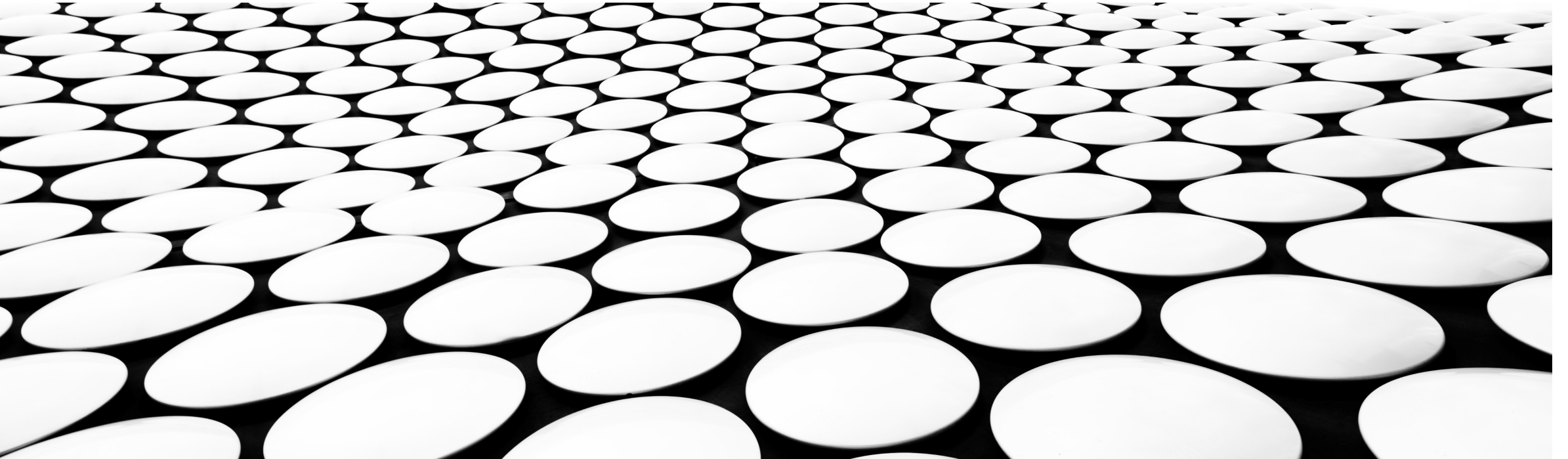# AIRBNB PROJECT

YINKA OLAIYA

BACKGROUND

- Airbnb is an online platform that allows people to rent short term accommodation. This ranges from regular people with a spare bedroom to property management firms who lease multiple rentals. On the one side, Airbnb enables owners to list their space and earn rental money. On the other side, it provides travelers easy access to renting private homes.

- Airbnb receives commissions from two sources upon every booking, namely from the hosts and guests. For every booking, Airbnb charges the guest 6-12% of the booking fee. Moreover, Airbnb charges the host 3% for every successful transaction.

OBJECTIVE

- To come up with a pricing model that can effectively predict the Rent for an accommodation and can help hosts, travelers, and also the business in devising profitable strategies.

  These would be achieved by:

1. Exploring and visualizing the dataset.

2. Building a linear regression model to predict the log of rental price

3. Generating a set of insights and recommendations that will help the business

DATA OVERVIEW

S/N	Code	Description
1	Id	Property ID
2	room_type	Type of Room in the property
3	accommodates	How many adults can this property accomodate
4	bathrooms	Number of bathrooms in the property
5	cancellation_policy	Cancellation policy of the property
6	cleaning_fee	This denotes whether the property's cleaning fee is included in the rent or not
7	instant_bookable	It indicates whether an instant booking facility is available or not
8	review_scores_rating	The review rating score of the property
9	bedrooms	Number of bedrooms in the property
10	beds	Total number of beds in the property
11	log_price	Log of the rental price of the property for a fixed period

OVERVIEW CONTD.

- There are 74,111 rows of data and 11 variables. Out of these 11, only 3 do not have missing values. We observe the variable with the highest number of missing values as reviews_score_rating with 16,722 missing values. Treating this would introduce a lot of bias to the data and we decided to drop it early in our analysis

- Skewness does not seem to be a problem for most of the variables, hence we will leave as is.

- We are seeing variables such as bedrooms, beds, bathrooms and log_price having records of zero. This is an anomaly that can be treated by further correct imputation that was yet to covered at the stage this project relates to. So we will leave as is at the moment.

