Prices, Response Times and Review Scores for Airbnb in Paris in December 2023

Predicting superhost status based on these factors

Boxuan Yi

28 February 2024

1 Introduction

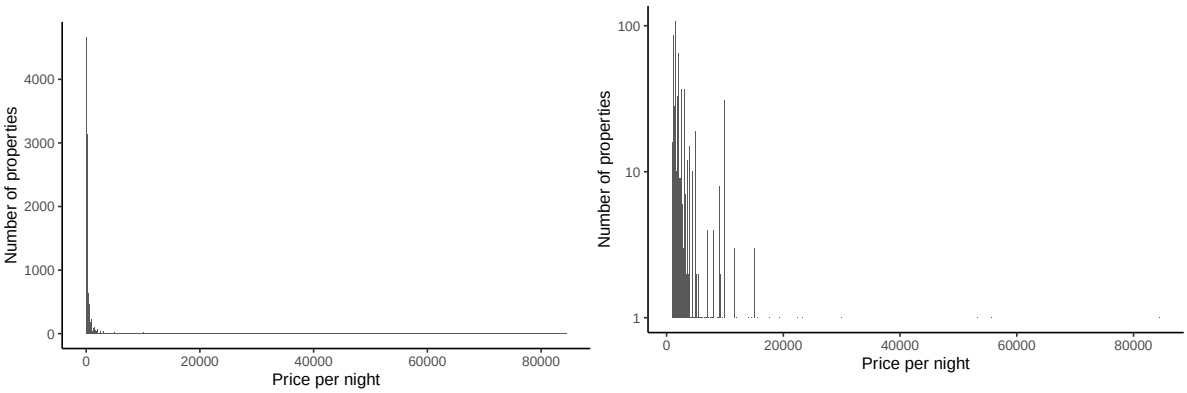
Paris stands as an iconic destination, renowned for its art, culture, and rich history. This exploratory data analysis (EDA) will use the Airbnb listings in Paris, as at 12 December 2023, to explore the distribution of prices, review scores, and hosts' response times, aiming to forecast whether a host is a superhost based on these factors. The dataset is obtained from Inside Airbnb (InsideAirbnb 2024).

The data will be processed and analyzed in R (R Core Team 2022) using packages readr (Wickham, Hester, and Bryan 2024), dplyr (Wickham et al. 2023), arrow (Richardson et al. 2024), stringr (Wickham 2023), ggplot2 (Wickham 2016), naniar (Tierney and Cook 2023), janitor (Firke 2021), modelsummary (Arel-Bundock 2022), and knitr (Xie 2014).

2 Data Analysis

Figure 1 (a) illustrates the distribution of nightly prices for Airbnb rentals in Paris in December 2023. The x-axis represents the price, and the y-axis shows the count of properties within each price range. Figure 1 (b) focus on the airbnb with price exceeding 1000. This graph uses logarithmic scale that helps to visualize and compare higher prices without completely compressing the lower values.

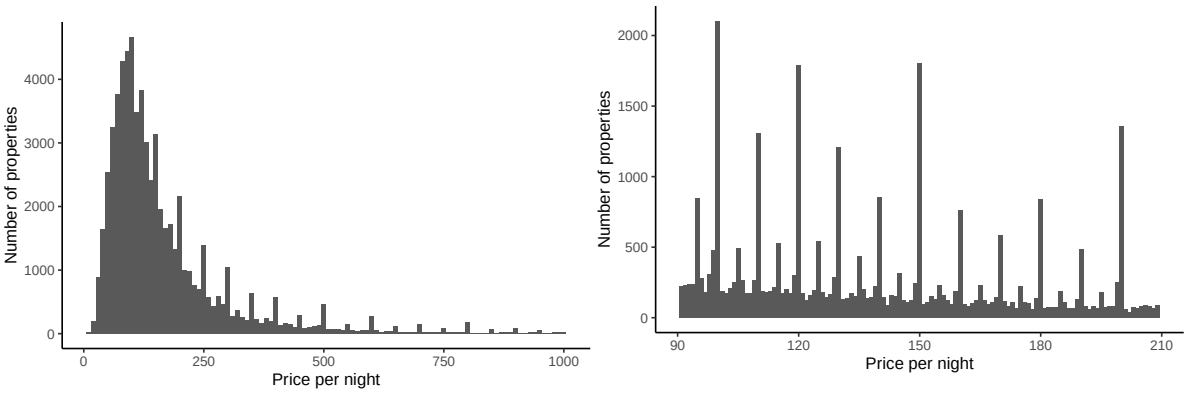
Focusing on prices that are less than 1000 dollars, we can see that most properties have a price less than $250 dollars per night from Figure 2 (a). The distribution of prices shows some bunching, indicating a tendency for prices to cluster around specific values rather than being uniformly spread across the entire range. For a closer inspection of this bunching phenomenon, Figure 2 only shows the distribution of prices between 90 and 210. There are noticeable concentrations of prices, underscoring the non-uniform distribution pattern.



