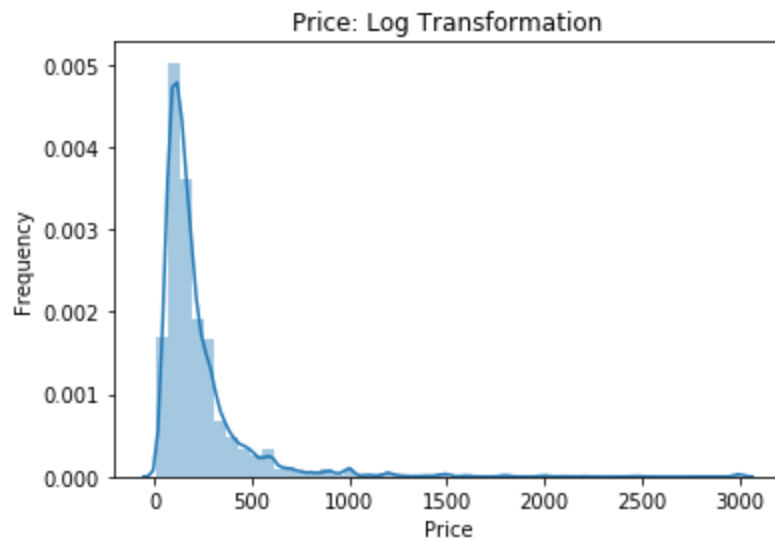
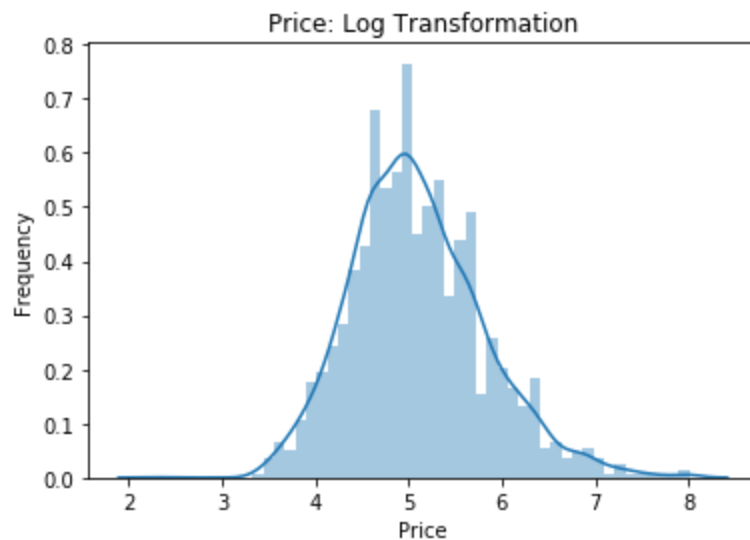


# Exploratory Data Analysis

## Price



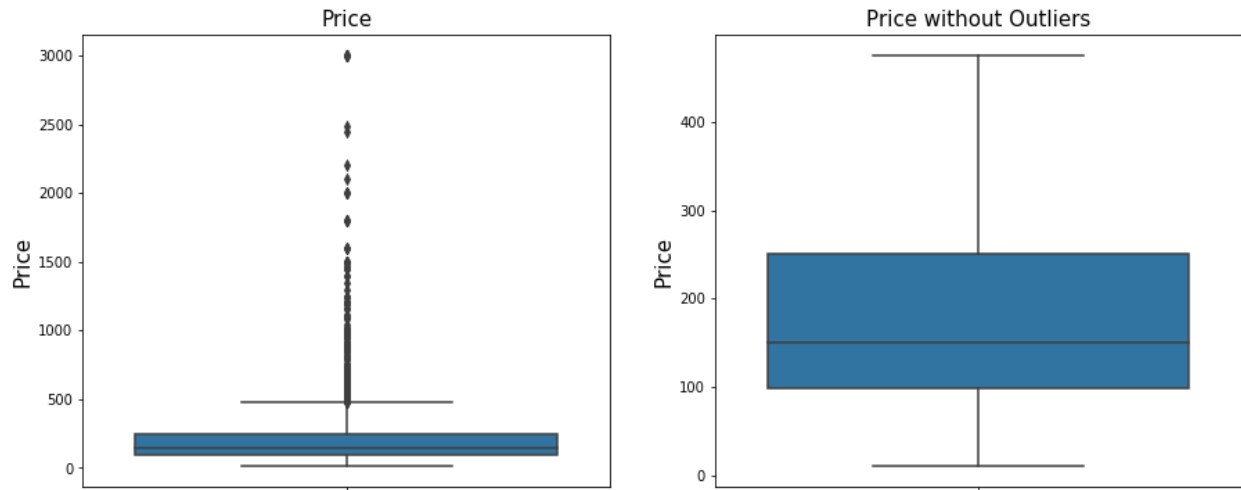
The price histogram shows strongly right-skewed data with a long tail. The majority of prices are very low in comparison to the few high-priced Airbnbs which skew the data.



