How Does Superhost Accreditation Benefit Airbnb Hosts in Chicago[1]

Yongfei Lu[2]

June 2020

Abstract

In order to obtain a superhost accreditation, an Airbnb host has to meet a series of strict requirements, including a 90% response rate, 80% 5-star reviews, honor confirmed reservations, etc. However, is it worth the effort to get such an accreditation? How does the superhost status benefit the hosts? Choosing Chicago as our target city, where Airbnb hosts have earned tens of millions of dollars over the 5 biggest weekends in 2019, we take advantage of the k-nearest neighbors(KNN) algorithm, decision tree and multilinear regression models to isolate the effects. Our findings confirm that superhosts do have significantly higher income in comparison with normal hosts. Nevertheless, such accreditation does not bring up their rental price. One critical channel through which superhosts earn more lies in their capacity to maintain a relatively higher occupancy rate, which is achieved mainly through better accommodation services. In addition, even though Chicago has a notorious reputation for high crime rates, safety issues neither appear to bother travelers nor reduce host revenues.

Key words: Airbnb superhost, price, revenue, KNN, decision tree, multilinear.

[1] I'd like to express my gratitude to Professor Richard Evans, who generously provides academic instructions and comments on my work in the past few months. I would also like to thank my classmates of MACS 3250, who unreservedly give me feedback and suggestions on improving my research.
[2] MACSS, University of Chicago, yongfeilu@uchicago.edu

1 Introduction

First proposed by Michael Spence in 1973, signaling theory has made a fundamental impact on our understanding of transactions in the market. According to him, educational credentials serve as reliable signals to employers when they are making hiring decisions in the labor market (Michael, 1973). Scholars further explore the mechanisms under the hood and find that adverse selection in the eBay online market caused by information asymmetries can be mitigated by rich information exchange through bandwidth and technology (Gregory, 2011). Recently, a new form of economy unrolls swiftly in online markets via the platforms like Airbnb, which is named 'sharing economy'. It has been claimed that one standard deviation increase in Airbnb listings is associated with a rise in asking rents of 0.4% in Boston (Keren and Mark, 2017). This, consequently, leads us to questions like how these establishers and lessors apply the signaling theory to build up their own businesses on the Internet. To emit positive signals to consumers, Airbnb has launched the Superhost program since 2019, which certificates those mostly dedicated to providing outstanding hospitality service as superhosts. The focus of the research, therefore, is to explore whether the superhost accreditation benefits Airbnb hosts in the city of Chicago and how the mechanisms work if it is the case.

2 Literature Review

A myriad of research has been done revolving the important attributes that make a room stand out in the market. Even though consumers tend to evaluate their living experience based on past hotel stays, their evaluation standard varies across different cities and differs from that of conventional hotels. Considering the complexity of the relationship between determinants and room price, non-parametric approaches, such as random forest and decision tree models, are employed to identify the critical factors (Manojit and Subrata, 2019). This inspires the researcher to use the decision tree and KNN models to explore what room features are mostly associated with price setting. Through sentiment analysis, ‘location’, ‘amenities’ as well as ‘host’ are the 3 important factors that contributes to a good living experience for tenants, while ‘noise’ is mostly likely to cause unpleasant comments (Mingming and Xin, 2019). As situations can vary across cities, the researcher would like to investigate how various room features impact travelers’ living experience in the City of Chicago. In particular, combined with Chicago’s crime data, the researcher is keen to check how safety issues concern short-term tenants.

In literature, there are contradictory opinions and assessments on the brand effect. Using the 13-month panel dataset on 1998-99 Internet shopping behavior, researchers show that information use for price comparison weakens the pull of brand by reducing shopping at concentrated branded retailers by approximately one tenth (Waldfogel and Chen, 2006). In the case of Airbnb, where price comparison is allowed and within easy reach, we can reasonably regard superhost accreditation as the special ‘brand’, or positive signal and explore its effects. In the hedonic price regression model, reputational aspects of the rooms, including rating scores and duration of membership, are found to be associated with economic reward (Timm et al., 2017). However, report from AirDNA shows that on the whole, superhosts charge less on a nightly basis but maintain higher occupancy rate (Scott, 2018). Another research, based on 33 cities listed on Airbnb.com, is contradictory to this and demonstrates that hosts with superhost status, more listings and verified identities usually set prices at a higher level since these are regarded as a kind of quality signals (Dan et al., 2017). As stated before, cases vary across locations. Such difference in research results may stem from use of distinct datasets for different cities. In this case, to develop a clear picture of Chicago, we will use the detailed listing dataset of Chicago.

