

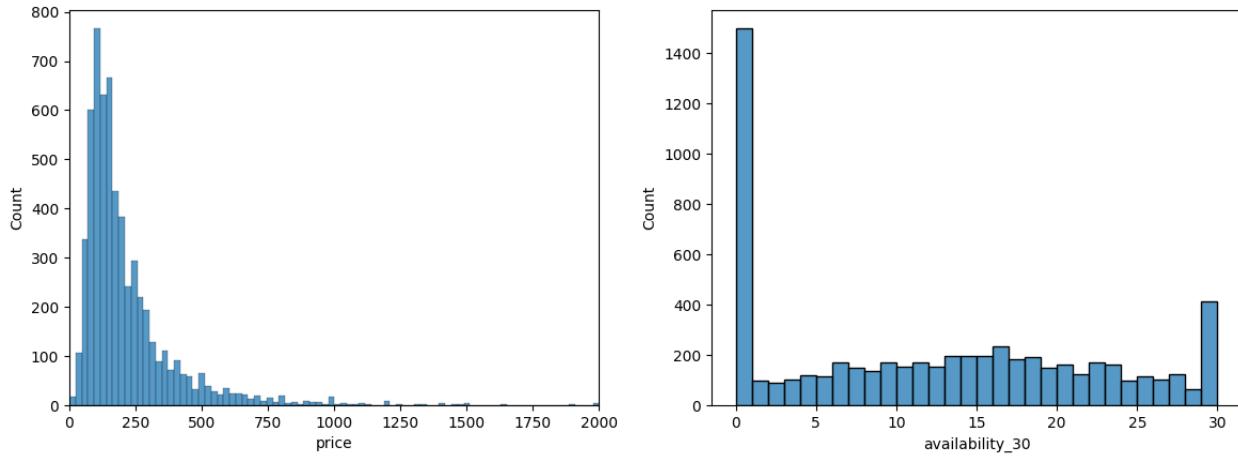
Information Visualisation Project

Exploring the New Orleans Airbnb Listings Data

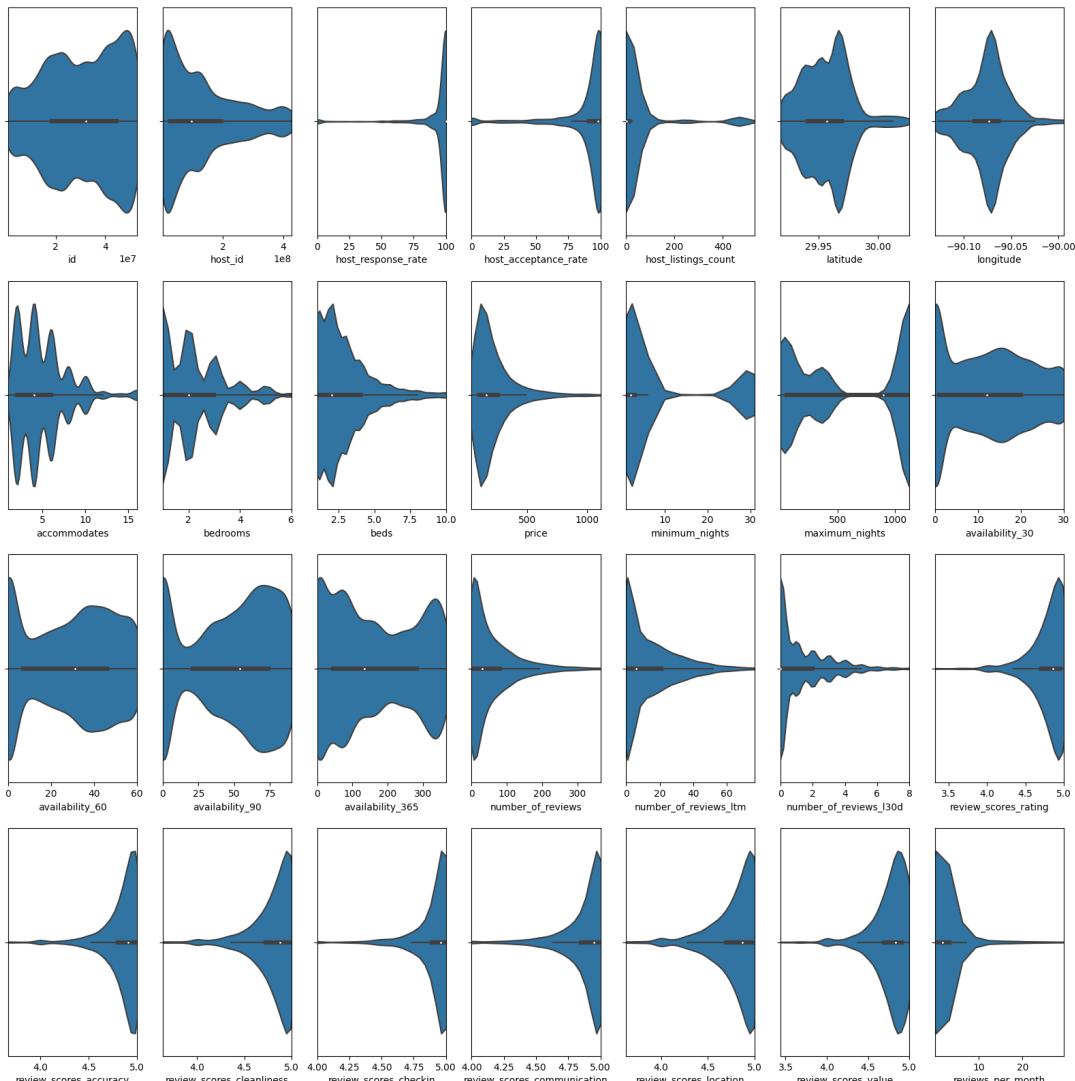
Zaharia Stefan-Tudor 511

Python Visualisation

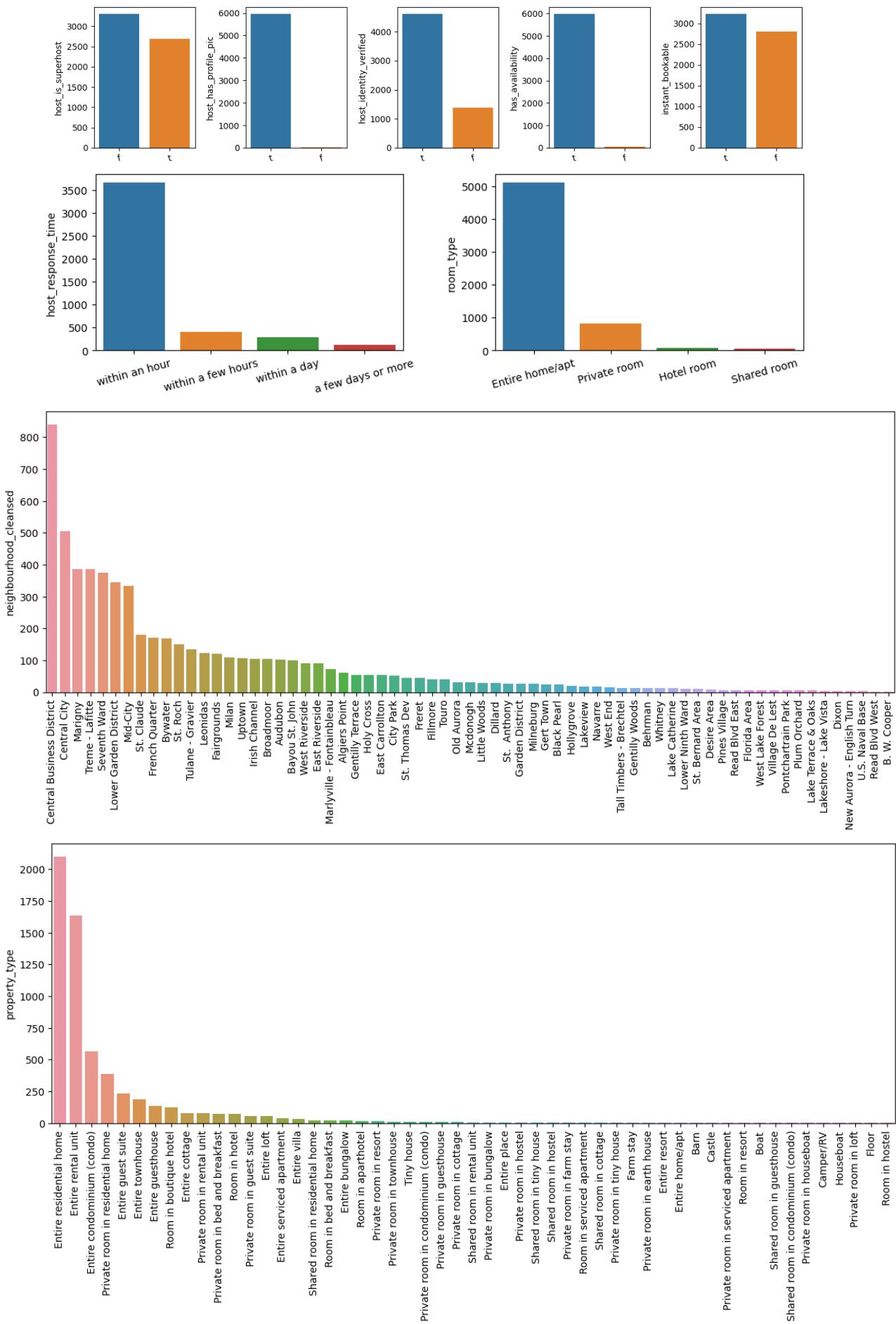
From the 49 columns providing details about each listing, we will consider as target value the **price** and the **availability** for the next 30 days (as a measurement of popularity).
A first step would be to plot the distribution of the target values as a histogram:

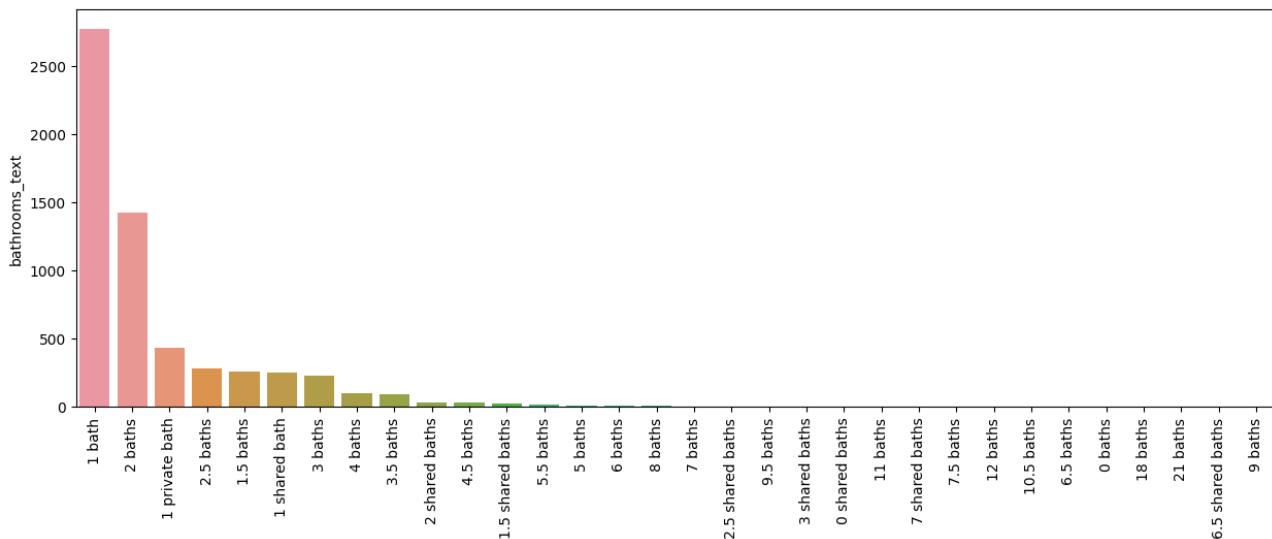


For simplicity, we can plot a violin plot for each numeric feature:



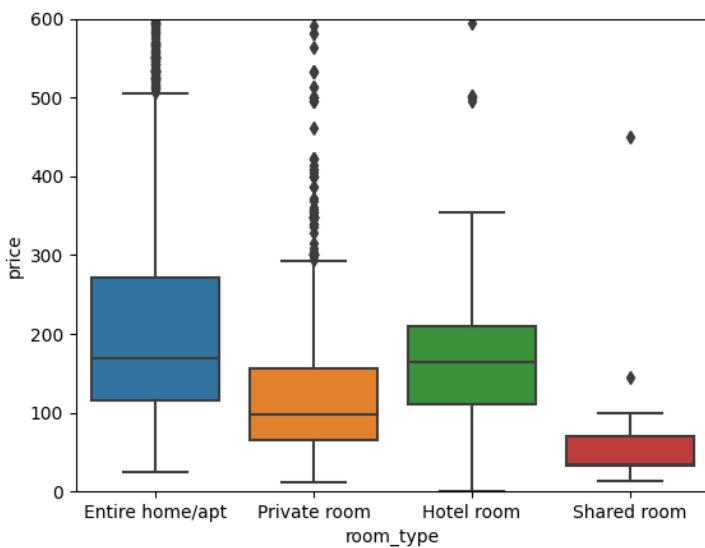
And for each categorical value, we can plot a barplot:





Questions we answered looking at the data:

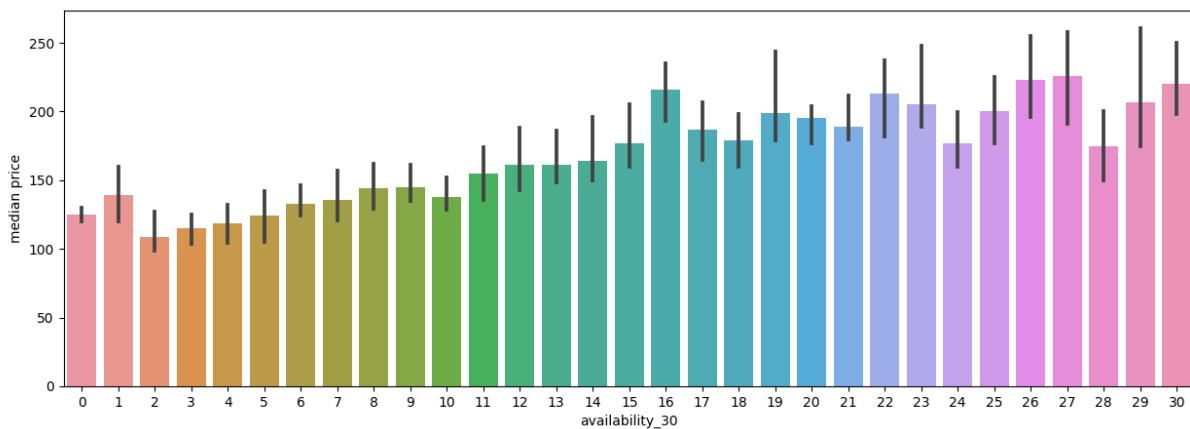
1. Is the price correlated with the room type?



Answer:

As expected, "Shared room" are the cheapest overall, followed by "Private room". The "Entire home/apt" and "Hotel room" have roughly the same median, but the higher prices tend to get bigger for the first type.

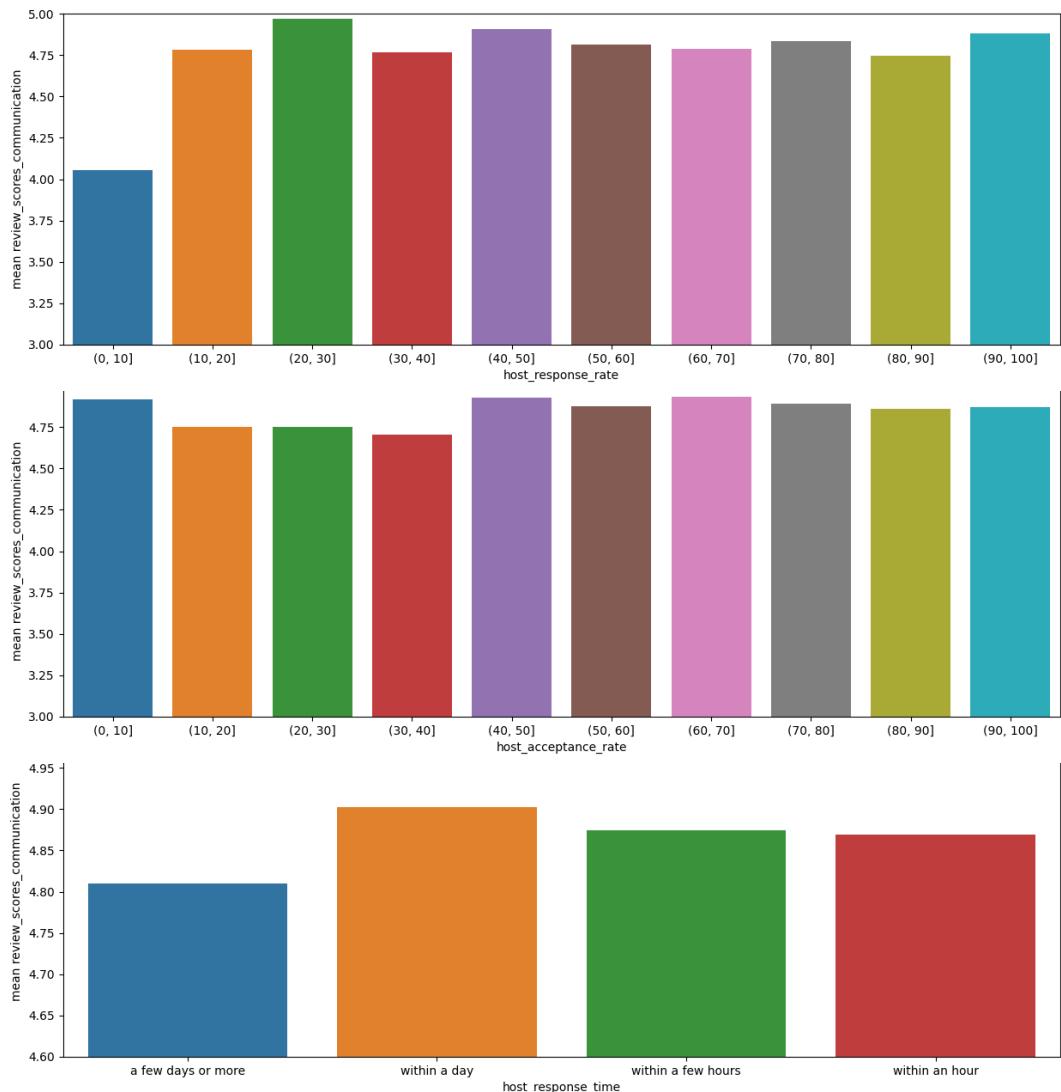
2. Does a lower price also lower the availability of a property?