(a) Distribution of price

(b) Distribution of Prices Using the Log Scale for Prices More Than $1,000

Figure 1: Distribution of prices of Paris Airbnb rentals in December 2023



(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 for Airbnb listings in Paris in December 2023

Limiting our focus to properties with complete information on whether the host is a superhost, Figure 3 visualizes the distribution of review score ratings. Each bar in the graph represents a range of ratings, with the height indicating the number of properties falling within those ranges. The majority of scores are over 4, and many of them receive a perfect score of 5. Figure 4 illustrates the distribution of average review scores for Airbnb listings in Paris, excluding properties with missing review scores. Similarly, the majority of scores exceed 4.

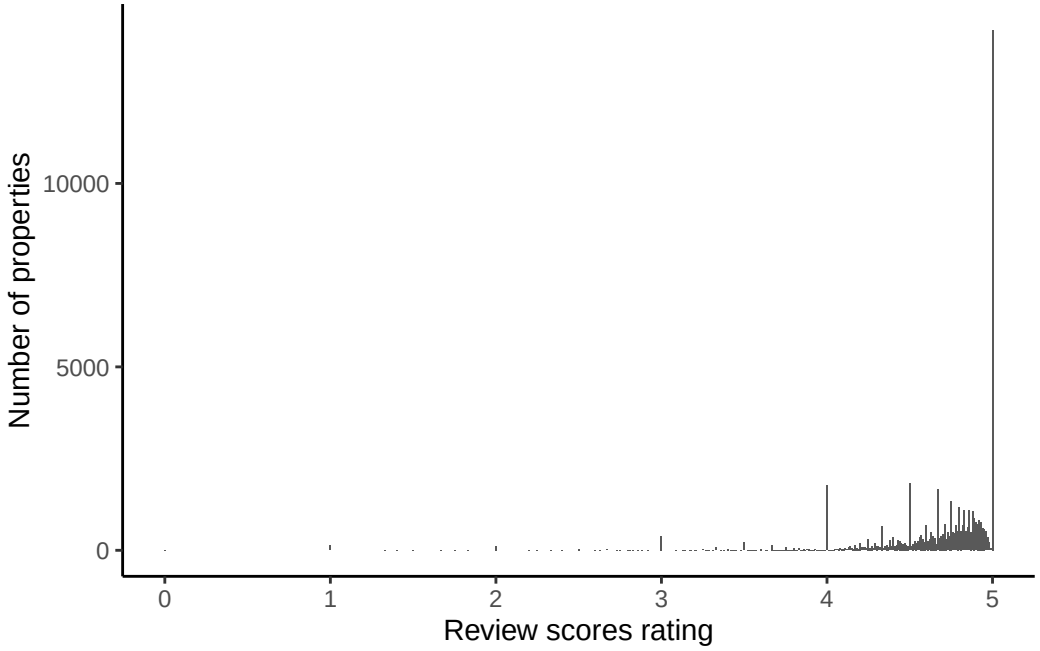


Figure 3: Distribution of review scores rating for Paris Airbnb rentals in December 2023

Out of the 51,978 properties with a review score and complete information about whether the host is a superhost, Table 1, a table of counts for different levels of the response time by hosts, reveals that over 22,000 properties have hosts responding within an hour, while 1,243 properties experience a response time of more than a day. Additionally, there are 16,533 properties with missing response time data. To check whether the absence of response time data is related to review scores, Figure 5 shows the distribution of review scores for properties with missing response time. This distribution is actually similar to the distribution of the review scores for all the Airbnb rentals in Paris, with the majority scoring over 4

Table 1: Distribution of Response Time by Hosts for Paris Airbnb in December 2023

host_response_time	n
a few days or more	1243
within a day	5297
within a few hours	6811

Table 1: Distribution of Response Time by Hosts for Paris Airbnb in December 2023

host_response_time	n
within an hour	22094
NA	16533

Figure 6 depicts the association between hosts' response time and the accuracy of review scores, distinguished by the status of missing (in red) or non-missing (in blue). Notably, the accuracy of review scores for missing values appears significantly lower comparing to the accuracy for non-missing values.