When the price is log-transformed, it becomes much more normally distributed.

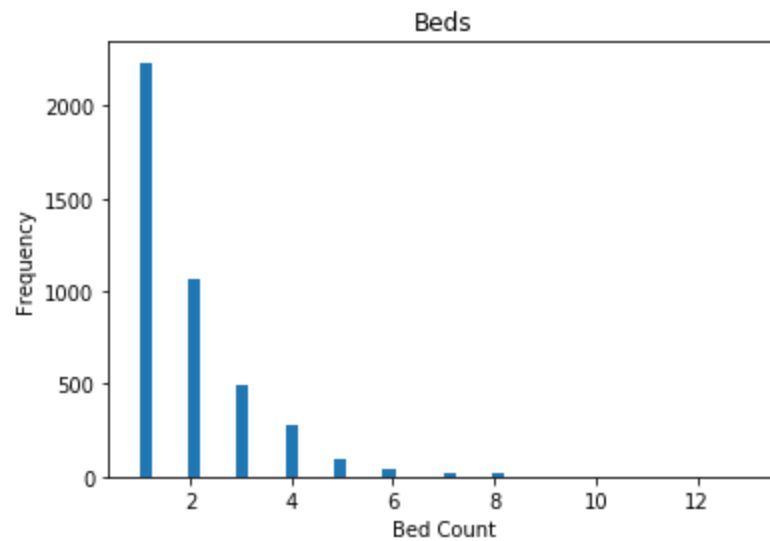
Normality test on logged price:

NormaltestResult(statistic=233.08008658455662, pvalue=2.43950817269063e-51)