Answer:

From the plot above, it seems that the median price for properties grouped by the availability for the next 30 days are positively correlated with the availability. As expected, cheaper listings are booked first.

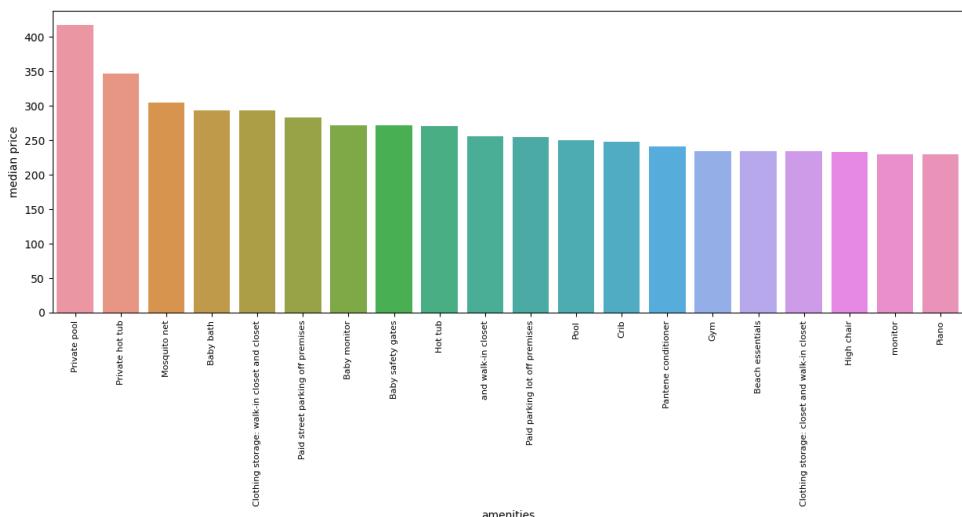
3. Are the reviews for communication correlated with the host response time, host response rate and host acceptance rate?

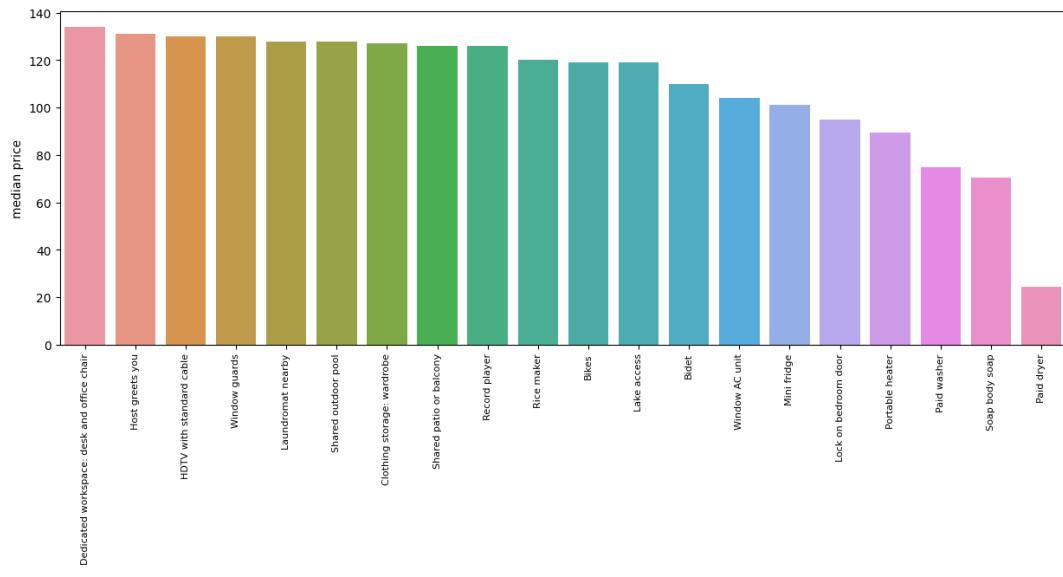


Answer:

It seems that the only conclusions we can make from this graph is that responding in a matter of days and not responding to any messages is associated with lower reviews for communication.

4. What amenities drive the price up? For this question, we selected only the amenities that appear at least in 10 or more listings.

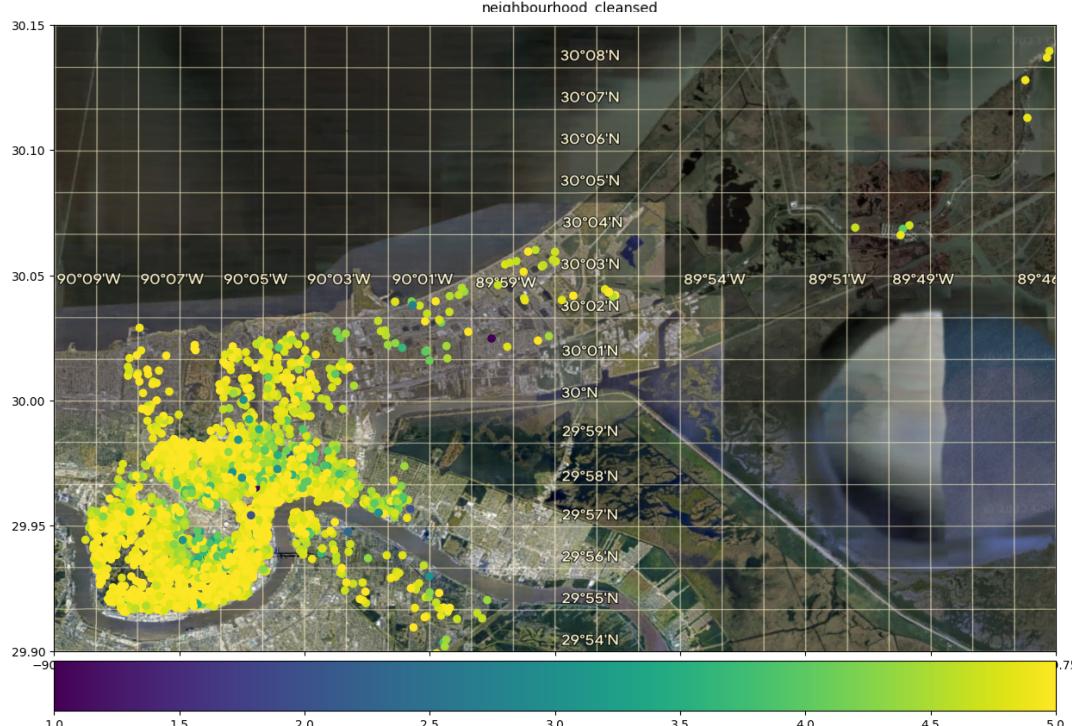
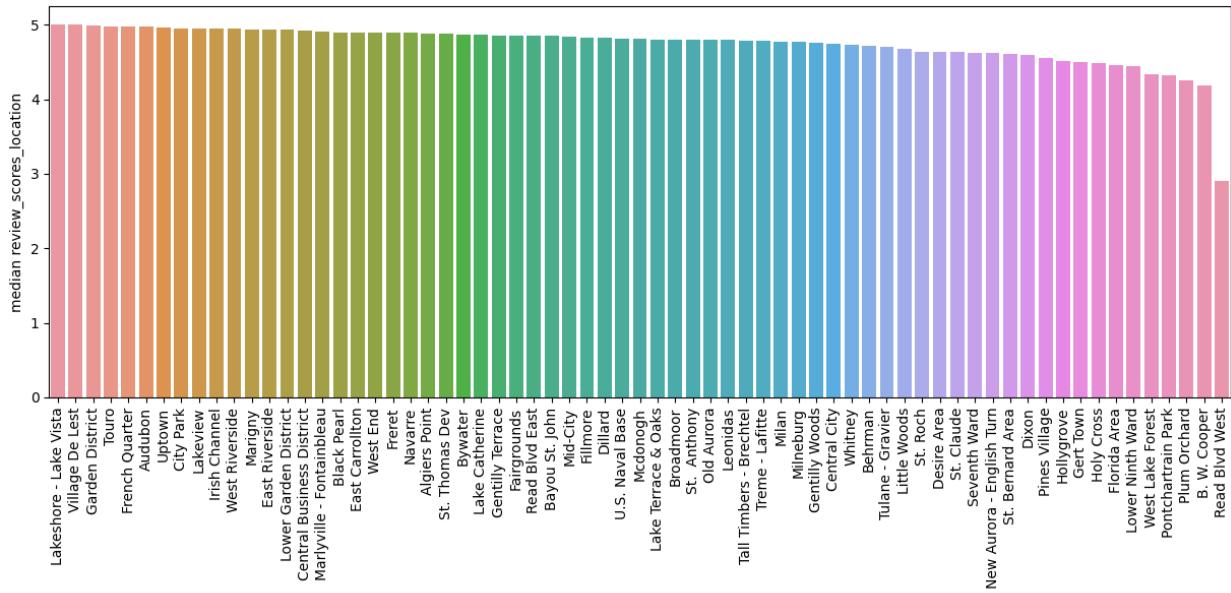




