

# Exploratory Data Analysis of Airbnb listings in Paris, France\*

Aryaman Sharma

March 5, 2024

We start by looking at the distribution and properties of individual variables.

## 1 Prices

We see in (Figure 1a), the distribution of prices. We see that we consider a log scale for (Figure 1b) in order to take into account the outliers.

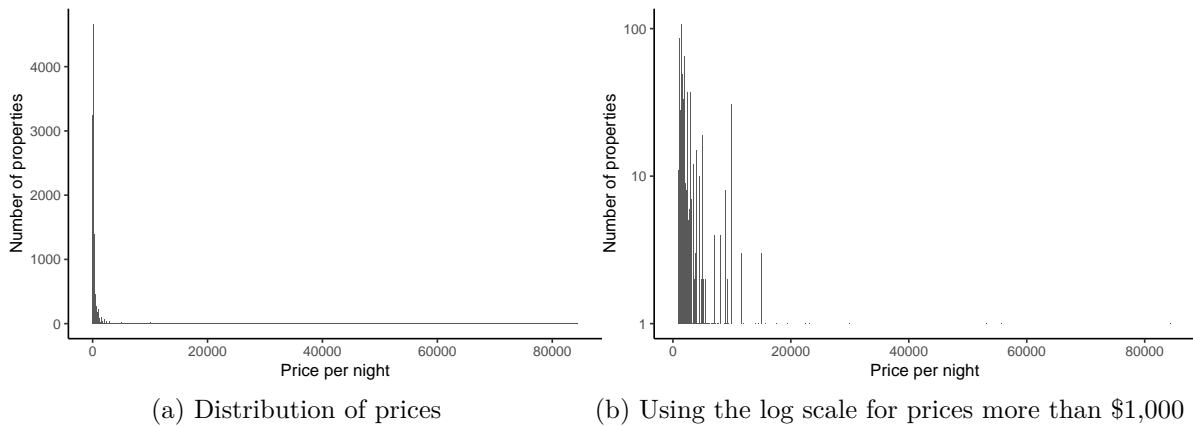
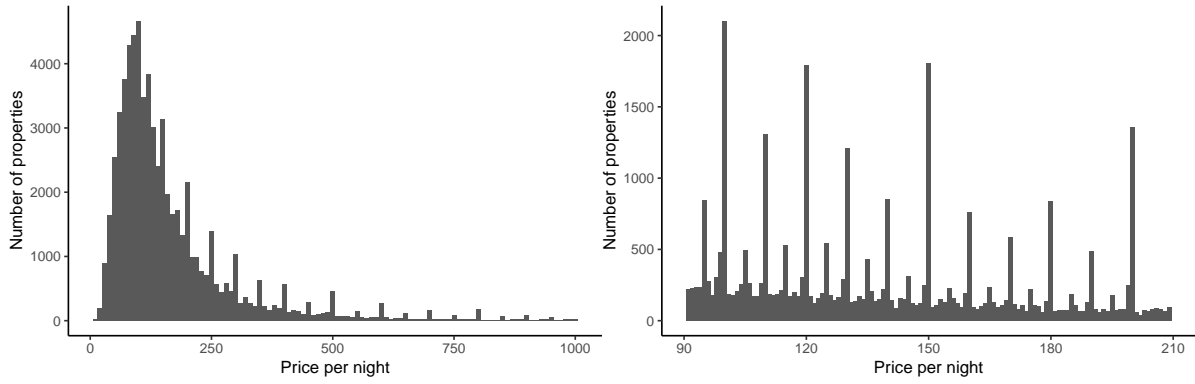


Figure 1: Distribution of prices of Paris Airbnb rentals

Moving on to (Figure 2a) we focus on prices lesser than \$1000, where we observe that most properties have a nightly price of less than \$250. In (Figure 2b) we specifically look at listing prices between \$90 and \$210.

---

\*Code and data are available at: <https://github.com/Ary4m3n/airbnb-paris.git>



(a) Prices less than \$1,000 suggest some bunching (b) Prices between \$90 and \$210 illustrate the bunching more clearly