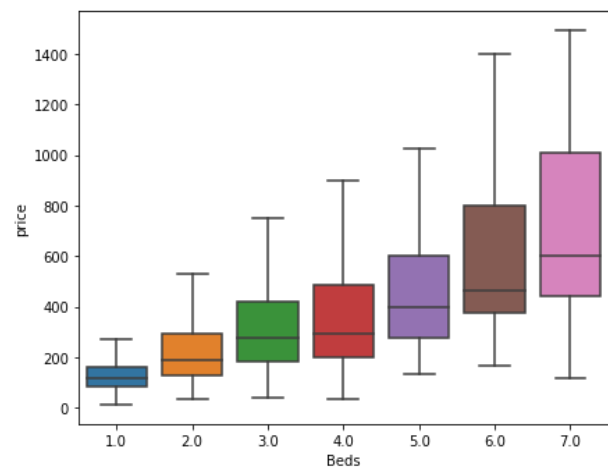
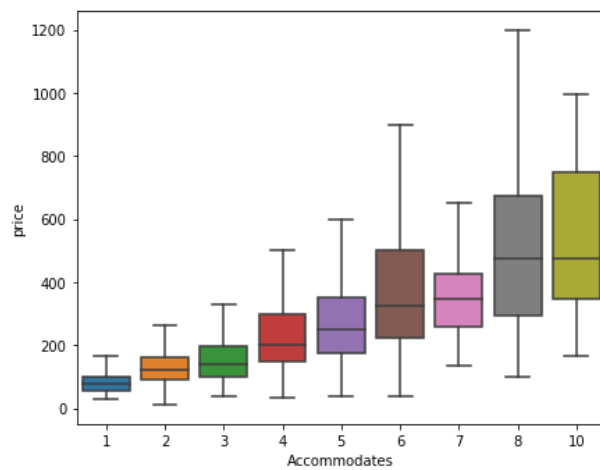
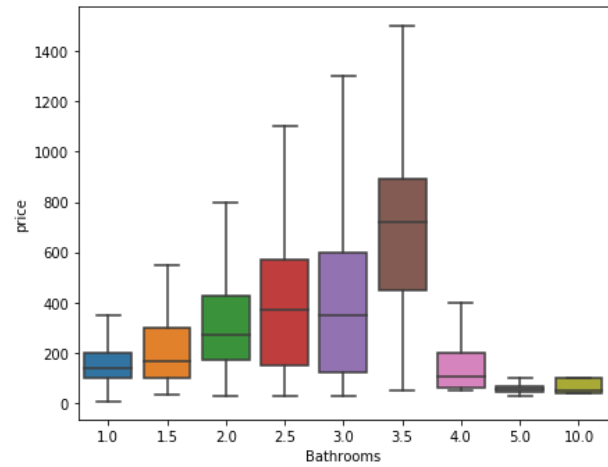
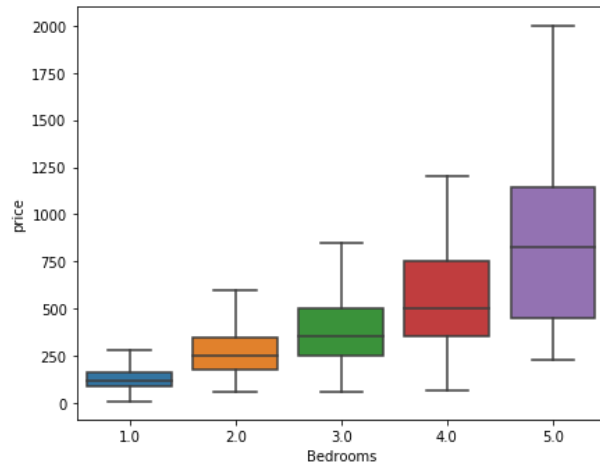


The boxplots above show how highly right-skewed the data is. The boxplot without outliers better depicts the price range between \$100 and \$250

