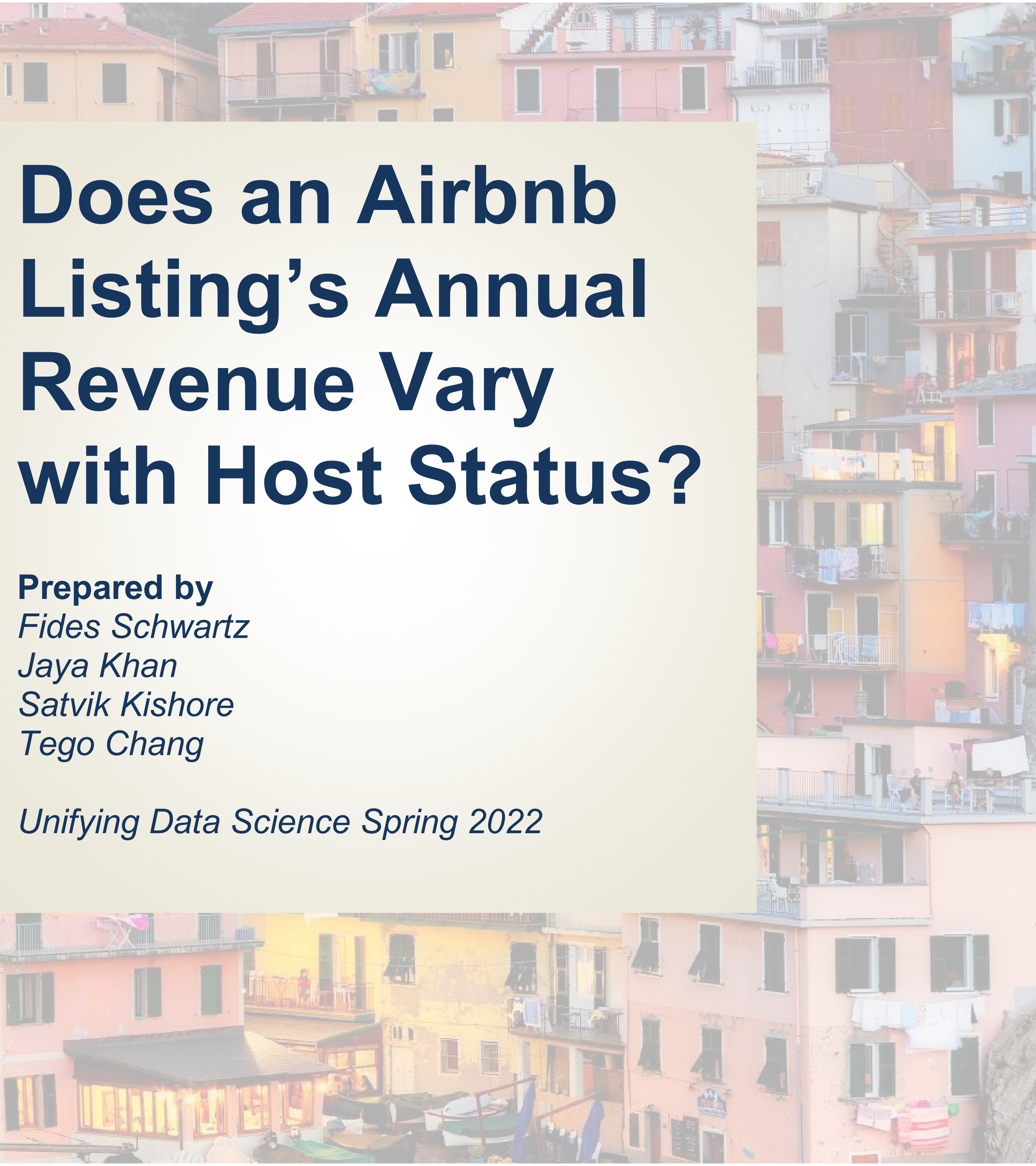


Does an Airbnb Listing's Annual Revenue Vary with Host Status?

Prepared by
Fides Schwartz
Jaya Khan
Satvik Kishore
Tego Chang

Unifying Data Science Spring 2022

Executive Summary

In recent years, Airbnb hosts have been generating revenue to supplement or replace their regular salaries.

As per airbnb.com (Airbnb 2022), Airbnb awards the label of superhost 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: "Does superhost status help Airbnb hosts generate more annual revenue?".

We find evidence that superhosts can generate more annual revenue per listing than regular hosts based on the Airbnb superhost label alone. On average, a superhost has an annual revenue of $3,807 per listing which is 2.53 times higher than the average revenue of a regular host with otherwise similar features.

