Duke UNIVERSITY

airbnb

Unified Data Science – IDS 701

# Does Airbnb listing's annual revenue vary by with host status?

**Prepared by:**

*Fides Schwartz, Jaya Khan, Satvik Kishore, Tego Chang*

*GitHub Repo: https://github.com/MIDS-at-Duke/uds-2022-ids-701-team-10*

## EXECUTIVE SUMMARY

In recent years, Airbnb hosts have been generating revenue to supplement or replace their regular salaries. As per airbnb.com (Airbnb 2022b), Airbnb awards the label of super host to a regular host who fulfils a certain set of criteria for specialized service throughout the booking process and a guest's stay. The question that immediately occurs is: *"Do super hosts help Airbnb generate more annual revenue?"* We find evidence that super hosts can generate more annual revenue per listing than regular hosts based on the Airbnb super host label alone.

*The average difference in annual revenue between listings by hosts who we observe as super hosts and listings by hosts who we observe as regular hosts in a world whether neither is a super host is $3,127.* Based on regression results, when we control for other explanatory factors, on average, a super host has 153% higher annual revenue, which is 2.53 times the average annual revenue of a regular host with otherwise similar features. We validate these findings based on data from two selected states – California and Florida – the states that generate the highest revenue on Airbnb annually (Dogru et al. 2020).

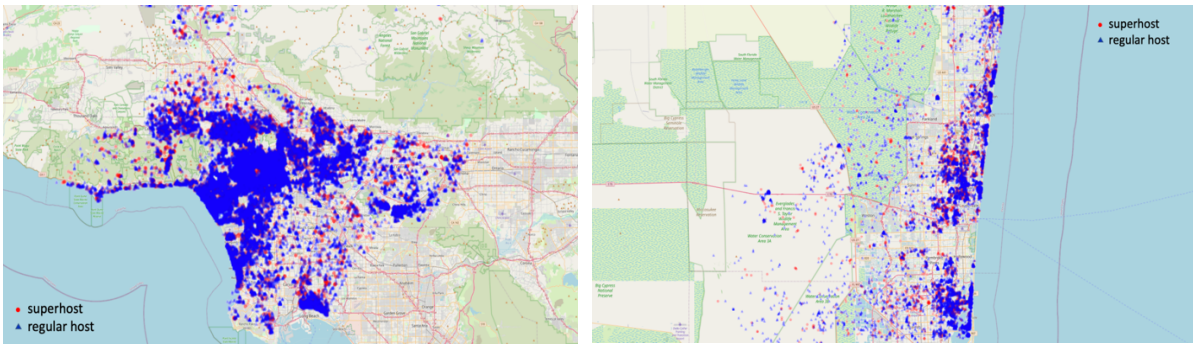



Figure 1.1: Shows similar distribution of listings belonging to super hosts (red circles) and listings belonging to regular hosts (blue triangles) in Los Angeles, California (left) and Broward County, Florida (*right*).

## I. INTRODUCTION

You are planning your first post-COVID family vacation and are trying to decide how to book accommodation for five people and two dogs quickly, because you left your decision making to the last minute, since Centres for Disease Control guidance is changing by the minute. While browsing the Airbnb listings for Wilmington, you find several beautiful looking houses, just a ten-minute drive from the beach and pet friendly with a separate room for the in-laws and a large kitchen to make meals and memories together. All the hosts offer instant booking and have been verified and active for years. But how do you make your decision now? Looking at the listings, only one of them has a super host label, so you end up booking their residence, because you expect better service than the regular hosts could provide.

This is, of course, a hypothetical scenario but one worth exploring through data analysis. Do people tend to book with Airbnb super hosts more frequently than with regular hosts, thus leading to higher annual revenue for super hosts over regular hosts with similar characteristics? Airbnb is one of the most prominent companies of the so-called "sharing economy" or "peer-to-peer markets" together with household names such as Uber and TaskRabbit, and it has had an impact on how people book holidays and the hotel industry in the markets it has established itself in (Zervas, Proserpio, and Byers 2017). Since its founding in 2008, approximately 500 million people have booked stays with Airbnb (Airbnb 2015), so the question of how to attract the largest number of these potential customers is relevant to a host's economic perspectives.

While Airbnb has changed the landscape for travellers looking for cheap or unique accommodation, it has also provided hosts with an increase in monthly income (increasing over the years (Management 2022; Poppick 2015)), and there are ~4 million global hosts with Airbnb listings (Management 2022; Lewis 2020), who are vying to take a slice of the huge holiday accommodation market. To understand what might drive decision making in customers, there has been some research on the influence of multi-property listings, as well as location and professionalism on Airbnb revenue in general and host revenue in particular (Chattopadhyay and Mitra 2020; Deboosere et al. 2019; Kwok and Xie 2019; Lane and Woodworth 2016; Xie, Heo, and Mao 2021; Xie and Mao 2019).

Our hypothesis is, that hosts who have the super host label awarded by Airbnb, can generate more annual revenue than regular hosts with similar characteristics but without the super host label, because customers elect to stay in their properties more frequently. InsideAirbnb.com makes some of Airbnb data available publicly, which makes it accessible to thorough data science analysis (Airbnb 2022a).

