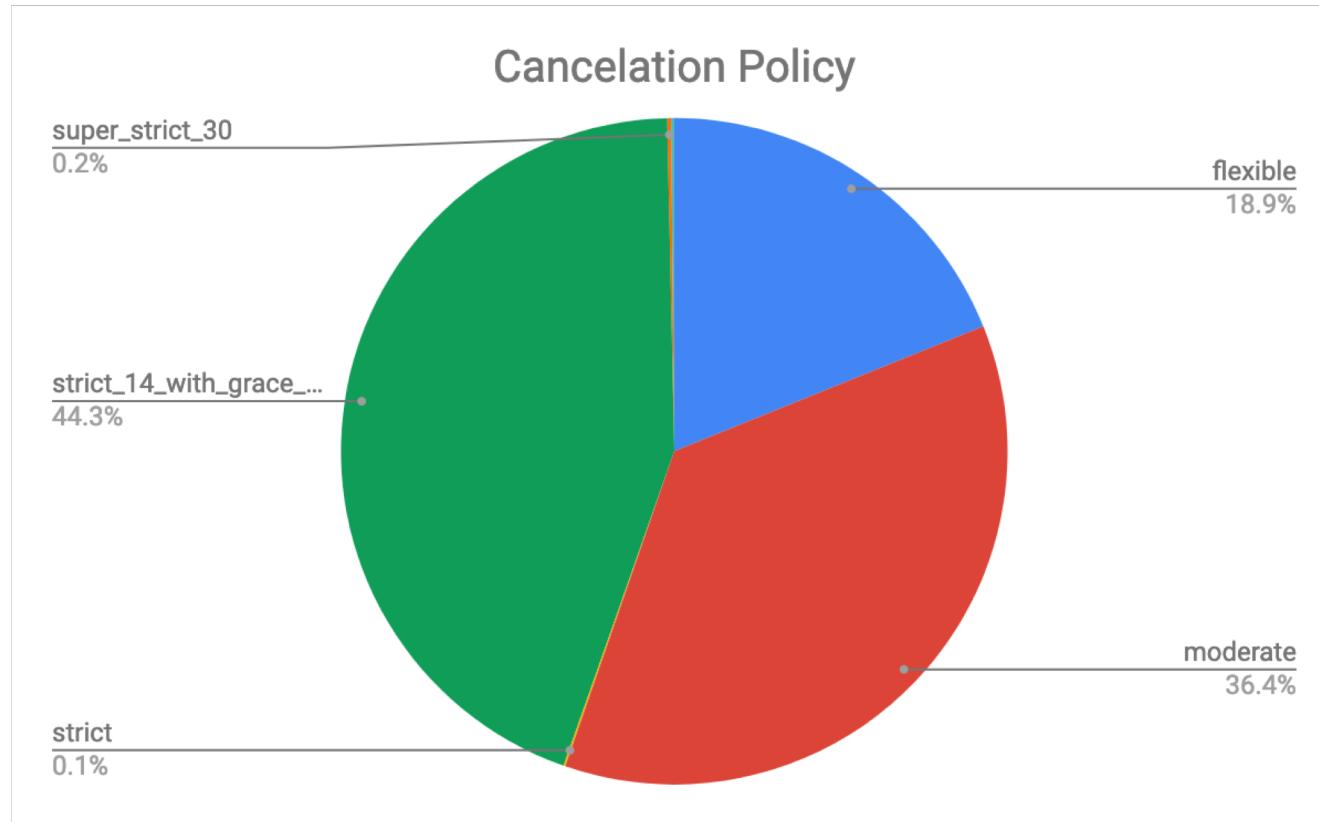
- Kitchen, Heating and Wifi are almost (~90%) guaranteed to be offered in a property.
- Pets Allowance, Pool, Kid Friendly, and Free Street Parking are much more less likely to be offered (~50%)

Hosts' Information

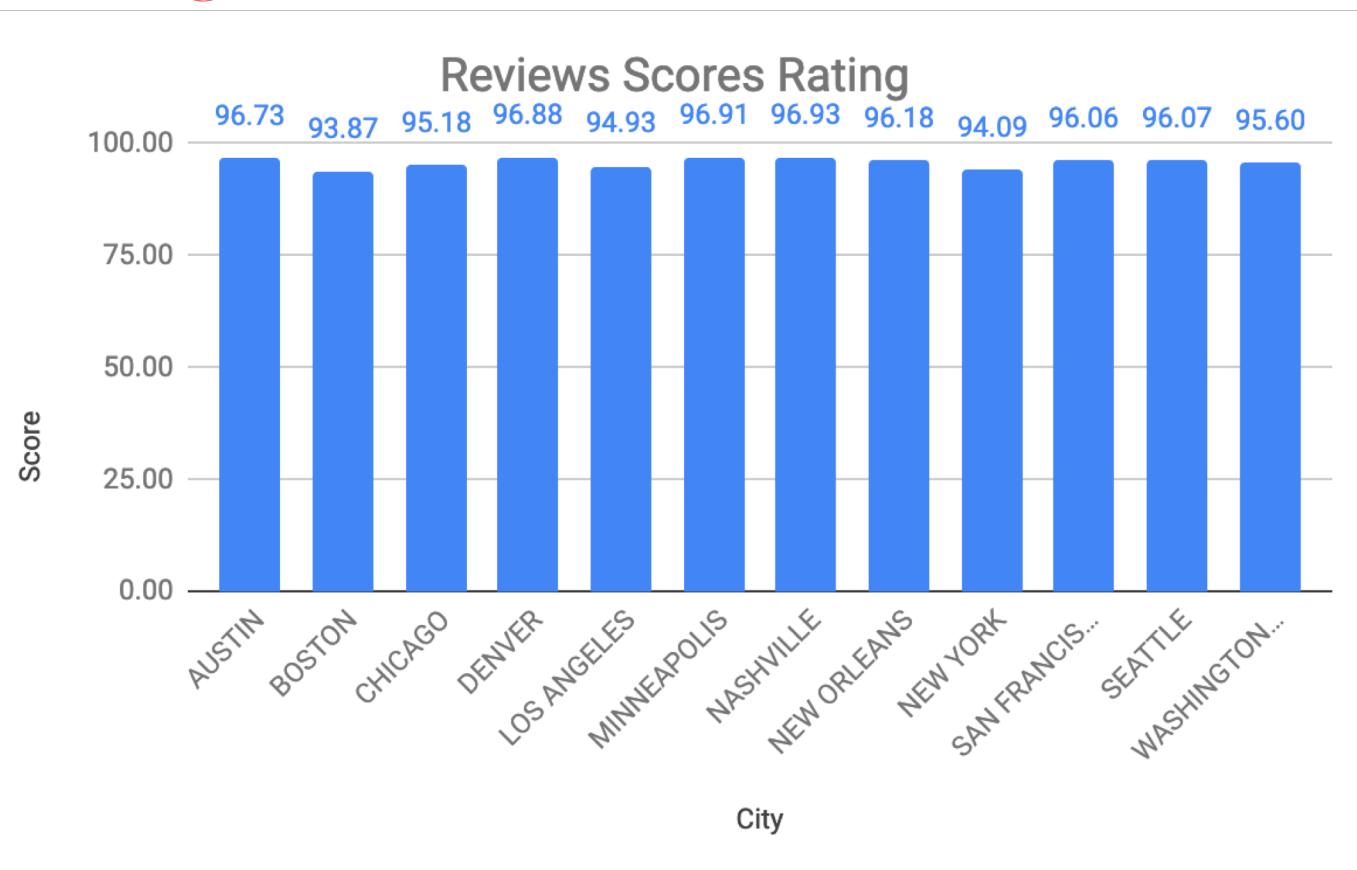


- 47.29% of hosts are superhost
- 46.91% of hosts' identity have been verified

Cancelation Policy



Review Ratings



Modeling & Analysis

- OLS Regression
 - Machine Learning
 - LASSO
 - Random Forest
 - KNN
-

Pricing Factors (OLS regression)