- For the missing values, we replaced all by the median of the number in each column having a missing values. However, there was an issue with columns having blanks instead of NaNs. Because these are few, we dropped the rows involves: 8,5 and 4, in cancelation policy, room type and cleaning fee columns respectively
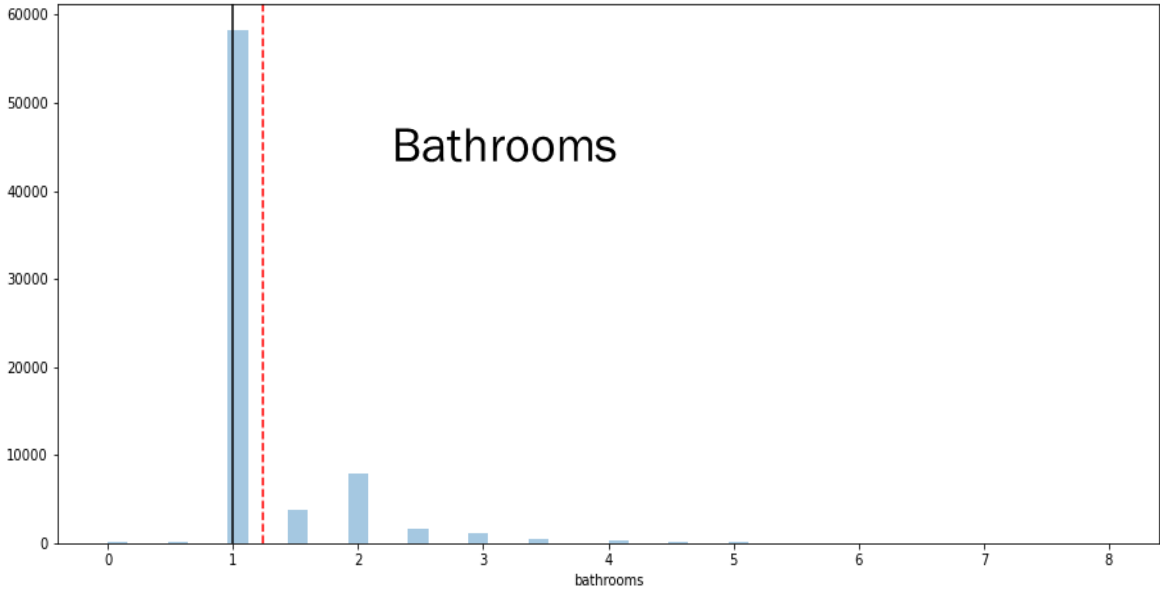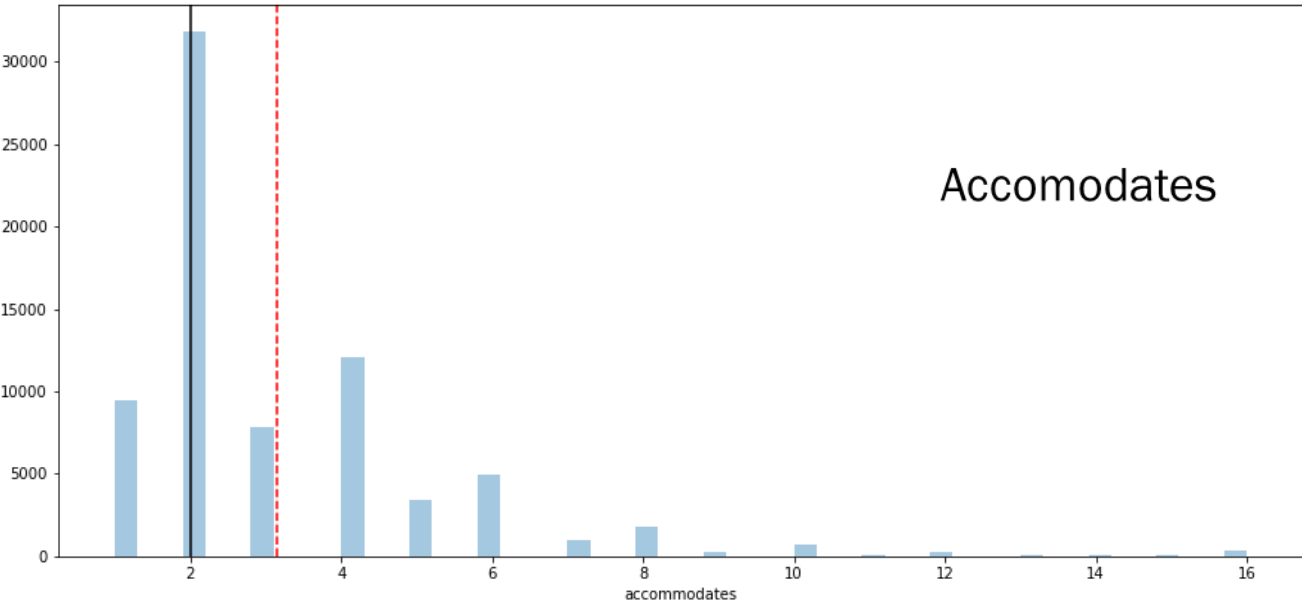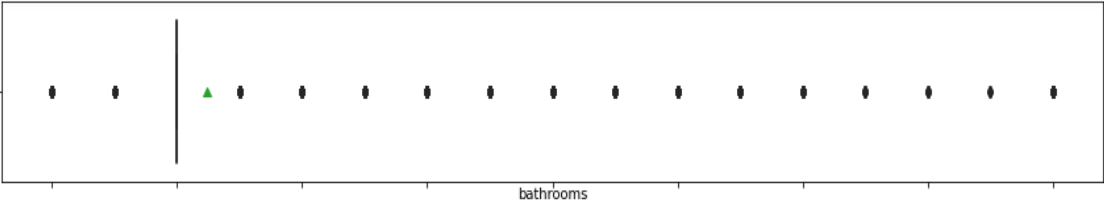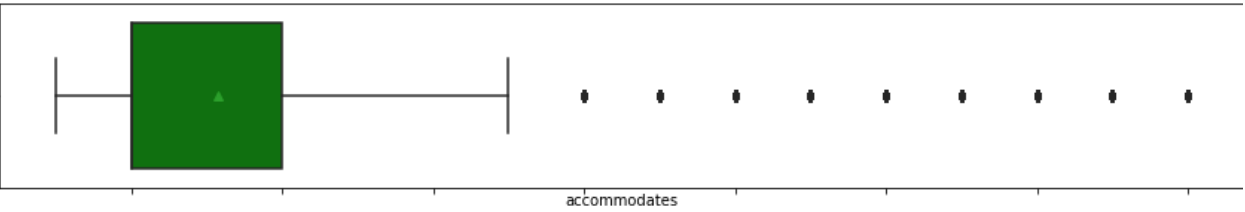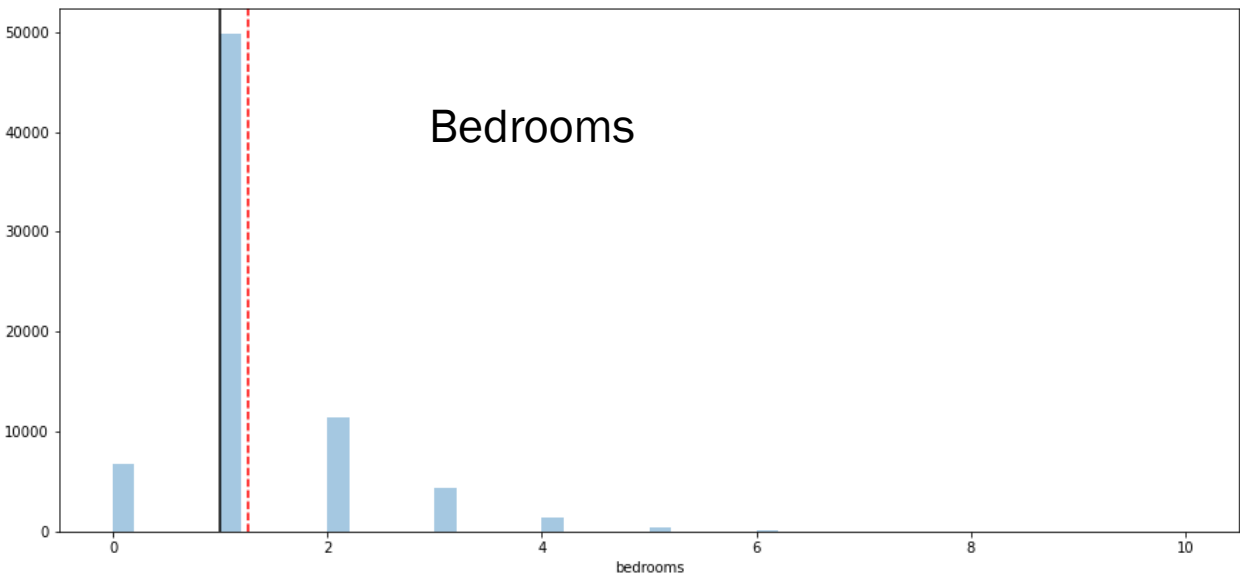
EXPLORATORY DATA ANALYSIS