## II. MOTIVATION FOR ANALYSIS

With this project, we hope to find solutions for the travel and hospitality industry. The same idea of assigning a special category to some accommodations can be applied to other home booking sites and hotels that advertise their rooms on rental websites such as Expedia, Sonder and Booking.com. We are interested in one major piece of information: *Is the label of super host helpful in generating more revenue?* Our response variable is thus the *estimated annual revenue* for listings grouped by hosts, and our predictors are (1) the Airbnb assigned super host label and (2) all other listing and host based variables.

## III. DATA

### Overview

We are using the data provided by Airbnb on the following website: http://insideairbnb.com/get-the-data.html. This data includes timepoints in March, June, September, and December of 2021 for 104 cities/regions that have Airbnb listings all over the world. Airbnb provides a data dictionary (https://tinyurl.com/y7h9m4nu) that includes 73 variables.

### Data Selection

Since there are vast amounts of data that would all need different pre-treatment, we consider datasets from two places for this project. Since California and Florida are the two highest ranking states for Airbnb revenue generation in the USA, *Los Angeles, California* and *Broward County, Florida* **(Figure 1.1)** were chosen (these were the two counties available for these states from Airbnb; (Dogru et al. 2020)). We decided to compare the difference between these two locations at the beginning of our research on the relationships between super host and the estimated annual revenue, because we wanted to include a locational component in the evaluation. Florida and California might attract very different types of travellers and those travellers may have disparate criteria for choosing to stay at an Airbnb listing.

### Data Cleaning

Once our data collection was complete, we went on to data processing and data wrangling. First, we excluded the columns that we assume to have no impacts on our response of interest – *estimated annual revenue per listing* – logically (e.g., *addresses of listings, host name, description*). Second, we decided to drop all columns that had duplicate information (e.g., one of '*bathrooms*' and '*bathroom_text*'). Third, we decided to drop the columns with the most missing data and to impute the rows with few missing data. To estimate annual revenue, we multiplied the *price of the listing* with *reviews in the last 12 months* because only 67% of travellers leave a review after their stay (Zervas, Proserpio, and Byers 2017).

---

*estimated annual revenue = price of listing * number of reviews in the last 12 months * 100/67*

---

## IV. MATCHING

To be able to analyse what influence the super host status has on the estimated annual revenue per listing, we needed to match all other factors that might influence revenue (e.g., location, size of property, star-rating) between the regular hosts and super hosts as closely as possible. To do this, we decided to use DAME-FLAME.

To prepare data for the matching process, several continuous variables had to be discretized and coarsened to limit the amount of detailed matching that would otherwise be attempted. An initial run with 50 iterations yielded a stark increase in prediction error before the $10^{th}$ iteration (see appendix), which led us to repeat the task with only 10 iterations, where

we saw a notable increase in prediction error between the 3rd and 4th iteration **(Figure 1.2**). Based on these diagnostics, we decided to use the third iteration of our matching process. This included 10,223 matches after the covariates pertaining to number of beds and bathrooms associated with a listing were dropped from the analysis. Out of these 10,223 matched grouped, there was a total of 4,218 listings owned by super hosts and 6,005 listings owned by regular hosts in the dataset.

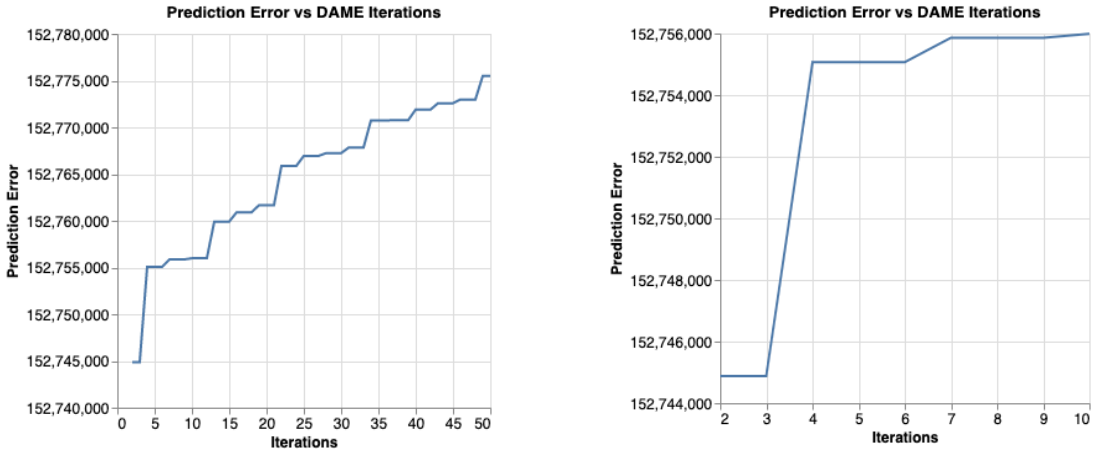



Figure 1.2: Shows the plot of DAME diagnostics at 50 (*left*) and 10 iterations (*right*). The prediction error increases markedly between the 3rd and 4th iteration.