### Accommodations

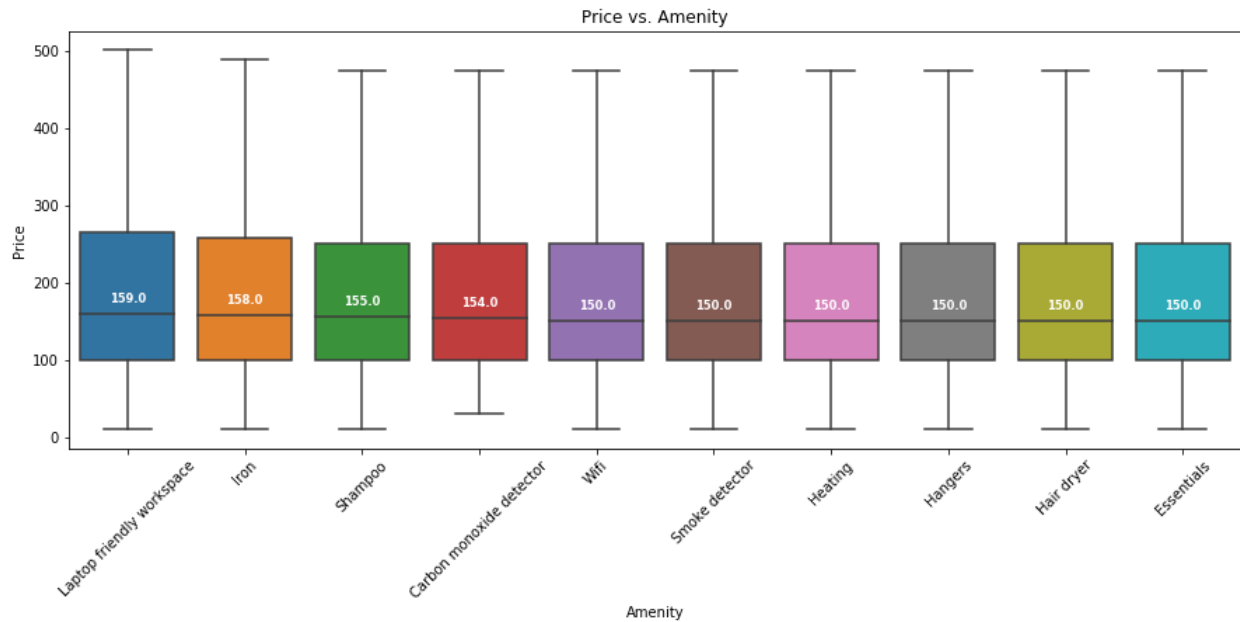


The vast majority listings offer one bed, again showing Airbnb strongly caters to solo and coupled guests. The amount of listings decreases as each bed count increases.



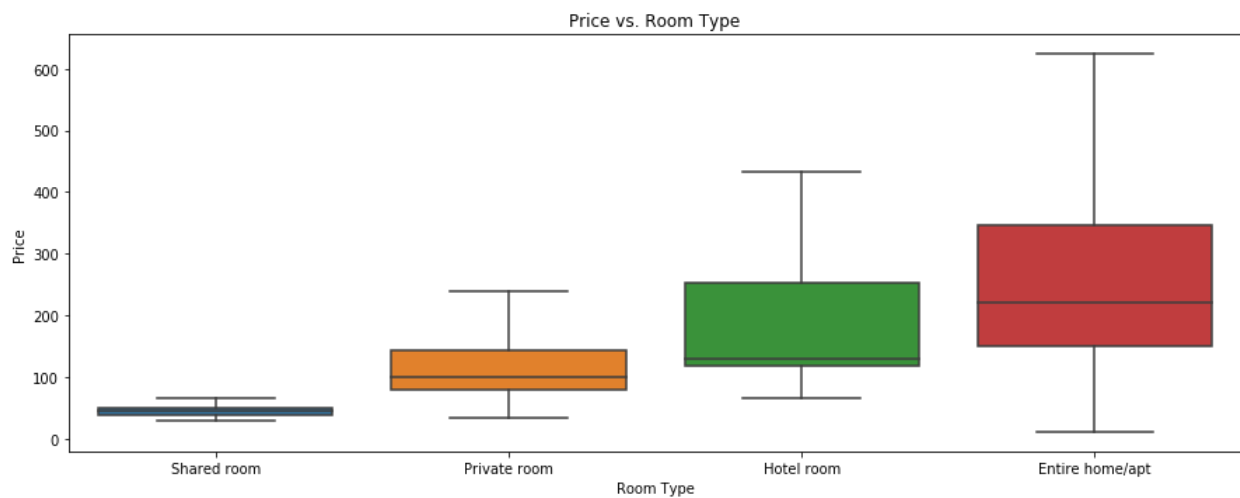
Bathrooms appear to be less correlated to price compared to bedrooms, accommodates and beds. All values with a sample size less than 20 were omitted from the boxplots. In addition, all features are significantly correlated with price based on the p-values obtained from the Point Biserial Correlation.

