1047904_Applied_Analytical_Statistics

January 15, 2021

1 Applied Analytical Statistics Summative

Candidate Number: 1047904

15 January, 2021

2 Good value for money? An explorative anaylsis of price determinants of Airbnb property listings in the city of Berlin

3 Abstract

Over the course of the last few years, peer-to-peer accommodation rental services like Airbnb have taken root as attractive source of income. Yet, hosts often grope in the dark when working out prices that adequately reflect the value of their properties. Based on a sample of 2,799 Airbnb rental properties for the city of Berlin, we investigate the effects of 13 explanatory variables spread across 4 categories – *roperty*, *management*, *reputation* and *host* – on price using ordinary least squares regression and quantile regression. Findings qualify property type as most important characteristic in determining listings prices followed by a set of capacity-related variables like number of bathrooms, bedrooms and guest capacity. Mixed effects arise for amenities featured by the apartment and reputation-related attributes such as number of reviews and user rating scores. Quantile regression results further elucidate that the magnitude of unearthed relationships changes for different parts of the price range. Results provide existing and future hosts with insights into the inner workings of price and unite to a more systematic framework for hosts to set revenue optimizing listings prices in the future.

4 Introduction

AirBnB has emerged as a showpiece of the sharing economy, a novel business paradigm which has become a well-established alternative to conventional business-to-customer transaction formats across many sectors (Zhang et al., 2017). Ever since its launch in 2008, the San Francisco based startup has administered over 800 million bookings and attracted over 4 million hosts onto its platform (Airbnb, 2021). In consequence, Airbnb has not only emerged as a serious contender for traditional hospitality and tourism businesses but also as a sustainable source of income for many private property owners around the world. The importance of sensible pricing decisions for long-term success in the hotel industry finds frequent mention in existing literature (e.g. Hung et al., 2010). Yet, recent research cautions against generalizing existing study findings to the sharing economy-based accommodation sector which structurally differs from commercially organized hospitality in many respects (Chang & Li, 2020; Wang & Nicolau, 2017). The importance of careful pricing does not weight less on Airbnb hosts. Yet, platform participants experience fairly little support from Airbnb to systematically appraise their listings (Hill, 2015). While researchers have begun to disentangle the dimensions and determinants Airbnb listings prices (e.g. Perez-Sanchez et al., 2018; Wang and Nicolau, 2017), more recent studies which highlight the role of city as important determinant of price (Chang & Li, 2020) provide a kind reminder not only to study price antecedents at scale but as localized phenomena. This is necessary to provide Airbnb hosts with the bespoke guidance they need to compete on an increasingly crowded platform (Gibbs 2020). With over 22,000 listings, the city of Berlin ranks among the 5 densest markets for Airbnb housing in Europe (Statista, 2019). Except for occasional considerations in larger studies (Wang & Nicolau, 2017), Berlin has not yet been made the center of attention by research when reasoning Airbnb pricing determinants at city-level.

Objective

In this article, we identify some of the key factors determining Airbnb listings prices for the city of Berlin using a dataset of 2,799 sets of observations. We examine whether and to which extent factors like host *room type*, *amenities* and *user rating scores* associate with listing price as well as to which extent these effects vary for different price ranges. By that, we aspire to draw Airbnb hosts' attention to promising pricing levers in order to get the most out of their properties.

5 Related Work

Scholarship acknowledges the utility of existing studies on hotel price determinants to frame research endeavours on equivalent sharing economy phenomena (Wang & Nicolau, 2017). By the same token, researchers have pointed out that peer-to-peer accommodation rental differs from traditional hospitality in many aspects (Chang & Li, 2020). Given that property selection on Airbnb proposes a novel user journey which does not resurface as such in studies on conventional hospitality (Gibbs et al., 2018), we decided to ground model and hypothesis development on the review of literature which has previously researched the kinds of progressive, fine-granular search criteria (e.g. instant booking options, number of guests, number of bedrooms, amenities) guests are exposed when discovering property listings on the Airbnb platform.

Existing research explored the determining factors of price for peer-to-peer accommodation services from various angles. In their study on Airbnb listings prices across 3 Chinese cities, Chang & Li (2020) explored the relevance of 23 attributes spread across 5 distinct categories. The researchers qualified the variables *room type*, *city*, *distance to points of interest* (e.g. airport, tourist attractions), *number of pictures posted* and *number of amenities* provided as most important price determinants.

For their part, Gibbs et al. (2018) verified the positive associations of size-related attributes like *number of bathrooms*, *number of bedrooms* and *guest capacity* with listings price. Extending Chang & Li 's (2020) investigation of amenities to specific property features, the researchers ascertain a positive relationship between price and the *availability of gym, pool* or *free parking* which appreciate a property's value perception. Yet, unhypothesised negative coefficients arose for the attributes, *instant booking*, *number of reviews* and *rating scores.* The researchers jointly reason these seemingly counterintuitive effects within a compelling narrative that presents Airbnb's instant booking feature as means for hosts to rapidly "fill the beds" of lower priced properties which raises both, guest turnover and the amount of "good value for money"-coined reviews. Gibbs et al. (2018) further inferred a novel attribute measuring *host professionalism* which turned out to positively associate with listings price. This complements Li et al.'s (2015) earlier finding that professional hosts who are in charge of multiple properties benefit from higher daily revenue and occupancy rates.

Zhang et al. (2017) reconfirm aforementioned negative associations between *number of reviews* and *rating scores* on price for an American regional sample comprising 974 listings. The researchers take these findings as evidence of low-price strategies pursued by hosts to rapidly gain traction in communities in which Airbnb only recently set foot.

Many of the aforementioned findings resonate with Wang and Nicolau's (2017) large-scale study on Airbnb pricing determinants across 33 cities. The researchers not only confirm a significantly negative association between the usage of Airbnb's *instant booking* feature and listing price but also prove pricing uplifts for *room types* "private room" and "entire home/apartment" when employing "shared room" as reference base. Such findings are believed to demonstrate guests' willingness to pay a premium for heightened comfort and security (Chang & Li, 2020) but also attest the ongoing significance of "classic" hospitality features in the era of sharing economy (Gibbs et al., 2018). The study further conforms with previous literature in that negative associations between *number of reviews* and listings price were uncovered but diverges by demonstrating a positive association between *user rating scores* and price.

Perez-Sanchez et al. (2018) confirm likewise discrepancies for a Spanish sample of Airbnb listings as far as property location is concerned. While the researchers agree with Gibbs et al. (2018) on the positive impact of size-related property characteristics such as *guest capacity* and *number of bathrooms*, the study contests the widely held opinion that increasing geographical distance between a property and city center coincides with lower price points (Chang & Li, 2020; Gibbs et al., 2018). By identifying positive associations between *distance* and price, the researchers draw attention to the uniqueness of different cityscapes as to where urban and economic, tourist appealing activities are happening. The aforementioned study provides a great testimony of the ongoing discrepancies in literature as to which factors prove most influential on listings prices in a given city and once more legitimates our study researching pricing determinants for Airbnb property listings in Berlin.