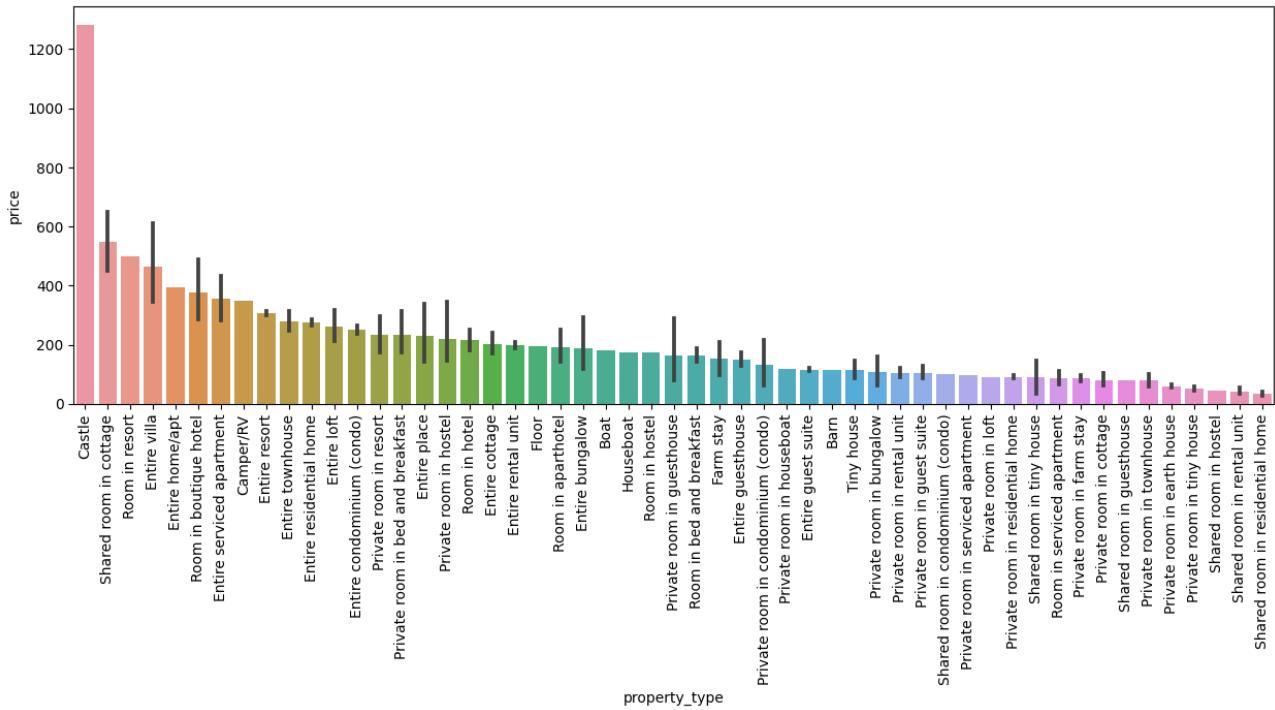
Answer:

As expected, luxury amenities such as private pools and hot tubs are associated with higher prices, while other amenities such as paid dryer and paid washer drive the price down (since you're already paying extra for those amenities)

5. What neighbourhoods have the best reviews for location?



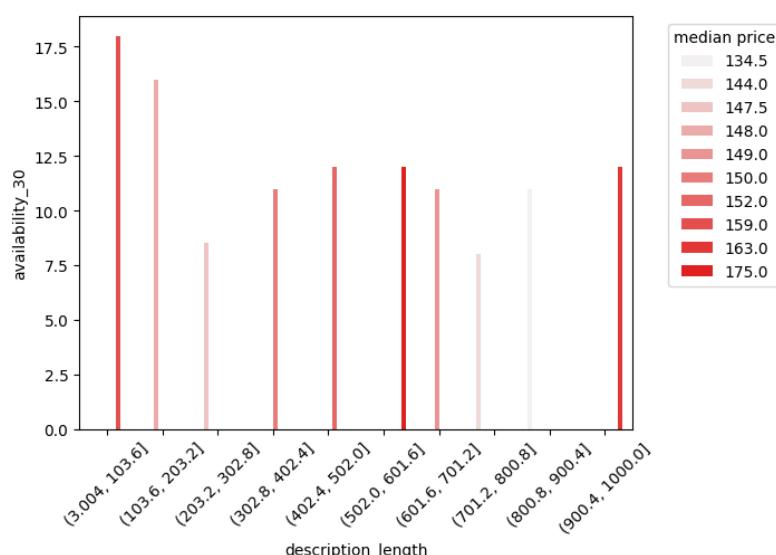
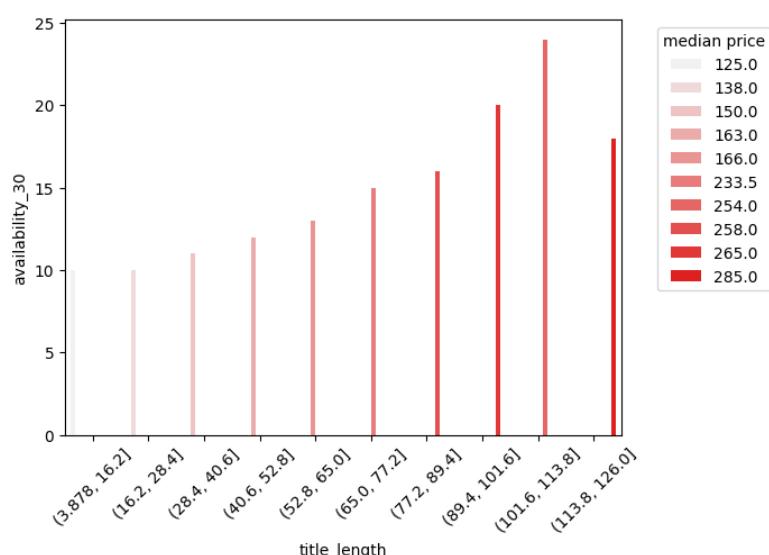
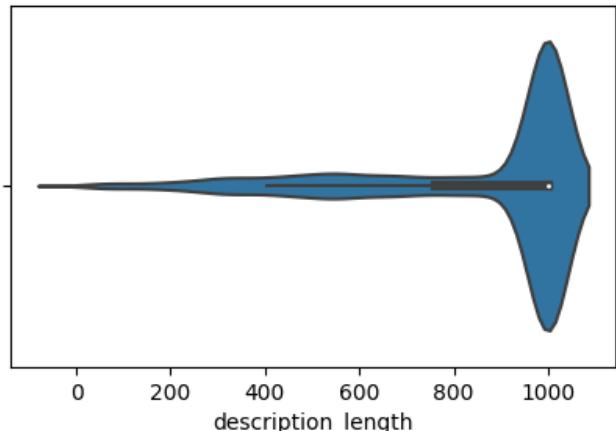
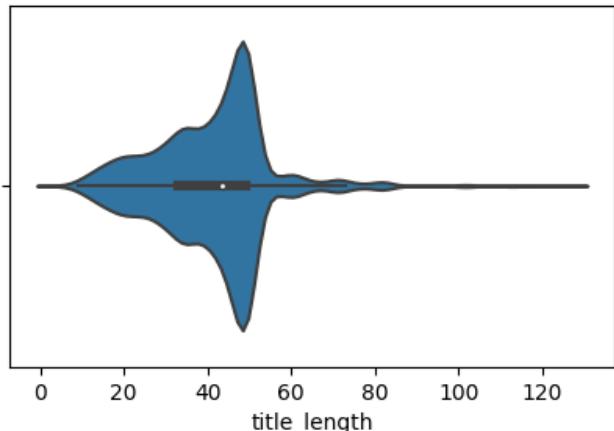
6. What are the property types with the highest median prices?



Answer:

As expected, the castle is the most expensive property type, while the shared rooms are at the bottom of the list in terms of price.

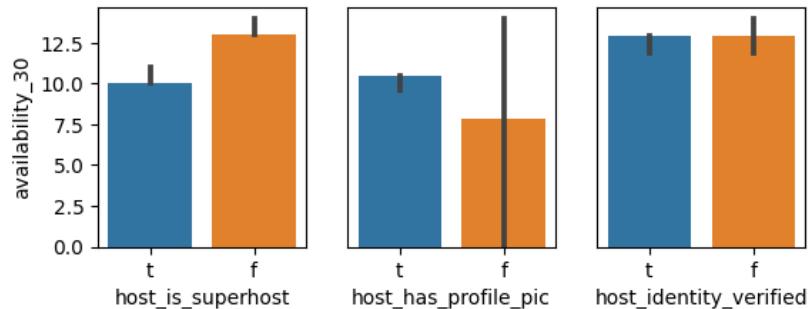
7. Could a certain description/title length in characters be more appealing and thus more popular (lowering the availability).