Figure 2: Distribution of prices of Paris Airbnb rentals

## 2 Reviews

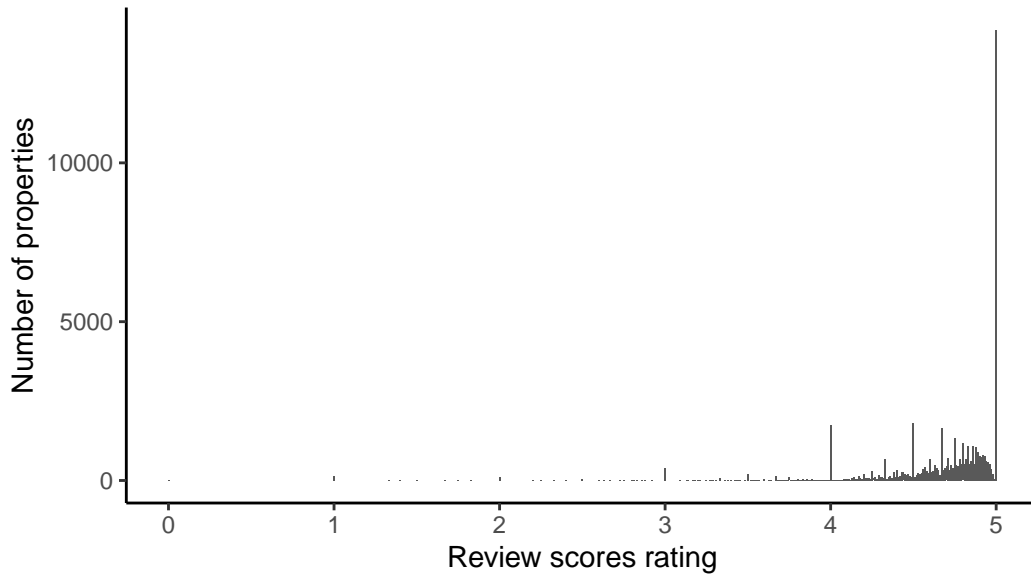


Figure 3: Distribution of review scores rating for Paris Airbnb rentals

Here we have just kept listing prices of under \$1000. On Airbnb, generally, guests can give 1 to 5 star ratings across a variety of different aspects, but when we look at the reviews in our dataset, as represented in (Figure 3) it is clear that it is effectively a binary, and almost entirely the case that either the rating is five stars or not.

We observe that there is a large number of listings that do not have a review rating yet and are this NA fields. In (Figure 4) we focus on listing reviews that are not missing and have a score.

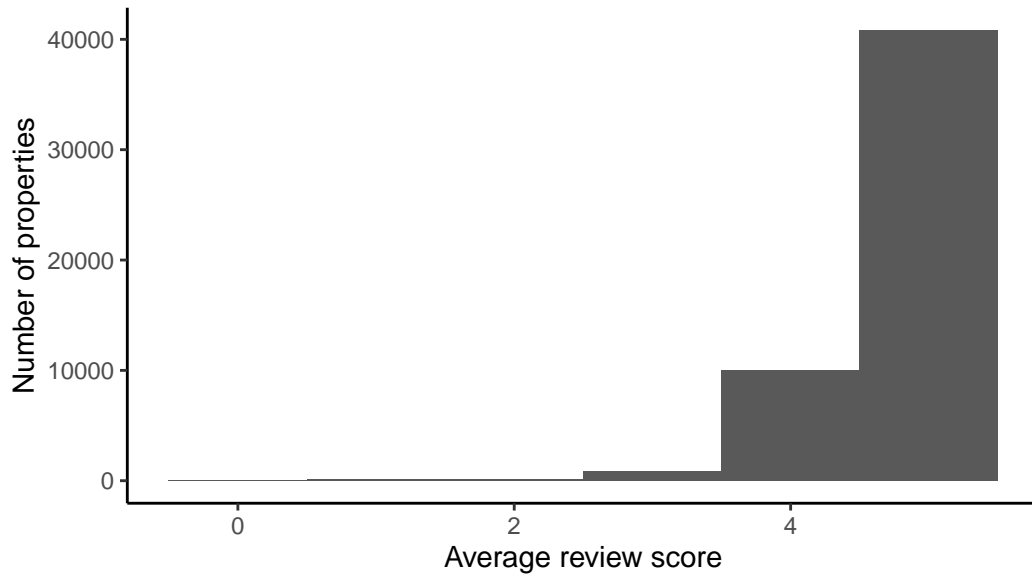


Figure 4: Distribution of review scores rating for Paris Airbnb rentals

We observe a large majority of 5 star ratings and generally no 0 star ratings in Figure 4.

### 3 Host Response Time

We observe a large amount of NA fields. In Figure 5 we explore the distribution of review scores for NA host response times for Paris Airbnb rentals.

Generally, missing values are dropped by our plots. Figure 6 includes them in the graph.