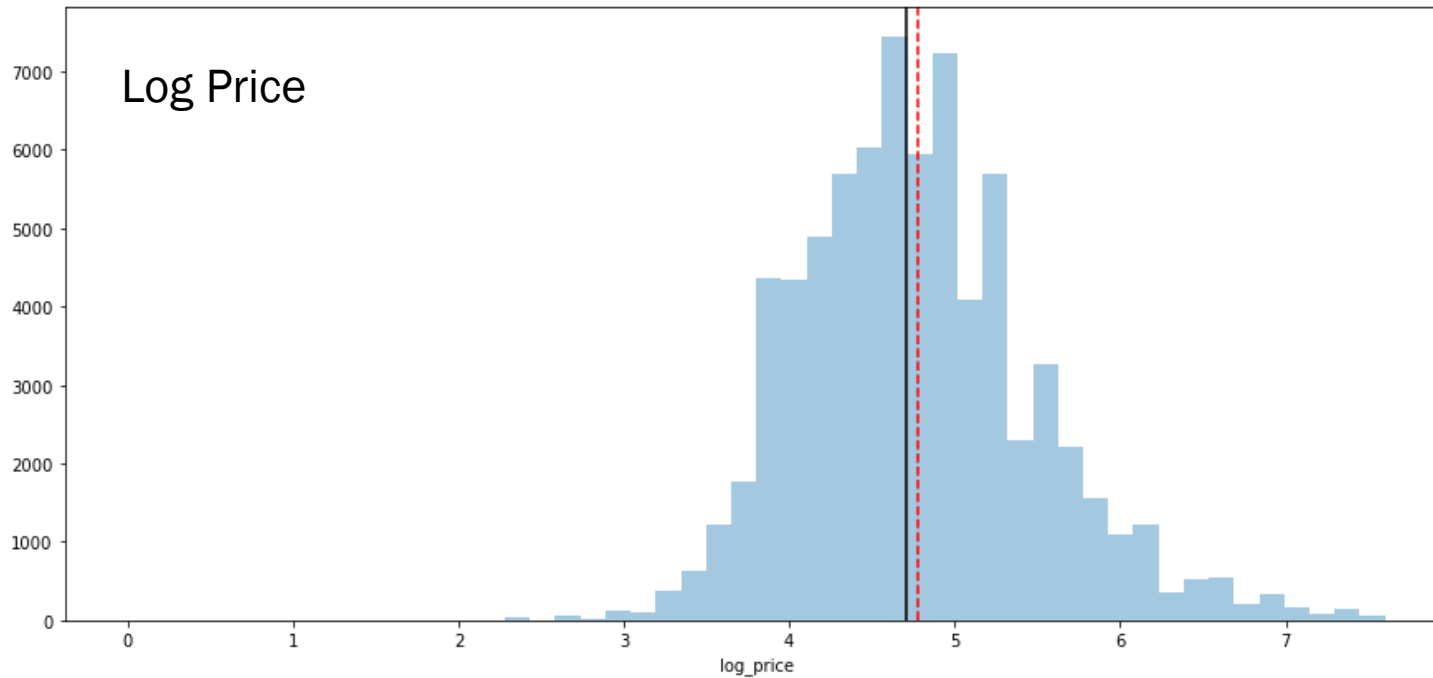
UNIVARIATE ANALYSIS
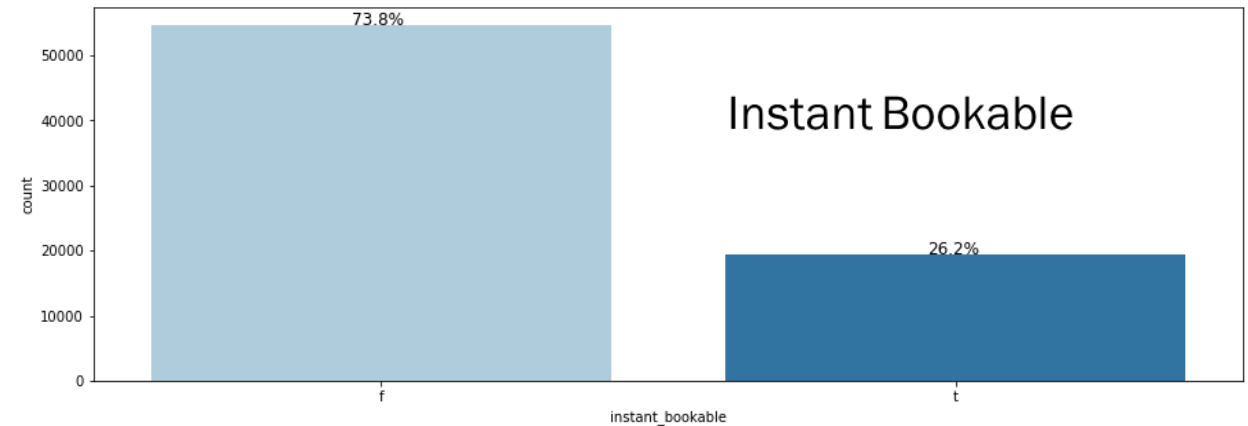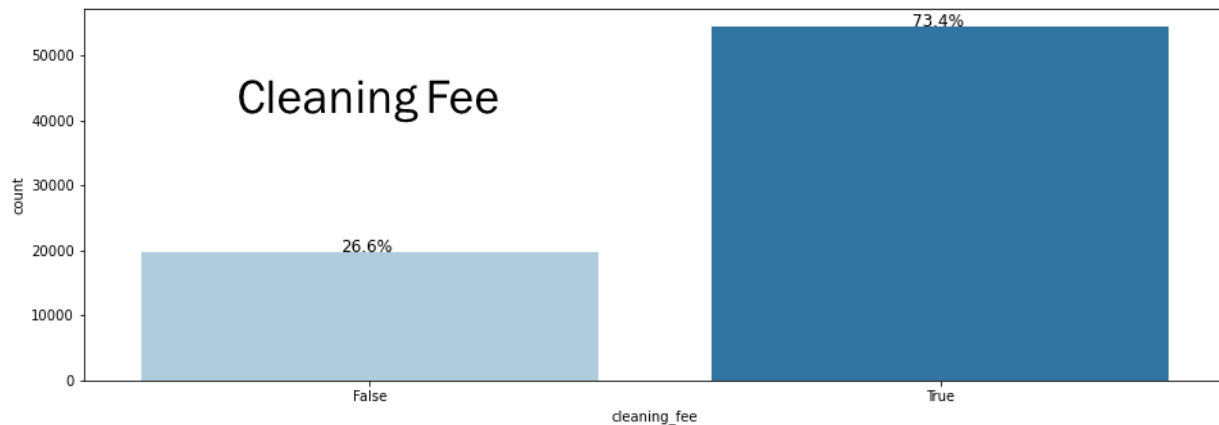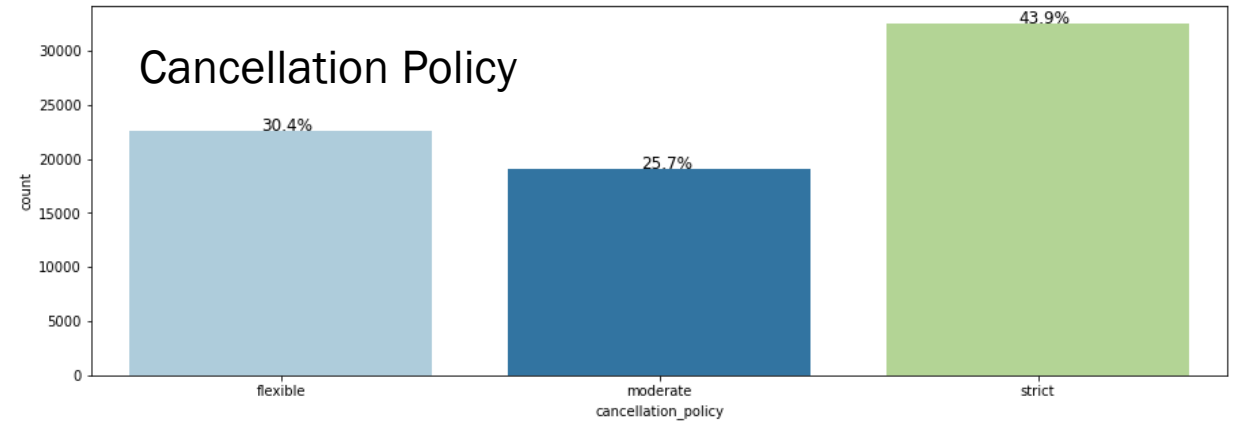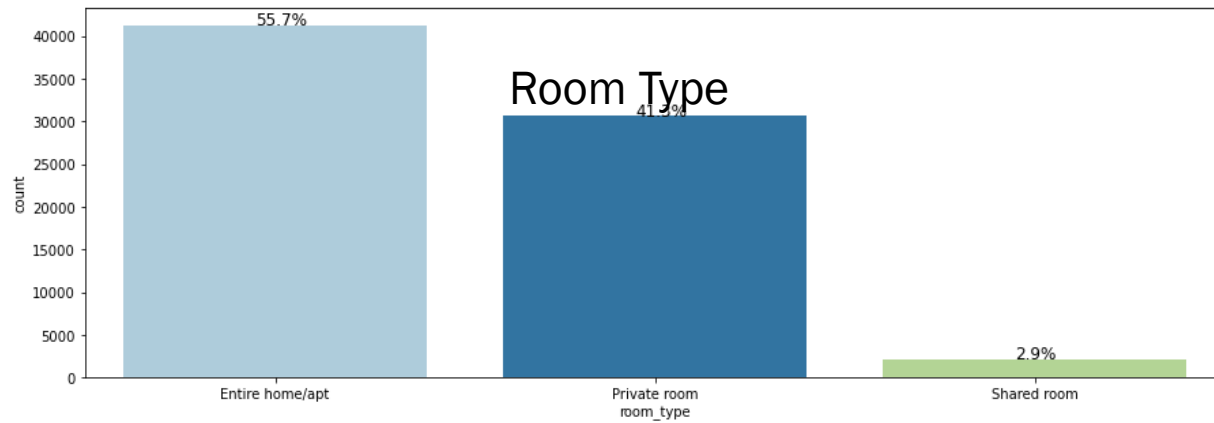
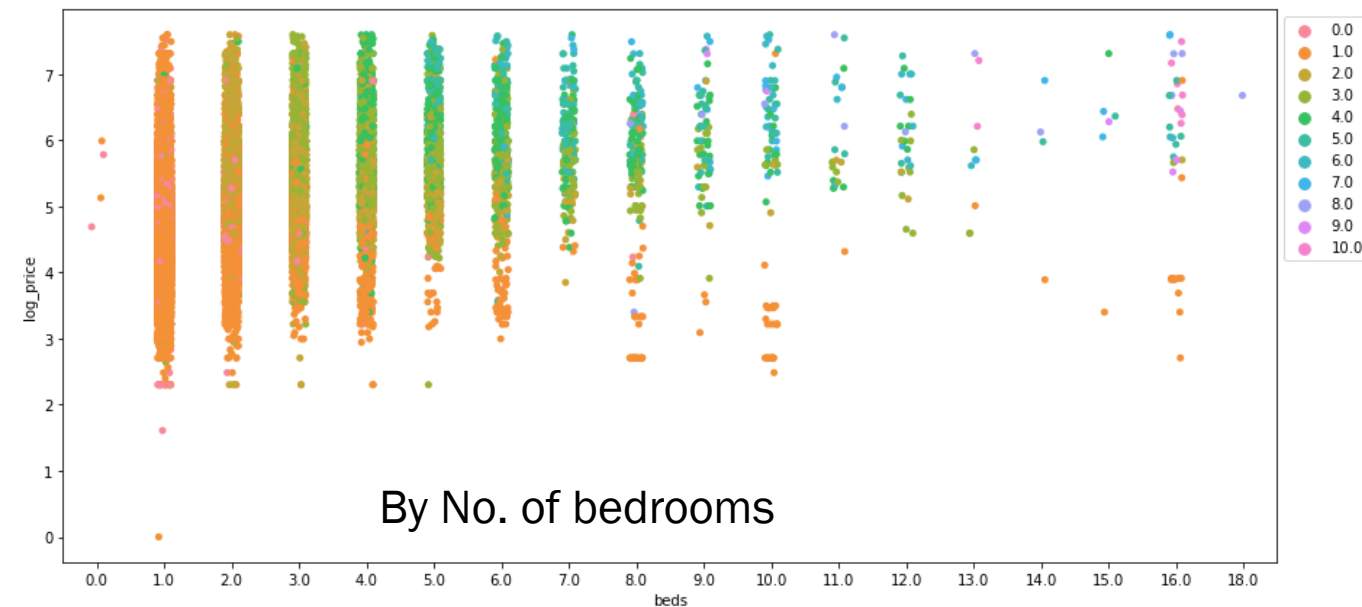# Univariate Analysis Contd.