## Amenities



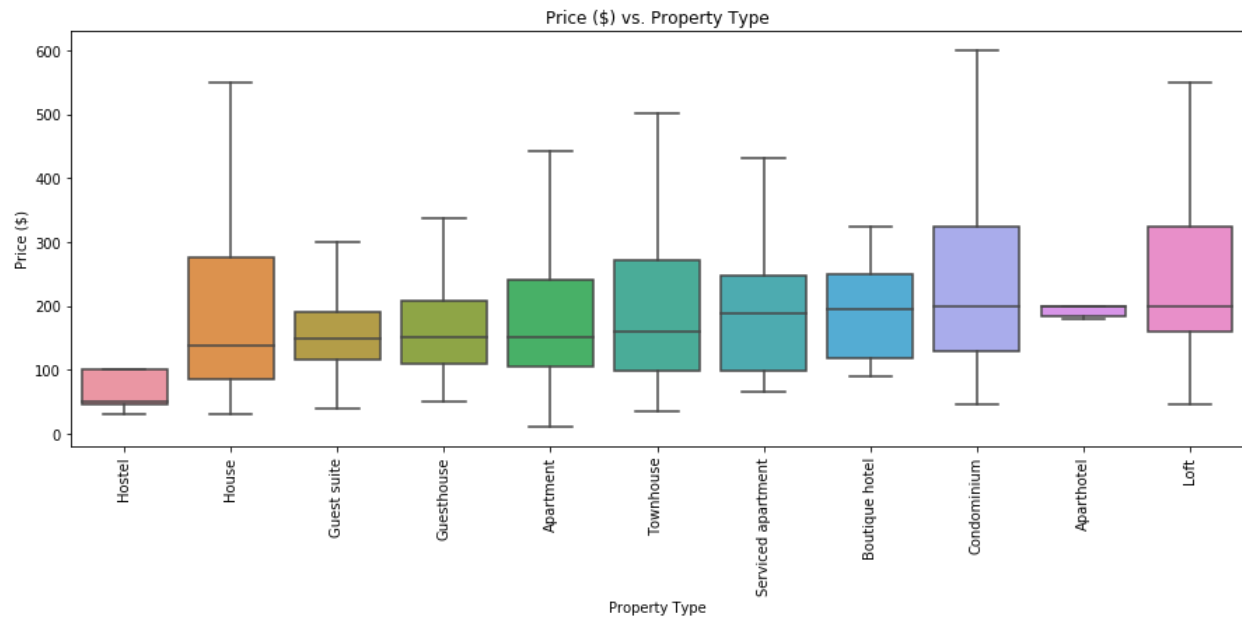
There is no visible difference between the top 10 amenities. All amenities except wifi and essentials are all significantly correlated with price based on the p-values obtained from the Point Biserrial Correlation.

## Room Type

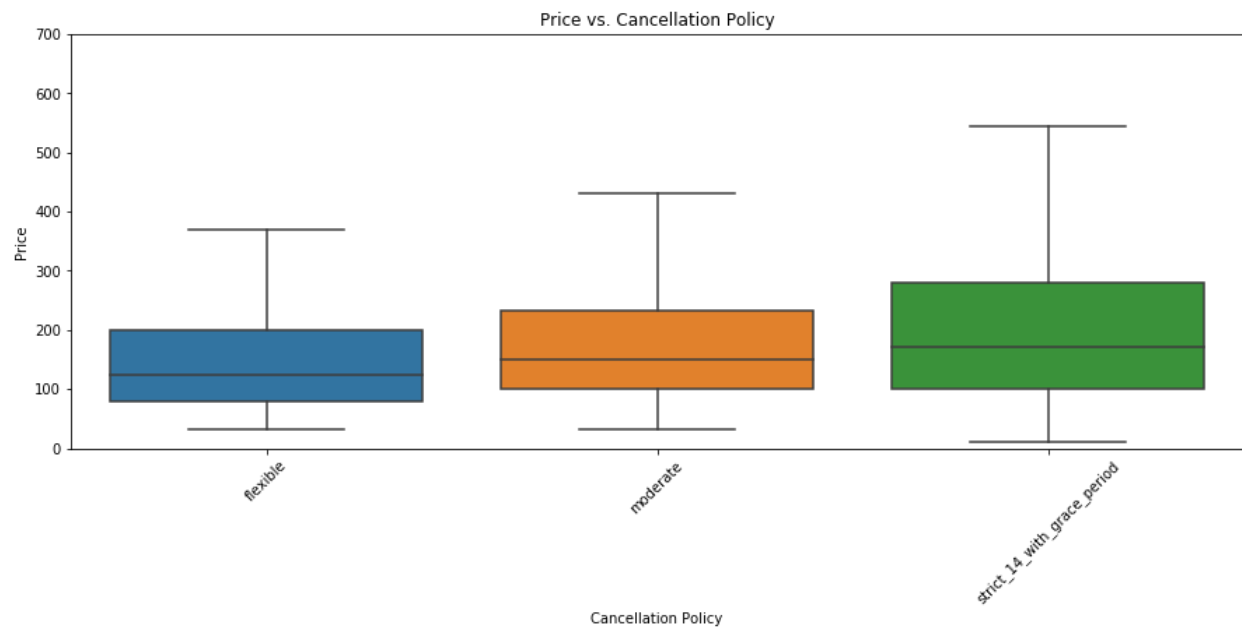


The size of room type appears to be highly correlated with price. Price increases as more privacy is given based on room type space.