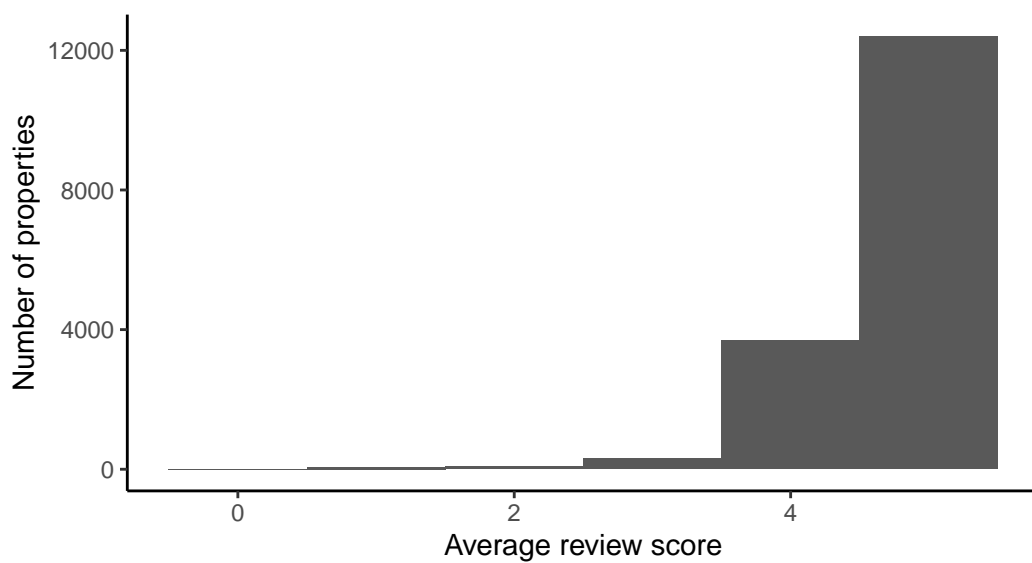


Figure 5: Distribution of review scores for properties with NA response time, for Paris Airbnb rentals

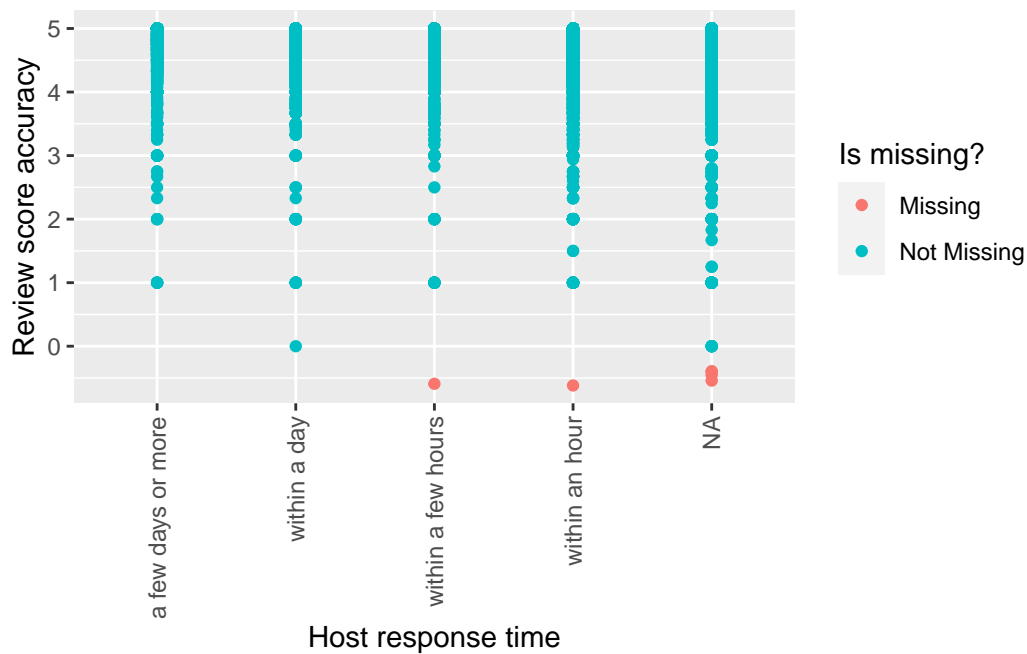
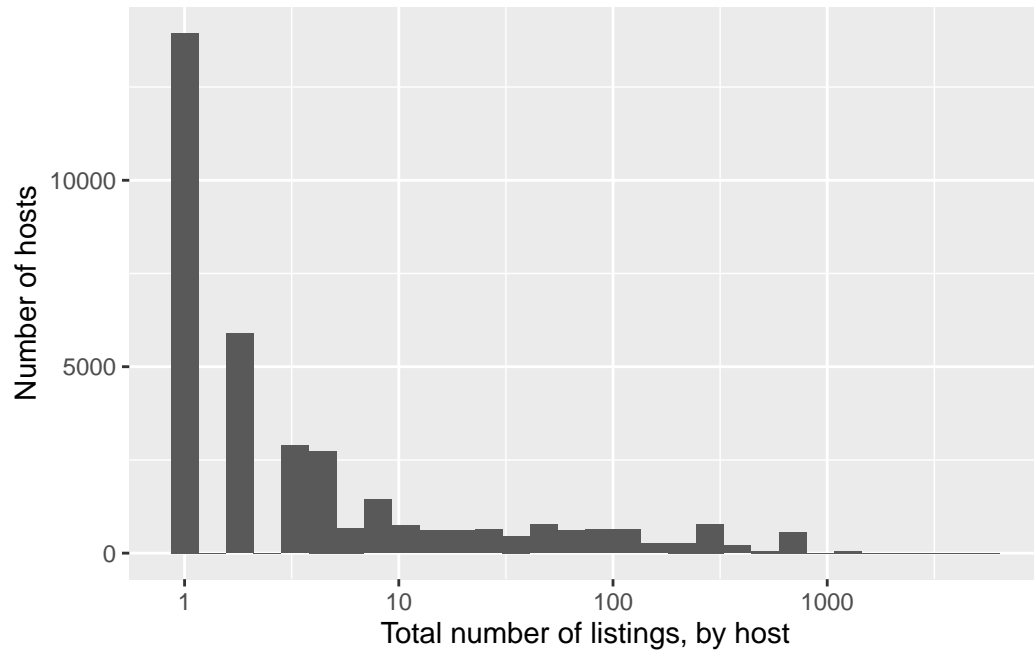