| | Categories | input | Coefficients (exp) | Significance Level | Percentage change in price | Note |
|---|--------------------|--------------------------------|-----------------------|-----------------------|----------------------------------|--|
| 1 | Housing Characters | room_typePrivate room | 0.631 | *** | -36.9% | vs. Entire home/apt vs. Entire home/apt |
| | | room_typeShared room | 0.341 | *** | -65.9% | |
| | | bedrooms | 1.164 | *** | 16.4% | |
| | | bathrooms | 1.102 | *** | 10.2% | |
| | | accommodates | 1.052 | *** | 5.2% | |
| 2 | Host Info | host_is_superhost (False) | 0.744 | | -25.6% | |
| | | host_identity_verified (False) | 0.995 | | -0.5% | |
| 3 | Reviews | review_scores_rating | 1.005 | *** | 0.5% | scale = 100 scale = 10 scale = 10 |
| | | review_scores_cleanliness | 1.027 | *** | 2.7% | |
| | | review_scores_location | 1.046 | *** | 4.6% | |
| 4 | Amenities | Pool | 1.102 | *** | 10.2% | |
| | | TV | 1.083 | *** | 8.3% | |
| | | Heating | 1.045 | *** | 4.5% | |
| | | Washer | 1.034 | * | 3.4% | |
| | | Hair_dryer | 1.023 | *** | 2.3% | |
| | | Workspace | 1.019 | *** | 1.9% | |
| | | Dryer | 1.019 | | 1.9% | |
| | | Family/kid_friendly | 1.016 | *** | 1.6% | |
| | | Pets | 1.006 | | 0.6% | |
| | | Kitchen | 0.993 | | -0.7% | |
| | | Wifi | 0.989 | | -1.1% | |
| | | Observations | 69,486 | | | |
| 5 | | R2 | 70.72% | | | |
| | | Adjusted R2 | 70.21% | | | |
| | | Residual Std. Error | 0.3614 | | *p<0.05; **p<0.01; ***p<0.001 | |

Specification:

$\log(\text{price}) \sim \beta_0 + \beta_1 \text{variables} + \beta_2 \text{zip_code} + \beta_3 \text{zip_code}$



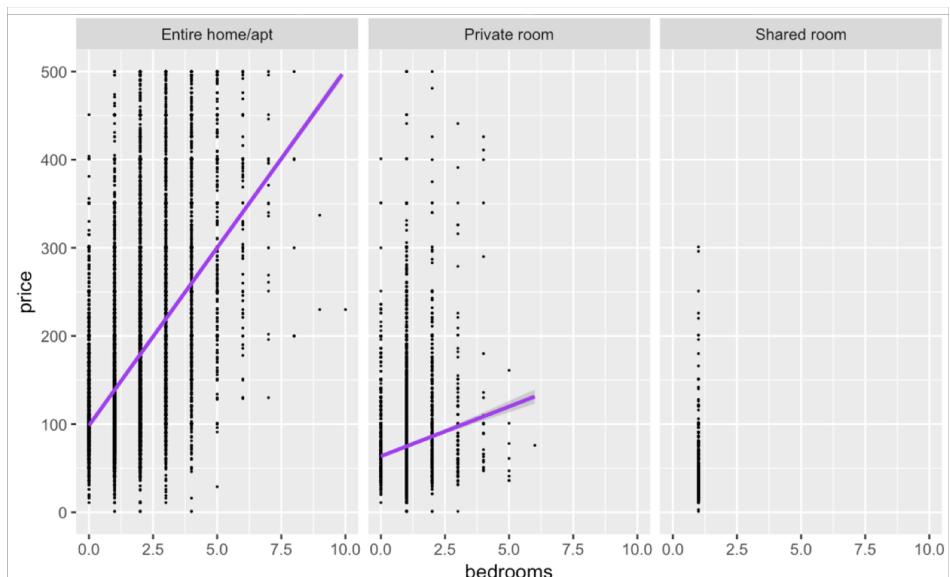
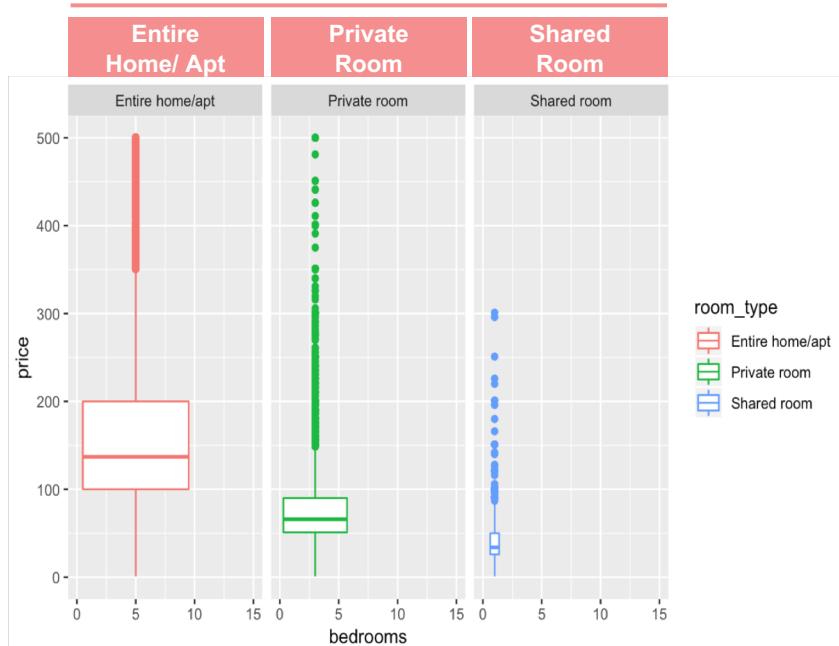
5

Variable Exclusion:

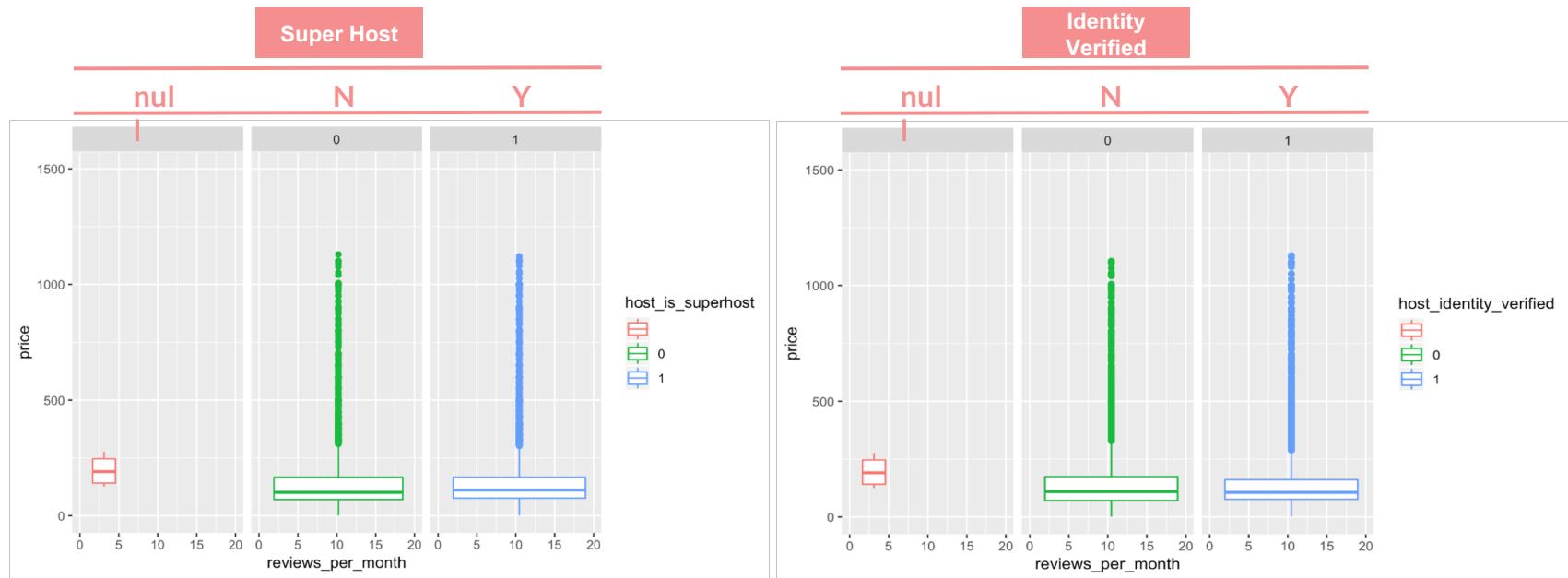
- Cancellation Policy
- Number of Reviews/ Review per Month
- ...