The literature review motivates the following research questions and associated hypotheses:

> RQ_1: What is relationship between property-related attributes and Airbnb listings prices in Berlin?
>
> H_1: Size-related attributes (*number of bedrooms*, *number of bathrooms*, *guest capacity*) *property type* and provided *amenities* (*number* and *specific types of amenities*) positively associate with listing price.
>
> RQ_2: What is relationship between property management-related attributes and Airbnb listings prices in Berlin?
>
> H_2: Using Airbnb's *Instant booking* feature negatively associates with listing price.
>
> RQ_3: What is the relationship between reputation-related attributes and Airbnb listings prices in Berlin?
>
> H_3: *Number of reviews* and *user rating scores* negatively associate with listing price.
>
> RQ_4: What is relationship between host-related attributes and Airbnb listings prices in Berlin?
>
> H_4: *Host professionalism* positively associates with listing price.

Lastly, several researchers estimate the effect of property characteristics for different parts of the price spectrum beyond the mere estimation of average effects (e.g. Perez-Sanchez et al., 2018; Wang & Nicolau, 2017). For instance, Perez-Sanchez et al., (2018) showed that the positive effect of *guest capacity* substantially diverges from the average effect estimate provided by ordinary least squares (OLS) regression for the upper and lower tails of the listings price distribution. These studies motivate the formulation of a 5th and final research question:

> RQ_5: Do the effects of price determinants vary for different Airbnb listings prices in Berlin?
>
> H_5: The effects of price determinants on price are not constant.

Previous studies provided actionable clues on the significance and coefficient directionality of selected features. Yet, material differences in coefficient sizes reported for similar attributes across studies discourage us from hypothesising the magnitude of suspected relationships.

Table 1 provides an overview of previously reviewed studies on price determinants for Airbnb listings motivating our hypotheses.

Category	**Attribute**	**Effect**	**Supporting Literature**
Property	number of bathrooms	Positive	Gibbs et al. (2018)
			Perez-Sanchez et al. (2018)
	number of bedrooms	Positive	Gibbs et al. (2018)
	guest capacity	Positive	Gibbs et al. (2018)
			Perez-Sanchez et al. (2018)
	number of amenities	Positive	Chang and Li (2020)
	specific amenities	Positive	Gibbs et al. (2018)
	room type	Positive	Chang and Li (2020)
			Wang and Nicolau (2017)
Management	instant booking	Negative	Wang and Nicolau (2017)
			Gibbs et al. (2018)
Reputation	number of reviews	Negative	Zhang et al. (2017)
			Gibbs et al. (2018)
			Wang & Nicolau (2017)
	user rating score	Negative	Zhang et al. (2017)
			Gibbs et al. (2018)
		Positive	Wang and Nicolau (2017
Location	distance	Positive	Perez-Sanchez et al. (2018)
		Negative	Gibbs et al. (2018)
			Chang & Li (2020
Host	host professionalism	Positive	Gibbs et al. (2018)
			Li et al. (2015)

Table 1: *Overview of price determinants investigated in related work*

6 Methods

6.1 Data Source

Detailed Airbnb listings data for the city of Berlin was sourced from watchdog website *Inside Airbnb.com*. The original dataset contains 20,255 observations and holds property, host and user review information (Inside Airbnb, 2020). The database was last updated on November 10th, 2019 and hence lends itself for the construal of contemporary claims.

In line with Zhang et al. (2017), we only retained listings affiliated with so-called "superhosts". Referred to by Airbnb (2020) as *"(...) experienced hosts who provide a shining example for other hosts(...)"*, a superhost-managed listings subset is expected to provide a reliable, relatively noise-free data foundation favouring meaningful analysis results. Second, listings with less than 3 reviews were precluded from the dataset. This filter serves a dual purpose. On the one hand, we anticipate effect distortions between price and review related attributes since Airbnb only publishes user ratings for properties that accrued a critical amount of feedback (Gutt & Hermann, 2015). On the other hand, we descoped listings associated with prices that might fail to translate in real-world transaction between host and guest. We drew on research confirming a significant relationship between user reviews and hotel sales (Ye et al., 2009) to consider listings prices meaningful approximation of the market equilibrium when backed up by a critical amount of reviews (Wang & Nicolau, 2017). In addition, we precluded observations of *room type* "Hotel room" to constrain our analysis to non-commercial offerings.

Beyond applying these conceptual filters, several outliers were removed through simple consistency checks. In this regard, normalizing prices over associated guest capacities and hence removed revealed several extreme observations. Per capita prices as small as 5 USD and above 300 USD were deemed unrealistic and hence removed from the dataset.Likewise, we precluded potentially erroneous instances affiliated with hosts for which a listings count of 0 was reported. Eventually, we imputed 220 missing values that existed for the variable *number of bedrooms* with the mean conditional on guest capacity to avoid unintentional distortions.

In consequence, 2,799 samples were retained for subsequent analyses. *Figure 1* summarizes the approach underlying the targeted exclusion of sample observations.

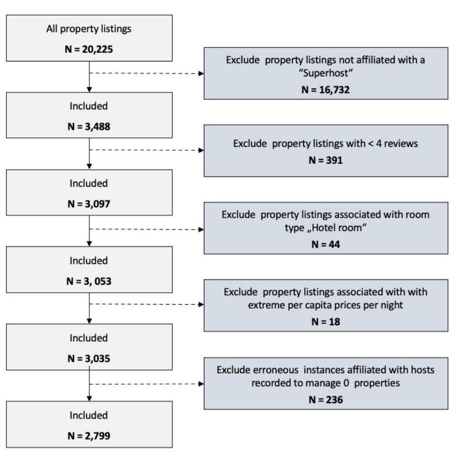


Figure 1: *Diagram summarizing data exclusion measures applied*

Research questions and derived hypotheses motivated the extraction of 14 attributes from the available variable spectrum. Several variables had to be modified prior to model inclusion. Applied transformations adhere closely to conventions set forth in related studies. In this manner, we constructed a binary variable *is professional* which qualifies a host as professional if he/she manages two or more listings (see Gibbs et al., 2018). We further made use of detailed amenity item lists to derive the *number of amenities* provided for each listing as well as four binary variables denoting the availability of specific items *Wifi*, *Free Parking*, *Elevator* and *Coffee Maker*. While *Wifi* and *Free Parking* were explicitly investigated in previous studies by Gibbs et al. (2018) and Wang and Nicolau (2017), we decided to include two additional variables that - to our knowledge - have not yet found consideration in preceding studies. *Elevator* and *Coffee Maker* were deliberately selected to extend existing research by one structural amenity and one commodity item whose provisioning lies entirely in the hands of the host.