Figure 6: Missing values in Paris Airbnb data, by host response time

## 4 Number of Properties by a host on Airbnb

Figure 7 represents the distribution of the number of properties a host has listed on Airbnb for Paris Airbnb rentals.



## 5 Relationship Between Variables

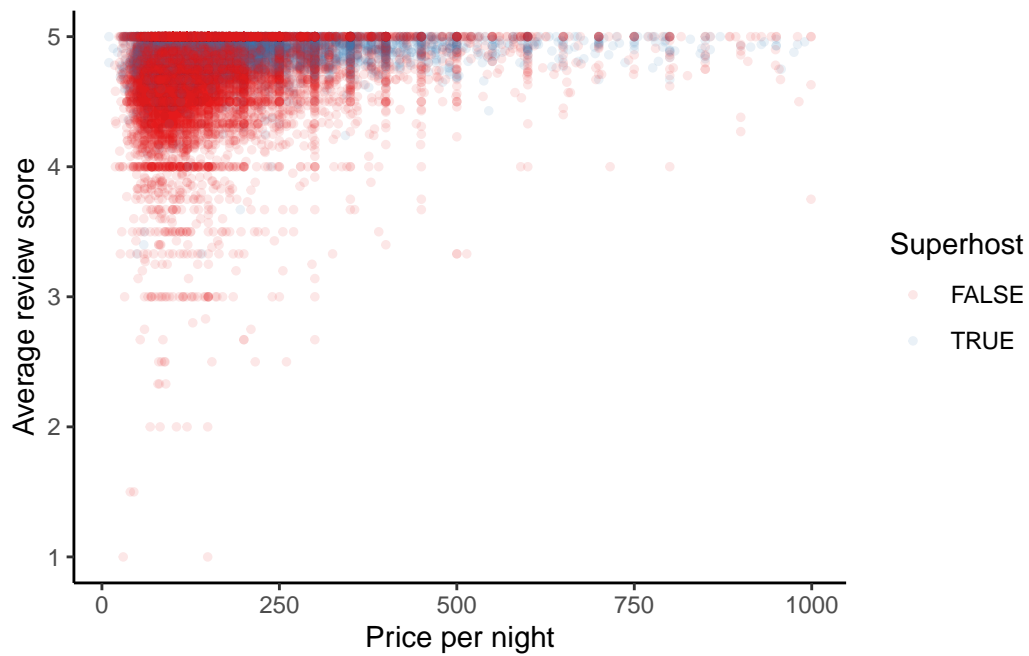


Figure 8: Relationship between price and review and whether a host is a superhost, for Paris Airbnb rentals

In Figure 8, we look at the relationship between price and reviews, and whether they are a super-host, for rental properties with more than one review.

We now run our model and present our results and model summary in Table 1.

Table 1: Explaining whether a host is a superhost based on their response time

	(1)
(Intercept)	−16.262 (0.481)
host_response_timewithin a day	2.019 (0.211)
host_response_timewithin a few hours	2.695 (0.210)
host_response_timewithin an hour	2.972 (0.209)
review_scores_rating	2.624 (0.089)
Num.Obs.	22 047
AIC	24 165.0
BIC	24 205.0
Log.Lik.	−12 077.507
RMSE	0.43