Group Project:
Predicting AirBnB Price in San Francisco

Group A8:
Aditi Lapasia, Diego Alfaro Legonia, Edgar Pon, Ivy Liu, Teen Ho Nicholas Ching, Roberto Keaney, Susan Wang

Executive Summary

(1) Examining the raw dataset and extracting meaningful observations from it
(2) Our summary statistics for variables of interest and how we chose them
(3) Superhosts and verified hosts as indicators of price
(4) Correlation matrix and key observations
(5) Summary table and comparing our models
(6) Final model

After conducting our analysis, what we found to be the strongest indicators of Airbnb prices in San Francisco were: bedrooms, property type, and the number of reviews. This can be quantitatively shown by the very low p-values in our Final Model, which we shall see later. This is consistent with what we may reasonably have expected: the more the number of bedrooms, and the more luxurious the property type the higher the price, and the greater the number of reviews, the lower the price.

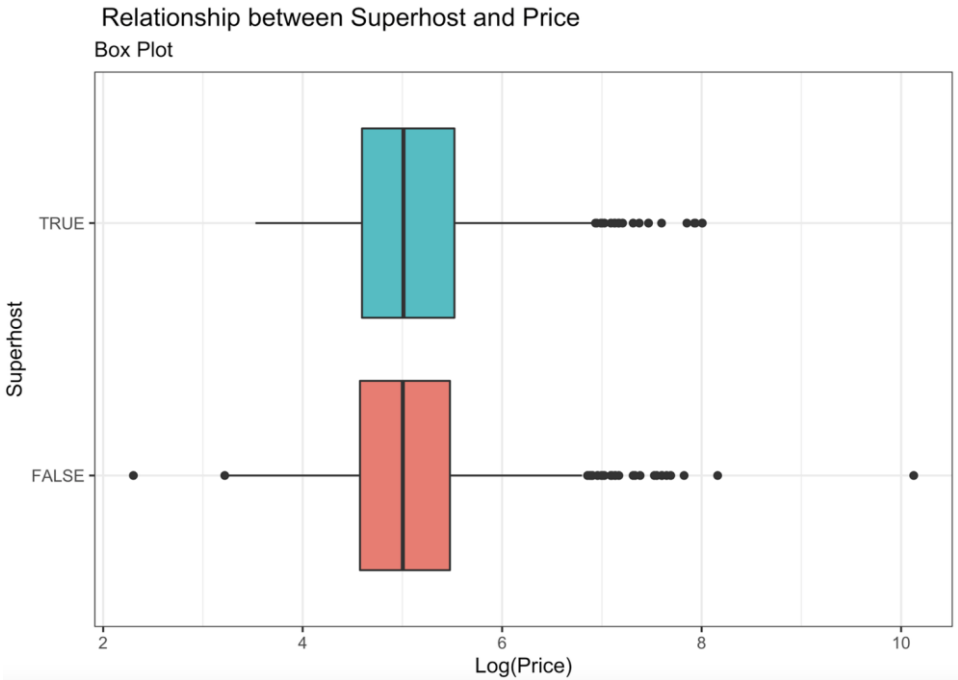
Examining the raw dataset and extracting meaningful observations from it

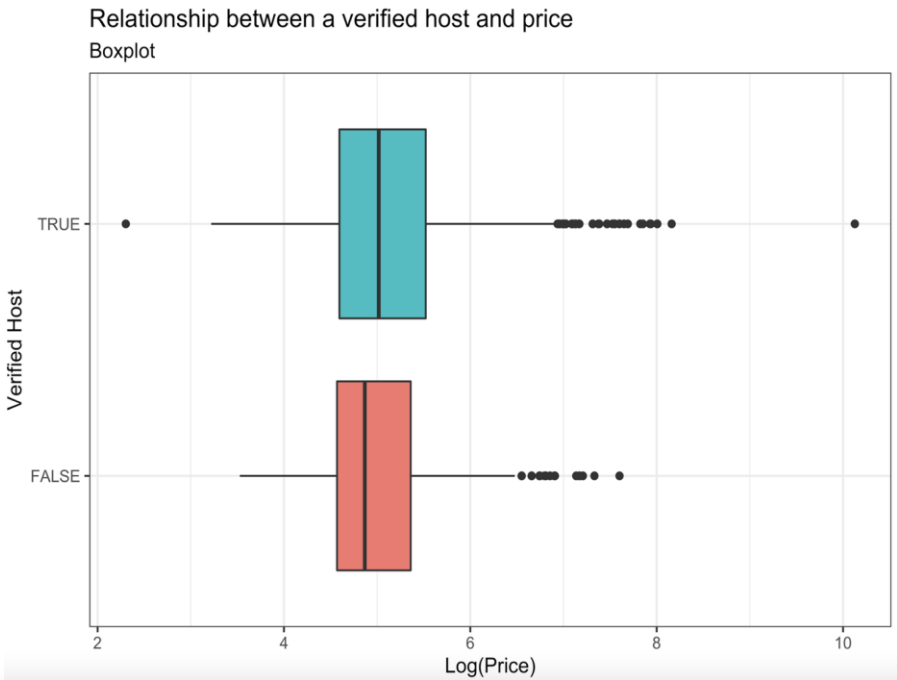
- The importance of categorical/factor variables
- Specific variables of interest to conduct the regression analysis:
    - accommodates, bedrooms, beds, number_of_reviews, review_scores_rating, review_scores_value, minimum_nights, and maximum_nights