Since the listings price distribution in our sample turned out to be heavily right-skewed, we applied a log transformation to meet the normality assumption of OLS. In this context,box-plots alluded to the presence of outliers. Flagged instances yet pertained multi-guest properties and were hence deemed conceptually sound. *Appendix 1* presents the sample distribution of listings prices before and after the log transformation.

6.2 Data Analysis

To build a preliminary understanding of the dataset and assess how listings prices vary across different attribute values, several descriptive analyses were performed. In addition, we ran a series of correlation analyses to grasp pairwise relationship strengths across the variable set prior to hypothesis testing. Hypothesis testing was carried out by regressing the log-transformed property listings price *log_price* on all 13 predictor variables. We used two models to investigate how listings prices change in response to changes in the predictors.

First, an OLS regression was used to quantify the relationship between dependent variable *log_price* and predictors $X_i ... X_{13}$. In this regard, regression coefficient β_i represents the average effect (directionality and magnitude) of a unit increase of X_i on *log_price* holding all other predictors constant. In general terms, the OLS regression model can be written as follows:

$$log\_price = \beta_0 + \sum_{i=1}^{13} \beta_i X_i + e$$

In acknowledgement of modelling trends set forth in recent studies by Perez-Sanchez et al. (2018) and Wang and Nicolau (2017), we specified a second model using Quantile Regression (QR) to assess the relationship between each predictor X_i and *log_price* at different sections of the response variable distribution. QR can hence be thought of as extension to OLS as it exposes potential variabilities in the predictor-response variable relationship, contrary to OLS which distills the relationship into a single "summary" coefficient. In general terms, the QR equation for the *jth* quantile can be written as follows:

$$Q_j(log\_price) = \beta_0(j) + \sum_{i=1}^{13} \beta_i(j) X_i + e$$

Both approaches taken together invite to a more nuanced and comprehensive analysis of hypothesised effects.

Considering the susceptibility of OLS to numerous data-related issues, we ran a range of tests to ensure popular assumptions were met. We used a residual plot to assess non-linearity, performed a *variance inflation factor* (VIF) assessment to assert the absence of collinearity and watched for high leverage points based on *Cook's Distance*.

7 Results

7.1 Exploratory Data Analysis

Table 2 presents summary statistics for each variable of interest. On average, an Airbnb listing is priced at 70.25 USD, receives approximately 71 reviews and a user rating score of 97 out of 100. While *user rating scores* appear to clutter at the upper bound of the value spectrum, high standard deviations for *price* and *number of reviews* allude to considerable variability between listings. The middle 50% of listings have 1 bedroom, 1 bathroom and accommodate between 2 to 4 guests. Notable differences between the interquartile range (IQR) and associated maximum values for capacity-related variables hint at the presence of several larger sized listings. 47% of properties affiliate with professional hosts. Concerning certain amenities, almost all listings are equipped with *Wifi* whereas just a quarter offer *elevator access*.

Variable	**Statistic**	**Value**
price	*mean (std)*	70.25 (56.26)
Property		
guest capacity	*25%,75%,(max)*	2,4 (16)
number of bedrooms	*25%,75%,(max)*	1,1 (12)
number of bathrooms	*25%,75%,(max)*	1,1 (7)
room type - "entire house"	*N(%)*	1627(58.13)
room type - "private room"	*N(%)*	1165(41.62)
room type - "shared room"	*N(%)*	7(0.25)
number of amenities	*mean (std)*	24.08 (8.45)
has wifi	*mean (std)*	0.98 (0.14)
has elevator	*mean (std)*	0.24 (0.43)
has free parking	*mean (std)*	0.57 (0.50)
has coffee maker	*mean (std)*	0.68 (0.47)
Reputation		
number of reviews	*mean (std)*	71.37 (79.21)
user rating score	*mean (std)*	97.22 (2.38)
Management		
instant bookable	*mean (std)*	0.39 (0.49)
Host		
is professional	*mean (std)*	0.47 (0.50)

Table 2: *Overview of selected summary statistics*

Figure 2 points to correlations between price and several attributes like *guest capacity* (0.63) or *number of bedrooms* (0.61). Notable associations between several capacity-related attributes such as *number of bedrooms* and both, *guest capacity* (0.75) and *number of bathrooms* (0.58) call for additional investigations assessing for multicollinearity.

Figure 2: Pairwise correlations across model variables

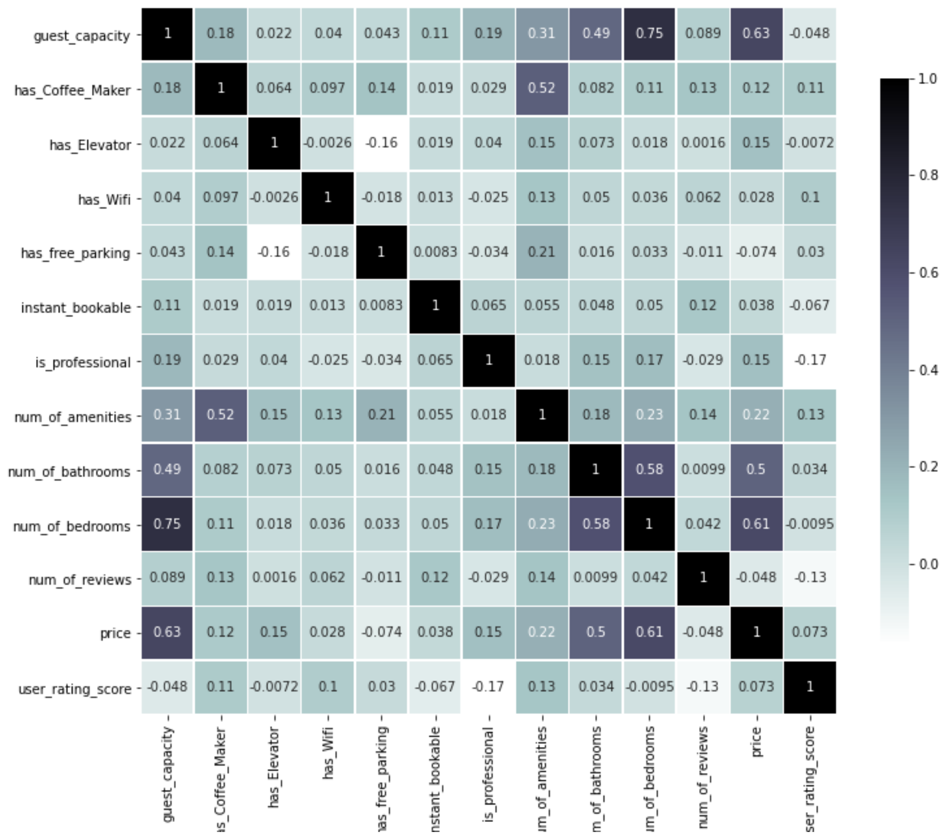
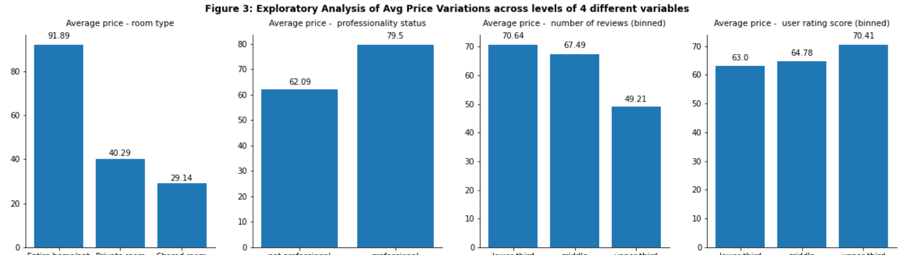