Bedrooms
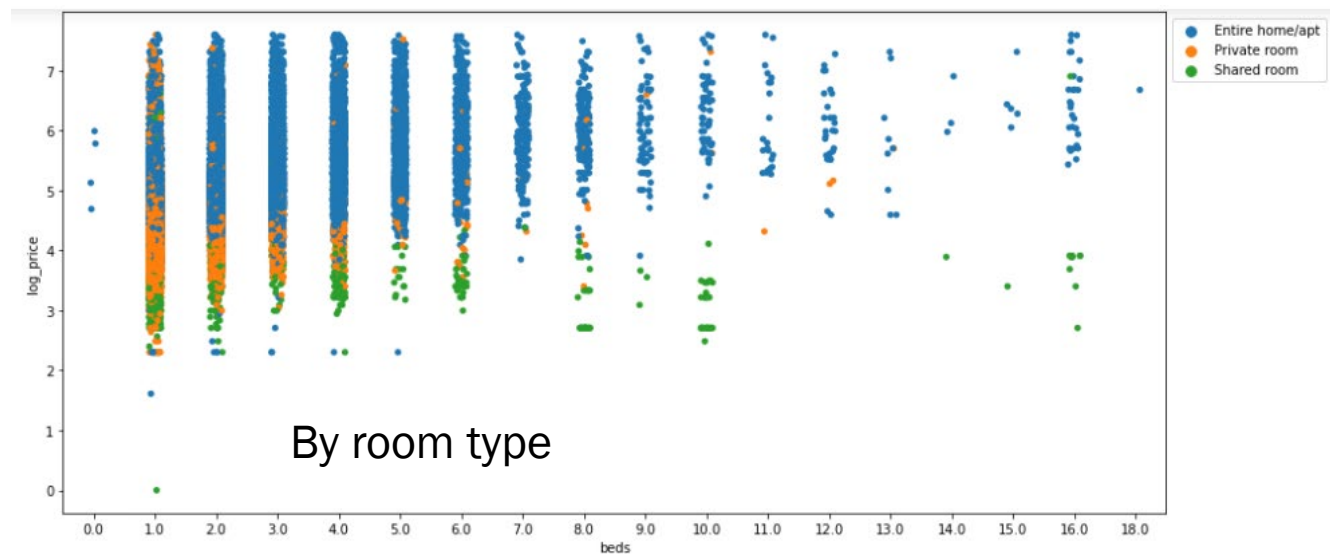
Beds

Log Price
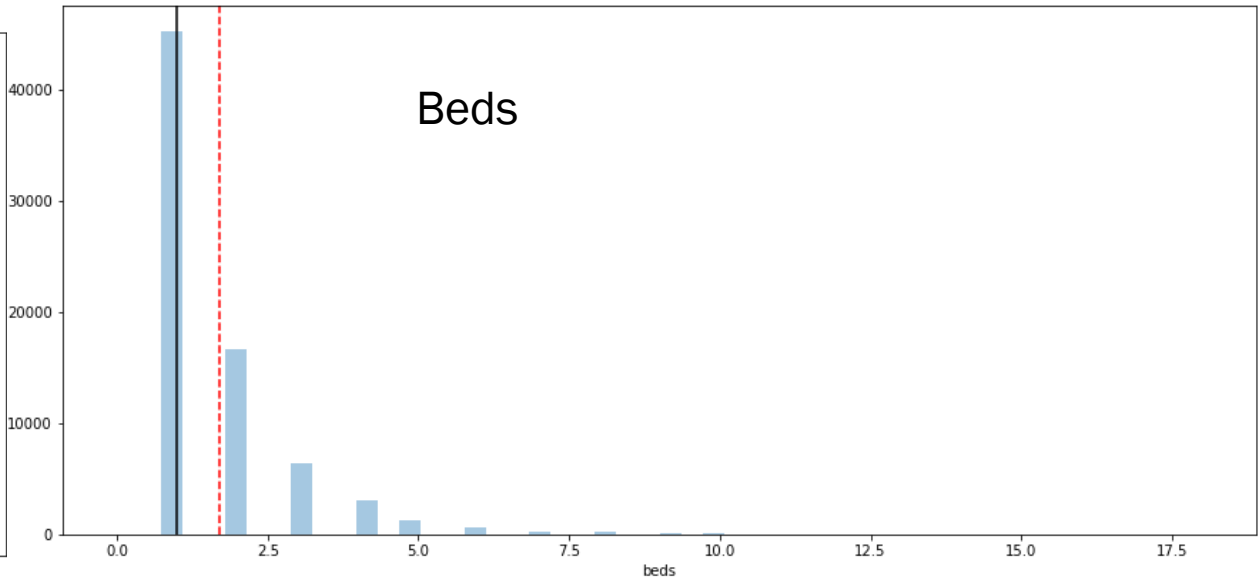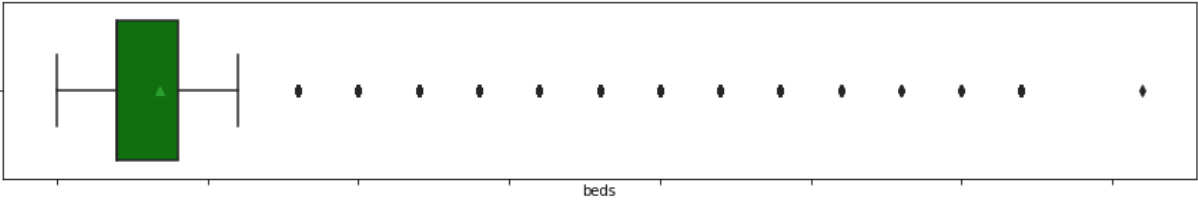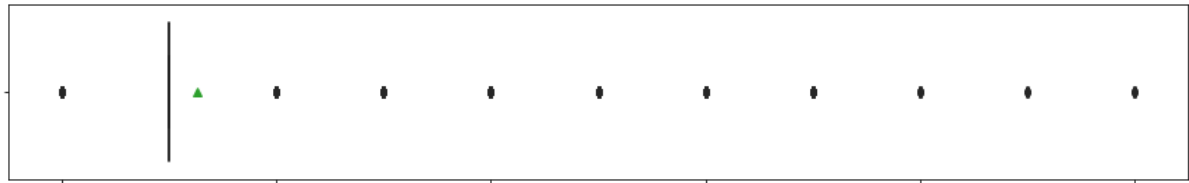
Univariate Analysis Observations:

- We have over 55% of our data in the entire home/apt category
- Most of the accommodations house only 2 people and they mostly have 1 bathroom
- About 44% of the accommodations available have a strict cancellation policy. Over 73% have cleaning fee included in the rent and the about same have instantly bookable facility
- The largest majority of the accommodations available have just 1 bedroom and 1 bed
- Rental price is mostly within the range of 3.8 and 5,5. We do not know the unit associated with this price
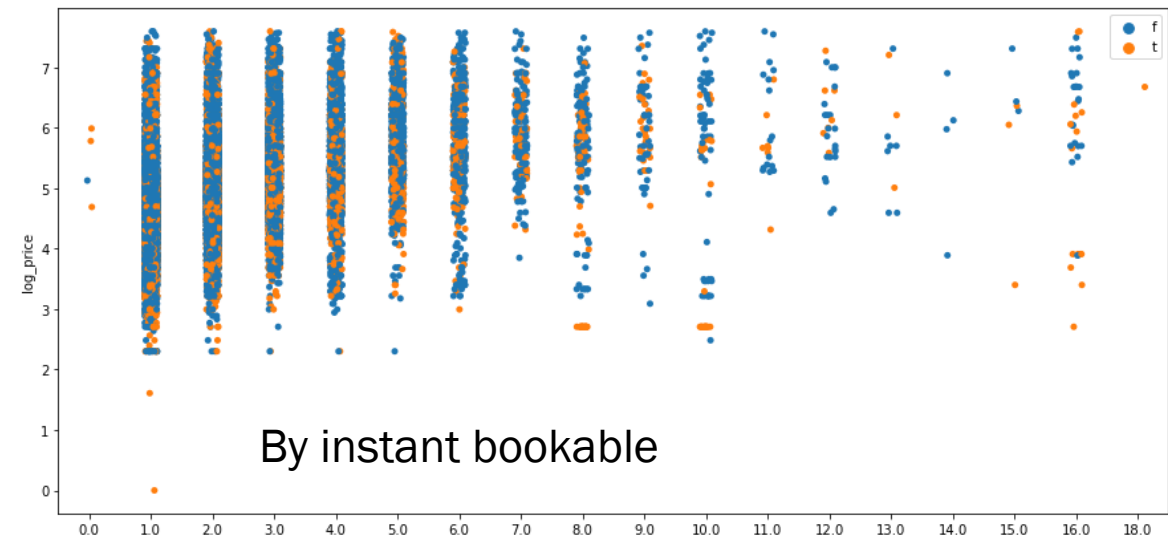
MULTIVARIATE ANALYSIS