1. Housing Characters

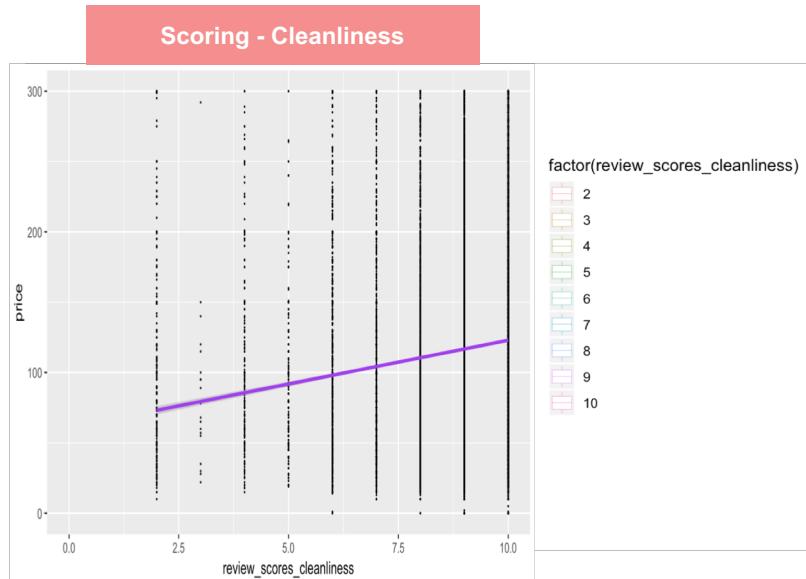
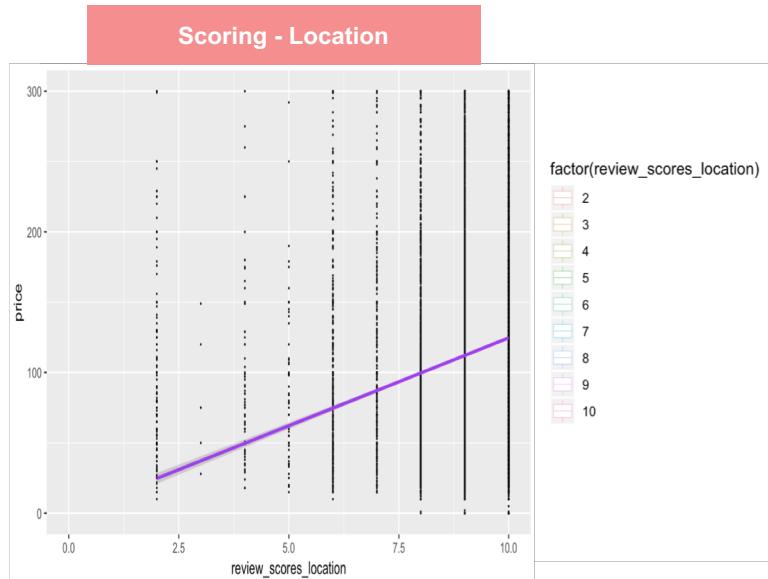
Room Type



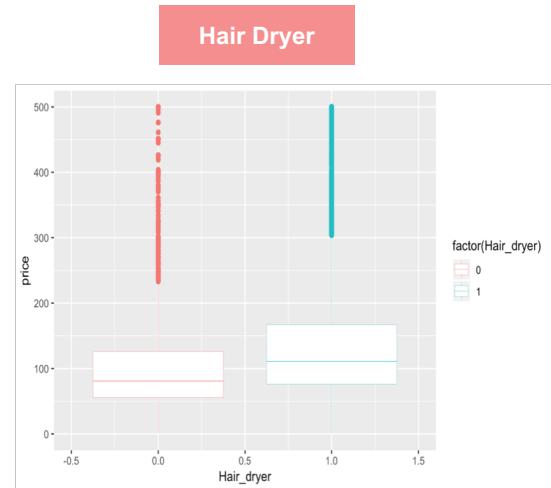
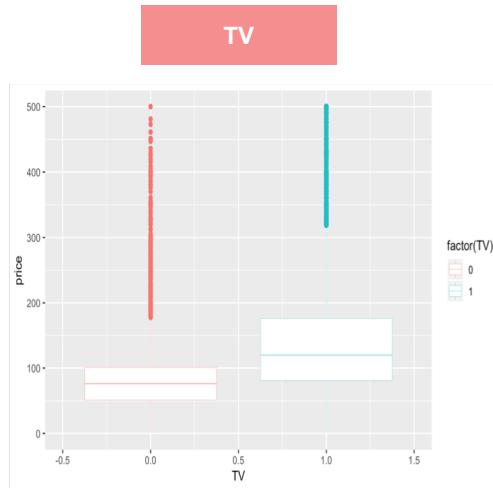
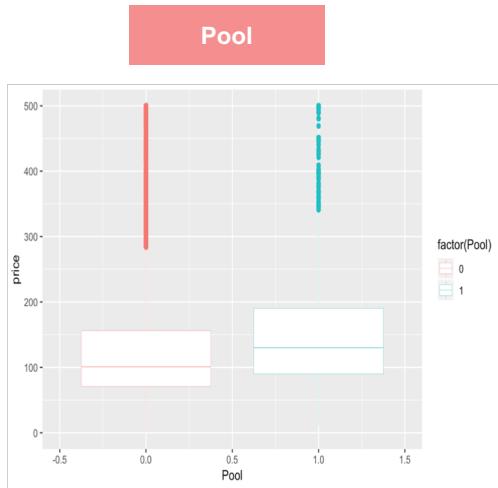
2. Host Info - Super host & Identity verified