Figure 3 reveals that listings prices considerably vary for different values of variables *room type* and *is professional* and *number of reviews.* Properties marketed as “entire home/apartment” associate with higher average prices (~92 USD) compared to private rooms (~40 USD) or shared rooms (~29 USD) whereas hosts classified as professional appear to charge higher prices (~80 USD) than casual hosts (~62 USD). Lower prices appear to coincide with increasing number of reviews. With respect to *user rating score*, properties with higher ratings seem to relate with higher prices relative to properties situated at the middle and bottom third of the user ratings distribution.

Figure 3: Exploratory Analysis of Avg Price Variations across levels of 4 different variables



7.2 OLS Regression

OLS regression results are summarized in *Table 3*.

Since the residual sum of squares always decreases as more variables are included into the model (James et al. 2013), we refer to the *Adjusted* R^2 to evaluate model goodness-of-fit. Results show that our model is able to explain 62.2% of the variability in the response variable *log price*. Previously discussed correlation matrix results (*Figure 2*) raised concerns of multicollinearity which might impede our model to separate out individual predictor effects. VIF results show consistently low values below 3 across the entire predictor set (Appendix 2) which appeases previous concerns in consideration of commonly applied thresholds of 5 or 10 (James et al. 2013). A residual plot relating fitted values with model residuals does not yield a salient pattern *(Appendix 3)*. In result, non-linearity concerns do not arise. Yet, Breusch-Pagan test results confirm the ongoing presence of heteroscedasticity and thus non-constant variance of error terms despite the preliminary log transformation of the response variable *(Appendix 4)*.

An inquiry of influential leverage points motivated the removal of two observations *(Appendix 5)*. Considering Cooks and Weisberg's (1999) rule of thumb to scrutinize observations with distances > 0.5, no additional removals were performed.

Variable	**Coefficent**	**% Effect**	**S.E.**
R^2	0.624		
Adjusted R^2	0.622		
Property			
guest capacity	0.077***	7.71%	0.006
number of bedrooms	0.118***	11.84%	0.015
number of bathrooms	0.141***	14.09%	0.023
room type - "private room"	-0.544***	-41.97%	0.016
room type - "shared room"	-0.715***	-51.10%	0.136
number of amenities	0.003**	0.28%	0.001
has wifi	0.098	10.34%	0.05
has elevator	0.147***	15.81%	0.016
has free parking	-0.106***	-10.05%	0.014
has coffee maker	0.001	0.09%	0.017
Reputation			
number of reviews	-0.001***	-0.05%	0.000
user rating score	0.026***	2.58%	0.003
Management			
instant bookable	-0.029*	-2.84%	0.014
Host			
is professional	0.021	2.10%	0.014

Table 3: *OLS Regression Results (*p < 0.05; ** p < 0.01, *** p < 0.001)*

Results reveal that 11 out of the 13 predictors have a significant influence on price. Given the log transformation of the dependent variable, regression coefficient values constitute semi-elasticities indicating the relative change in price provided a predictor variable varies by 1. While this convention holds true for continuous variables witout additional intervention, we follow Halvorsen and

Palmquist's (1980) proposal to transform coefficients associated with dummy encoded variables by $(e^{(\beta)} - 1)$ to quantify the relative discontinuous effect of the variable in presence of the factor expressed by this variable. *Table 3* hence also displays the percentage change in price caused by a unit change in any of the model predictors.

With regards to property-related attributes, *guest capacity* (7.7%, $p < 0.001$) , *number of bedrooms* (11.8%, $p < 0.001$) and number of bathrooms (14.1%, $p < 0.001$) have a significantly positive effect on price . In addition, property types "Private room" and "Shared room" were found to have significantly negative impact when "entire home /apartment" is chosen as reference base. More specifically, listings of type "Private room" and "Shared room" coincide with price drops amounting to ~42% ($p < 0.001$) and ~51% ($=p < 0.001$) relative to stand-alone properties. While *number of amenities* has a significant positive influence on price, associated effect size is miniscule. An additional amenity coincides with an average price increase of 0.28%. Inspecting coefficients for specific types of amenities, *has elevator* positively affects price (14.7%, $p < 0.001$). To our surprise, *has free parking* associates with a 10.05% decrease in price on average ($p < 0.001$). No significant effects were found for *has Wifi* ($p = 0.054$) and *has coffee maker* ($p = 0.956$). In light of these findings, H_1 is partially supported.

In support of H_2, attribute *instant bookable* negatively affects price (-2.84%, $p < 0.05$).

Concerning reputation-related attributes, *number of reviews* has a significantly negative, although marginal effect on price (-0.05%, $p < 0.001$) while *user rating scores* – contrary to our hypothesis - positively affect price (2.52%, $p < 0.001$). We hence find partial support for H_3.

In contradiction to H_4, perceived variable *is professional* was not found to significantly influence listings price ($p = 0.141$).

7.3 QM Regression

Table 4 presents the quantile regression (QR) results. Regression coefficient estimates are outlined for the 10th, 25th, 50th, 75th, and 90th quantile, together with an indication of statistical significance. With changes in effect sizes rather than the magnitude of the effect in the spotlight, efforts to transform QR coefficients for additional interpretability were omitted. In addition, *Figure 4* visualizes regression coefficient sizes for different quantiles along the dependent variable distribution (red line) for those predictors found to exercise significant effects on listings price across all quantiles surveyed. Predictor-wise plots were further overlaid with associated OLS regression coefficient estimates, along with 95% confidence interval (CI) boundaries pertaining to both, QR (red dotted lines) and OLS estimates (blue dotted lines) to set changes in predictor coefficients in context.