## Property Type

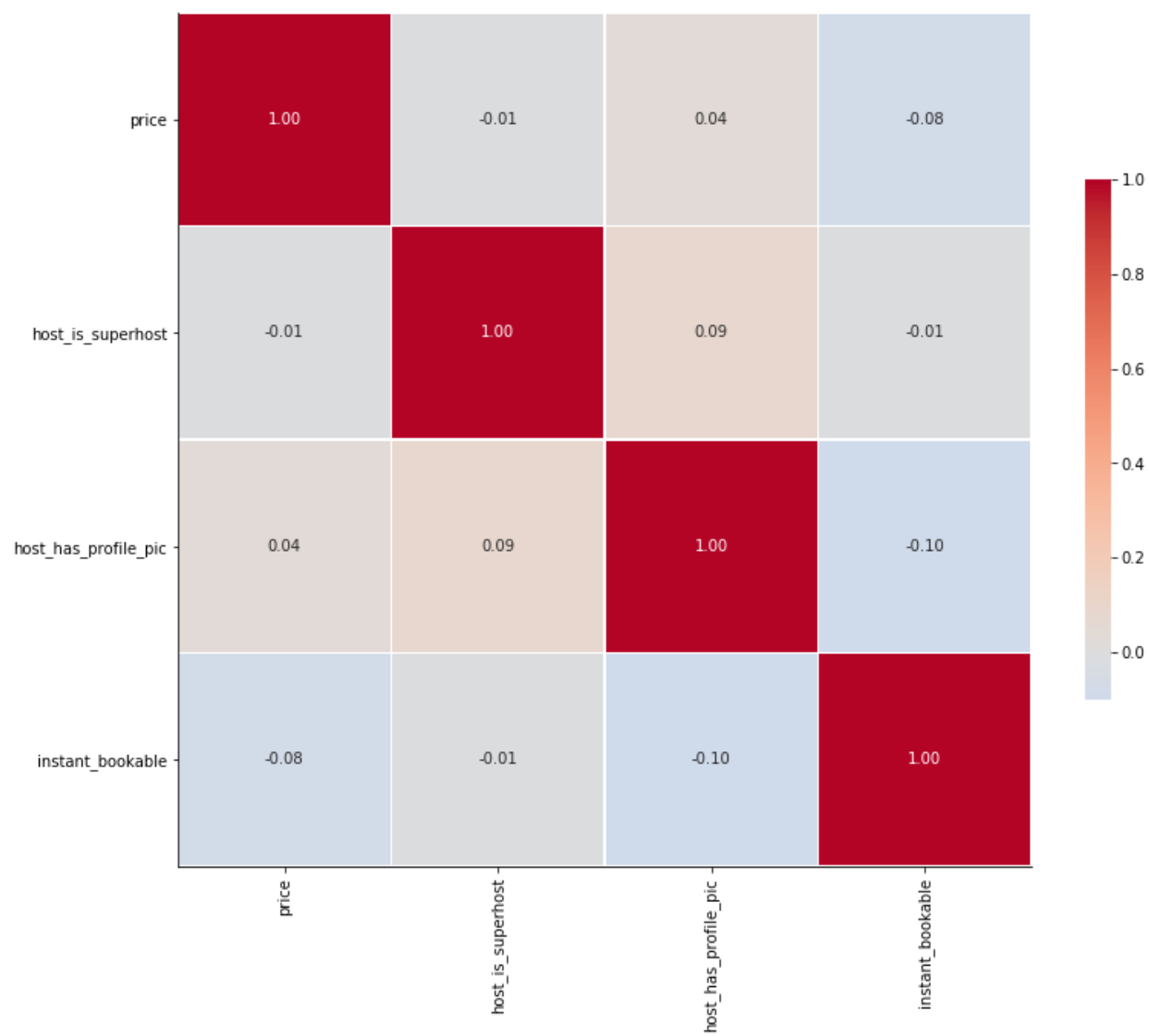


## Cancellation Policy

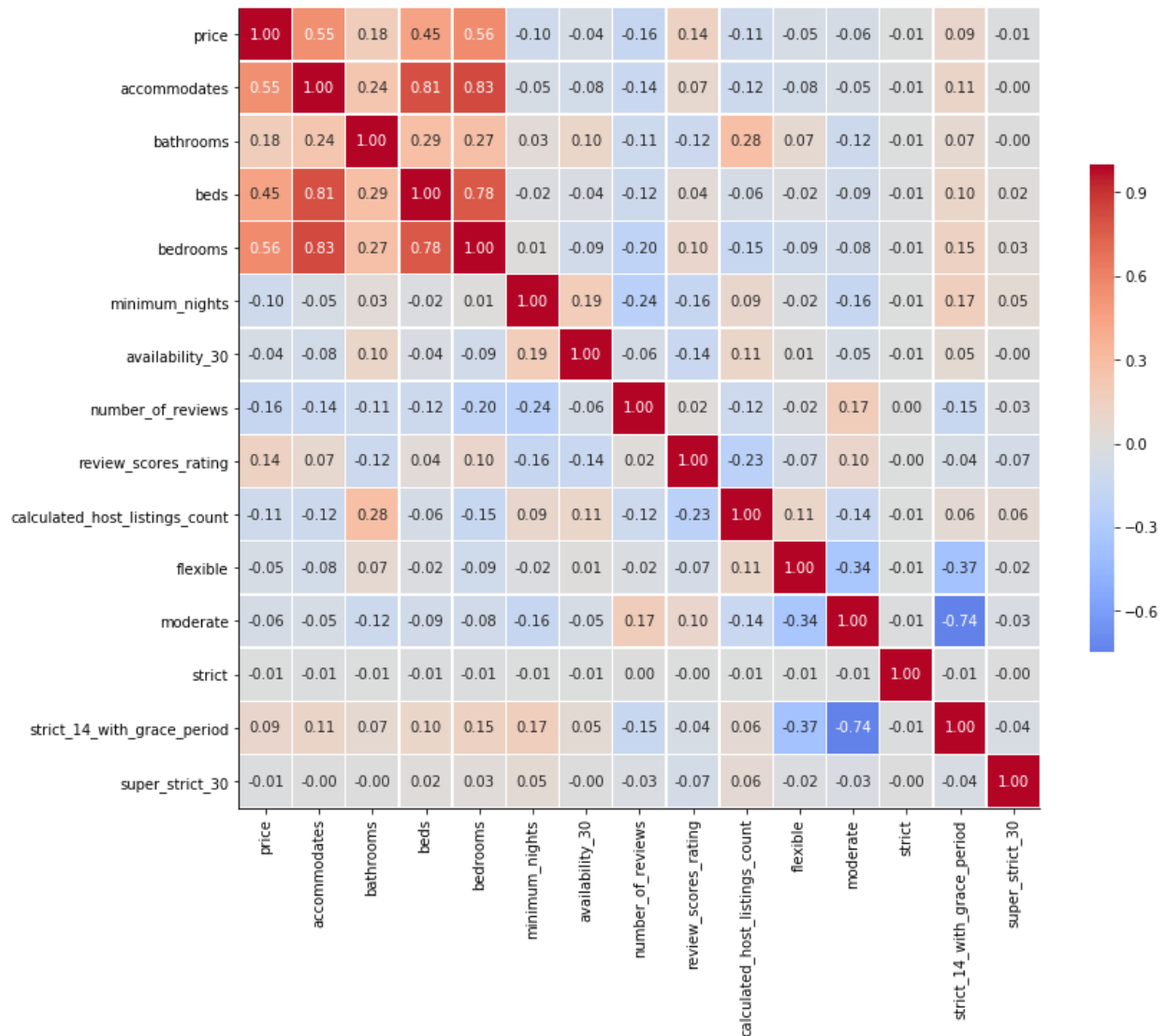


There appears to be a direct correlation between the price and level of cancellation policy, with prices increasing with a stricter policy. There is also a steadily increasing IQR as the cancellation policy becomes more strict.