## V. REGRESSION

Once the matching process was complete, we were able to run the multiple linear regressions. Our regressions were based on the matched dataset provided by DAME-FLAME output. We first ran a regression that included our response variable grouped by host status and all predictors (except the one used in revenue calculation that was part of our analysis after data cleaning). Assessment of this model violated the assumption of linearity and normality with a Q-Q-plot that suggested an exponential distribution (**Figure 1.3**), which led us to run a second regression on log-transformed data. This second regression satisfied the assumptions for linear regressions **(Figure 1.4),** and we proceeded to use it four our statistical analysis.

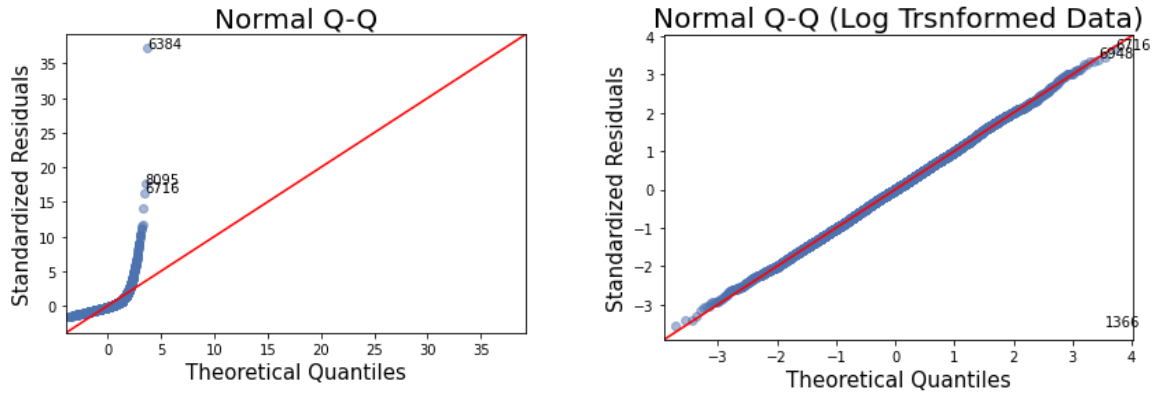



Figure 1.3: Shows Q-Q-plots generated to test the assumption of normality in our data. The shape of the plot on right suggests that a log transformation of the data could improve model performance down the line.

| | |
|---|---|
| *log(Annual Revenue) ~* | *C(super-host) + C(county) + C(room type) + C(accommodates) + C(bathrooms) + C(bedrooms) + C(beds) + C(essentials) + C(other amenities) + C(instant booking available) + C(minimum nights) + C(maximum nights) + C(review score on accuracy) + C(review score on check-in) + C(review score on cleanliness) + C(review score on communication) + C(review score on location) + C(review score on value) + C(overall rating) + C(host identity verified) + C(listings owned by hosts) + C(host acceptance rate) + C(host response time)* |

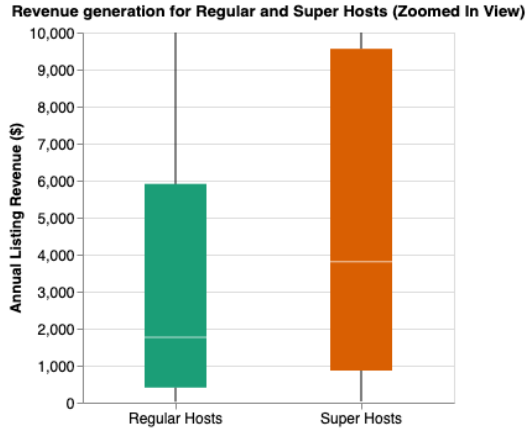
Figure 1.4: Shows the final regression equation which includes all the discretized predictors from the DAME output.

## VI. SUMMARY STATISTICS

### Average Treatment Effect

After removing the baseline differences between the two groups of hosts through our matching process, we found that our causal inference yields the following result: *The average difference in annual revenue between listings by hosts who we observe as super hosts and listings by hosts who we observe as regular hosts in a world whether neither is a super host is $3,127.*

### Regression Results




On average, super host makes **2.53** times higher annual revenue than a regular host when we control for all other predictors.

Median of Annual Revenue generation for regular hosts: $1765.67

Median of Annual Revenue generation for super hosts: $3807.46

Figure 1.5: Shows the output from regression model: annual revenue generation for super hosts and regular hosts (zoomed in view).

To explore which other variables might be relevant for revenue generation of an Airbnb listing, we generated a plot that visualizes both error bars and confidence intervals **(Table 1.1).** We summarized the variables that seem to influence the revenue a listing generates, apart from the super host label awarded by Airbnb in **Table 1.*2.***