In conclusion, the literature stated above gives an inspiration for the researcher to explore the specific situation in the city of Chicago. We will firstly apply the KNN model to check whether superhosts charge more. We’ll then explore whether the superhost status exerts a significant influence on revenue and investigate other critical revenue-influencing factors through the decision tree model. On top of this, we’ll further explore the underlying channels by studying what attributes separate superhosts from common ones. The machine learning methods subtly circumvent the limitation of using linear regression models and thus can more precisely capture the relationship between superhost status and revenue in the City of Chicago. Finally, as the robustness check, we'll use the multilinear regression to check whether the relationship explored above still holds.

3 Data

Our main source of data comes directly from insideairbnb.com, an independent, non-commercial and widely used website which provides monthly scraped data from Airbnb. We select the Chicago listing data of August, 2019 to do our research, when huge amounts of income was generated due to high concentration of various festivals. We also merge the dataset with the crime rate data released by Chicago Police Department. After cleaning the raw data, we obtain 8646 observations in our dataset. Moreover, we create a new variable occupancy rate based on the initially-existent variable reviews per month. The variable is calculated based on following equation, which is supported by the research on the effects of short-term rentals on San Francisco housing market, conducted by Budget and Legislative Analyst's Office in 2015. Here, we take 0.5 for review rate, and 3 for average length of stay. The two estimated values are provided on the insideairbnb website.

Figure 3.1 gives the correlation heat map for the main variables of our dataset. In our research, we mainly focus on variables, including revenue, price, host_is_superhost, avg_yearly_cirmes and others indicating accommodation service quality. From the graph, it's obvious that there is a positive correlation between revenue and occupancy

$$occupancy\ rate = \frac{reviews\ per\ month}{review\ rate} \times \frac{average\ length\ of\ stay}{30}$$

rate, while price doesn't seem to be correlated with the superhost status. Price appears to be influenced by cleaning fees and other variables demonstrating how suitable the room is for living, including bedrooms, and other amenities. Besides, even though superhost status doesn't seem to be correlated with rent rate, it is positively correlated various review rating scores and revenue, indicating the potential difference between superhosts and normal ones. On the other hand, Figure 3.2 sheds light on the possible effect of crime rate on host revenues. Somewhat counterintuitively, we haven't seen any negative effects of average yearly crimes per thousand residents on host revenues. Overall, we've had a big picture of potential relationships among the key variables, then we'll take advantage of different models to isolate our target effects.