3. Reviews - Location & Cleanliness



4. Amenities



N

Y

N

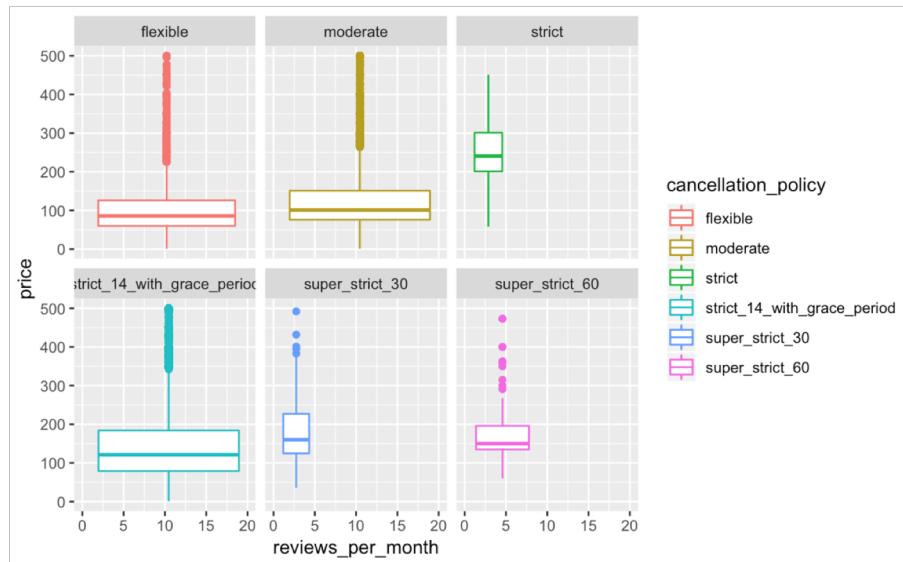
Y

N

Y

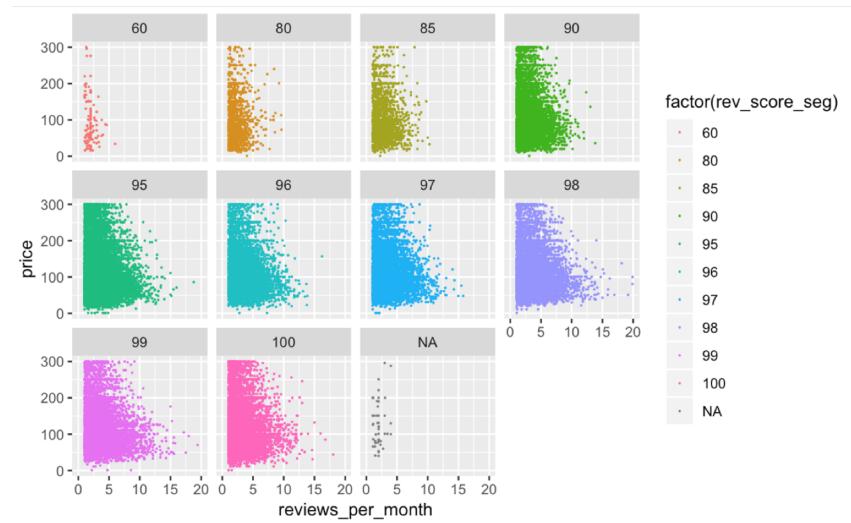
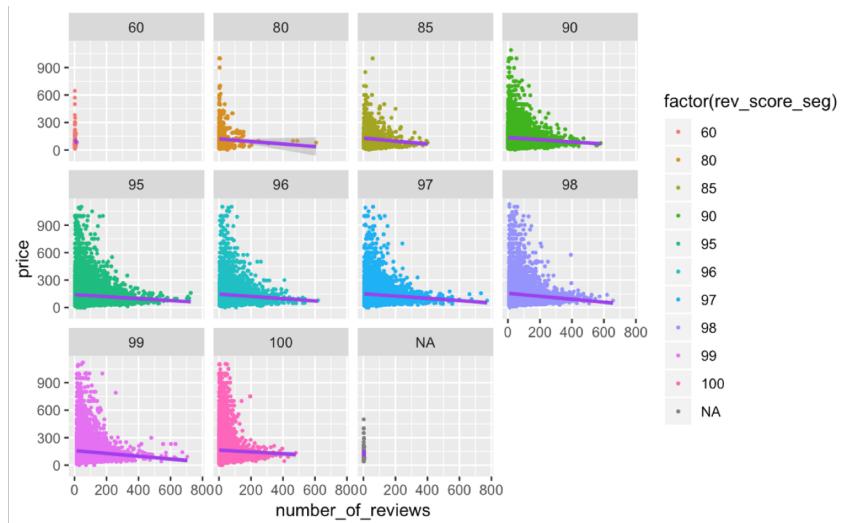
5. Variable Exclusion

- Cancelation Policy
- Number of Reviews/ Review per Month



5. Variable Exclusion

- Cancelation policy
- Number of Reviews / Review per Month



Pricing Factors (OLS regression)

| Categories | input | Coefficients (exp) | Significance Level | Percentage change in price | Note |
|--------------------|--------------------------------|-----------------------|-----------------------|----------------------------------|--|
| Housing Characters | room_typePrivate room | 0.631 | *** | -36.9% | vs. Entire home/apt vs. Entire home/apt |
| | room_typeShared room | 0.341 | *** | -65.9% | |
| | bedrooms | 1.164 | *** | 16.4% | |
| | bathrooms | 1.102 | *** | 10.2% | |
| | accommodates | 1.052 | *** | 5.2% | |
| Host Info | host_is_superhost (False) | 0.744 | | -25.6% | |
| | host_identity_verified (False) | 0.995 | | -0.5% | |
| Reviews | review_scores_rating | 1.005 | *** | 0.5% | scale = 100 |
| | review_scores_cleanliness | 1.027 | *** | 2.7% | scale = 10 |
| | review_scores_location | 1.046 | *** | 4.6% | scale = 10 |
| Amenities | Pool | 1.102 | *** | 10.2% | |
| | TV | 1.083 | *** | 8.3% | |
| | Heating | 1.045 | *** | 4.5% | |
| | Washer | 1.034 | * | 3.4% | |
| | Hair_dryer | 1.023 | *** | 2.3% | |
| | Workspace | 1.019 | *** | 1.9% | |
| | Dryer | 1.019 | | 1.9% | |
| | Family/kid_friendly | 1.016 | *** | 1.6% | |
| | Pets | 1.006 | | 0.6% | |
| | Kitchen | 0.993 | | -0.7% | |
| | Wifi | 0.989 | | -1.1% | |
| | Observations | 69,486 | | | |
| | R2 | 70.72% | | | |
| | Adjusted R2 | 70.21% | | | |
| | Residual Std. Error | 0.3614 | | | *p<0.05; **p<0.01; ***p<0.001 |

Specification:

$\log(\text{price}) \sim \beta \cdot \text{x variables} | \text{zip_code} | 0 | \text{zip_code}$

Housing characters plays the determinant roles on price.