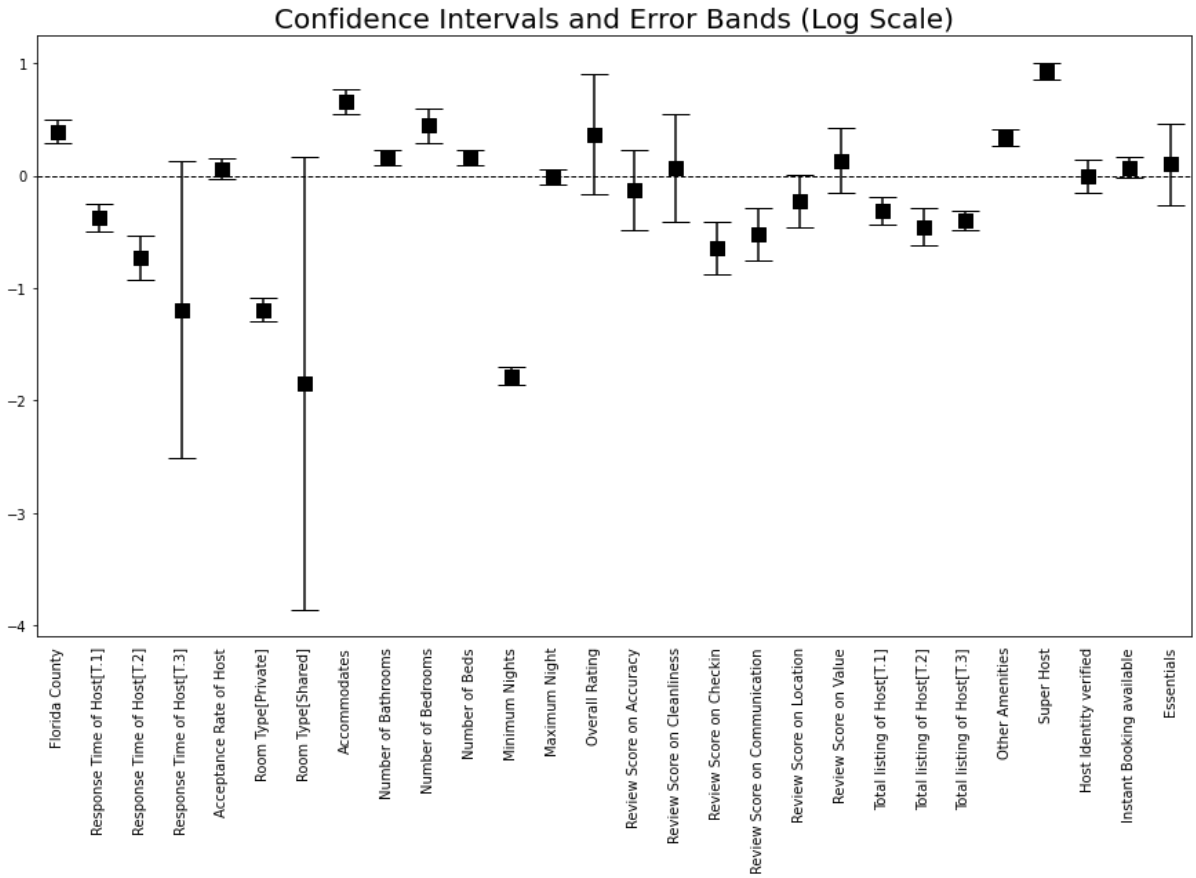



Table 1.1: Shows statistical significance of super host variable which is way above the '0' level boundary and have negligible error bands.

| | *log* (estimated Annual Revenue) |
|---|---|
| Intercept | 11.137*** |
| Super host | 0.930*** |
| Florida County | 0.394*** |
| Room Type [Private] | -1.193*** |
| Room Type [Shared] | -1.843* |
| Accommodates | 0.662*** |
| Number of bathrooms | 0.162*** |
| Number of bedrooms | 0.447*** |
| Number of beds | 0.162*** |
| Other Amenities | 0.344*** |

| | |
|---|---|
| Minimum Nights | -1.782*** |
| Review Score on Accuracy | -0.643*** |
| Review Score on Communication | -0.518*** |
| Review Score on Location | -0.225* |
| Total listing of Host [T.1] | -0.315*** |
| Total listing of Host [T.2] | -0.457*** |
| Total listing of Host [T.3] | -0.394*** |
| Response Time of Host [T.1] | -0.369*** |
| Response Time of Host [T.2] | -0.728*** |
| Response Time of Host [T.3] | -1.193* |
| $R^2$ | 0.432 |
| Adjusted $R^2$ | 0.430 |

Table 1.2: Shows regression output on log transformed response variable

## VII. DISCUSSION

In conclusion, we found that being a super host does have an impact on overall estimated annual revenue, i.e. on average, super host status helps generate 153% more annual revenue than just a regular host status when we control for other explanatory factors. Other variables that we found to influence annual revenue are *room type* (entire residences create more revenue), *number of bathrooms* (more bathrooms create more revenue), *number of bedrooms* (more bedrooms create more revenue), *accommodates* (higher the accommodates higher the revenue), *number of beds* (higher the number of beds higher the revenue), *other amenities* (more amenities create more revenue), *minimum night stays* (less the number of minimum night policy higher the revenue), *review score on check-in*, *communication, location* (surprisingly, these have a negative effect on overall revenue), *host listings count* (less listings create more revenue), and *host response time* (less time creates more revenue). We also noted that county in Florida potentially has makes more revenue than the county in California.

### Limitations

There were a few caveats to our analysis. The first is that we did not use all the data available to us through InsideAirbnb and chose only to focus on two states from the United States. This analysis can be expanded to include more locations in a follow-up study. The second is that we were limited by the quality of the data scraped by InsideAirbnb, along with potential selection biases arising through the scraping methodology. The third is that since we do not have longitudinal data, our effect estimates rely on the matching algorithm, and thus our results are dependent on the quality of the matches generated.