Variable	QR=0.1	QR=0.25	QR=0.5	QR=0.75	QR=0.9
$Pseudo-R^2$	0.33	0.37	0.41	0.42	0.42
Property					
guest capacity	0.068^a	0.064^a	0.078^a	0.084^a	0.09^a
number of bedrooms	0.111^a	0.135^a	0.1^a	0.103^a	0.111^a
number of bathrooms	0.118^a	0.128^a	0.142^a	0.165^a	0.188^a
room type - "private room"	-0.532^a	-0.544^a	-0.553^a	-0.574^a	-0.56^a
room type - "shared room"	-0.592^c	-0.698^a	-0.732^a	-0.618^a	-0.741^b
number of amenities	0.002	0.002	0.002^c	0.003^b	0.06^a
has wifi	0.45^a	0.173^b	0.063	0.015	-0.368
has elevator	0.099^a	0.135^a	0.146^a	0.151^a	0.158^a
has free parking	-0.087^a	-0.1^a	-0.115^a	-0.101^a	-0.116^a
has coffee maker	0.021	0.009	0.013	0.00	-0.025
Reputation					
number of reviews	-0.0	-0.0^b	-0.0^a	-0.001^a	-0.001^a
user rating score	0.024^a	0.023^a	0.025^a	0.024^a	0.025^a
Management					
instant bookable	-0.023	-0.049^b	-0.045^b	-0.041^c	-0.035
Host					
is professional	-0.029	-0.009	0.01	0.043^c	0.088^a

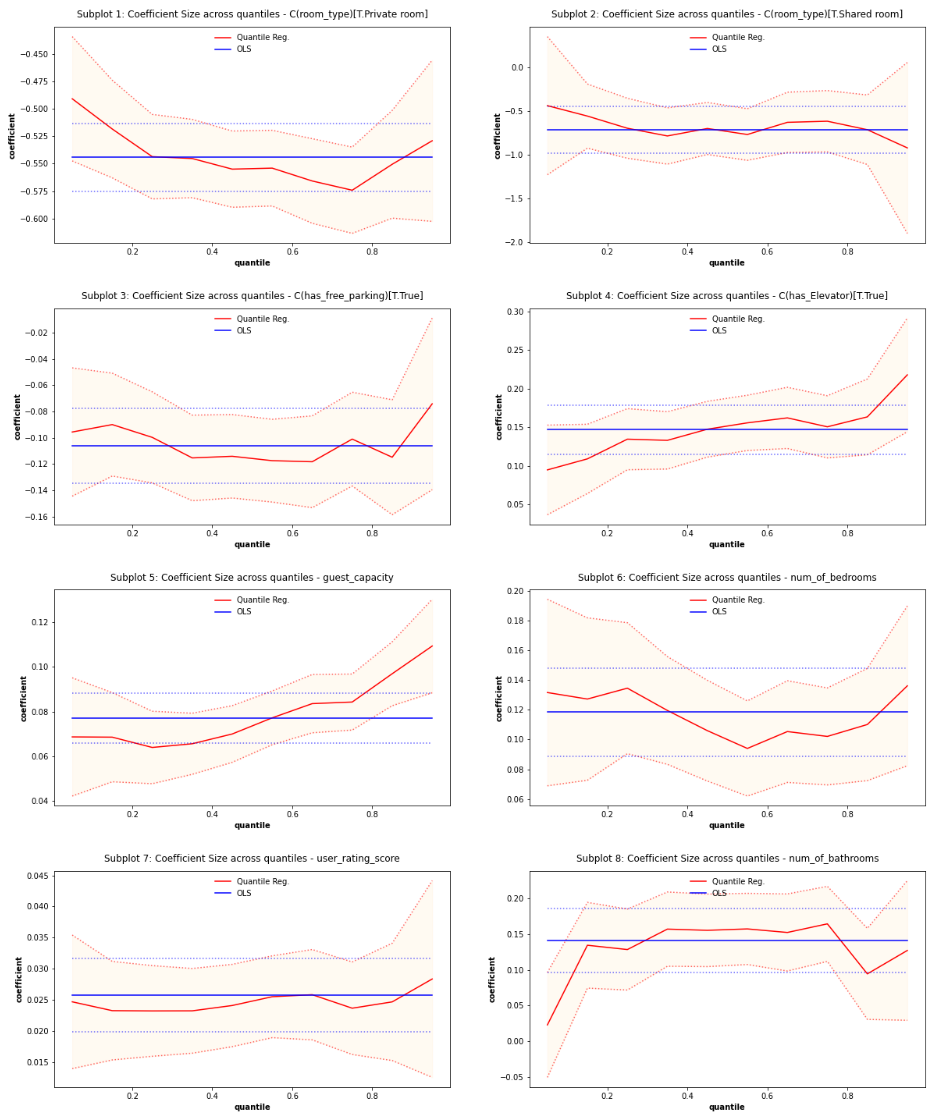
Table 4: *QM Regression Results (β^c: $p < 0.05$; β^b: $p < 0.01$*; *β^a*: *$p < 0.001$)*

In line with OLS results, we observe significantly positive regression coefficients for all capacity-related variables *number of bedrooms*, *number of bathrooms* and *guest capacity* across the entire quantile range. Yet effect sizes notably vary for different parts of the listings price distribution. The positive effect of *number of bedrooms* is most acute for lower and higher prices respectively and less pronounced for the middle part of the distribution *(Figure 4, Subplot 6)*. Conversely, the positive effect of *number of bathrooms* on price increases with increasing listing prices and even exceeds the average effect anticipated by OLS for mid-range priced listings between the 25th and 90th quantile *(Figure 4, Subplot 8)*. The significantly positive effect of *guest capacity* on price increases quasi-monotonically beyond the 25th quantile. Increasing capacity by 1 hence coincides with a 9% increase in price at the 90th quantile as opposed to a 6.4% increment at the 25th quantile *(Table 4)*. Likewise, the significantly positive effect of *has elevator* on price becomes more influential with rising listings prices *(Figure 4, Subplot 4)*.

Reconfirming OLS outcomes, *has free parking* exerts a significantly negative impact on price while *has coffee maker* proves insignificant across all quantiles in scope *(Table 4)*. Results for *has Wifi* marginally contrast with OLS in that the availability of Wifi has positive influence on prices between the bottom 10th and 25th quantiles. All things equal, no significant effects emerge for the remainder of the quantile range. A similar finding arises for attribute *number of amenities*. While OLS ascribes a marginal yet significantly positive contribution of *number of amenities* to price *(Table 3)*, quantile regression results only accredit significance to this relationship for prices beyond the 50th quantile *(Table 4)*. With respect to *room type*, QR coefficients for categories "Private room" and "Shared room" have a significantly negative effect on price across all quantiles when category "entire home/apartment" is chosen as reference base. In this context, coefficient sizes prove relatively invariant for category "Private room" as opposed to "shared room" which coincides with a 44.6%

$(e^{(-0.592)} - 1)$ decrease in price for the 10th quantile compared to a 52.3% $(e^{(-0.741)} - 1)$ decrease for the 90th quantile *(Table 4)*. A similar narrative emerges for variable *review scores_rating* with small yet significantly negative effect sizes closely revolving around the average effect estimated by OLS *(Figure 4, Subplot 7)*.