Answer:

There is a positive correlation between title_length, availability_30 and price. Since availability_30 and price are already correlated, a fair assumption would be that hosts with expensive listings tend to increase the title length.

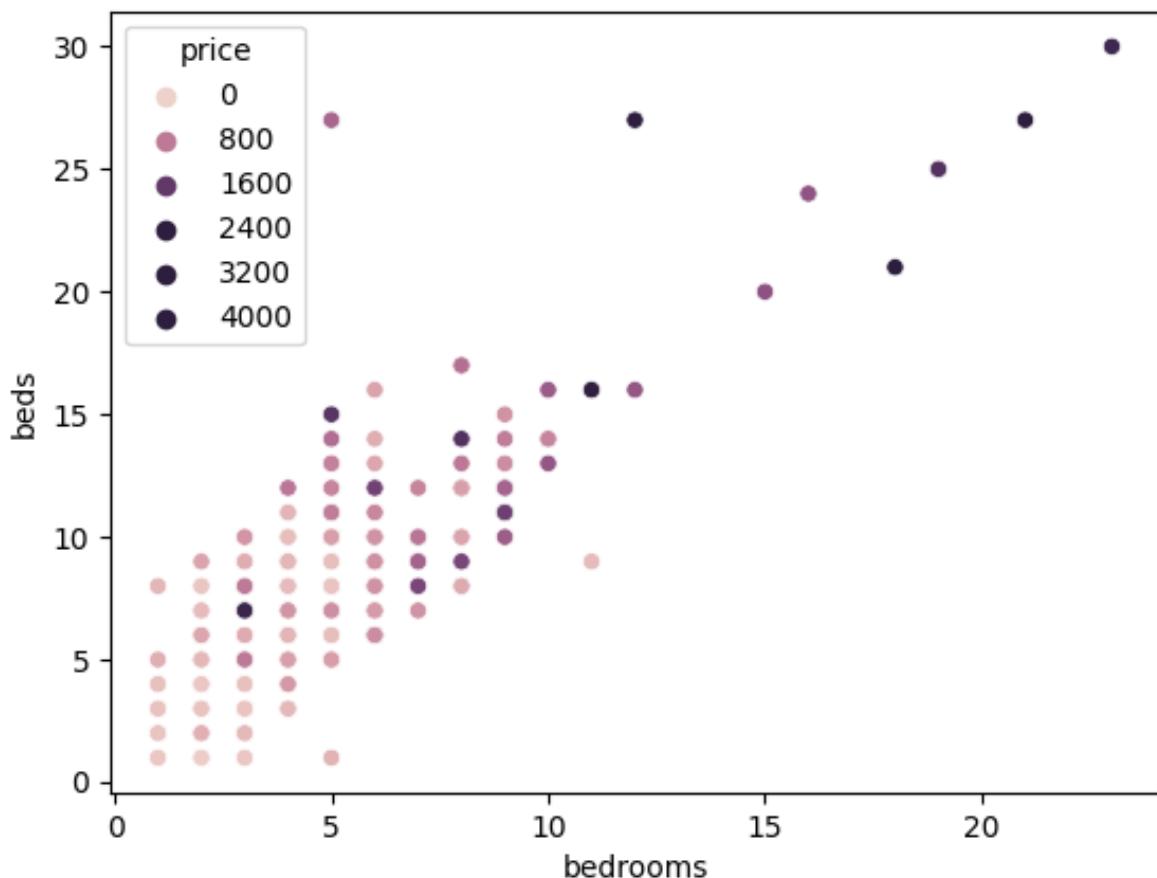
8. Will a profile picture, verified identity and the status of super host increase the popularity (lower availability) of the listing?



Answer:

- host_is_superhost → more popular listings
- host_has_profile_pic → less popular listings (majority of hosts don't have a profile pic)
- host_identity_verified → no impact

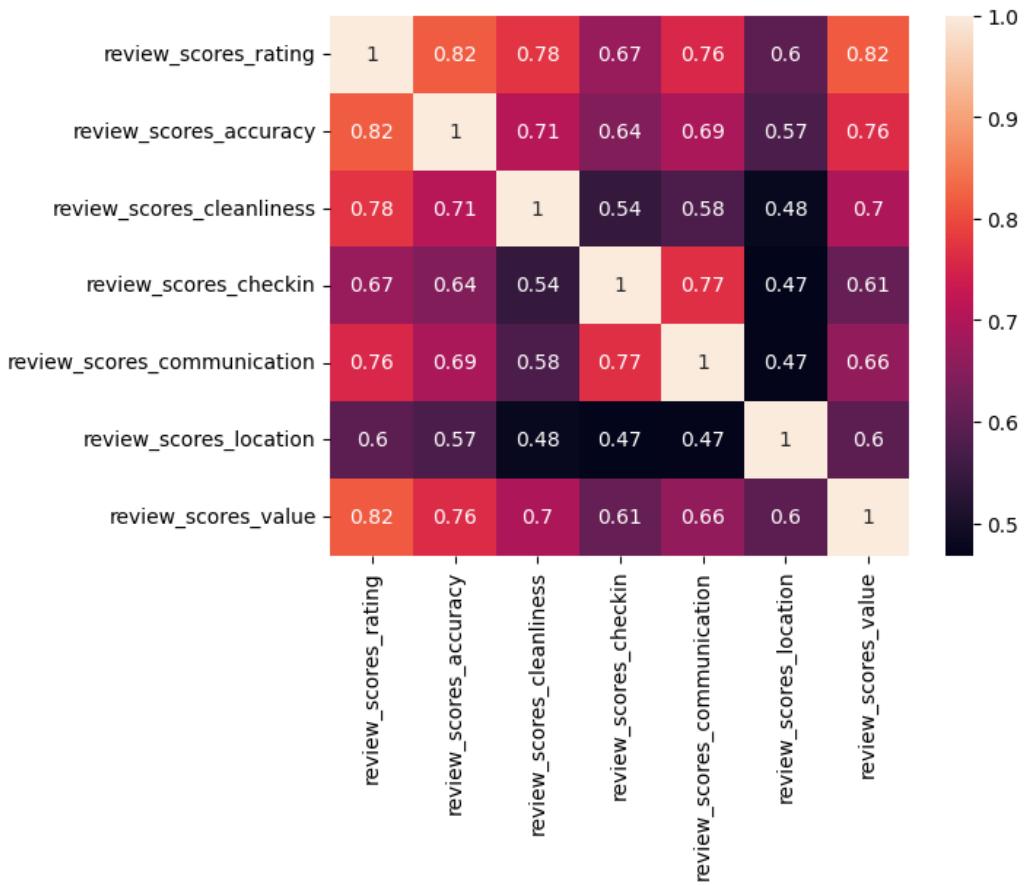
9. Does the number of bedrooms and beds increase the price?



Answer:

Obvious expectations with obvious results.

10. (Extra question) Is there a correlation between the types of reviews?



Answer:

Yes, but some types are better correlated than others, such as checkin with communication and accuracy with rating.

Also, location is the least correlated type, which makes sense since it shouldn't depend on other factors.

Python Pros:

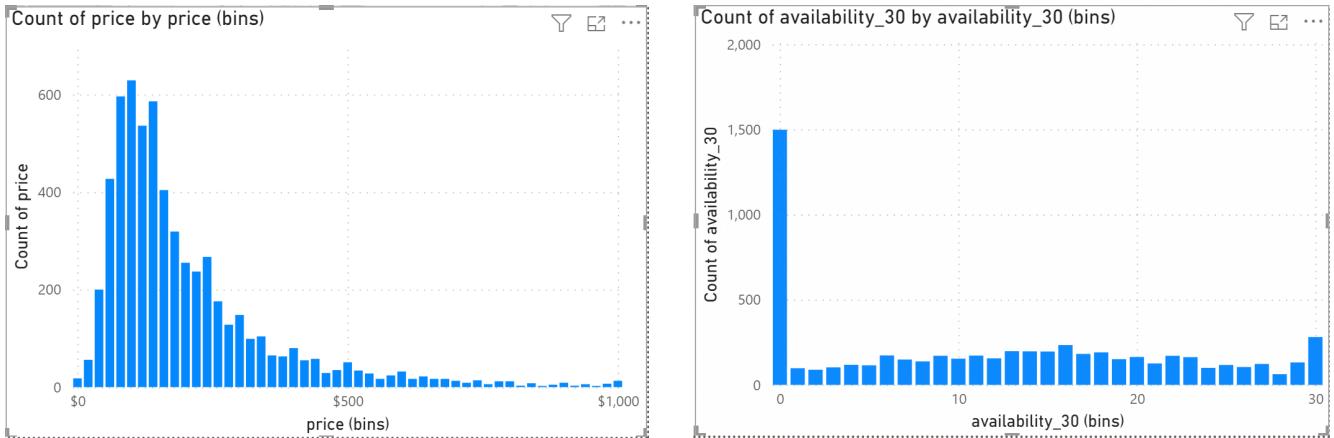
- very flexible (answering the amenities and title/description length questions)
- can automate tasks (plotting multiple visualisations -> the violin plot for every feature)