We validate these findings based on data from two select states - California and Florida - the states that generate the highest revenue on Airbnb annually (Dogru, Mody et al. 2020).

SUPERHOSTS HAVE HIGHER ANNUAL REVENUE

Airbnb superhosts make an average of $2,041 more per listing annually than regular hosts

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 superhost label, so you end up booking their residence, because you expect better service than the other hosts could provide.
This is, of course, a hypothetical scenario but one worth exploring through data analysis. Do people tend to book with Airbnb superhosts more frequently than with regular hosts, thus leading to higher annual revenue for superhosts 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 et al. 2017). Since its founding in 2008, approximately 500 million people have booked stays with Airbnb (Airbnb 2015), so the

AIRBNB SUPERHOSTS

★ 4.8+
Star Rating

★ 10+
Completed stays in the last year

★ <1%
Cancellation rate

★ 90%
Response rate

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 (Poppick 2015, Management 2022)), and there are ~4 million global hosts with Airbnb listings (Lewis 2020, Management 2022), 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 (Lane and Woodworth 2016, Deboosere, Kerrigan et al. 2019, Kwok and Xie 2019, Xie and Mao 2019, Chattopadhyay and Mitra 2020, Xie, Heo et al. 2021).

Our hypothesis is, that hosts who have the superhost label awarded by Airbnb, can generate more annual revenue than other hosts with similar characteristics but without the superhost label, because customers elect to stay in their properties more frequently.

Airbnb makes some of its data available publicly, which makes it accessible to thorough data science analysis (Airbnb 2022).

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 and Booking.com.

We are interested in one major piece of information:
Is the label of superhost helpful in generating more revenue?
Our response variable is thus the estimated annual revenue for listings clustered by hosts, and our predictors are (1) the Airbnb assigned superhost status and (2) all other listing-based variables with statistical significance.

HYPOTHESIS

Superhosts generate more revenue per listing than regular hosts

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* were chosen (these were the two counties available for these states from Airbnb; (Dogru, Mody et al. 2020)).

We decided to compare the difference between these two locations at the beginning of our research on the relationships between superhost and the estimated 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.

COUNTIES

LA

Inhabitants:
4 million

Superhosts:
9,442

Broward

Inhabitants:
2 million

Superhosts:
3,187

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., URL-addresses of listings, id of the data scrape). 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.

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 et al. 2017).

An initial run with 100 iterations yielded a stark increase in prediction error before the 10th iteration (**Appendix 1**), 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**).

Based on these diagnostics, we decided to use the third iteration of our matching process. This included 10,223 matches after the variables pertaining to number of beds and bathrooms associated with a listing were dropped from the analysis.
This iteration yielded a total number of 1,617 matched groups.

$$Estimated\ Revenue = Price\ of\ Listing * \ Minimum\ Average\ Night\ Value * (\ Reviews\ in\ the\ last\ 12\ months * \frac{100}{67})$$

Matching

To be able to analyse what influence the superhost status has on the estimated annual revenue per listing, we need 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.

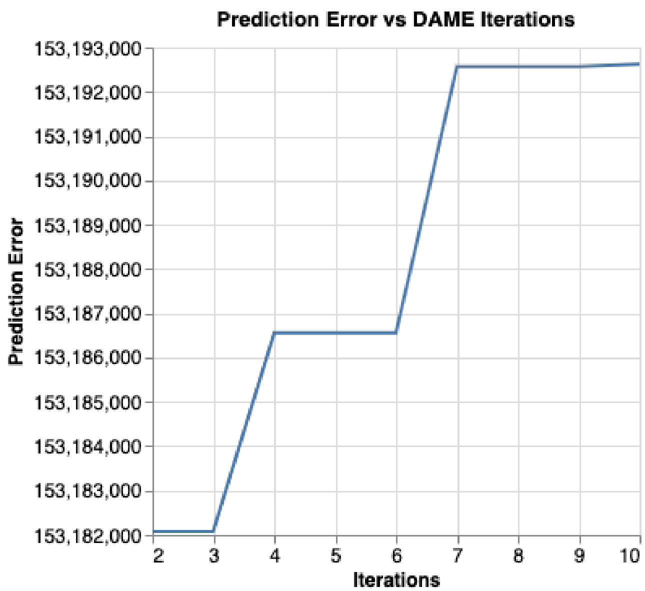


***Figure 1**: Shows the plot of DAME diagnostics. The prediction error increases markedly between the 3rd and 4th iteration.*

Linear Regression

Once the matching process was complete, we were able to run the linear regressions using fixed effects and clustering by host. Our regressions were based on the DAME-FLAME output.