Figure 4: QR coefficient changes across listings price quantiles



As far as predictors *number of reviews*, *instant bookable* and *is professional* are concerned, QR once again nuances corresponding OLS estimates. The miniscule negative effect of *number of reviews* on price qualified by OLS *(Table 3)* is only found to be statistically significant beyond the 10th quantile *(Table 4)*. While OLS did not qualify *is professional* as significant predictor of price *(Table 3)*, QR asserts a significantly positive impact on price for higher-priced listings beyond the 75th quantile *(Table 4)*. Similarly, QR narrows the significantly negative association between *instant bookable* and price qualified by OLS *(Table 3)* down to the middle part of the listings price distribution with no significant effects found for the 10th and 90th quantile *(Table 4)*.

The results summarized in *Table 4* confirm that predictor effects on price vary in size across quantiles. While *Figure 3* visualizes that recorded changes represent noticeable departures from static OLS estimates at times (e.g._Figure 3, Subplots 4 and 5_), it becomes evident that QR CI bounds consistently enclose the OLS estimate for 6 out of the 8 predictor-response relationships visualized, not to speak of the fact that the confidence intervals associated with OLS and QR are not found to segregate at any time across predictors.

With the diversity of just-discussed findings in mind, we consider exploratory hypothesis H5 as partially supported.

8 Discussion

Research has only started to decipher the factors determining prices of sharing economy-based accommodation services. Based on a sample comprising 2799 property listings for the city of Berlin, we have analyzed the potential effects of 13 attributes on price using regular OLS and Quantile regression. The combination of both modeling approaches allowed us to compute both, average effects and changes in effect sizes for different ranges of the conditional distribution of the response variable.

OLS results present *room type* as most important price determinant. When renting out private rooms or shared rooms as opposed to entire homes/apartments, listings prices undergo a steep decline of 42% and 51% respectively on average. In contrast to our study, previous research has investigated this attribute with respect to different reference bases ("shared room" in Wang and Nicolau, (2017); "private room" in Chang and Li (2020)). Yet, previous findings of positive price effects as one traverses from shared room to private room up to entire "home/apartment" are congruent with our findings. Similarly, property attributes indicative of size such as *number of bathrooms*, *number of bedrooms* and *guest capacity* coincide with higher prices. Quantile regression results find these effects to remain in force across the listings price distribution. Confirming previous study outcomes, our findings hence show that conventional hospitality characteristics, while receiving a digital makeover by sharing economy unicorn Airbnb, continue to pull the strings (Gibbs et al., 2018). While private rooms partially alleviate potential security and privacy concerns among guests in connection with shared room rentals, only entire houses will guarantee exclusive access to all rooms and amenities (Chang & Li, 2020). However, convenience and piece-of-mind come at a price.

A small positive, yet significant, effect between number of amenities and price resonates with existing literature (Chang & Li, 2020). While identified effect significances beyond the 50th quantile motivate the conclusion that offering "more" justifies higher prices in the eyes of the guest, the miniscule size of the effect suggests that it might not be the sheer volume but the specific types of amenities provided that justify a premium. Identified effects across amenity-specific variables reinforce this narrative. Non-significant OLS regression results for variables *has wifi* and *has coffee maker* let us conclude that such amenities are either taken for granted or are perceived as too mundane to move the needle. With 98% of sample listings being equipped with WIFI, the more nuanced QR regression finding which attests significant and positive price effects among lower-priced properties only up to the 25th quantile dwell at an "all-necessities-inclusive" expectation among guests that buy into mid and high-priced Airbnb properties. A similar storyline emerges for *has elevator* found to have a stronger positive effect for listings at the upper end of the price range where value-conscious guests might be deliberately wanting to buy into an additional increment of convenience.

To our surprise, *has free parking* was found to negatively associate with price. While previous literature portrayed the availability of free parking as welcomed perk (e.g. Wang & Nicolau, 2017), QR regression results find significantly negative effects across all sections of the listings price distribution. Two interpretations arise: First, free parking might not be a scarce resource in the city of Berlin. At any given price range, more rudimentary, cheaper listings might recourse to stating the availability of free parking to artificially inflate otherwise sparse value propositions whereas hosts managing more generously equipped listings might not consider it worth-mentioning. Second, our choice to model the availability of free parking based on amenity items "Free Street Parking" & "Free Parking on Premises" might have confounded the coefficient estimate in retrospect. While Street parking might be more common among lower priced listings, parking on premises might be a given for premium properties. This subtlety yet evaporates in the process of binarization. All things equal, additional efforts should be made to reason out this unexpected finding.

Contrary to our hypothesis, the effect of *user rating score* on price turned out to be significantly positive. Rather than being the result of lower-priced properties setting easy-to-exceed expectations among budget travelers (Zhang et al., 2017), positive price effects observed across all quantile ranges portray positive user feedback as pricing lever among lower-and higher-priced properties alike. Yet, effect sizes remain vanishingly small along the listings price distribution. Considering the very high, largely invariant *user rating scores* within our sample, QR regression findings resonate with Gibbs et al.'s (2018) claim that the predominance of high user ratings raises doubt whether individual hosts can leverage such credentials to their advantage.

Hypothesized, significantly negative effects for variables *number of reviews* and *instant bookable* coincide with existing literature presenting both attributes as input and output of a low-price strategy which aims at high guest turnover and leads to large number of "good value for money"-coined reviews in return (Gibbs et al., 2018). Eventually, OLS did not attribute significance to *is professional* while QR regression qualifies a significantly positive price effect for listings prices beyond the 75th percentile. Findings hence tentatively affiliate with existing literature that found professional hosts to be savvier in setting adequate prices (Li et al., 2015). We conclude that host professionalism might be a quality assurance characteristic guests eying at more expensive listings are willing to pay a premium for. We further acknowledge that non-significant OLS regression results on this variable might stem from our decision – motived by existing literature (Gibbs et al., 2018; Li et al., 2015) – to qualify hosts in charge of two or more listings as professional. Existing literature raises the issue of information loss as one of the serious consequences of variable dichotomization (Fedorov et al., 2009). We motivate future research to experiment with more suitable thresholds to create binary or multi-category representations to model host professionalism.

Limitations and avenues for future research

Eventually, our study comes with several limitations which open up multiple avenues for future research.

First, with 61.8% of the variability in the response variable explained, our model leaves room for improvement. Future research might experiment with additional attributes to build a more comprehensive understanding of the inner workings of Airbnb listings prices. Considering the urban diversity of a city like Berlin, one avenue might be to account for potential price heterogeneities in neighborhoods using multi-level modeling.