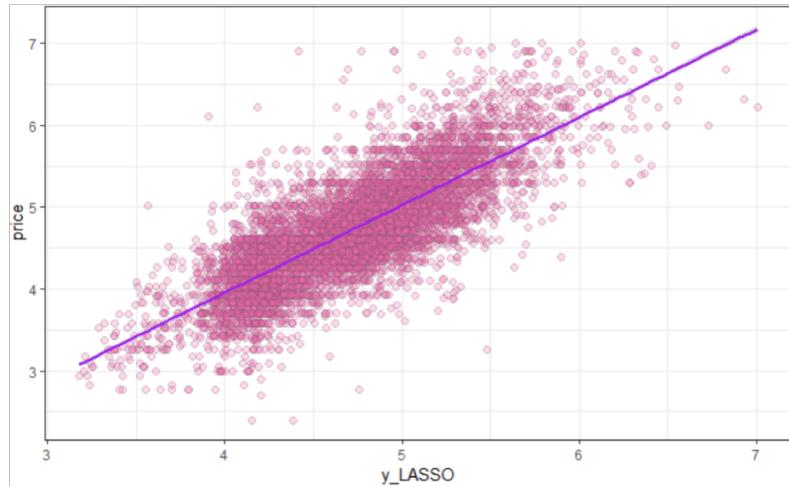
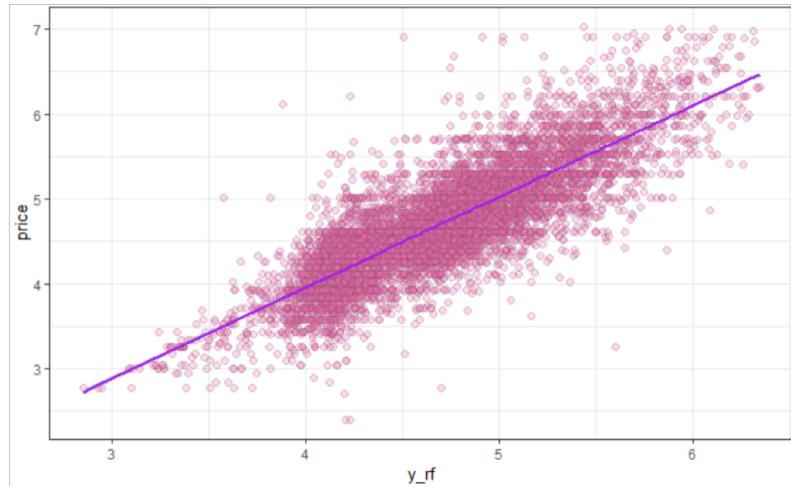
Host Info are not significant.

Location & Cleanliness are important.

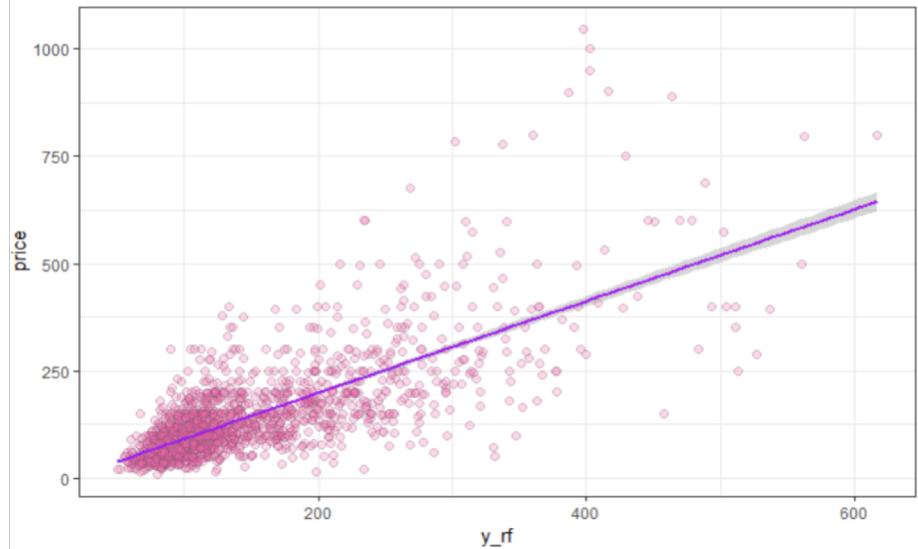
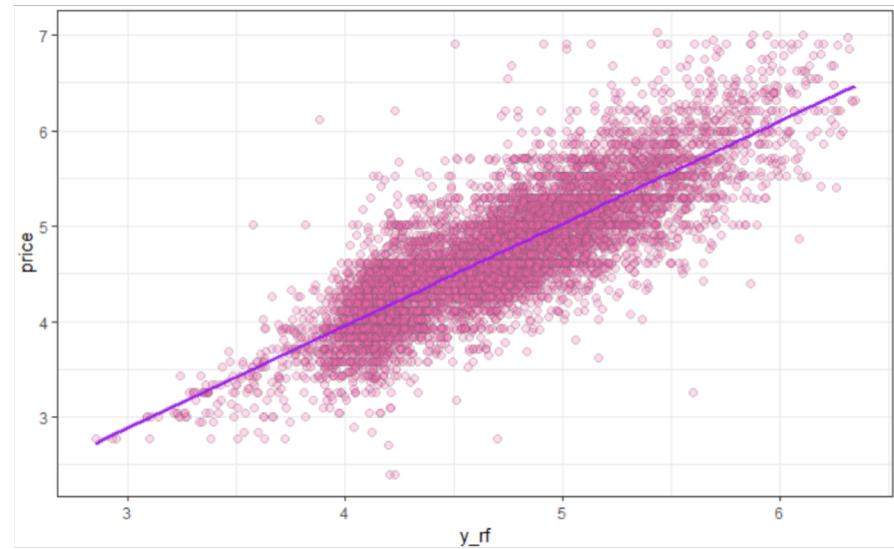
Travelers are willing to pay more for certain amenities (eg. pool, TV, hair dryer, etc.) ; some amenities might be taken for granted (eg. wifi, kitchen)

Machine Learning Results

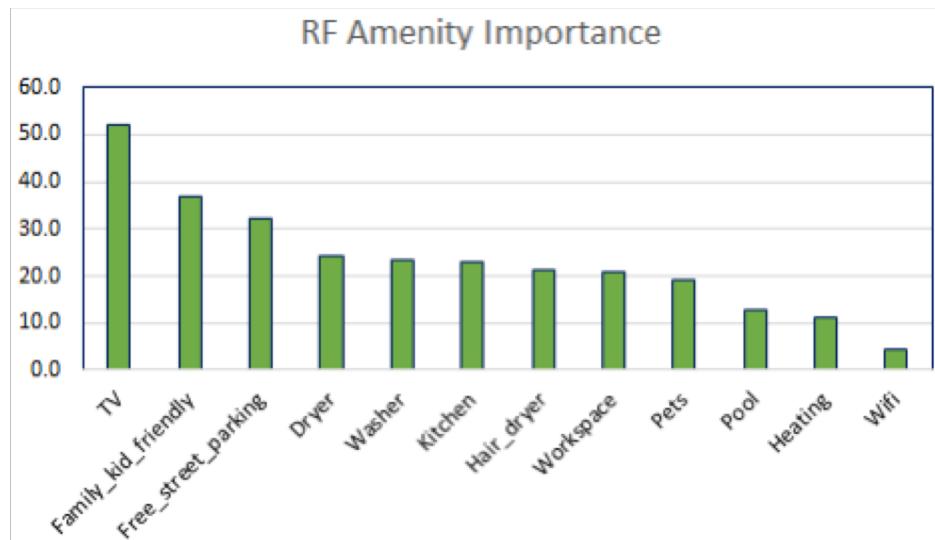
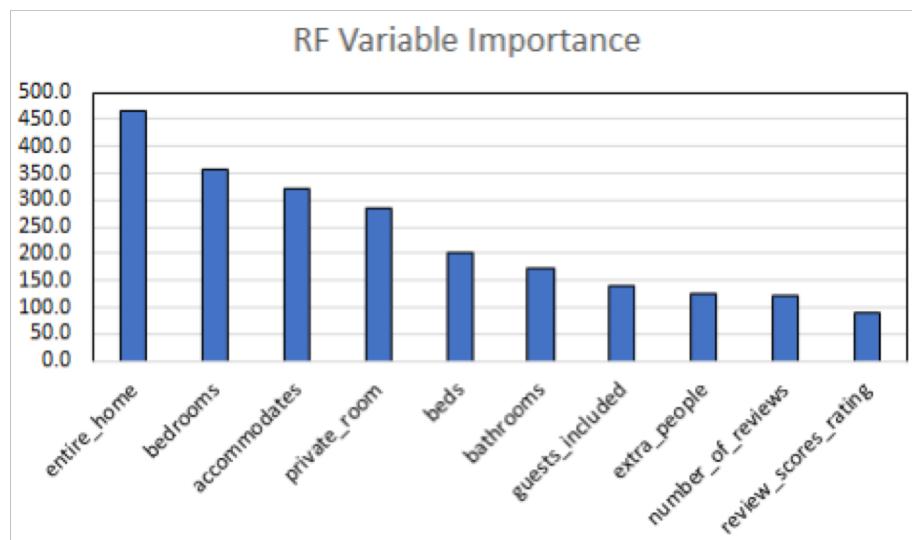
| Method | RF | LASSO |
|---------|---------|----------|
| OOS MSE | 0.13694 | 0.158763 |