## References


[1] Airbnb. (2015). "Airbnb Summer Travel Report 2015." Retrieved 04/03/2022, from https://blog.atairbnb.com/wp-content/uploads/2015/09/Airbnb-Summer-Travel-Report-1.pdf.Airbnb. (2022). "Get the Data." Retrieved 04/03/2022, from http://insideairbnb.com/get-the-data.html.

[2] Airbnb. (2022). "Superhost: Recognizing the best in hospitality." Retrieved 04/03/2022, from https://www.airbnb.com/d/superhost.

[3] Chattopadhyay, M. and S. K. Mitra (2020). "What Airbnb Host Listings Influence Peer-to-Peer Tourist Accommodation Price?" Journal of Hospitality & Tourism Research44(4): 597-623.

[4] Deboosere, R., D. J. Kerrigan, D. Wachsmuth and A. El-Geneidy (2019). "Location, location and professionalization: a multilevel hedonic analysis of Airbnb listing prices and revenue." Regional Studies, Regional Science6(1): 143-156.

[5] Dogru, T., M. Mody, C. Suess, N. Line and M. Bonn (2020). "Airbnb 2.0: Is it a sharing economy platform or a lodging corporation?" Tourism Management78: 104049.

[6] Kwok, L. and K. L. Xie (2019). "Pricing strategies on Airbnb: Are multi-unit hosts revenue pros?" International Journal of Hospitality Management82: 252-259.

[7] Lane, J. and R. M. Woodworth (2016). "The sharing economy checks in: An analysis of Airbnb in the United States." CBRE Hotel's Americas Research.

[8] Lewis, T. (2020). "Airbnb Statistics (Growth, Revenue, Hosts + More!)." Retrieved 04/03/2022, from 8.https://hostsorter.com/airbnb-statistics/.Management, i. (2022). "Airbnb Statistics." from https://ipropertymanagement.com/research/airbnb-statistics.

[9] Poppick, S. (2015). "Airbnb Says Renting Your Place Is Like Getting a Big Raise." Retrieved 04/03/2022, from https://money.com/airbnb-raise-income-report/.Xie, K., C. Y. Heo and Z. E. Mao (2021). "Do professional hosts matter? Evidence from multi-listing and full-time hosts in Airbnb." Journal of Hospitality and Tourism Management47: 413-421.

[10] Xie, K.and Z. Mao (2019). "Locational strategy of professional hosts: Effect on perceived quality and revenue performance of Airbnb listings." Journal of Hospitality & Tourism Research43(6): 919-929.

[11] Zervas, G., D. Proserpio and J. W. Byers (2017). "The Rise ofthe Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry." Journal of Marketing Research54(5): 687-705.

# APPENDIX

## 1.1 Without Log Transformed Regression Output

| | | | | | | |
|---|---|---|---|---|---|---|
| Intercept | 4446.5209 | 1142.905 | 3.891 | 0.000 | 2206.202 | 6686.840 |
| C(county)[T.1] | 1328.8406 | 303.843 | 4.373 | 0.000 | 733.248 | 1924.433 |
| C(host_response_time)[T.1] | -1219.4307 | 346.137 | -3.523 | 0.000 | -1897.927 | -540.934 |
| C(host_response_time)[T.2] | -2687.0573 | 557.718 | -4.818 | 0.000 | -3780.294 | -1593.821 |
| C(host_response_time)[T.3] | -2852.0052 | 3767.087 | -0.757 | 0.449 | -1.02e+04 | 4532.226 |
| C(host_acceptance_rate)[T.1] | -752.2059 | 258.505 | -2.910 | 0.004 | -1258.927 | -245.485 |
| C(room_type)[T.2] | -2136.8971 | 305.386 | -6.997 | 0.000 | -2735.514 | -1538.280 |
| C(room_type)[T.3] | -4728.4908 | 5749.763 | -0.822 | 0.411 | -1.6e+04 | 6542.175 |
| C(accommodates)[T.1] | 2547.2895 | 307.149 | 8.293 | 0.000 | 1945.218 | 3149.361 |
| C(bathrooms_text)[T.1] | 1576.5118 | 199.068 | 7.919 | 0.000 | 1186.299 | 1966.724 |
| C(bedrooms)[T.1] | 425.6043 | 448.105 | 0.950 | 0.342 | -452.769 | 1303.978 |
| C(beds)[T.1] | 1576.5118 | 199.068 | 7.919 | 0.000 | 1186.299 | 1966.724 |
| C(minimum_nights)[T.1] | -3823.5832 | 230.633 | -16.579 | 0.000 | -4275.669 | -3371.498 |
| C(maximum_nights)[T.1] | 576.2506 | 199.775 | 2.884 | 0.004 | 184.652 | 967.849 |
| C(review_scores_rating)[T.1] | 1757.4957 | 1527.708 | 1.150 | 0.250 | -1237.112 | 4752.103 |
| C(review_scores_accuracy)[T.1] | -565.9764 | 1022.295 | -0.554 | 0.580 | -2569.875 | 1437.922 |
| C(review_scores_cleanliness)[T.1] | -832.1037 | 1359.674 | -0.612 | 0.541 | -3497.332 | 1833.124 |
| C(review_scores_checkin)[T.1] | -2262.8792 | 662.864 | -3.414 | 0.001 | -3562.223 | -963.535 |
| C(review_scores_communication)[T.1] | -1536.5589 | 675.322 | -2.275 | 0.023 | -2860.324 | -212.794 |
| C(review_scores_location)[T.1] | -913.5854 | 664.774 | -1.374 | 0.169 | -2216.673 | 389.502 |
| C(review_scores_value)[T.1] | 459.7578 | 827.029 | 0.556 | 0.578 | -1161.381 | 2080.897 |
| C(calculated_host_listings_count)[T.1] | -807.7889 | 348.004 | -2.321 | 0.020 | -1489.944 | -125.633 |
| C(calculated_host_listings_count)[T.2] | -1321.5017 | 477.667 | -2.767 | 0.006 | -2257.823 | -385.181 |
| C(calculated_host_listings_count)[T.3] | -822.0900 | 244.021 | -3.369 | 0.001 | -1300.420 | -343.760 |
| C(other_amenities)[T.1] | 1028.8056 | 215.863 | 4.766 | 0.000 | 605.671 | 1451.940 |
| host_is_superhost | 3035.8533 | 205.377 | 14.782 | 0.000 | 2633.274 | 3438.433 |
| host_identity_verified | 297.4074 | 412.308 | 0.721 | 0.471 | -510.797 | 1105.612 |
| instant_bookable | 69.2405 | 262.332 | 0.264 | 0.792 | -444.982 | 583.463 |
| essentials | 615.3525 | 1027.561 | 0.599 | 0.549 | -1398.870 | 2629.575 |

