How Does Superhost Accreditation Benefit Airbnb Hosts in Chicago

Yongfei Lu

May 26, 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 them? 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 relatively higher occupancy rate, which is achieved mainly by better accommodation services. In addition, even though Chicago has a notorious reputation for high crime rate, safety issues neither appear to bother travelers nor reduce host revenues.

1. 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 were 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.

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

Figure 1.1 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 other variables indicating host service quality. From the graph, it's clear that there is positive correlation between revenue and occupancy rate, while price doesn't seem to be correlated with whether the host is superhost or not. Price appear to be influenced by cleaning fees and other variables demonstrating how suitable

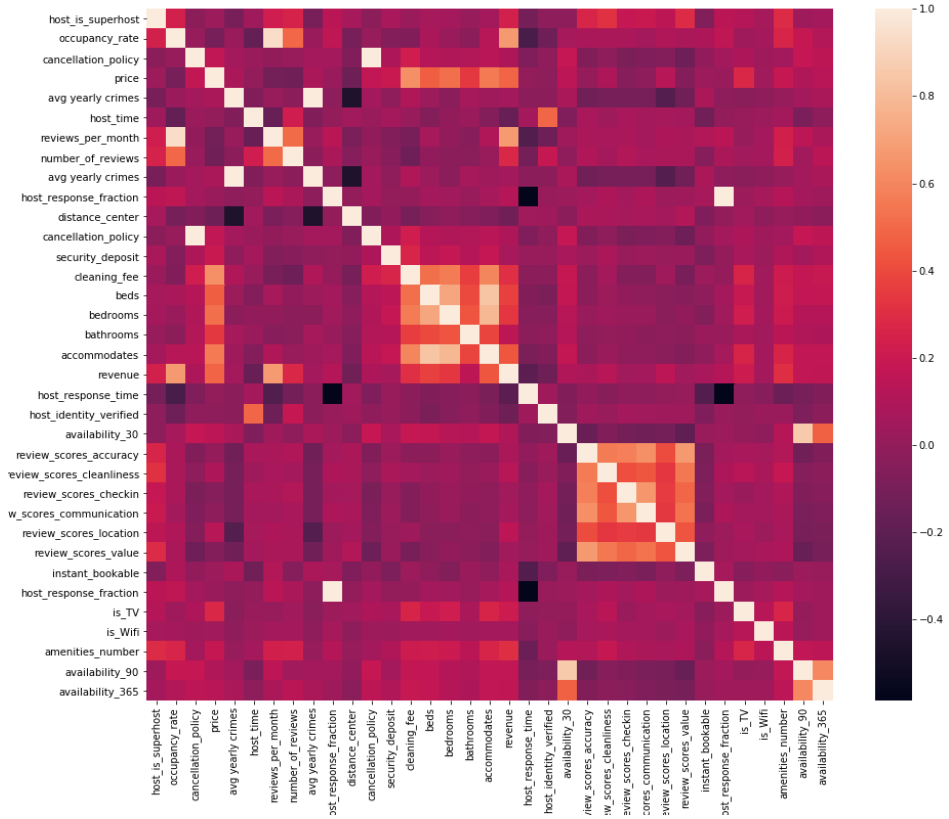

Figure 1.1 heat map

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. On the other hand, Figure 1.2 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.