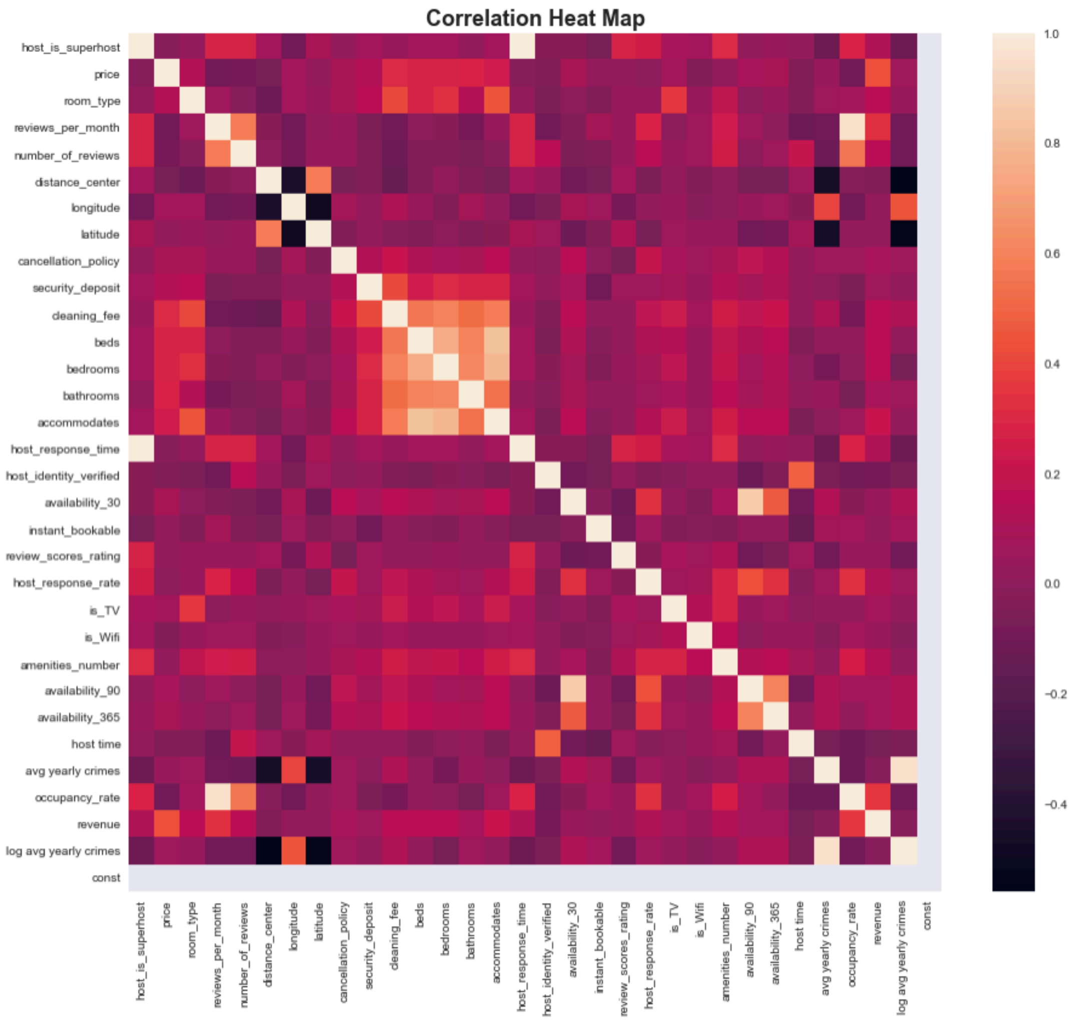


Figure 3.1 Correlation Heat Map

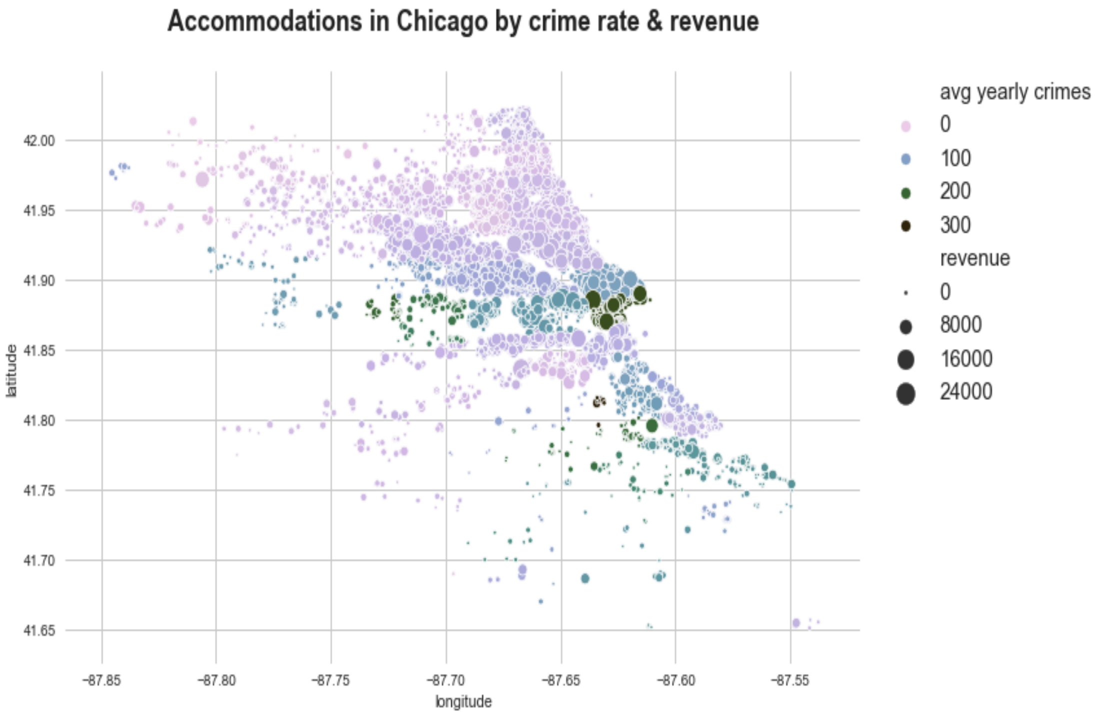


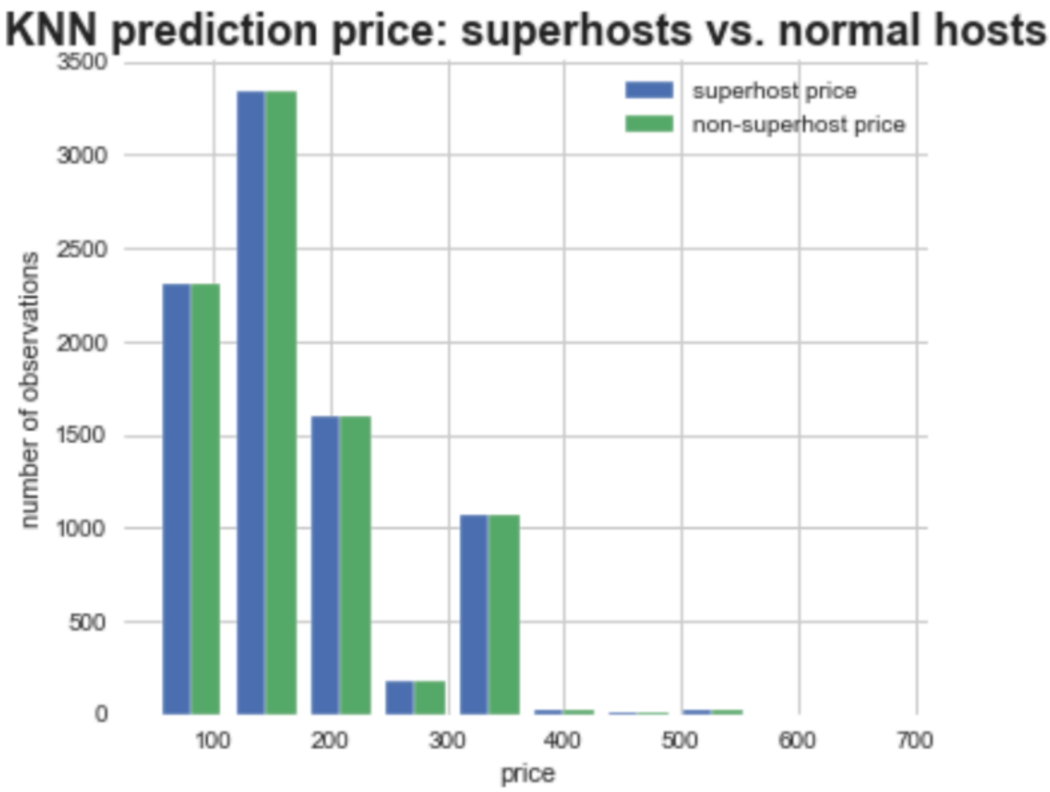
Figure 3.2 Accommodation Graph

4 Methods & Result

4.1 KNN: do Chicago superhosts get a ‘brand premium’?

Based on signaling theory, we assume that superhost accreditation sends positive signals to travelers that those with the certificates are mostly dedicated to providing outstanding hospitality services, thereby exerting more market power and raising their prices. To verify the assumption, we choose the KNN model to predict room prices, in which we use Euclidean distance to construct the similarity metric, and take host_is_superhost along with other important price-influencing variables as predictors. In the process, we randomly choose 75% of the sample to train the model, which excludes the observations whose price is exorbitantly high (over 1000 dollars per night), and obtain one with a root mean squares error (RMSE) of 89.24. Such a low RMSE value ensures that our model can reliably predict prices for different rooms. Then we counterfactually create two samples by setting the host_is_superhost for all the observations of the original dataset to 1 (yes) and 0 respectively and maintaining the original values of the other predictors. We use the trained model to predict the prices for these two samples whose only difference is the value of the host_is_superhost variable. Afterwards, we compute the Welch’s t-test[3] statistics to determine whether the predicted prices for the two samples are significantly different, and it turns out that superhosts don’t set a significantly higher price than normal hosts. We get a p-value of 0.97. The result is shown in Figure 4.1.