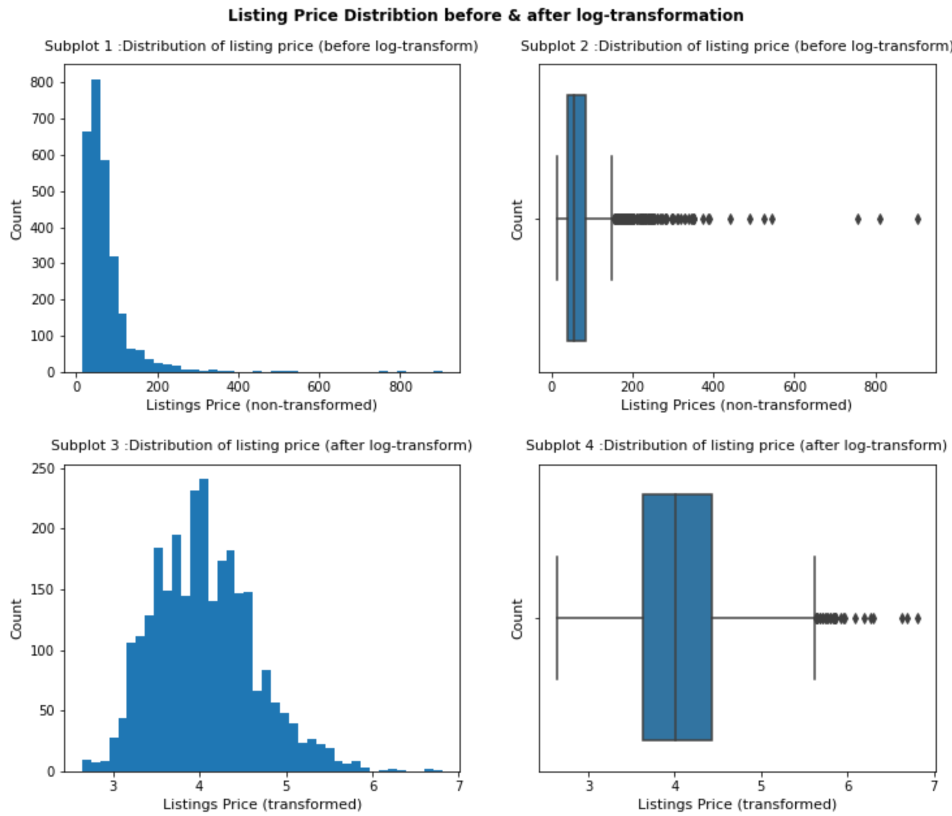
Second, Breusch-Pagan test results raised concerns about heteroskedasticity in our model which might have informed erroneous claims about predictor significances. Following the example of Wang & Nicolau (2017), additional efforts could be made to correct coefficient standard errors using the Eicker-White procedure and hence reconfirm ascribed significances under more refined circumstances. All things equal, QM is considered more robust to data-related issues than OLS (Baum, 2013) and hence emerged as welcomed extension to accentuate OLS results in light of heteroscedasticity concerns being raised in our study.

Third, we encourage readers to scrutinize quantile regression results for practical significance. Quantile regression uncovered conditional significances across 5 predictors for which OLS either categorically denied or asserted statistically significant relationships. Complementing OLS with QR hence allowed for a richer characterization of predictor-response relationships beyond the conditional mean. Yet we cannot unambiguously confirm the significance of observed variabilities without additional testing. While Koenker & Hallock (2001) note that QR coefficients beyond the CI boundaries of OLS allude at non-constant covariates, results caution us not to jump to conclusions. We did not only find that confidence intervals associated with estimates from each model overlap at all times, but also that QR CI boundaries almost always enclose the corresponding OLS estimate. Re-running the analysis on a larger sample as well as evaluating differences in coefficient sizes across quantiles for statistical significance are sound avenues to meaningfully extend our research.

In consideration of previous research arguing significant between city differences (Chang & Li, 2020), we discourage the generalization of study outcomes which ground on a regional sample for the city of Berlin to other geographies within and outside Germany. In the same vein, applied cleansing and filtering operations substantially reduced our sample in size. While this was necessary to assure data correctness and conceptual compliance with frequently cited research in this field, it calls sample representativeness into question. We hence encourage further city-specific studies, ideally on larger samples, to validate our findings in other settings.

9 Appendices

9.1 Appendix 1

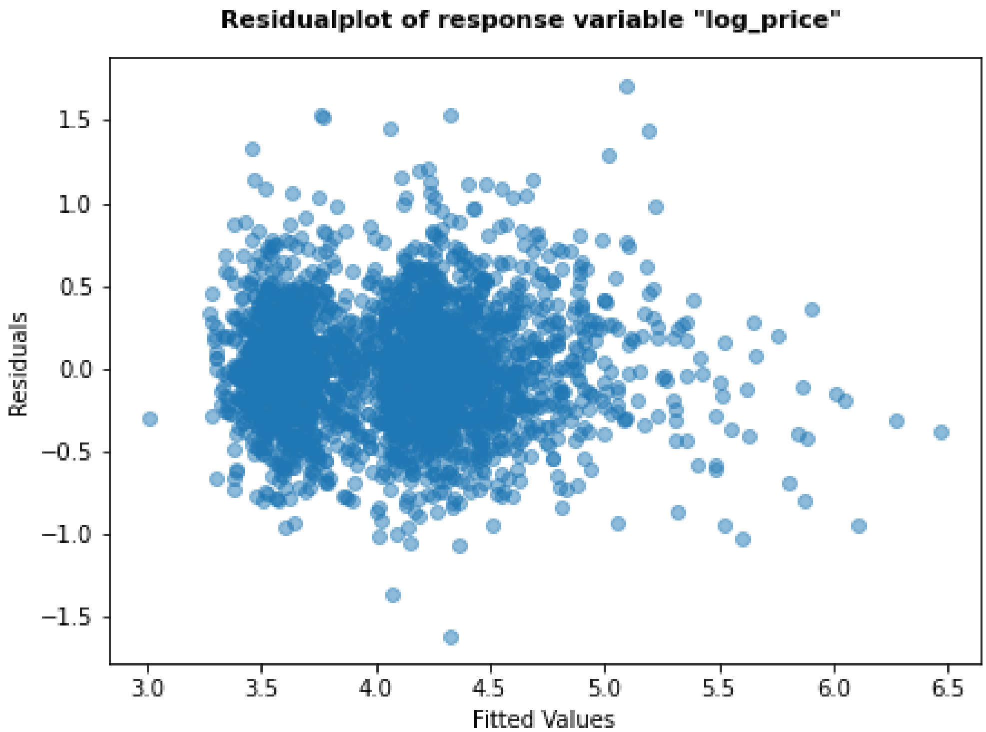


9.2 Appendix 2

	Predictor	VIF
1	C(room_type - “Private room”)	1.32497
2	C(room_type - “Shared room”)	1.01175
3	C(instant_bookable)	1.03355
4	C(is_professional)	1.08177
5	C(has_free_parking)	1.11491
6	C(has_Coffee_Maker)	1.39993
7	C(has_Wifi)	1.03597
8	C(has_Elevator)	1.07936
9	guest_capacity	2.82155
10	num_of_bedrooms	2.5748
11	num_of_reviews	1.07717
12	user_rating_score	1.10368
13	num_of_bathrooms	1.51172
14	num_of_amenities	1.61739