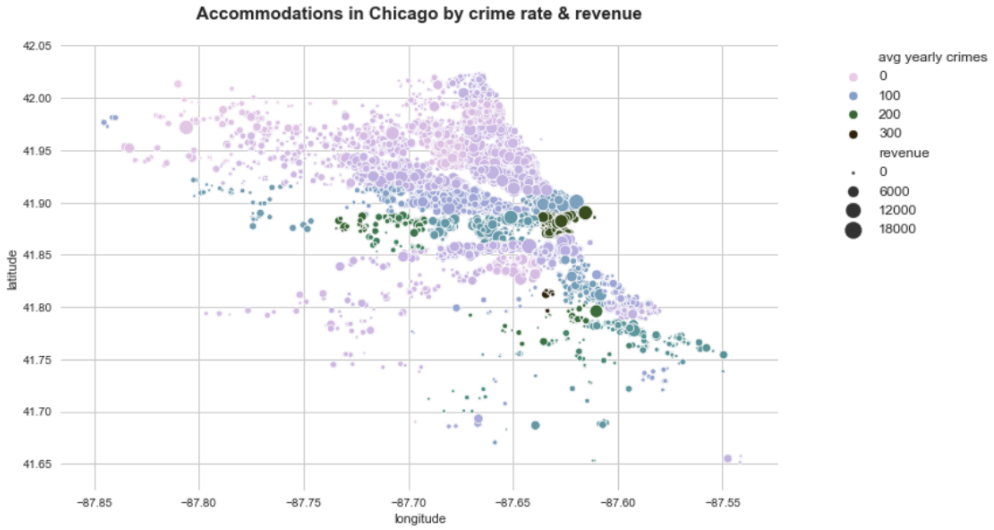

Figure 1.2 Accommodation Graph

2. Methods & Result

2.1 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 check 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 use 75% of the sample to the model, which is randomly selected and excludes the observations whose price is exorbitantly high (over 1000 dollars per night), and obtain one with a root mean squares error (RMSE) of 89.30. Such a low RMSE value indicates 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 our 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 Welsh's t-test[1] 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. Table 2.1 shows the result.

Table 2.1 KNN prediction result

```
T value: -0.03294459848962013
p value: 0.9737191522687281
```

	superhost price	non-superhost price
count	8548.000000	8548.000000
mean	158.841203	158.882218
std	81.391222	81.392298
min	49.200000	49.200000
25%	105.800000	105.800000
50%	117.200000	117.200000
75%	201.200000	201.200000
max	686.600000	686.600000

2.2 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,

[1] 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

host_is_superhost, occupancy_rate, price and other variables related to the room equipment as predictors. As shown by Figure 2.2, the decision tree indicates that occupancy rate and price are the two most important predictors. This leaves us with the question whether superhosts have significant difference from normal hosts in these aspects. 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 1% significant level, as illustrated by Figure 2.2(2). Therefore, we conclude that superhosts tend to have higher occupancy rate, thus earning more.

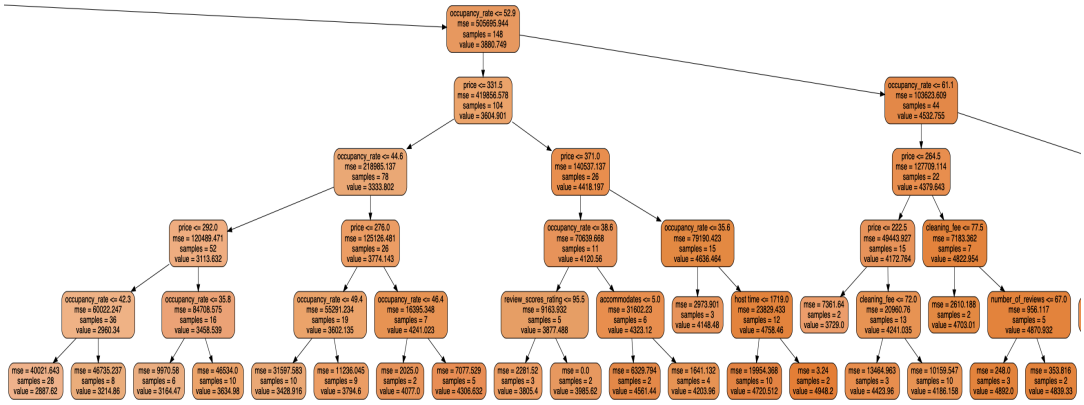

Figure 2.2 part of the decision tree (too large to view full graph)

Table 2.2 Hosts v.s. Superhosts

Variable	T value	P value
host_is_superhost	inf	0.000000e+00
review_scores_rating	27.894031	7.390617e-158
occupancy_rate	19.944346	8.869971e-86
number_of_reviews	19.934474	3.996356e-85
beds	5.409559	6.580762e-08
bedrooms	4.665044	3.151869e-06
accommodates	4.175125	3.020912e-05
room_type	2.610262	9.069459e-03

2.3 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, $\boldsymbol{Z}$ includes all price-influencing variables, and $\boldsymbol{X}$ includes all revenue-influencing factors.