[3] Welch's t-test, or unequal variances t-test, can be used to test the hypothesis that two populations have equal means. It’s more reliable when two samples have unequal variances or unequal sample sizes.




```
Welch's t-test statistics:
T value: -0.034583596067340294
p value: 0.9724121859140163
```

Figure 4.1 KNN Prediction Result

4.2 Decision Tree: Revenue Prediction

The previous section has shown that superhosts don't set higher prices than normal ones. We now further our exploration to check other channels through which superhosts could gain more benefits. This time we use the non-parametric supervised learning method, decision regression tree, to determine whether superhosts gain more, along with the possible mechanisms. Here, we take revenue as the predicted variable, host_is_superhost, occupancy_rate, price and other variables related to the room equipment as predictors. As shown by Figure 4.2, the decision tree indicates that occupancy rate and price are the two most important predictors. Therefore, if we can demonstrate that superhosts have significant difference from normal hosts in these aspects, then we can conclude that superhost accreditation does influence host revenues.

Similarly, we compute the Welsh's t-test statistics to identify the differences. Among these predictor variables, we find those (including occupancy rate) which are significantly different among superhosts and normal ones at the 5% significant level, as illustrated by Table 4.2. Therefore, we conclude that superhosts can maintain higher occupancy rates, thus earning more.

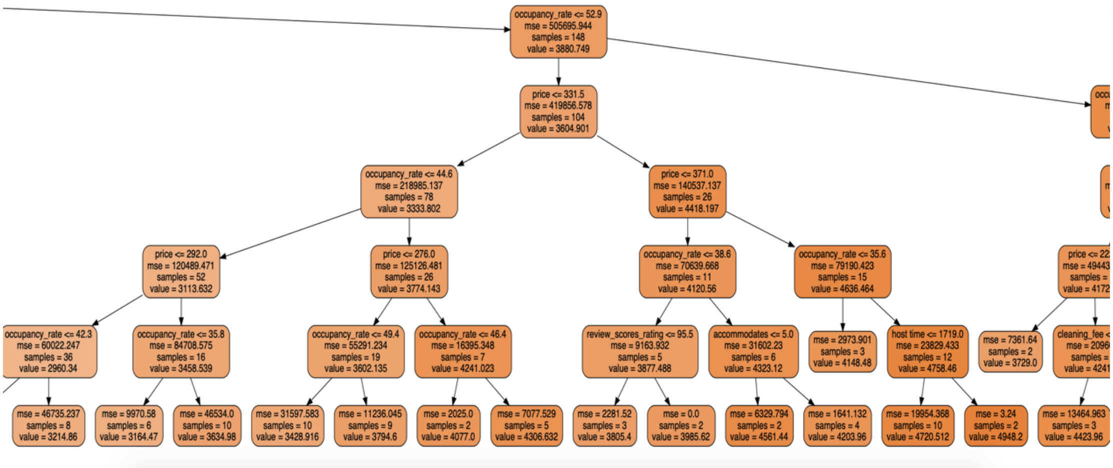


Figure 4.2 Decision Tree: Revenue Prediction

Variable	T value	P value
review_scores_rating	30.78394471	6.0413E-196
occupancy_rate	28.24623481	1.923E-166
number_of_reviews	24.73389822	5.2424E-128
bedrooms	7.370895682	1.87964E-13
accommodates	7.310573409	2.98256E-13
beds	6.943957388	4.21273E-12
cleaning_fee	3.762768068	0.00016926
bathrooms	2.446236554	0.014457674

Table 4.2 Welch's t-test statistics

4.3 Linear regression

At last, we use a multilinear regression model to do the robustness check. The models are specified as follows, where S is the superhost status, Z includes all price-influencing variables, and X includes all revenue-influencing factors.