We first ran a regression that included all confounding variables (except the ones used in revenue calculation) that were part of our analysis after data cleaning.

Assessment of this model violated the assumption of normality with a Q-Q-plot that suggested an exponential distribution (**Figure 2**), which led us to run a second regression on log-transformed data.

This second regression satisfied the assumptions for linear regressions (**Appendix 2**), and we proceeded to use it four our statistical analysis.

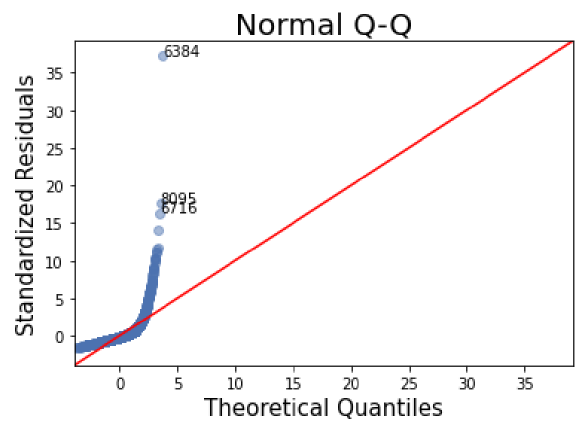


Figure 2: *Shows the initial Q-Q-plot generated to test the assumption of normality in our data. The shape of the plot suggests that a log transformation of the data could improve model performance down the line.*

$$\mathit{annual\ hotel\ revenue} \sim C(\mathit{host\_is\_superhost}) + C(\mathit{state}) + \mathit{host\_response\_time} + \mathit{host\_response\_rate} + \mathit{host\_acceptance\_rate} + C(\mathit{host\_has\_profile\_pic}) + C(\mathit{host\_identity\_verified}) + C(\mathit{room\_type}) + \mathit{accommodates} + \mathit{bathrooms\_text} + \mathit{bedrooms} + \mathit{beds} + C(\mathit{has\_availability}) + \mathit{number\_of\_reviews\_ltm} + \mathit{review\_scores\_rating} + C(\mathit{instant\_bookable}) + \mathit{caculated\_host\_listings\_count} + \mathit{essentials} + C(\mathit{other\_amenities}) + \mathit{host\_experience}$$

IV. SUMMARY STATISTICS

There was a total of 10,179 listings owned by superhosts and 15,591 listings owned by regular hosts in our dataset.

The locational distribution of regular hosts and super hosts was similar for each location (**Figures 3 and Appendix 3**).

Causal 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 actually observe as superhosts and listings by hosts who we actually observe as regular hosts in a world whether neither is a superhost is $2,041 (**Appendix 4**).