For now, anyone with missing values in their response time will be excluded. Based on Figure 7, we can see there is a large number of hosts owning 1 to 10 properties, and the majority of hosts have only one property, with an expected long tail in the distribution.

The relationship between price, review scores, and superhost status, for properties with more than one review, is shown in Figure 8, with red representing the host is not a superhost, blue representing the superhost. There are a lot more non-superhosts than superhosts, and superhosts have a higher average review, regardless of property prices per night, than hosts that are not superhosts. Interestingly, Figure 8 shows the prices of properties without a superhost predominantly cluster under $250, while properties with a superhost display a more evenly distributed price range.

I construct a model predicting whether someone is a superhost based on response times and review scores, and the summary of the model is presented in Table 2. Each level of these factors has a positive correlation with the likelihood of being a superhost. Furthermore, having a host that responds within an hour emerges as the most crucial factor contributing to superhost designation in this dataset

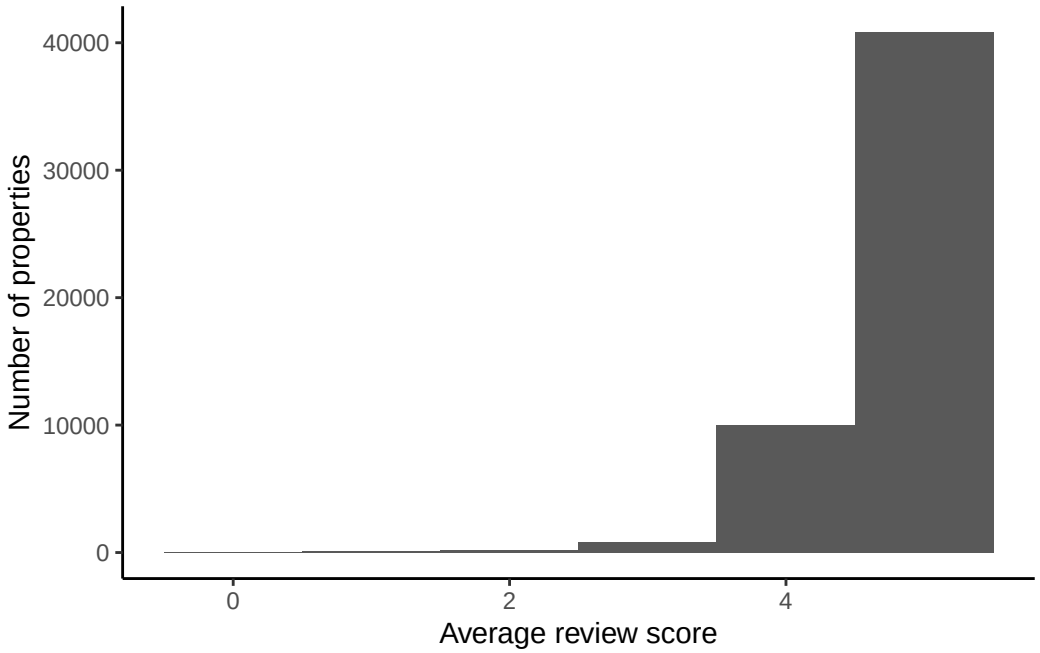


Figure 4: Distribution of review scores for Paris Airbnb rentals in December 2023