$$Price = \theta + \gamma S + \beta X + u$$

$$Revenue = \alpha + \lambda S + \beta X + \mu$$

(1) Superhosts don't set up higher prices

Table 4.3(1) - price OLS Regressions

	Model 1	Model 2	Model 3
const	-16.57	-111.71***	-183.06***
	(12.18)	(25.67)	(21.06)
host_is_superhost	-2.63	3.51	-4.96*
	(2.94)	(3.20)	(2.58)
log avg yearly crimes	39.26***	33.06***	25.92***
	(2.79)	(2.76)	(2.28)
occupancy_rate		-0.67***	-0.66***
		(0.05)	(0.04)
host time		-0.01***	-0.01***
		(0.00)	(0.00)
host_response_rate		44.03***	4.80
		(4.57)	(3.80)
number_of_reviews		-0.14***	-0.03
		(0.03)	(0.03)
review_scores_rating		1.32***	1.15***
		(0.23)	(0.19)
cancellation_policy			8.45***
			(1.34)
cleaning_fee			0.41***
			(0.03)
security_deposit			0.02***
			(0.00)
beds			-8.71***
			(1.12)
bedrooms			6.48***
			(1.80)
bathrooms			30.53***
			(2.04)
accommodates			17.63***
			(0.86)
R-squared	0.02	0.07	0.40
	0.02	0.07	0.40
R-squared	0.02	0.07	0.40
No. observations	8548	8548	8548

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

From the regression result, we can see that the effect of superhost status on price setting is not significant in Model 1 and 2. While the effect in Model 3 is significant at the 10% significant level, it is negative, which means that superhosts tend to set lower prices ($4.96 lower than that of normal hosts). The number of bedrooms, bathrooms, and cancellation policy exert significantly positive effects on price. The stricter the cancellation policy, the higher the price.

Table 4.3(2) - revenue OLS Regressions

	Model 4	Model 5	Model 6
const	-974.21***	-2056.02***	-3438.09***
	(124.53)	(262.08)	(242.89)
host_is_superhost	325.63***	310.84***	268.35***
	(30.86)	(32.68)	(28.97)
occupancy_rate	35.95***	38.00***	36.64***
	(0.45)	(0.56)	(0.50)
price	1.08***	1.06***	0.53***
	(0.04)	(0.04)	(0.04)
log avg yearly crimes	192.23***	182.03***	155.51***
	(28.22)	(28.30)	(25.60)
host time		-0.04**	-0.06***
		(0.02)	(0.02)
number_of_reviews		-2.78***	-1.98***
		(0.32)	(0.28)
review_scores_rating		11.57***	9.76***
		(2.34)	(2.08)
host_response_rate		142.11***	-120.82***
		(46.54)	(42.65)
cancellation_policy			100.19***
			(15.14)
security_deposit			0.13***
			(0.04)
cleaning_fee			1.71***
			(0.32)
beds			-109.85***
			(11.78)
bedrooms			89.27***
			(20.03)
bathrooms			76.77***
			(22.72)
accommodates			165.19***
			(9.05)
room_type			368.64***
			(28.86)
R-squared	0.48	0.49	0.60
	0.48	0.49	0.60
R-squared	0.48	0.49	0.60
No. observations	8559	8559	8559

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

(2) Superhosts get more revenues by maintaining a higher occupancy rate

From the comparison in section 4.2, we know that superhosts maintain significantly higher occupancy rates. Now the regression model indicates that one percent increase in occupancy rate leads to an average increase of $36.64 in host revenues. We therefore conclude that superhosts can benefit from their accreditation by maintaining a relatively higher occupancy rate. Moreover, setting stricter cancellation policies is an effective way to increase income. It's also worth mentioning that renting an entire apartment brings more revenues than renting private or publicly sharing rooms.

(3) Crime rate is not a big concern plaguing the business of Airbnb in Chicago

Even though Chicago has a notorious reputation for high crime rate, our results show that one percent increase in average yearly crimes per thousand residents leads to a $0.2592 increase in price, and $1.5551increase in host revenues. Though significant, the trivial economic effects verify our initial observation of Figure 3.2 that safety issues haven't become a big concern for travelers impairing Airbnb business in Chicago.