Python Cons:

- slower to implement
- longer learning curve —> reading documentation/online resources required

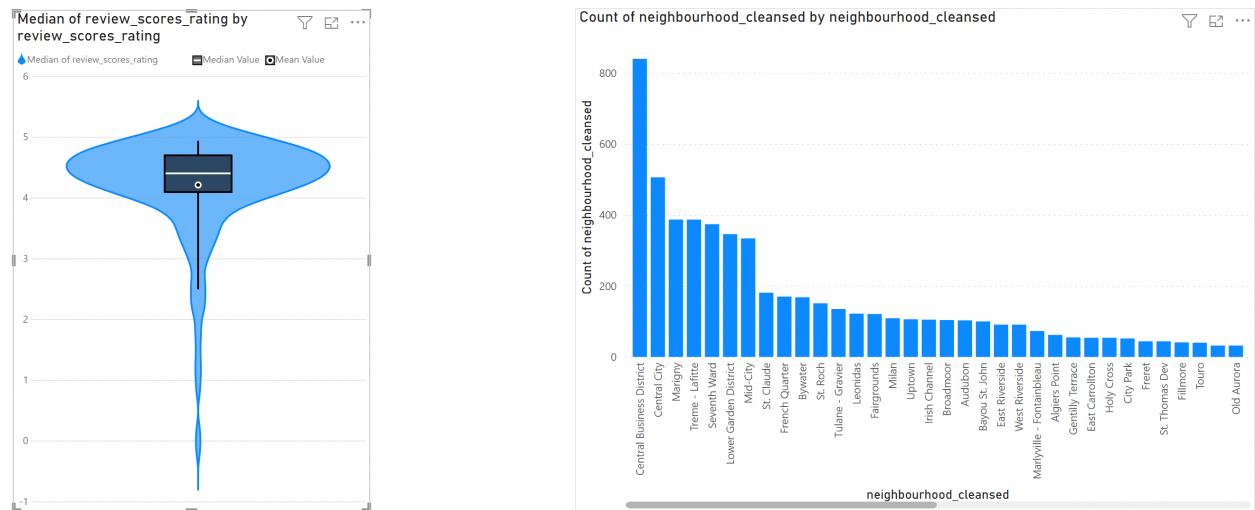
PowerBI Visualisation

Plotting the histograms for price and availability:



Individual features examples:

- Violin plot for the review_scores_rating
- Bar plot for the neighbourhood distribution

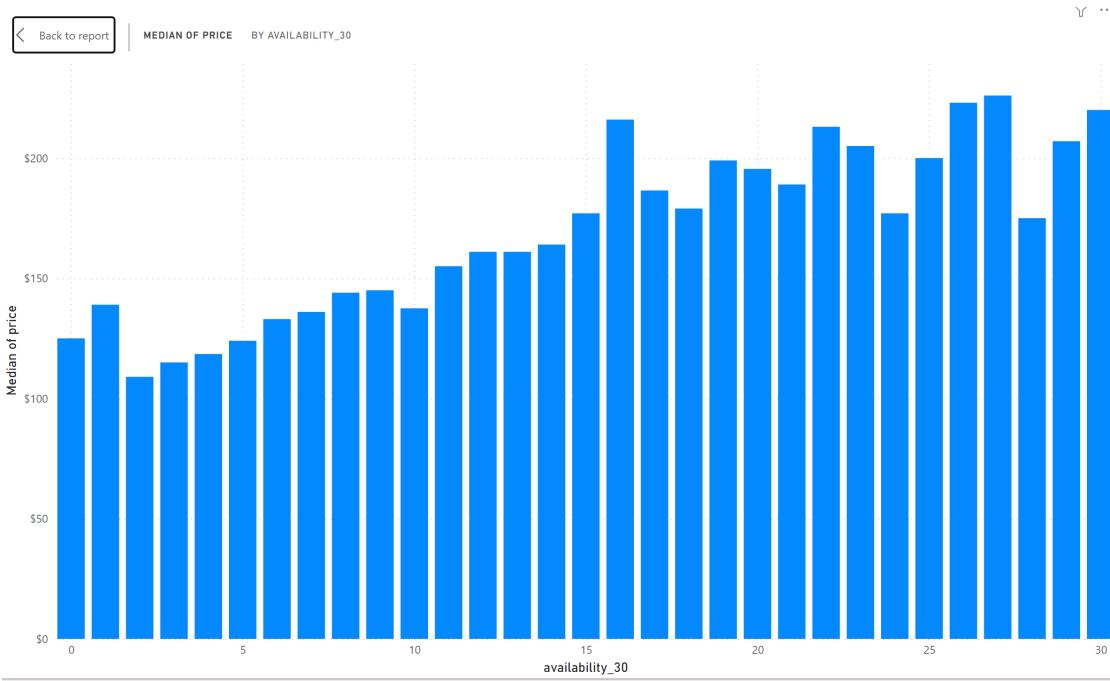


Questions answered:

1. Is the price correlated with the room type?

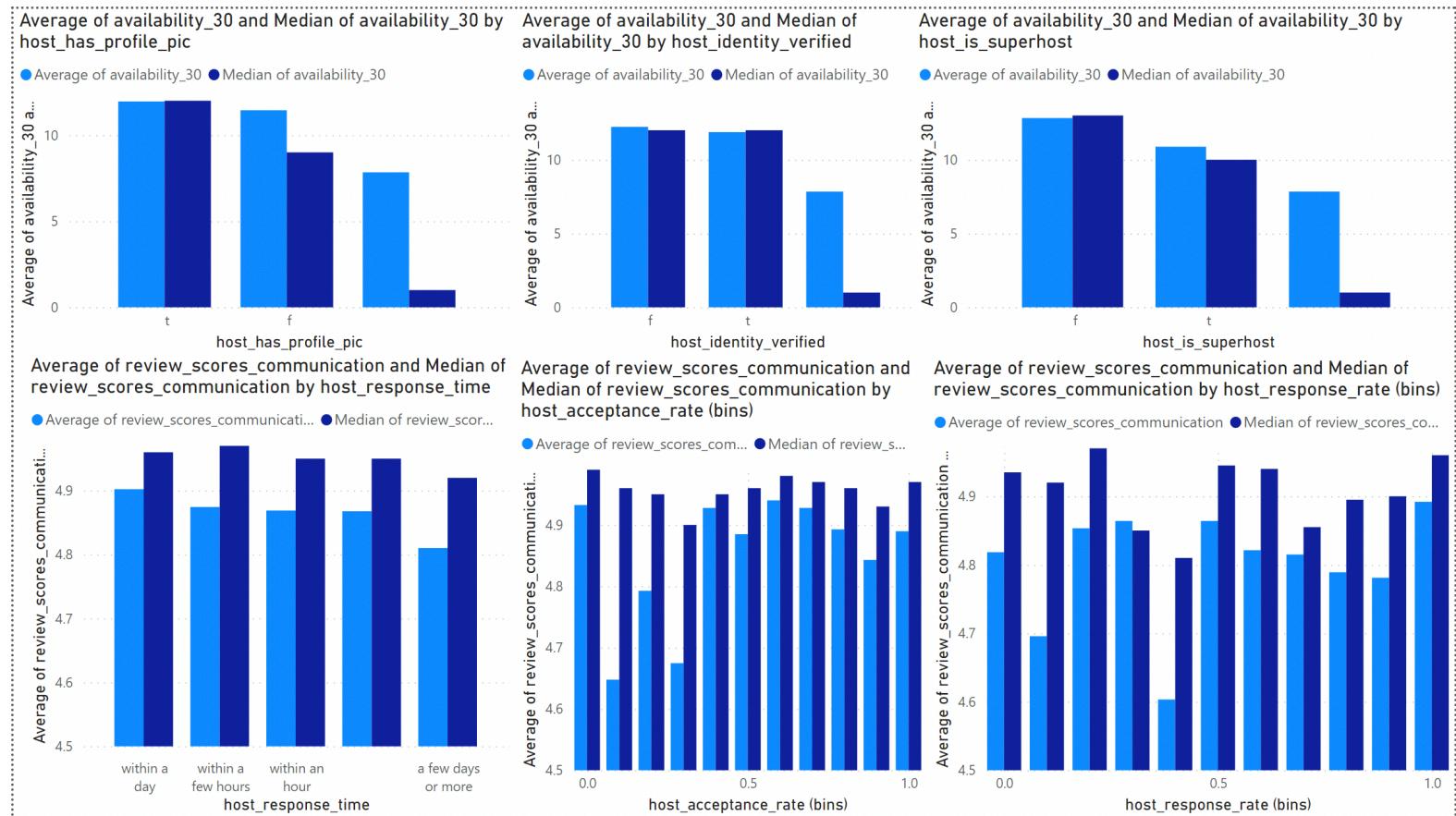


2. Does a lower price also lower the availability of a property?



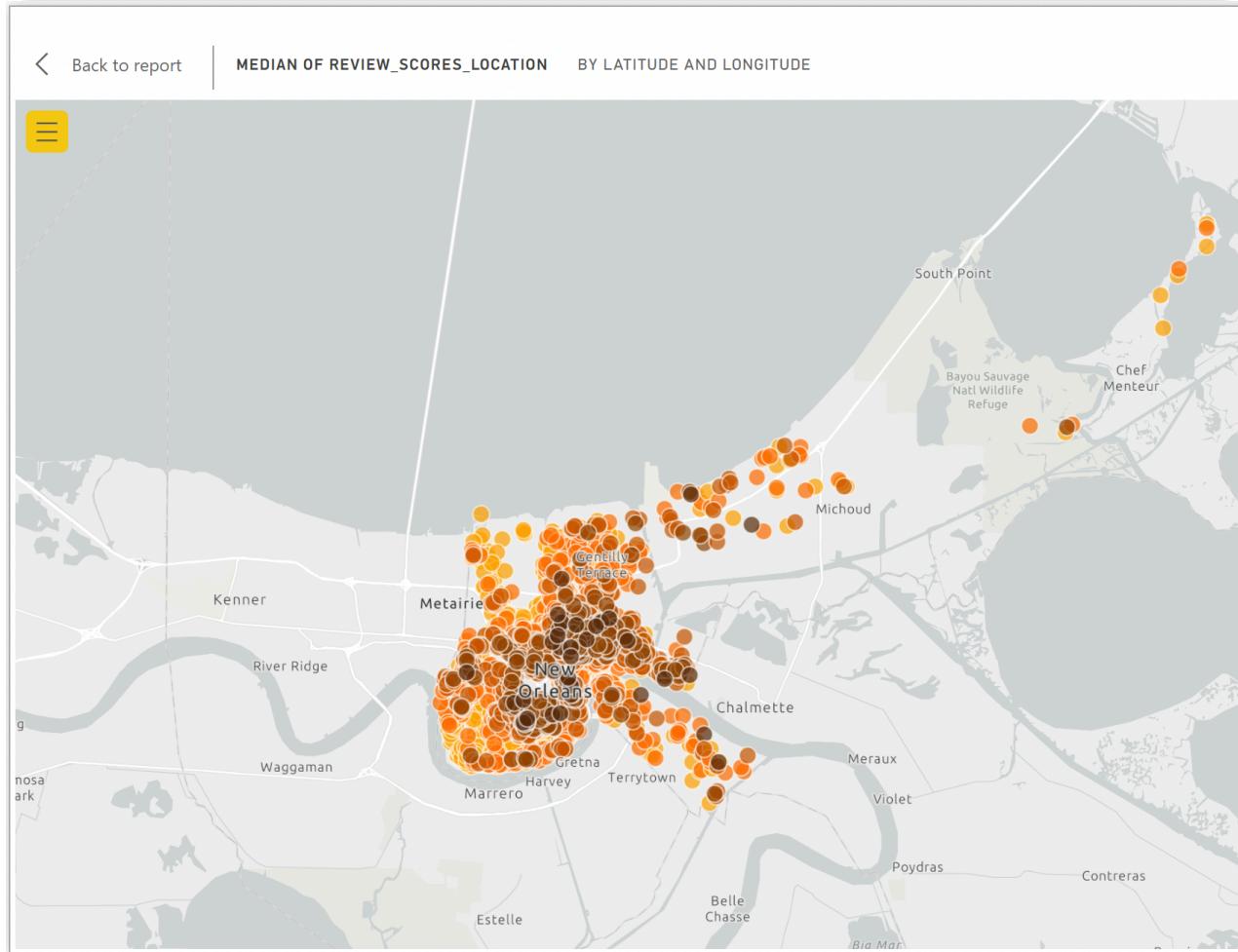
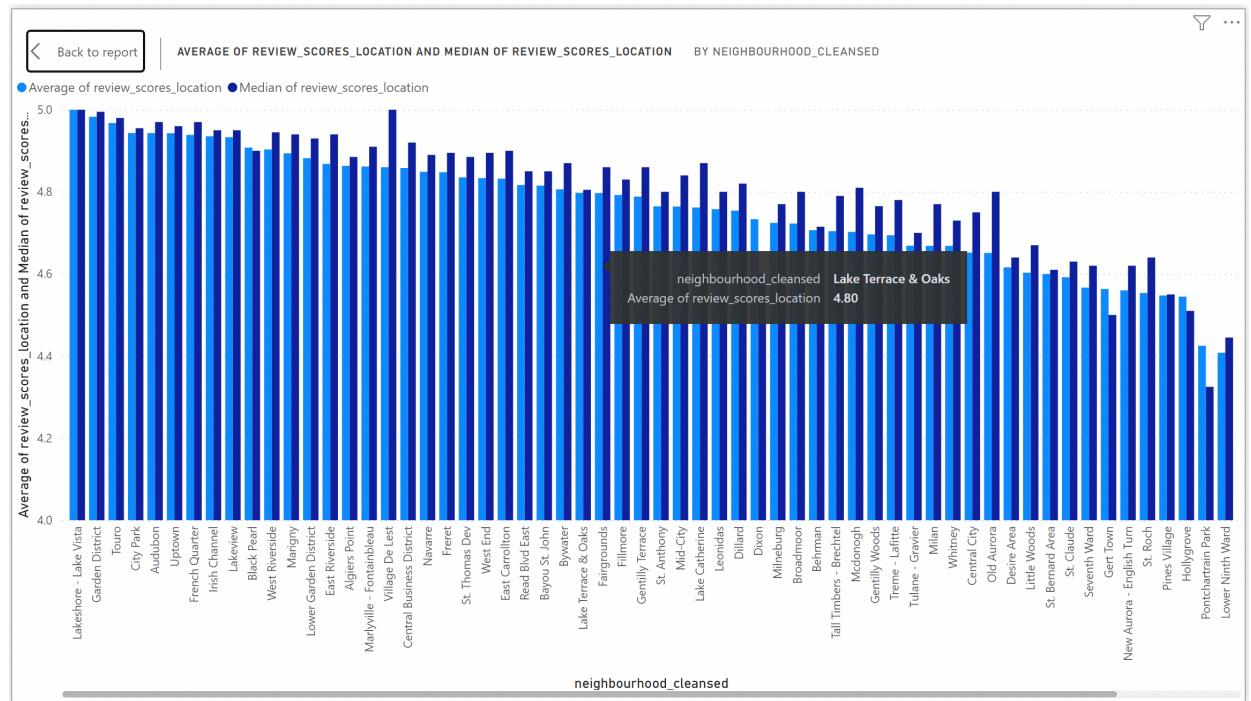
3&8. Are the reviews for communication correlated with the host response time, host response rate and host acceptance rate?

Will a profile picture, verified identity and the status of super host increase the popularity (lower availability) of the listing?



4. What amenities drive the price up? For this question, we selected only the amenities that appear at least in 10 or more listings. → Too specific for PowerBI

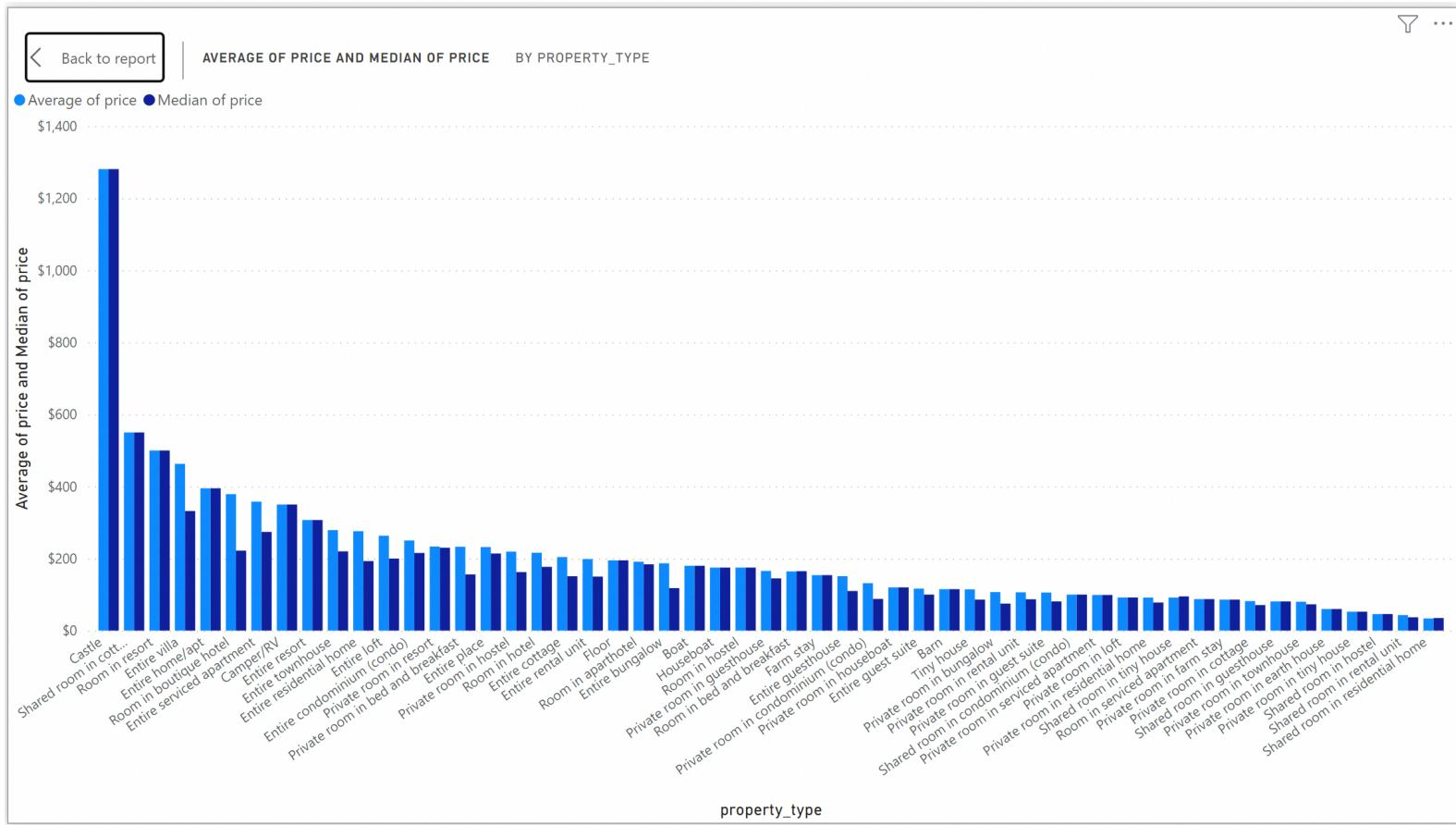
5. What neighbourhoods have the best reviews for location?