Table: *Variance Inflation Factor results across all predictors*

9.3 Appendix 3

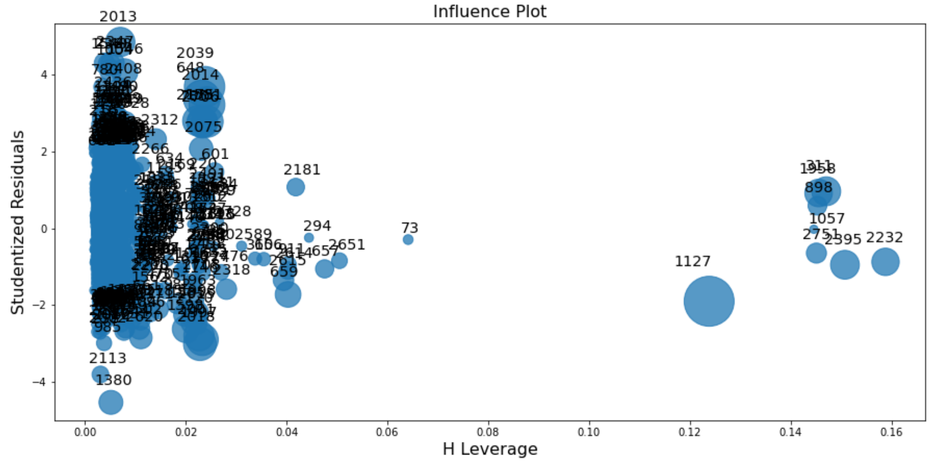


9.4 Appendix 4

Test Statistic	Values
Lagrange multiplier statistic	109.258
p-value	0
f-value	8.07786
p-value (F-statistic)	0

Table: *Variance Inflation Factor results across all predictors (H_0: Homoscedasticity is present)*

9.5 Appendix 5



Suplementary Note: Observations with indices *1127* and *2039* were dropped due comparatively high cook's distance in reference to all other sample observations. Subsequent to removal, cooks distances were extracted for the sample remainder and sorted in descending order (see table below) to confirm the absence of additional concerning instances

Observation	leverage	cooks distance	standardized residuals	studentized residuals
648	0.0232235	0.0182367	3.39198	3.3984
2013	0.0242253	0.0169912	3.20404	3.20939
2017	0.0228319	0.0140105	-2.99906	-3.00338
1996	0.0233142	0.0131966	-2.87968	-2.88347
1580	0.0243319	0.0127332	2.76743	2.77075
2000	0.022676	0.0122897	-2.81873	-2.82225
2173	0.0224314	0.0120845	2.81064	2.81413
785	0.0233302	0.0119718	2.74183	2.74505
2704	0.0233642	0.0118208	2.72245	2.72559
2012	0.00734093	0.0113713	4.80259	4.82176

Table: *Top 10 observations sorted by cook's distance after removal of 2 suspicious instances*

10 References

Airbnb (2021, January 13)._About Us - Airbnb Newsroom_. news.airbnb.com. https://news.airbnb.com/about-us/

Airbnb (2021, January 13)._What is a Superhost? - Airbnb Help Centre_.airbnb.com. https://www.airbnb.co.uk/help/article/828/what-is-a-superhost

Baum,C. F.(2013, Spring)._EC 823: Applied Econometrics: Quantile Regression_.Boston College.https://fmwww.bc.edu/EC-C/S2013/823/EC823.S2013.nn04.slides.pdf

Cook, R. D., & Weisberg, S. (1999). *Applied regression including computing and graphics* .(Vol. 488). John Wiley & Sons.

Chang, C., & Li, S. (2020). Study of Price Determinants of Sharing Economy-Based Accommodation Services : Evidence from Airbnb . com. Theoretical and Applied Electronic Commerce Research, 16, 584–601.

Fedorov, V., Mannino, F., & Zhang, R. (2009). Consequences of dichotomization. Pharmaceutical Statistics, 8(1), 50–61. https://doi.org/10.1002/pst.331

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., & Goodwill, A. (2018). Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings. Journal of Travel and Tourism Marketing, 35(1), 46–56. https://doi.org/10.1080/10548408.2017.1308292

Gutt, D., & Hermann, P. (2015). Sharing Means Caring? Hosts' Price Reaction to Rating Visibility. European Conference on Information Systems (ECIS) Research-in-Progress Papers, 1–13.

Halvorsen, R., & Palmquist, R. (1980). The Interpretation of Dummy Variables in Semilogarithmic Equations. The American Economic Review, 70(3), 474–475.

Hill, D. (2015). How much is your spare room worth? IEEE Spectrum, 52(9), 32–58. https://doi.org/10.1109/MSPEC.2015.7226609

Hung, W. T., Shang, J. K., & Wang, F. C. (2010). Pricing determinants in the hotel industry: Quantile regression analysis. International Journal of Hospitality Management, 29(3), 378–384. https://doi.org/10.1016/j.ijhm.2009.09.001

Inside Airbnb(2020,November 10)._Inside Airbnb: Adding Data to the debate_.[Data set].http://insideairbnb.com/get-the-data.html

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* .(Vol. 112, p. 18). New York: springer.

Koenker, R., & Hallock, K. (2001). Quantile regression. Journal of Economic Perspectives, 15(4), 143–156. https://doi.org/10.1002/9781118752685.ch2

Lawani, A., Reed, M. R., Mark, T., & Zheng, Y. (2019). Reviews and price on online platforms: Evidence from sentiment analysis of Airbnb reviews in Boston. Regional Science and Urban Economics, 75(September 2017), 22–34. https://doi.org/10.1016/j.regsciurbeco.2018.11.003

Li, J., Moreno, A., & Zhang, D. (2015). Pros vs Joes: Agent Pricing Behavior in the Sharing Economy. Ssrn, 1298. https://doi.org/10.2139/ssrn.2708279

Perez-Sanchez, V. R., Serrano-Estrada, L., Marti, P., & Mora-Garcia, R. T. (2018). The what, where, and why of airbnb price determinants. Sustainability (Switzerland), 10(12), 1–31.

https://doi.org/10.3390/su10124596

Statista (2019, November 19)._Number of Airbnb listings in selected European cities as of 2019_. statista.com. https://www.statista.com/statistics/815145/airbnb-listings-in-europe-by-city/

Wang, D., & Nicolau, J. L. (2017). Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. International Journal of Hospitality Management, 62, 120–131. https://doi.org/10.1016/j.ijhm.2016.12.007

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. International Journal of Hospitality Management, 28(1), 180–182. https://doi.org/10.1016/j.ijhm.2008.06.011

Zhang, Z., Chen, R. J. C., Han, L. D., & Yang, L. (2017). Key factors affecting the price of Airbnb listings: A geographically weighted approach. Sustainability (Switzerland), 9(9), 1–13. https://doi.org/10.3390/su9091635