5 Conclusion

Based on signaling theory, the research intends to explore the mechanisms through which Airbnb hosts benefit from the superhost accreditation. While research results vary across locations, we focus on the case of the City of Chicago, where Airbnb business flourishes in the context of relatively higher crime rates. We first use KNN prediction model to check whether supherhosts set higher prices, and then apply decision tree to extract important revenue-influencing factors. In comparison with traditional parametric approaches, these machine learning methods can more precisely capture our target relationship. Finally, we use multilinear regression, which is also good at explaining the effects, to do a robustness check. Our research confirms that superhosts don’t ask a higher price, but instead decrease their price a little bit and maintain a significantly higher occupancy rate, thereby increasing their revenues. In addition, the city’s reputation of high crime rates does not impair the Airbnb short-term rental business.

While the research yields consistent results, it does have limitations that need further disposal. As we calculate the occupancy rate based on certain already-existent variables, we may introduce error to our estimation. Besides, we only focus on the period in a year when huge amounts of revenues are generated, therefore may omit some other important patterns. In the future, we may need to develop a more precise model to estimate the occupancy rate, and use more extensive dataset to conduct our investigation.

APPENDIX

A-1 Descriptive Statistics

	N	Mean	St.Dev	min	p25	p75	max	t-value
index	8646	4326.32	2496.39	0	2165	6488	8651	161.14
revenue	8646	1780.82	4288.51	0	204	2241	187200	38.61
host is superhost	8646	.39	.49	0	0	1	1	74.58
occupancy rate	8646	38.91	34.18	0	7.6	66	100	105.84
price	8646	176.64	397.15	0	69	189	10000	41.36
log avg yearly cri~s	8646	4.26	.51	2.83	3.84	4.66	5.57	772.98
host time	8646	1382.67	791.9	-16	752	1981	4002	162.35
number of reviews	8646	39.11	57.72	0	3	51	588	63.01
room type	8646	2.66	.52	1	2	3	3	473.78
review scores rating	8646	95.16	6.33	20	94	99	100	1397.23
host response rate	8646	.86	.33	0	.98	1	1	241.02
cancellation policy	8646	2.23	.88	1	2	3	6	236.17
security deposit	8646	163.58	360.65	0	0	250	5000	42.18
cleaning fee	8646	59.09	58.97	0	19	90	1500	93.18
beds	8646	2.26	2.06	0	1	3	50	101.55
bedrooms	8646	1.61	1.17	0	1	2	24	128.28
bathrooms	8646	1.36	.77	0	1	1.5	21	165.69
accommodates	8646	4.34	3	1	2	6	38	134.41

References

Michael Spence, “Job Market Signaling,” The Quarterly Journal of Economics, 1973, 87(3), 355-374.

Gregory Lewis, “Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors,” American Economic Review, 2011, 101(4).

Keren Horn, Mark Merante, “Is home sharing driving up rents? Evidence from Airbnb in Boston,” Journal of Housing Economics, 2017, 38, 14-24.

Manojit Chattopadhyay, Subrata Kumar Mitra, “Do airbnb host listing attributes influence room pricing homogenously?” International Journal of Hospitality Management, 2019, 81, 54-64.

Mingming Cheng, Xin Jin, “What do Airbnb users care about? An analysis of online review comments,” International Journal of Hospitality Management, 2019, 76, 58-70.

Waldfogel, J. and Chen, L., “Does information undermine brand? Information intermediary use and preference for branded web retailers,” The Journal of Industrial Economics, 2006, 54, 425-449.

Teubner, Timm and Hawlitschek, Florian and Dann, David, “Price Determinants on Airbnb: How Reputation Pays Off in the Sharing Economy,” Journal of Self-Governance and Management Economics, 2017, 5, 53-80.

Scott Shatford, “What is Airbnb’s Superhost Status Really Worth?” 2018. https://www.airdna.co/blog/airbnb_superhost_status

Dan Wang, Juan L. Nicolau, “Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com,” International Journal of Hospitality Management, 2017, 62, 120-131.

San Francisco Planning Department, ‘Amendments Relating to Short-Term Rentals’, 2014.

Budget and Legislative Analyst's Office, 'Analysis of the impact of short-term rentals on housing', Policy Analysis Report, 2015.