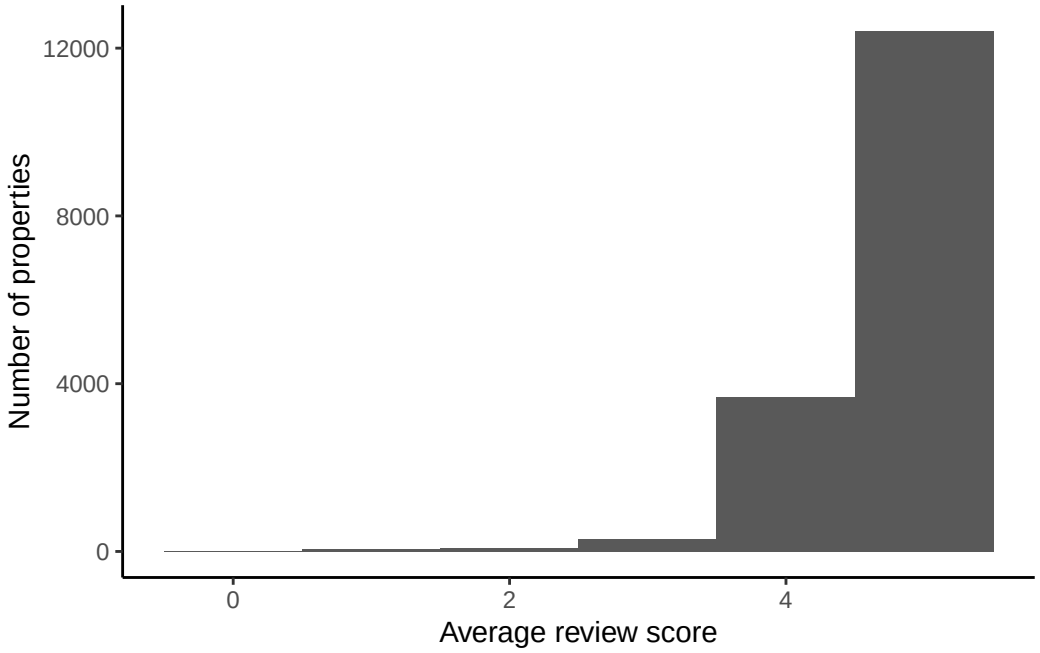


Figure 5: Distribution of review scores for properties with missing response time, for Paris Airbnb rentals in December 2023