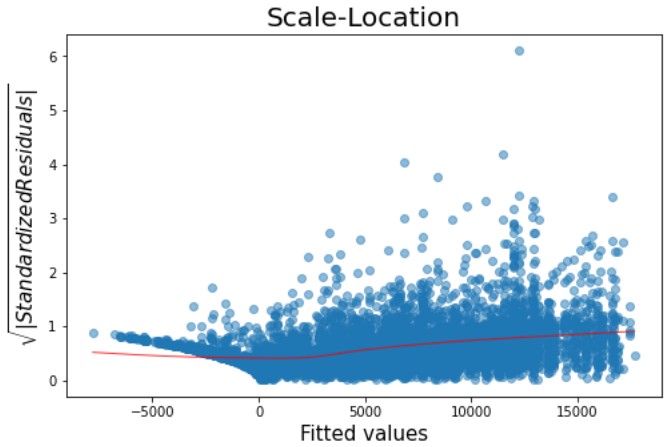
## 1.2 With Log Transformed Regression Output

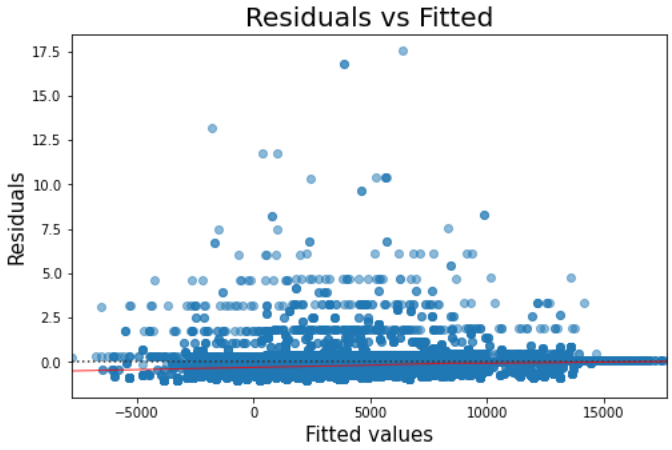
| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | 11.1366 | 0.255 | 43.719 | 0.015 | 7.900 | 14.373 |
| C(county)[T.1] | 0.3944 | 0.151 | 2.611 | 0.233 | -1.525 | 2.313 |
| C(host_response_time)[T.1] | -0.3692 | 0.106 | -3.487 | 0.178 | -1.715 | 0.976 |
| C(host_response_time)[T.2] | -0.7281 | 0.155 | -4.701 | 0.133 | -2.696 | 1.240 |
| C(host_response_time)[T.3] | -1.1932 | 0.472 | -2.529 | 0.240 | -7.187 | 4.801 |
| C(host_acceptance_rate)[T.1] | 0.0623 | 0.047 | 1.319 | 0.413 | -0.538 | 0.663 |
| C(room_type)[T.2] | -1.1926 | 0.184 | -6.484 | 0.097 | -3.530 | 1.145 |
| C(room_type)[T.3] | -1.8429 | 0.517 | -3.563 | 0.174 | -8.416 | 4.730 |
| C(accommodates)[T.1] | 0.6615 | 0.067 | 9.879 | 0.064 | -0.189 | 1.512 |
| C(bathrooms_text)[T.1] | 0.1622 | 0.112 | 1.449 | 0.385 | -1.260 | 1.584 |
| C(bedrooms)[T.1] | 0.4471 | 0.255 | 1.756 | 0.329 | -2.787 | 3.682 |
| C(beds)[T.1] | 0.1622 | 0.112 | 1.449 | 0.385 | -1.260 | 1.584 |
| C(minimum_nights)[T.1] | -1.7816 | 0.069 | -25.778 | 0.025 | -2.660 | -0.903 |
| C(maximum_nights)[T.1] | -0.0067 | 0.062 | -0.108 | 0.931 | -0.792 | 0.779 |
| C(review_scores_rating)[T.1] | 0.3657 | 0.023 | 15.944 | 0.040 | 0.074 | 0.657 |
| C(review_scores_accuracy)[T.1] | -0.1273 | 0.050 | -2.568 | 0.236 | -0.757 | 0.502 |
| C(review_scores_cleanliness)[T.1] | 0.0696 | 0.187 | 0.371 | 0.774 | -2.312 | 2.452 |
| C(review_scores_checkin)[T.1] | -0.6433 | 0.063 | -10.197 | 0.062 | -1.445 | 0.158 |
| C(review_scores_communication)[T.1] | -0.5178 | 0.009 | -56.632 | 0.011 | -0.634 | -0.402 |
| C(review_scores_location)[T.1] | -0.2251 | 0.212 | -1.064 | 0.480 | -2.914 | 2.464 |
| C(review_scores_value)[T.1] | 0.1340 | 0.149 | 0.897 | 0.534 | -1.763 | 2.031 |
| C(calculated_host_listings_count)[T.1] | -0.3149 | 0.307 | -1.024 | 0.492 | -4.221 | 3.591 |
| C(calculated_host_listings_count)[T.2] | -0.4569 | 0.240 | -1.901 | 0.308 | -3.511 | 2.597 |
| C(calculated_host_listings_count)[T.3] | -0.3943 | 0.282 | -1.400 | 0.395 | -3.972 | 3.183 |
| C(other_amenities)[T.1] | 0.3443 | 0.115 | 2.990 | 0.205 | -1.119 | 1.807 |
| host_is_superhost | 0.9297 | 0.037 | 25.193 | 0.025 | 0.461 | 1.399 |
| host_identity_verified | -0.0020 | 0.015 | -0.133 | 0.916 | -0.195 | 0.191 |
| instant_bookable | 0.0707 | 0.069 | 1.018 | 0.494 | -0.812 | 0.954 |
| essentials | 0.1041 | 0.022 | 4.667 | 0.134 | -0.179 | 0.387 |