Correlation Matrix between Price and Features:

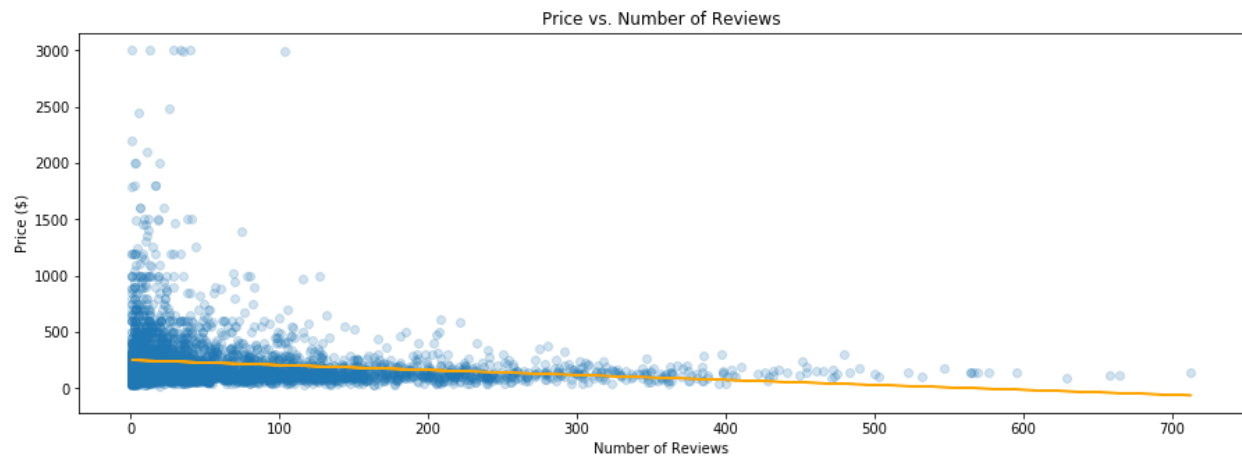


Not seeing a high correlation between these variables

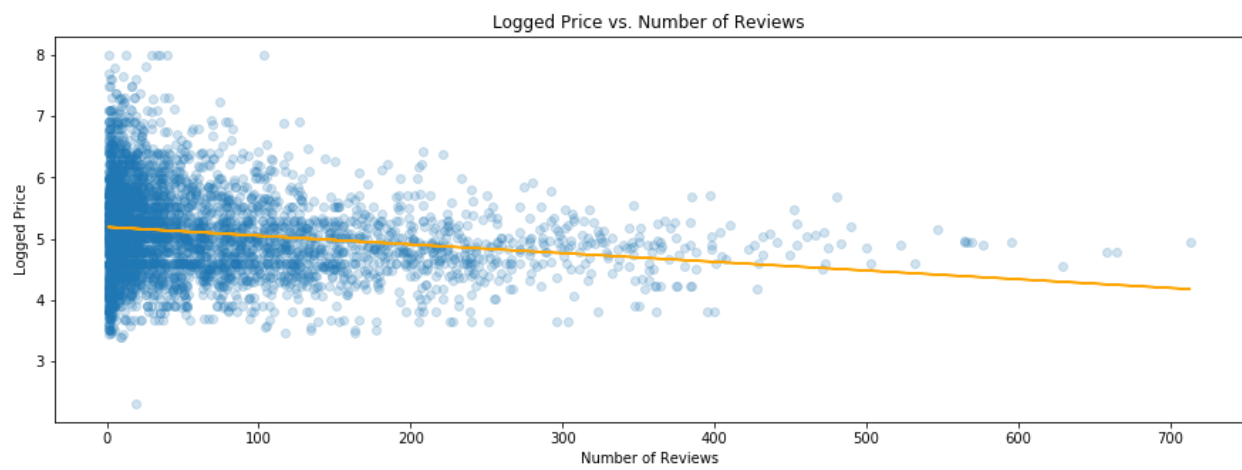


Accommodates, bathrooms, beds and bedrooms are all highly correlated with each other. The cancellation policies (flexible, moderate, strict, strict\_14\_with\_grace\_period and super\_strict\_30) are categorical data (hence correlation isn't technically correct), but included in the heatmap to get an overall sense of the differences between features and price.

## Number of Reviews



The scatter plot above shows the number of reviews are highest for listings lowest in price.



Here is the same data depicted with logged pricing.