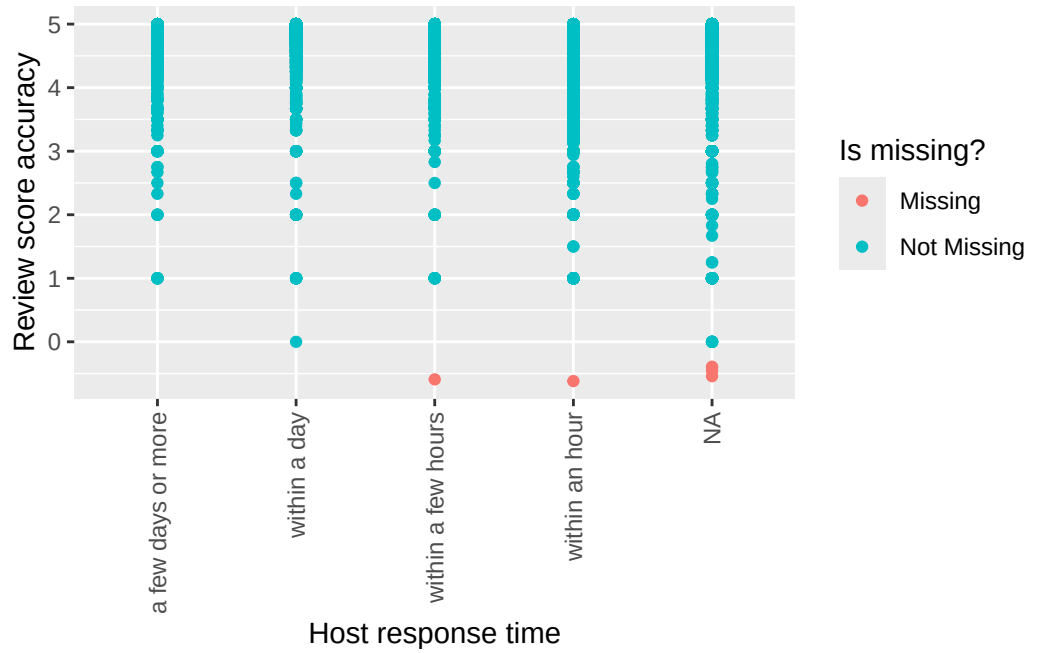


Figure 6: Missing values in Paris Airbnb data, by host response time

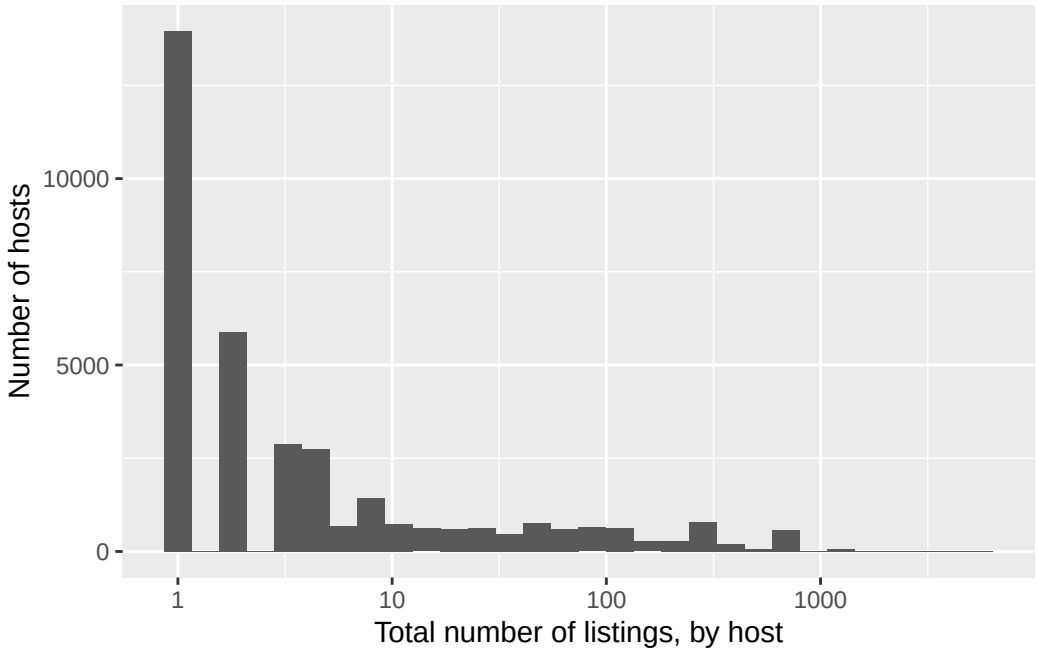


Figure 7: Distribution of the number of properties a host has on Airbnb, for Paris Airbnb rentals in December 2023

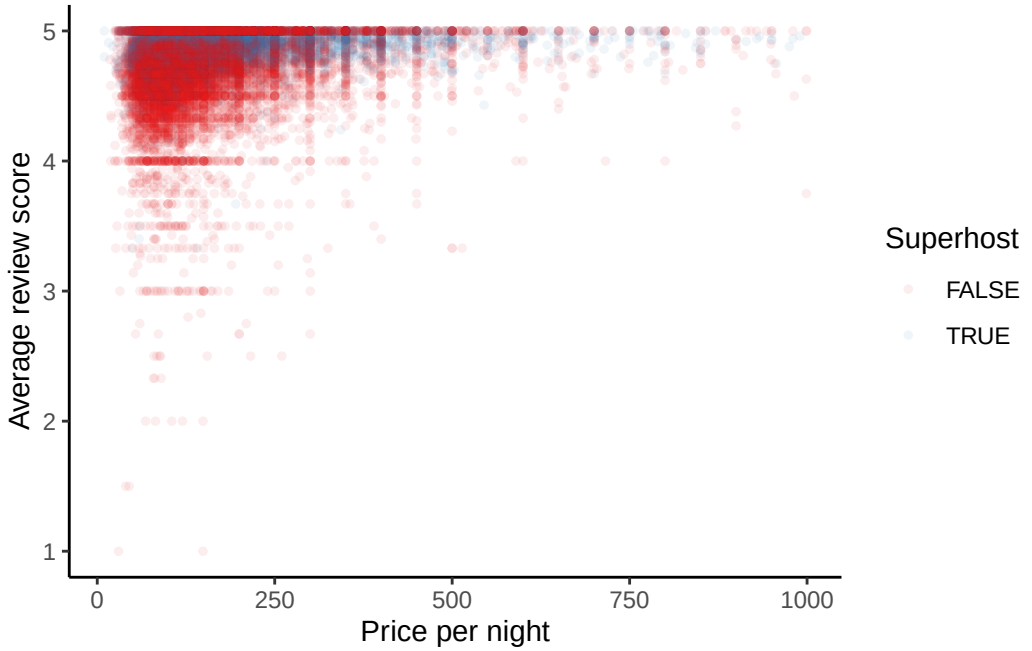


Figure 8: Relationship between price and review and whether a host is a superhost, for Paris Airbnb rentals in December 2023

Table 2: Explaining whether a host is a superhost based on their response time and review scores

	(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

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

InsideAirbnb. 2024. "Get the Data — Insideairbnb.com." http://insideairbnb.com/get-the-data/.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://github.com/apache/arrow/.

Tierney, Nicholas, and Dianne Cook. 2023. "Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations." *Journal of Statistical Software* 105 (7): 1–31. https://doi.org/10.18637/jss.v105.i07.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2023. *Stringr: Simple, Consistent Wrappers for Common String Operations*. https://stringr.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://readr.tidyverse.org.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.