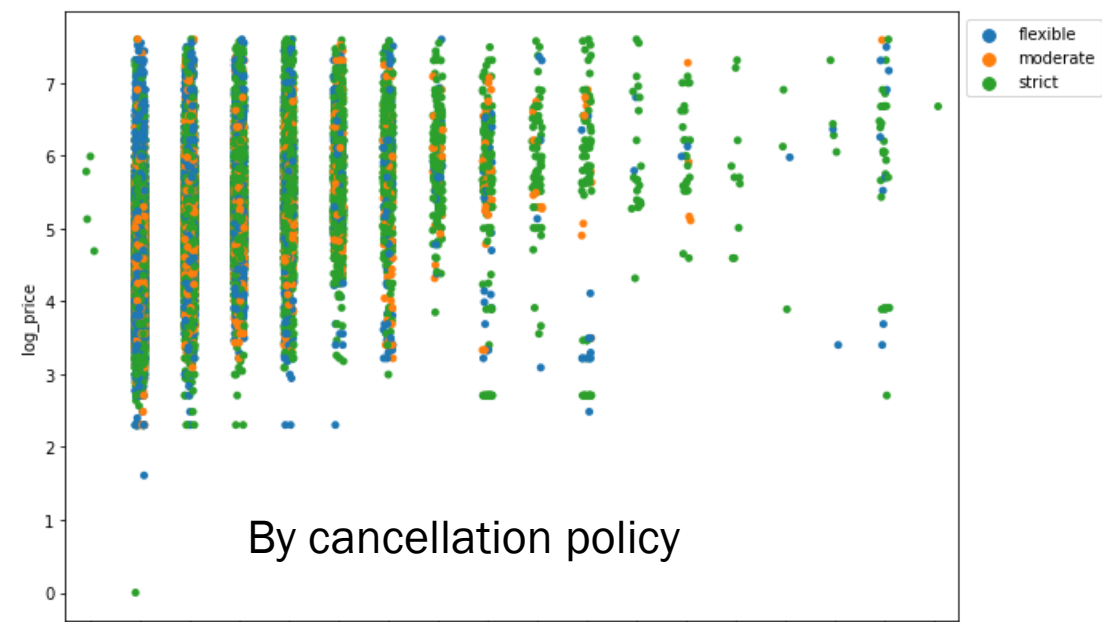## No. of Beds vs Log Price against other variables
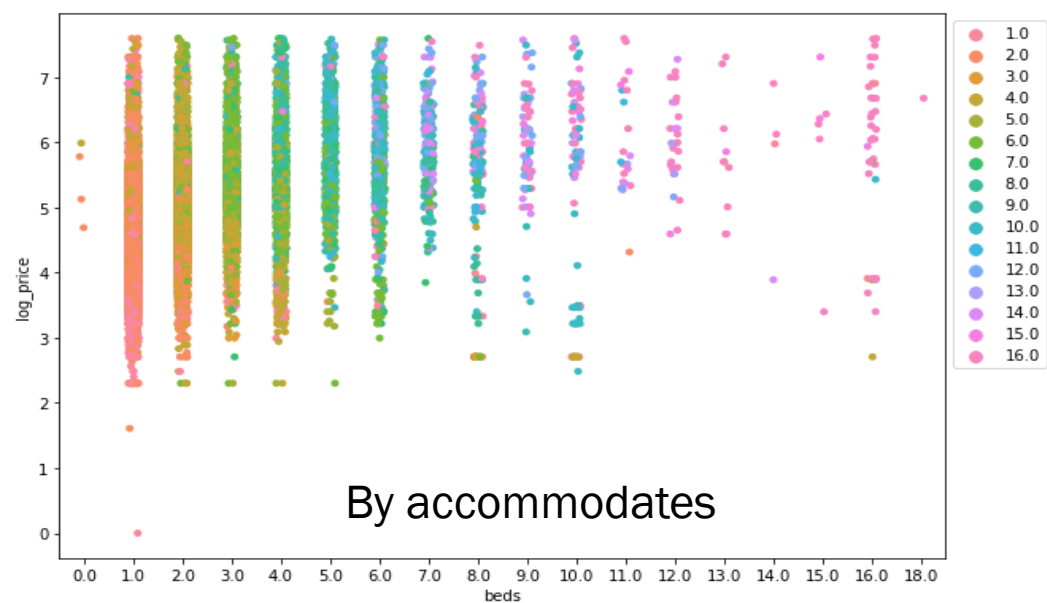

By No. of bedrooms

By instant bookable

By room type

By cancellation policy

By cleaning fee

By No. of bathrooms

By accommodates

Instant bookable vs log price vs cancellation policy

Cancellation policy vs log price vs instant bookable

Cleaning fee vs log price vs no. of bathrooms

No. of beds vs bathrooms vs instant bookable

Accommodates vs log price vs cleaning

No. of beds vs accommodates vs bathrooms

No. of beds vs bedrooms vs bathrooms

No. of bedrooms vs beds vs instant bookable

Multivariate Analysis Observations:

- The cheapest of all the accommodations is in one bedroom and one bed. We see some high-ends in some with one bedroom but missing the number of beds, i.e. those showing zero beds.
- The spread seem to indicate that the accommodations in our data is mostly those with one bed per bedroom. We see that in the change of color as the number of beds increases
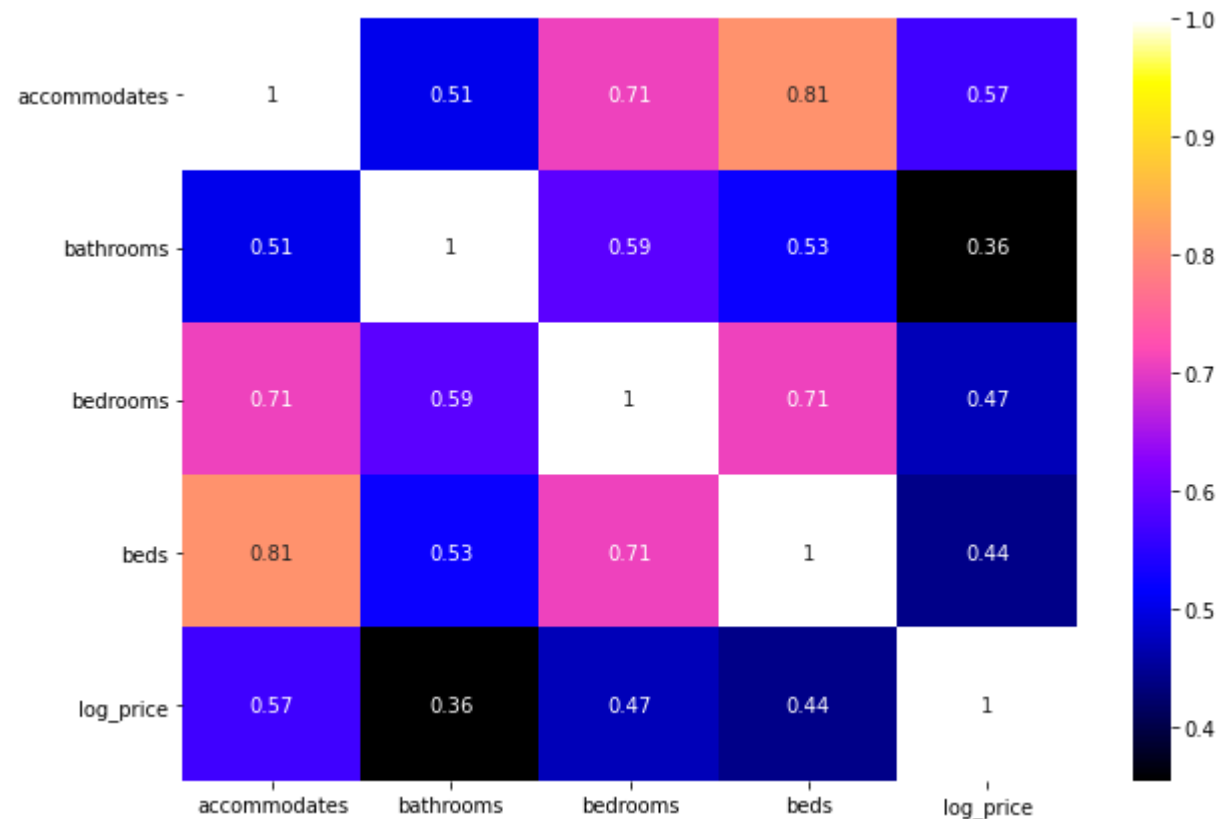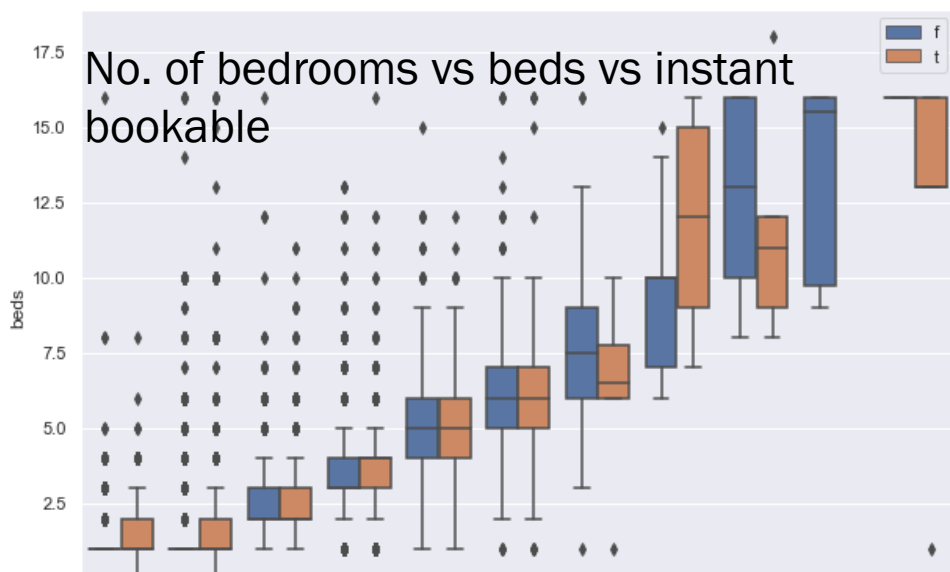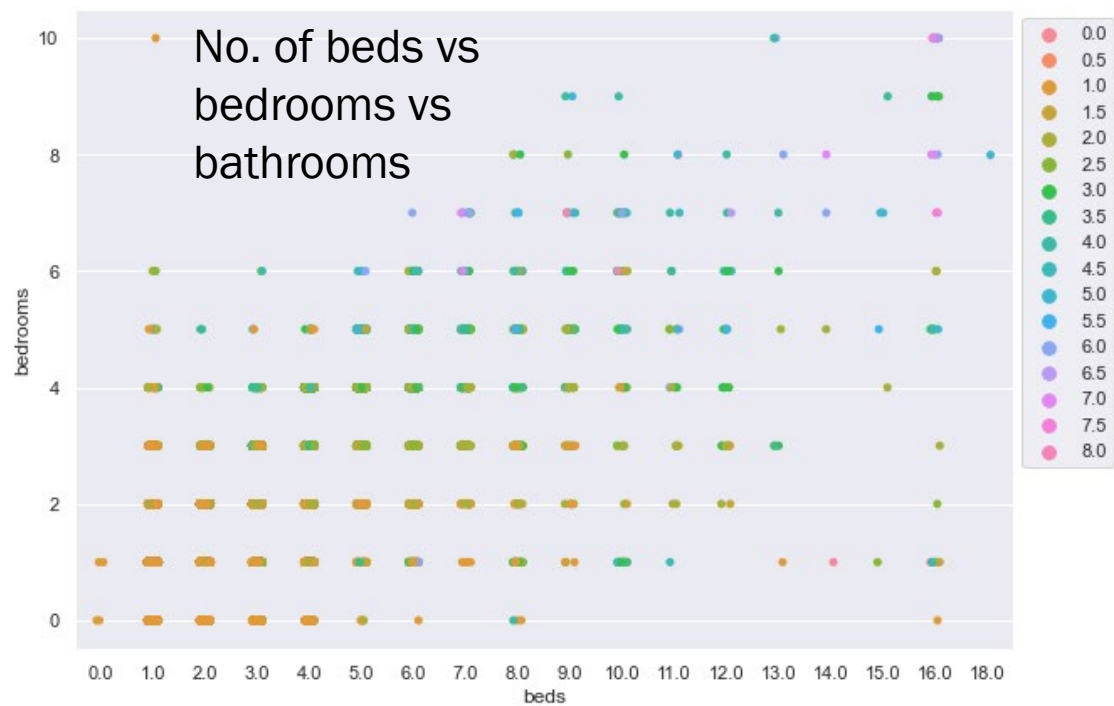- We mostly have accommodations renting out entire homes/apts, just a sprinkle in homes with multiple beds where private rooms are rented out. It is also fascinating to see that almost all the homes/apts have shared rooms, except having 11, 12 and 13 beds
- Most of the rentals we have are not instantly bookable and strict cancellation policy is the most prevailing.
- We see that cleaning fee is included in the rent for most of the accommodations
- The number of bathrooms available does not match the number of beds in some instances
- We see that some rentals with just one bed indicates that it can accommodate up to 16 adults. That indicates there is a problem with our data.
- Rentals with 9 bedrooms are entirely not instantly bookable
- Rentals that not instantly bookable have a little more with strict cancelation policy that those that can be booked instantly. The difference seen is very little.
- The highest correlation is seen in the relationship between the number of people a house or room can accommodates and number of beds it has. This is followed closely by number of bedrooms and number of beds available.
- Log price is mostly correlated to the number of people a room or house can accommodate.
- The least correlation we can see here is in the relationship between log price and number of bathrooms available

BUSINESS INSIGHTS AND RECOMMENDATIONS

- There is need to do a better job at getting adequate data for good analysis to be made. We can see some anomalies in the data we worked on, where there were zero bedrooms, bathrooms and beds. This is definitely not a possibility in this scenario. If they all happen to be zero for all three variables on the same row, a possible assumption would be that the rental is not currently available or taken off-market by the owner but that is not the case we are dealing with here.

- The type of rental provided is the strongest indication of how price can be set. The more private, the better it is to set a higher price for the rental.

- Cancellation policies also play a vital role in determining the price of a rental, the more flexible it is, the more it is attractive and can be a pull for setting higher prices that could be possibly dictated by demand