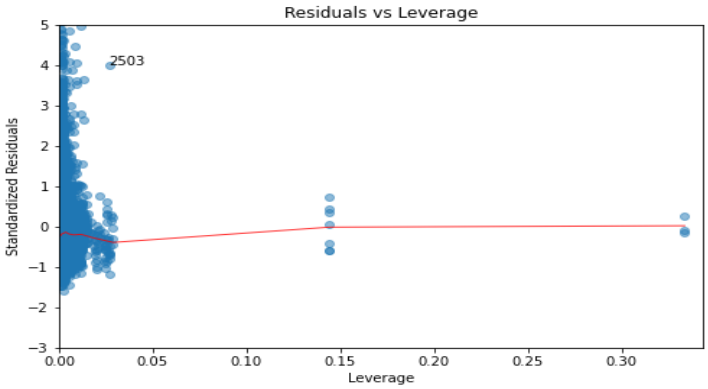
1.3 With Log Transformed Regression Output (clustered by hosts)

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | 11.1366 | 0.204 | 54.536 | 0.000 | 10.736 | 11.537 |
| C(county)[T.1] | 0.3944 | 0.054 | 7.265 | 0.000 | 0.288 | 0.501 |
| C(host_response_time)[T.1] | -0.3692 | 0.062 | -5.970 | 0.000 | -0.490 | -0.248 |
| C(host_response_time)[T.2] | -0.7281 | 0.100 | -7.307 | 0.000 | -0.923 | -0.533 |
| C(host_response_time)[T.3] | -1.1932 | 0.673 | -1.773 | 0.076 | -2.513 | 0.126 |
| C(host_acceptance_rate)[T.1] | 0.0623 | 0.046 | 1.349 | 0.177 | -0.028 | 0.153 |
| C(room_type)[T.2] | -1.1926 | 0.055 | -21.857 | 0.000 | -1.300 | -1.086 |
| C(room_type)[T.3] | -1.8429 | 1.027 | -1.794 | 0.073 | -3.857 | 0.171 |
| C(accommodates)[T.1] | 0.6615 | 0.055 | 12.054 | 0.000 | 0.554 | 0.769 |
| C(bathrooms_text)[T.1] | 0.1622 | 0.036 | 4.560 | 0.000 | 0.092 | 0.232 |
| C(bedrooms)[T.1] | 0.4471 | 0.080 | 5.585 | 0.000 | 0.290 | 0.604 |
| C(beds)[T.1] | 0.1622 | 0.036 | 4.560 | 0.000 | 0.092 | 0.232 |
| C(minimum_nights)[T.1] | -1.7816 | 0.041 | -43.234 | 0.000 | -1.862 | -1.701 |
| C(maximum_nights)[T.1] | -0.0067 | 0.036 | -0.187 | 0.851 | -0.077 | 0.063 |
| C(review_scores_rating)[T.1] | 0.3657 | 0.273 | 1.340 | 0.180 | -0.169 | 0.901 |
| C(review_scores_accuracy)[T.1] | -0.1273 | 0.183 | -0.697 | 0.486 | -0.485 | 0.231 |
| C(review_scores_cleanliness)[T.1] | 0.0696 | 0.243 | 0.286 | 0.775 | -0.407 | 0.546 |
| C(review_scores_checkin)[T.1] | -0.6433 | 0.118 | -5.432 | 0.000 | -0.875 | -0.411 |
| C(review_scores_communication)[T.1] | -0.5178 | 0.121 | -4.291 | 0.000 | -0.754 | -0.281 |
| C(review_scores_location)[T.1] | -0.2251 | 0.119 | -1.895 | 0.058 | -0.458 | 0.008 |
| C(review_scores_value)[T.1] | 0.1340 | 0.148 | 0.907 | 0.365 | -0.156 | 0.424 |
| C(calculated_host_listings_count)[T.1] | -0.3149 | 0.062 | -5.064 | 0.000 | -0.437 | -0.193 |
| C(calculated_host_listings_count)[T.2] | -0.4569 | 0.085 | -5.353 | 0.000 | -0.624 | -0.290 |
| C(calculated_host_listings_count)[T.3] | -0.3943 | 0.044 | -9.043 | 0.000 | -0.480 | -0.309 |
| C(other_amenities)[T.1] | 0.3443 | 0.039 | 8.926 | 0.000 | 0.269 | 0.420 |
| host_is_superhost | 0.9297 | 0.037 | 25.336 | 0.000 | 0.858 | 1.002 |
| host_identity_verified | -0.0020 | 0.074 | -0.027 | 0.978 | -0.146 | 0.142 |
| instant_bookable | 0.0707 | 0.047 | 1.509 | 0.131 | -0.021 | 0.163 |
| essentials | 0.1041 | 0.184 | 0.567 | 0.571 | -0.256 | 0.464 |