Log vs. Not Log Price



Random Forest



K-Nearest Neighbors

- For continuous dependent variable: predicts an average value of the nearest K-observations
- Optimal K = 9
- RMSE of 0.4075 = MSE of 0.166
- Downside: no model output, computationally expensive

```
k-Nearest Neighbors

34733 samples
 49 predictor

Pre-processing: centered (49), scaled (49)
Resampling: Cross-Validated (10 fold, repeated 3 times)
Summary of sample sizes: 31259, 31261, 31259, 31261,
31259, 31261, ...
Resampling results across tuning parameters:

      k    RMSE     Rsquared     MAE
      5   0.4185314  0.6031767  0.3095908
      7   0.4106618  0.6171437  0.3046711
      9   0.4074679  0.6233735  0.3024636

RMSE was used to select the optimal model using
the smallest value.
The final value used for the model was k = 9.
```

Compare ML Lasso to OLS Model

| Manually Specified OLS | LASSO Model | | Manually Specified OLS | LASSO Model | |
|---------------------------|-------------------------------|-------------------------|---------------------------|-------------------------|-------------------------|
| Property | | | Property cont'd | | |
| bedrooms | bedrooms | value score (-) | | communication score | |
| bathrooms | bathrooms | checkin score (-) | | value score | |
| accommodates | accommodates | communication score (-) | | superhost | |
| overall review score | overall review score | | | number of host listings | |
| cleanliness score | cleanliness score | | Amenity | | |
| location score | location score | | hair dryer | hair dryer | Free Street Parking (-) |
| private/shared room (-) | entire home / private room | shared room (-) | heating | heating | |
| | verified host identity | | pool | pool | |
| | extra people | | TV | TV | |
| | guests included | | washer | washer | |
| | check-in score | | workspace | workspace | |
| | BOS/SF/NYC | MINN/DEN/CHI/AUS (-) | family/kid friendly | family/kid friendly | |

Conclusion

- Business Recommendation
 - Limitation & Future Direction
-

Recommendation

Overall

- A better location (4.6%).
- For every one more bedroom (16.4%) and bathrooms (10%).
- Having a pool (10%).

Actionable

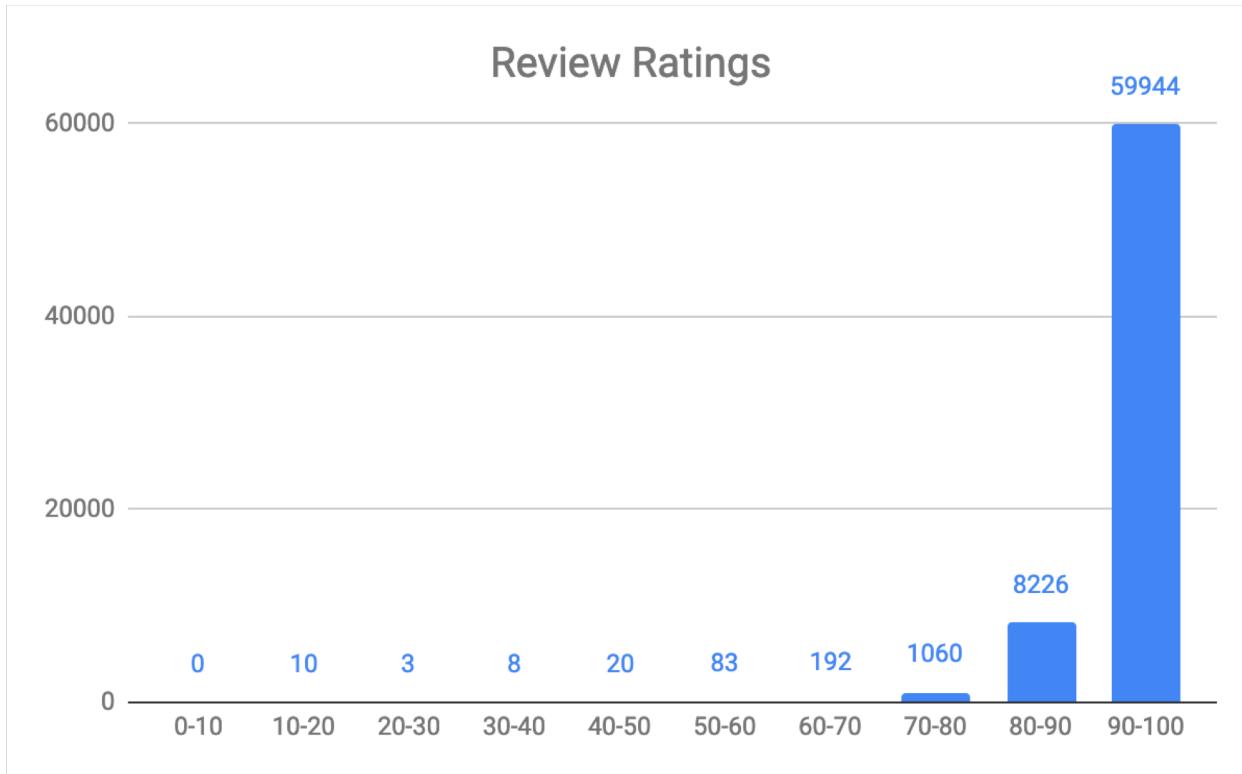
- Increase the accommodation ability, i.e. bunk beds, sofa beds, etc. (>5%).
- Make sure the house is clean (2.3%).
- Offer hair dryer, heating, TV, washer, and workspace (i.e. a nice desk) and open to families and kids (1.5% ~ 8%) per item.
- Renting out the entire house.

Limitations & Future Direction

| Limitation | Future Direction |
|---|---|
| <ul style="list-style-type: none">• Limited skills on text mining and image processing.• Model may not accurately predict high priced listings.• OLS is not a causal model. | <ul style="list-style-type: none">• Texting mining and text analysis on verbal listing & review descriptions.• Image processing on property photos listed on Airbnb app. |