Bright → High review scores for location

Dark → Low review scores for location

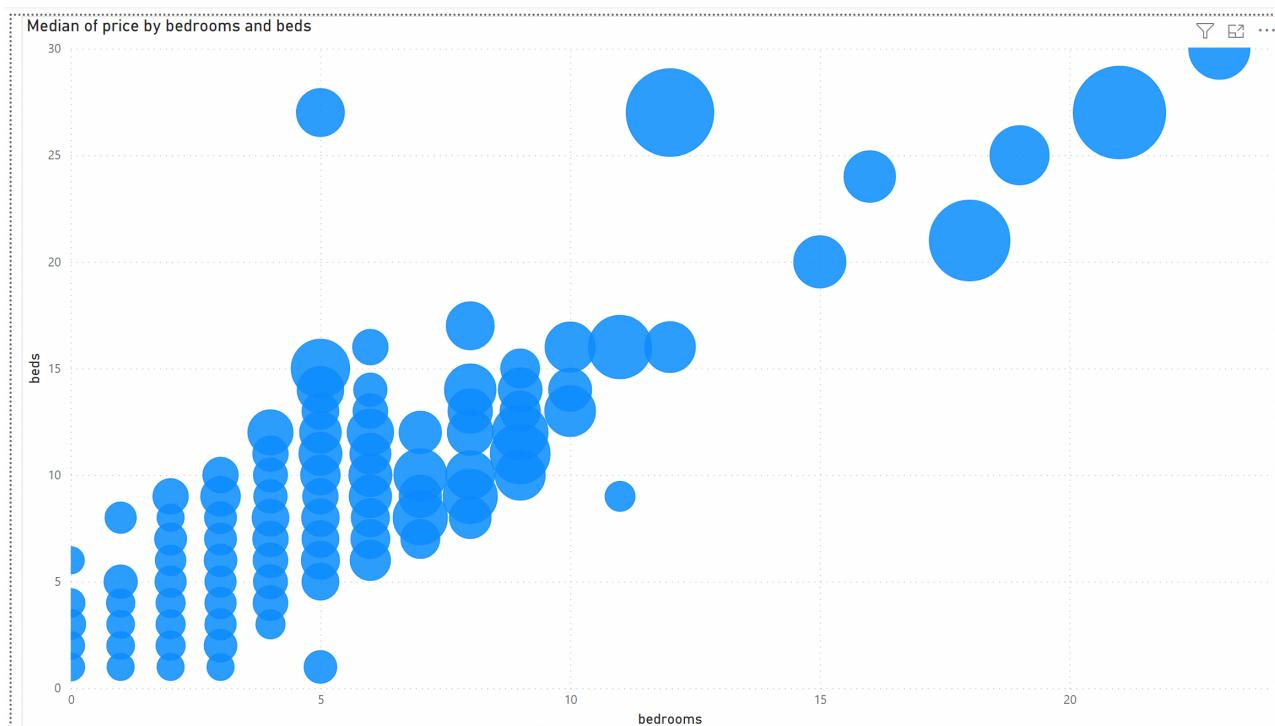
6. What are the property types with the highest prices?



7. Could a certain description/title length in characters be more appealing and thus more popular (lowering the availability).—> Too specific for PowerBI

8. See question no. 3

9. Does the number of bedrooms and beds increase the price?

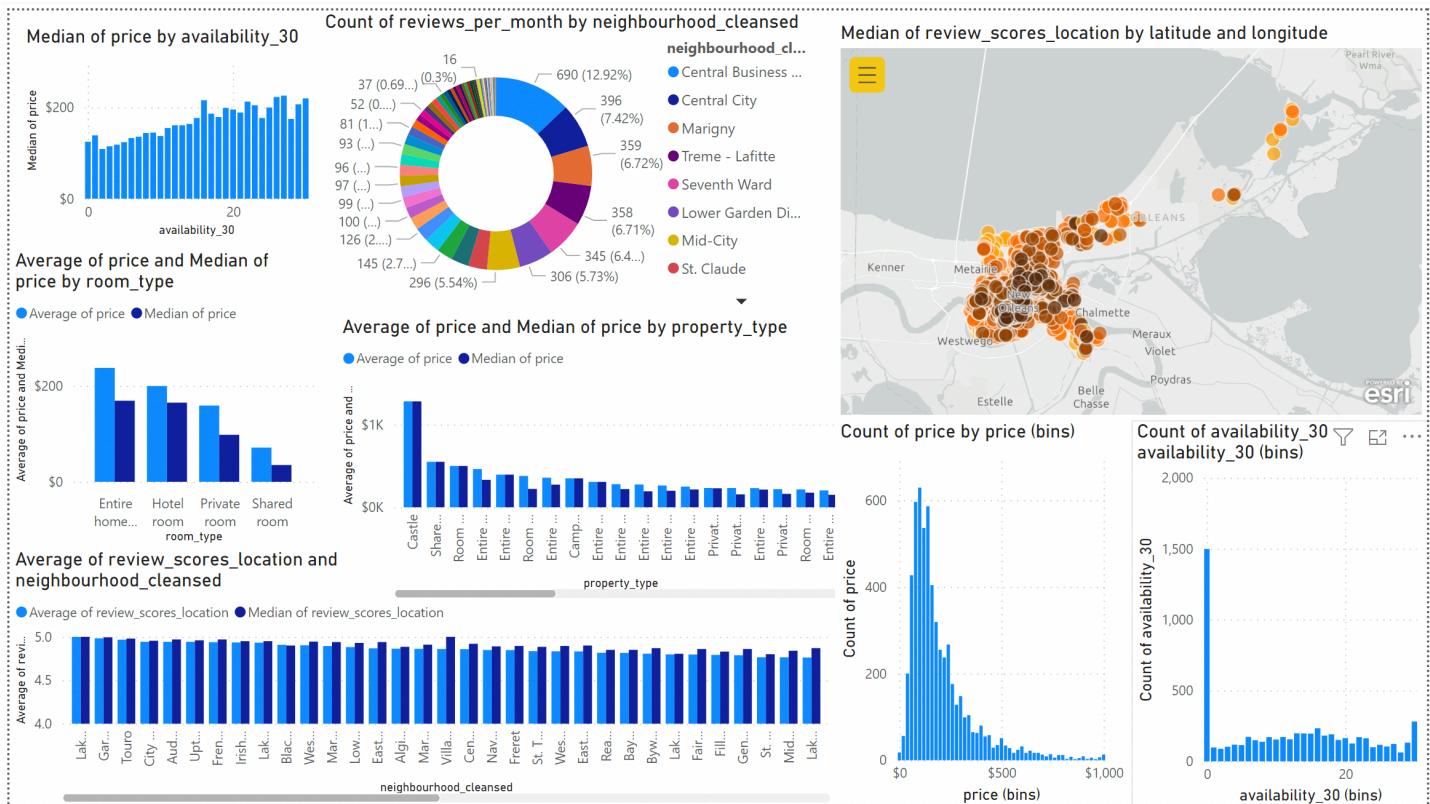


10. (Extra question) Is there a correlation between the types of reviews? —> Couldn't find the correlation matrix visualisation

PowerBi pros:

- The ability to make a system of plots that interact with each other (**MAJOR PRO**)

Initial dashboard:



Dashboard after selecting a specific category (Central City neighbourhood):



- Easy to use: drag and drop interface
- Library of many downloadable visuals
- More built-in data types (percent(%), price(\$)) —> no need for data transformation

PowerBi cons:

- Limited capabilities (the unanswered questions)
- Confusing on the first time

And the winner is... both

Python is the choice for specific problems, while PowerBI is better at visualising data and the relations between the features in a fast manner.