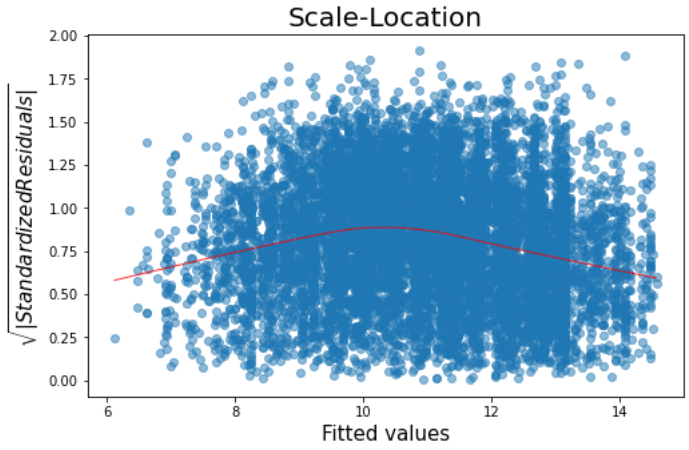
1.4 Model Evaluation – Without Log transformed output

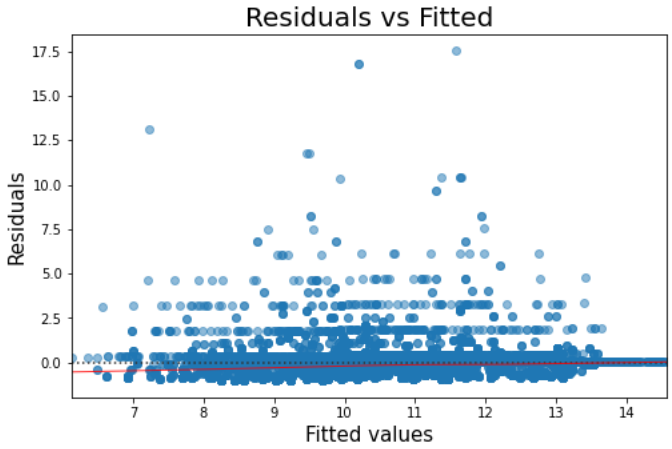


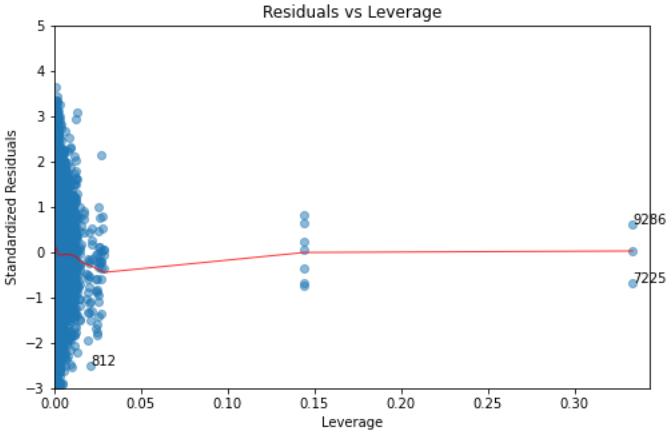




1.5 Model Evaluation – With Log transformed output


Scale-Location
$\sqrt{|StandardizedResiduals|}$
Fitted values


Residuals vs Fitted
Residuals
Fitted values


Residuals vs Leverage
Standardized Residuals
Leverage
9286
7225
812

**Revenue generation for Regular and Super Hosts (Log Scale)**