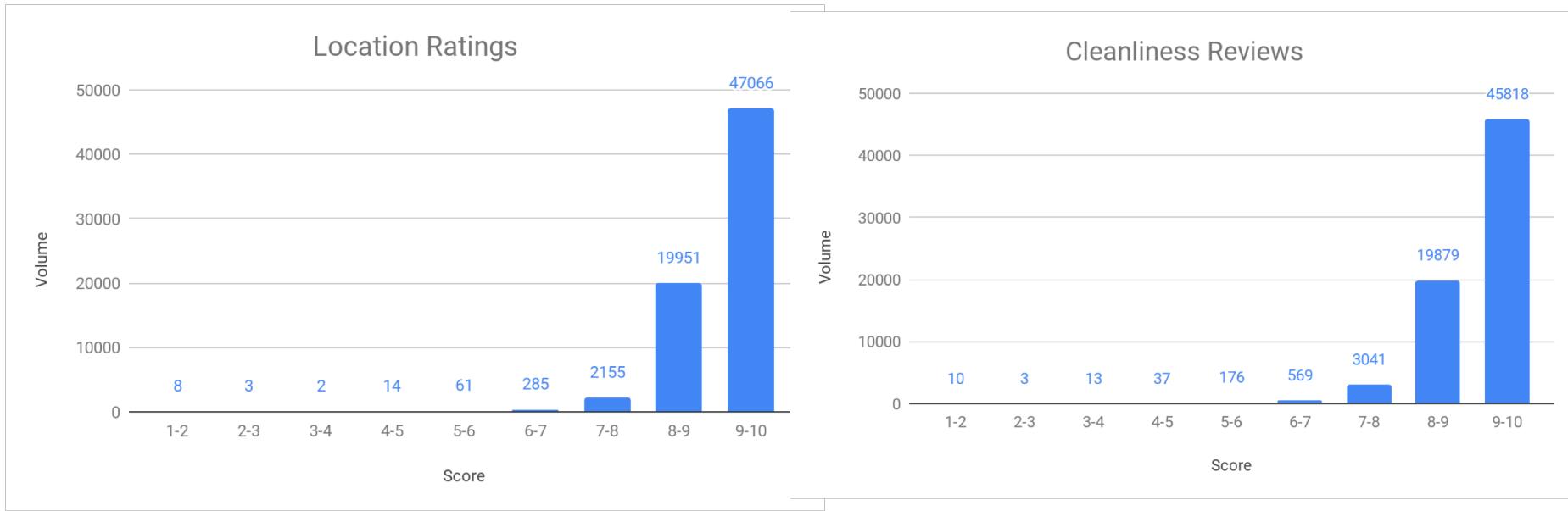
Q&A

Review Ratings



Most Airbnb customer are satisfied, scores are inflated

Location & Cleanliness Score



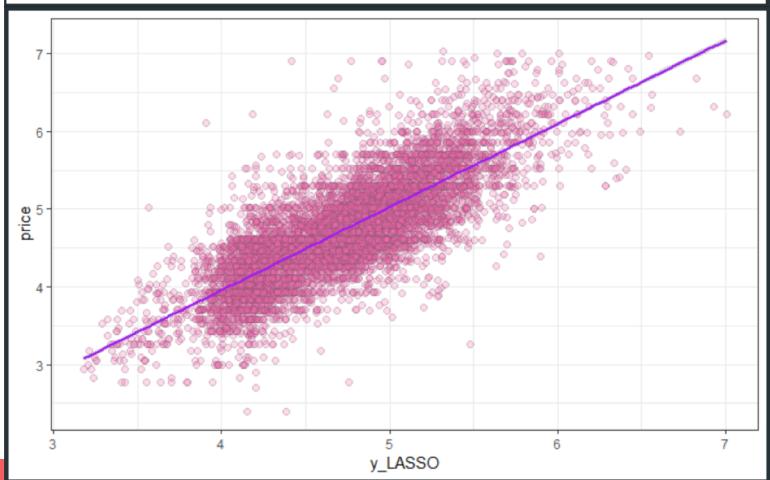
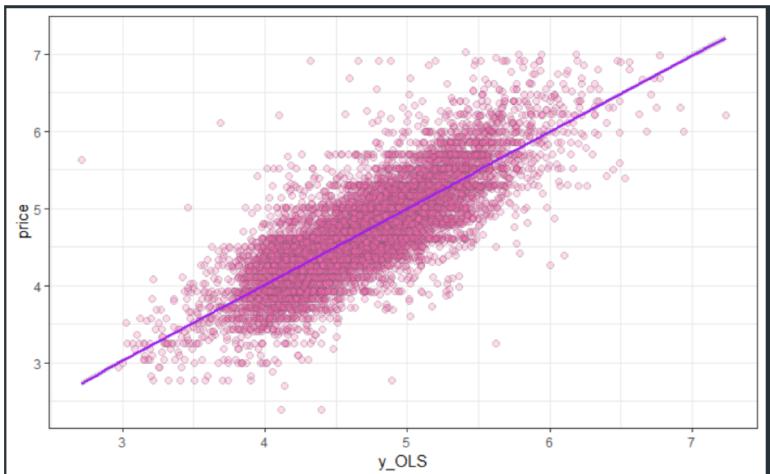
Appendix 3

Cleaned Data

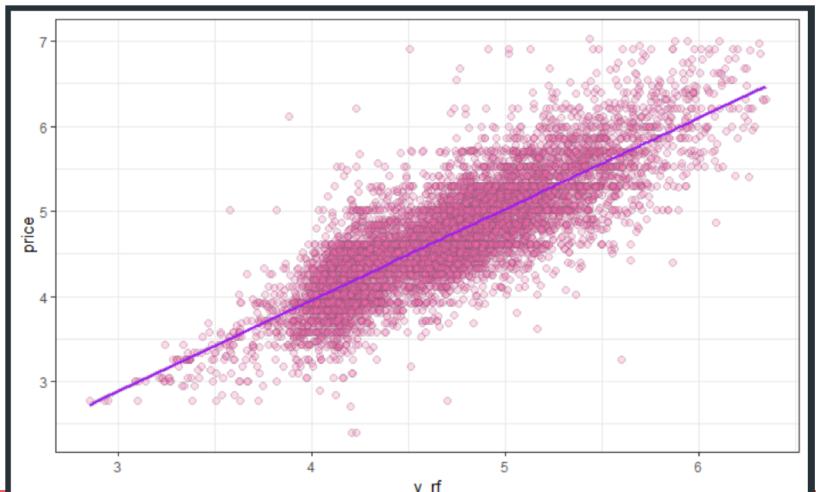
- Most variables have enough data, except for square feet
- Created dummies for the amenities