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

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

(1) Superhosts don't set up higher prices

From the regression result, we can see that the effect of superhost status on price setting is not significant. The number of beds, bedrooms, bathrooms, and cancellation policy exert significantly positive effects on price. The stricter the cancellation policy, the higher the price.

Table 2.3(1) - price OLS Regressions

	Model 1	Model 2	Model 3
const	-4.35	-60.10	-151.70**
	(36.54)	(78.48)	(75.10)
host_is_superhost	-3.49	14.37	-3.31
	(8.80)	(9.74)	(9.18)
log avg yearly crimes	42.80***	34.38***	20.44**
	(8.38)	(8.45)	(8.13)
occupancy_rate		-1.09***	-0.72***
		(0.17)	(0.16)
host time		-0.02***	-0.03***
		(0.01)	(0.01)
host_response_rate		0.29**	-0.47***
		(0.14)	(0.14)
number_of_reviews		-0.17*	0.04
		(0.10)	(0.09)
review_scores_rating		1.44**	1.06
		(0.70)	(0.66)
cancellation_policy			20.59***
			(4.78)
cleaning_fee			1.09***
			(0.10)
security_deposit			0.02
			(0.01)
beds			21.25***
			(3.67)
bedrooms			29.49***
			(6.32)
bathrooms			65.07***
			(7.06)
accommodates			-12.01***
			(2.70)
R-squared	0.00	0.01	0.13
	0.00	0.01	0.13
R-squared	0.00	0.01	0.13
No. observations	8646	8646	8646

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

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

With the comparison in section 2.2, we know that superhosts maintain a significantly higher occupancy rate. Now the regression model further indicates that one percent increase in occupancy rate leads to an average increase of $41.35 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

private or publicly sharing rooms brings more revenues than renting an entire apartment.

Table 2.3(2) - revenue OLS Regressions

	Model 4	Model 5	Model 6
const	-1670.68***	-1585.84***	-1586.00***
	(98.17)	(244.12)	(305.44)
host_is_superhost	177.87***	214.60***	235.80***
	(22.63)	(25.53)	(28.83)
occupancy_rate	38.58***	39.68***	41.35***
	(0.34)	(0.45)	(0.53)
price	13.35***	14.28***	14.65***
	(0.15)	(0.16)	(0.25)
log avg yearly crimes	-22.75	-33.41	-28.77
	(22.88)	(24.47)	(27.87)
host time		-0.08***	-0.09***
		(0.02)	(0.02)
number_of_reviews		-0.25	-0.35
		(0.22)	(0.26)
review_scores_rating		2.21	3.35
		(1.84)	(2.18)
host_response_fraction		-331.92**	-356.11**
		(135.86)	(171.75)
cancellation_policy			54.55***
			(16.46)
security_deposit			-0.05
			(0.04)
cleaning_fee			-0.63
			(0.45)
beds			55.05***
			(15.59)
bedrooms			5.80
			(22.55)
bathrooms			-105.09***
			(24.57)
accommodates			13.72
			(11.04)
room_type			-133.89***
			(34.80)
R-squared	0.76	0.76	0.78
	0.76	0.76	0.78
R-squared	0.76	0.76	0.78
No. observations	6303	5611	4464

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

(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 an $0.2 increase in price, and that no significant effects of crime rate has been discerned on host revenues. The trivial effects confirm our initial observation of Figure 2 that safety issues haven't become a hurdle impairing Airbnb business in Chicago.

3. Reference

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.