Data summary

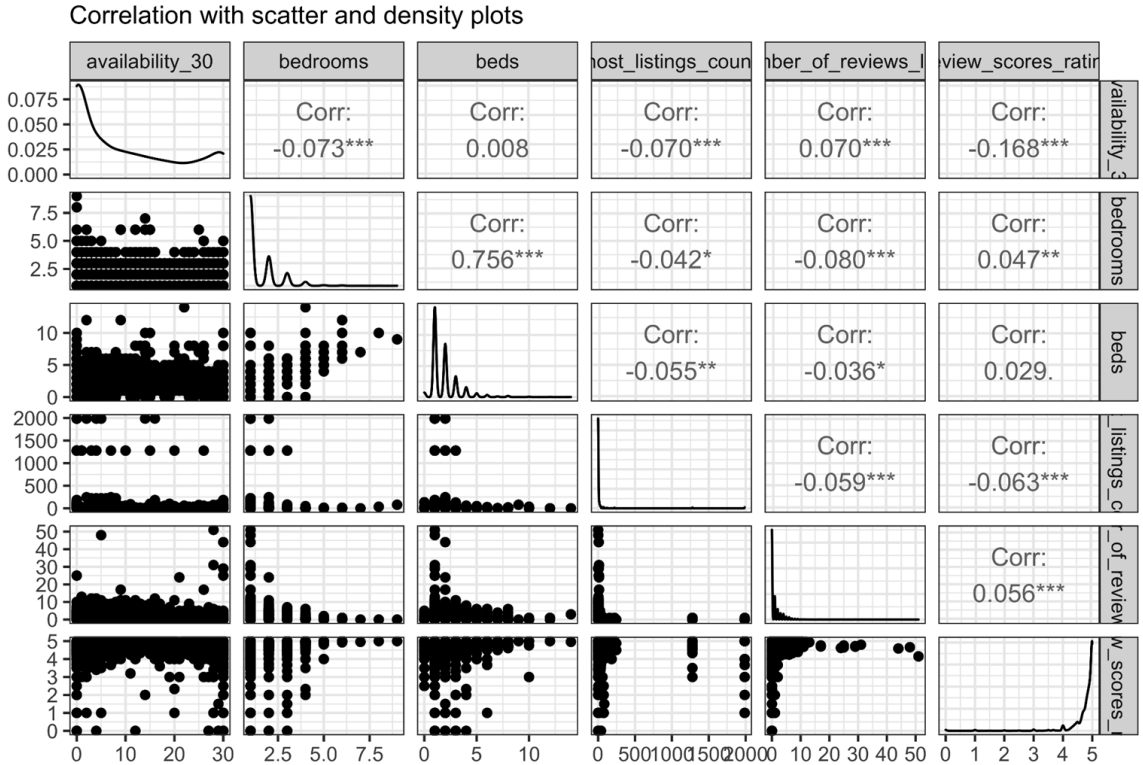
Name	listings
Number of rows	6566
Number of columns	74
Column type frequency:	
character	24
Date	5
logical	8
numeric	37

Superhosts and verified hosts





Relationship between selected variables: correlation matrix



The final model

- Model 1: included property type, number of reviews, review scores rating
- Model 2: added room type (Hotel, Private, Shared) to model 1
- Model 3: included bathrooms number, bedrooms, beds, and accommodates
- Combined Model: included variables in the 3 models, and added availability, reviews per month, neighborhood, instant bookable, superhost, and number of review l30d
- Final model: based on model 2, we added bedrooms, availability, neighbourhood, and instant bookable
- Log(price_4_nights)=6.20 + 0.0034*availability_30 + 0.13*neighbourhood_simplified_Prime - 0.13*instant_bookable_TRUE + 0.32*bedrooms - 0.16* proptype_ Entire rental unit - 0.18* proptype_ Entire residential home - 0.34* proptype_ Other - 0.63*proptype_ Private room in residential home + 0.0008*number_of_reviews + 0.08*review_scores_rating + 0.33*room_typeHotel room - 0.20*roomtype_Private room -1.37*roomtype_Shared room + error

Comparison of models					
	Model 1	Model 2	Model 3	Combined Model	Final Model
Number of observations	1,631	1,631	1,606	1,606	1,631
Adj. R Squared	0.416	0.497	0.390	0.590	0.600
Residual SE	0.494	0.458	0.493	0.405	0.409

Thank you for listening

Group A8