| Descriptive Statistics | | | | | | | | | |
|--------------------------------|--------|--------------|--------------|------|--------------|--------------|--------------|---------------|--|
| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Median | Pctl(75) | Max | |
| V1 | 69,594 | 82,168.3 | 48,506.4 | 2 | 38,344.2 | 78,119.5 | 127,243.8 | 166,174 | |
| id | 69,594 | 17,901,718.0 | 8,207,060.0 | 360 | 12,451,508 | 19,872,616.0 | 24,487,407.0 | 30,068,089 | |
| host_id | 69,594 | 64,688,062.0 | 63,622,945.0 | 46 | 10,999,790.0 | 40,413,100.0 | 107,463,957 | 225,403,958 | |
| host_response_rate | 52,846 | 1.0 | 0.1 | 0.0 | 1.0 | 1.0 | 1.0 | 1.0 | |
| host_listings_count | 69,590 | 16.2 | 117.4 | 0.0 | 1.0 | 2.0 | 4.0 | 2,491.0 | |
| accommodates | 69,594 | 3.9 | 2.6 | 1 | 2 | 3 | 5 | 40 | |
| bathrooms | 69,551 | 1.3 | 0.6 | 0.0 | 1.0 | 1.0 | 1.0 | 17.0 | |
| bedrooms | 69,578 | 1.4 | 1.0 | 0.0 | 1.0 | 1.0 | 2.0 | 12.0 | |
| beds | 69,574 | 2.1 | 1.7 | 0.0 | 1.0 | 2.0 | 3.0 | 32.0 | |
| square_feet | 580 | 791.8 | 717.9 | 0.0 | 350.0 | 650.0 | 1,000.0 | 7,000.0 | |
| price | 69,594 | 140.4 | 118.2 | 0 | 72 | 105 | 165 | 1,129 | |
| security_deposit | 54,902 | 213.1 | 361.1 | 0.0 | 0.0 | 106.0 | 300.0 | 5,100.0 | |
| cleaning_fee | 63,620 | 65.0 | 53.1 | 0.0 | 25.0 | 50.0 | 90.0 | 750.0 | |
| guests_included | 69,594 | 2.1 | 1.8 | 1 | 1 | 1 | 2 | 20 | |
| extra_people | 69,594 | 16.0 | 23.5 | 0 | 0 | 10 | 25 | 300 | |
| minimum_nights | 69,594 | 2.7 | 15.4 | 1 | 1 | 2 | 2 | 3,000 | |
| maximum_nights | 69,594 | 34,526.0 | 8,158,076.0 | 1 | 30 | 1,000 | 1,125 | 2,147,483,647 | |
| number_of_reviews | 69,594 | 59.5 | 66.2 | 2 | 15 | 38 | 80 | 773 | |
| review_scores_rating | 69,546 | 95.3 | 5.1 | 20.0 | 93.0 | 97.0 | 99.0 | 100.0 | |
| review_scores_accuracy | 69,546 | 9.8 | 0.5 | 2.0 | 10.0 | 10.0 | 10.0 | 10.0 | |
| review_scores_cleanliness | 69,546 | 9.6 | 0.7 | 2.0 | 9.0 | 10.0 | 10.0 | 10.0 | |
| review_scores_checkin | 69,545 | 9.9 | 0.4 | 2.0 | 10.0 | 10.0 | 10.0 | 10.0 | |
| review_scores_communication | 69,546 | 9.9 | 0.4 | 2.0 | 10.0 | 10.0 | 10.0 | 10.0 | |
| review_scores_location | 69,545 | 9.6 | 0.6 | 2.0 | 9.0 | 10.0 | 10.0 | 10.0 | |
| review_scores_value | 69,545 | 9.6 | 0.6 | 2.0 | 9.0 | 10.0 | 10.0 | 10.0 | |
| calculated_host_listings_count | 69,594 | 6.7 | 24.8 | 1 | 1 | 2 | 4 | 323 | |
| reviews_per_month | 69,594 | 3.3 | 2.1 | 1.0 | 1.8 | 2.8 | 4.2 | 191.3 | |
| TV | 69,594 | 0.8 | 0.4 | 0 | 1 | 1 | 1 | 1 | |
| Wifi | 69,594 | 1.0 | 0.1 | 0 | 1 | 1 | 1 | 1 | |
| Internet | 69,594 | 0.3 | 0.5 | 0 | 0 | 0 | 1 | 1 | |
| Pets | 69,594 | 0.2 | 0.4 | 0 | 0 | 0 | 0 | 1 | |
| Pool | 69,594 | 0.1 | 0.3 | 0 | 0 | 0 | 0 | 1 | |
| Kitchen | 69,594 | 0.9 | 0.3 | 0 | 1 | 1 | 1 | 1 | |
| Washer | 69,594 | 0.6 | 0.5 | 0 | 0 | 1 | 1 | 1 | |
| Dryer | 69,594 | 0.6 | 0.5 | 0 | 0 | 1 | 1 | 1 | |
| Family_kid_friendly | 69,594 | 0.4 | 0.5 | 0 | 0 | 0 | 1 | 1 | |
| Free_street_parking | 69,594 | 0.5 | 0.5 | 0 | 0 | 1 | 1 | 1 | |
| Hair_dryer | 69,594 | 0.9 | 0.3 | 0 | 1 | 1 | 1 | 1 | |
| Heating | 69,594 | 1.0 | 0.2 | 0 | 1 | 1 | 1 | 1 | |
| Workspace | 69,594 | 0.8 | 0.4 | 0 | 1 | 1 | 1 | 1 | |
| estimated_occupancy_per_month | 69,594 | 6.5 | 4.2 | 2.0 | 3.6 | 5.6 | 8.4 | 382.6 | |

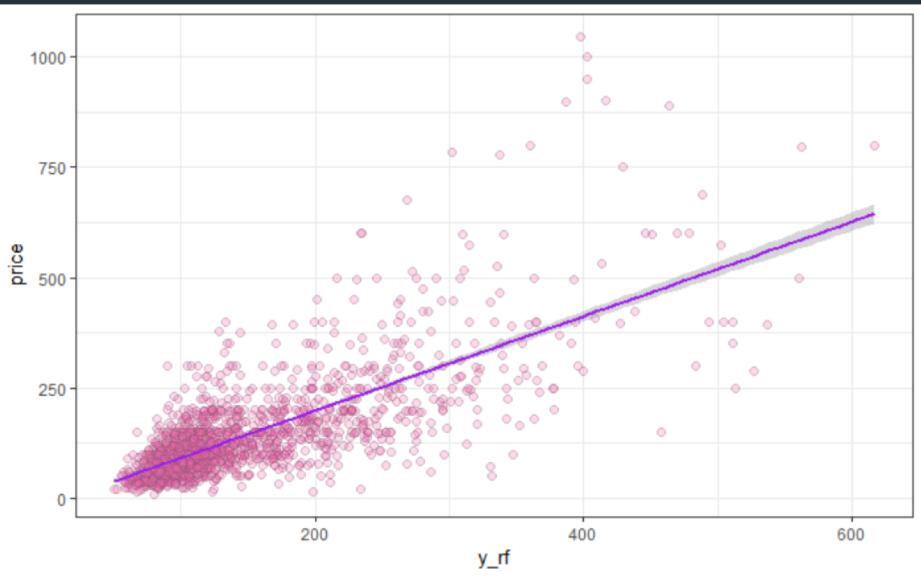
Machine Learning Results



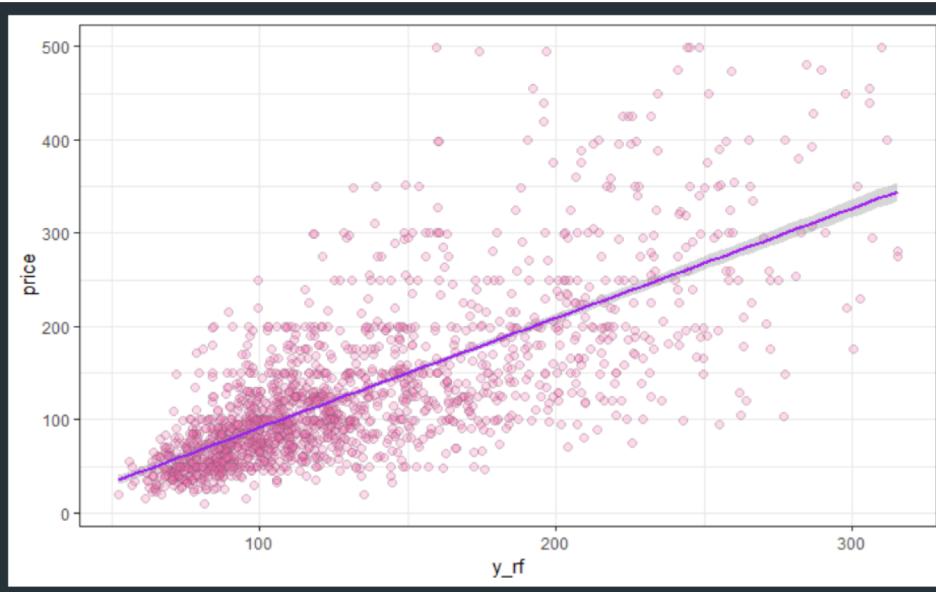
| Method | RF | OLS | LASSO |
|---------|---------|----------|----------|
| OOS MSE | 0.13694 | 0.153274 | 0.158763 |



Random Forest Model



Without Subset: MSE = 5164.36



Subset Prices < 500: MSE = 2778.17