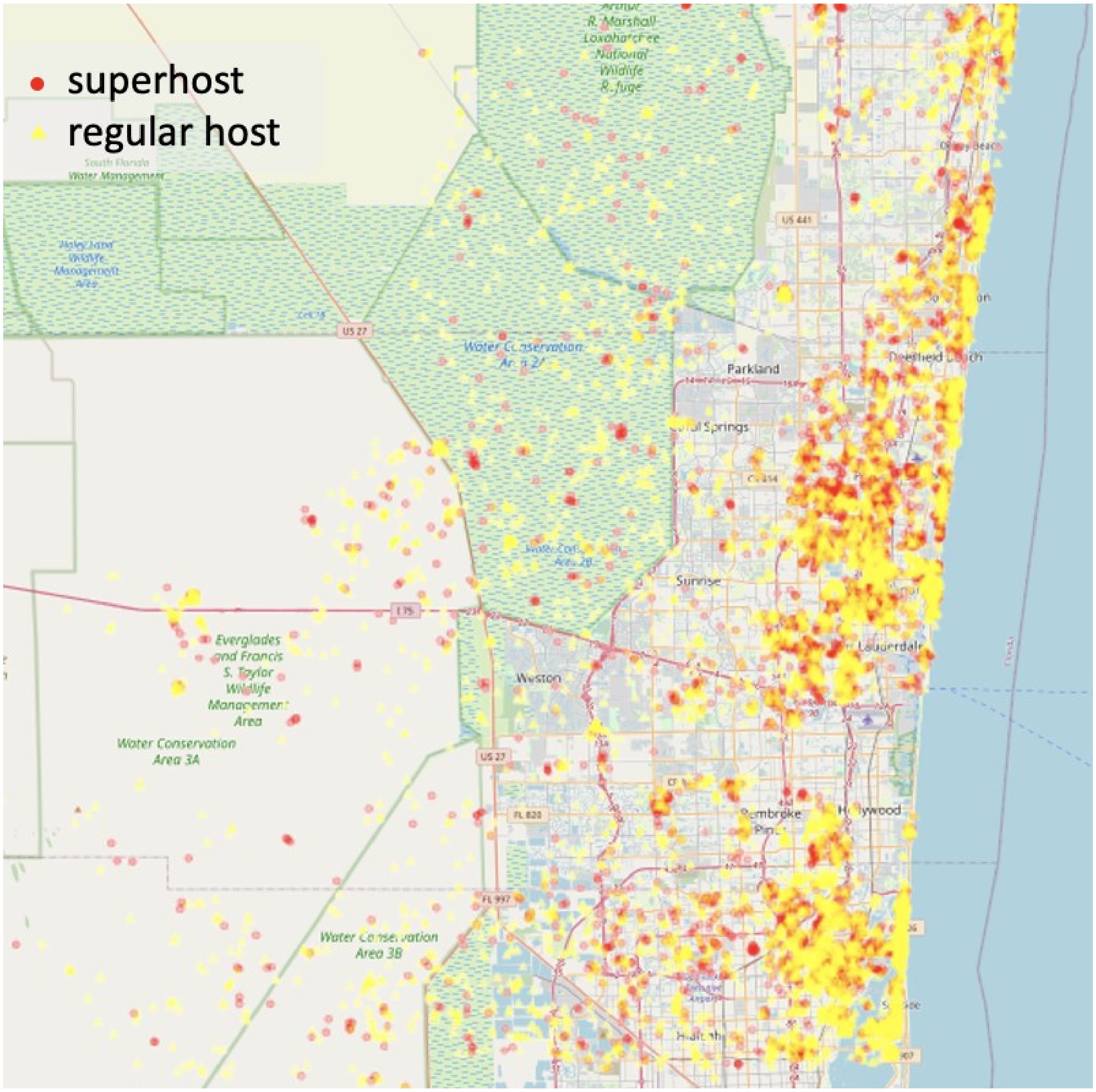


Figure 3: *Shows the distribution of listings belonging to superhosts (red circles) and listings belonging to regular hosts (yellow triangles) in Broward County, Florida.*

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 (**Appendix 5**). We summarized the variables that seem to influence the revenue a listing generates, apart from the superhost label awarded by Airbnb in **Table 1**.

***Table 1**: Shows the variables that showed a statistically significant influence on annual revenue with p-values of lower than 0.001 on log transformed data*

	Coefficient	***SE***
Intercept	11.137	0.204
Host is Superhost	0.930	0.037
Florida	0.394	0.054
Private room	-1.193	0.055
Accommodates	0.662	0.055
No. Bathrooms	0.162	0.036
No. Bedrooms	0.447	0.080
No. Beds	0.162	0.036
Other Amenities	0.344	0.039
Min nights	-1.782	0.041
Check-in score	-0.643	0.118
Communication score	-0.518	0.121
Listings/host [T1]	-0.315	0.062
Listings/host [T2]	-0.457	0.085
Listings/host [T3]	-0.394	0.044
Response time [T1]	-0.369	0.062
Response time [T2]	-0.728	0.100

V. DISCUSSION

We found that being a superhost does correlate with generating more estimated annual revenue per listing (*$2,041)* than being a non-superhost to the level of statistical significance.

The higher annual revenue generated by superhosts is in accordance with Airbnb's communication of the overall average amount of money that is made over a year by all hosts ($9,600), and the amount they say an "experienced host" can expect to make in one year ($10,000), though our estimate is lower than this $4,000 difference.

Limitations
We did not use the whole dataset that Airbnb makes available on listings worldwide but focused instead on the data from the two states generating the most revenue. The analysis could be extended to all US states and worldwide locations that data is available for.
In addition, our analysis is limited by the quality of the data that Airbnb provides and by the fact that our data is cross-sectional, thus relying on the quality of our matching process.
We could also consider a longitudinal analysis of hosts before and after they achieve superhost status.

In conclusion, we have found that the listings under the superhost label generate 2.53 times more revenue than ones under regular hosts and can thus be considered a relevant factor for revenue generation that hosts should aspire to.

VI. REFERENCES

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.

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

Chattopadhyay, M. and S. K. Mitra (2020). "What Airbnb Host Listings Influence Peer-to-Peer Tourist Accommodation Price?" Journal of Hospitality & Tourism Research **44**(4): 597-623.

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 Science **6**(1): 143-156.

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 Management **78**: 104049.

Kwok, L. and K. L. Xie (2019). "Pricing strategies on Airbnb: Are multi-unit hosts revenue pros?" International Journal of Hospitality Management **82**: 252-259.

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.

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.

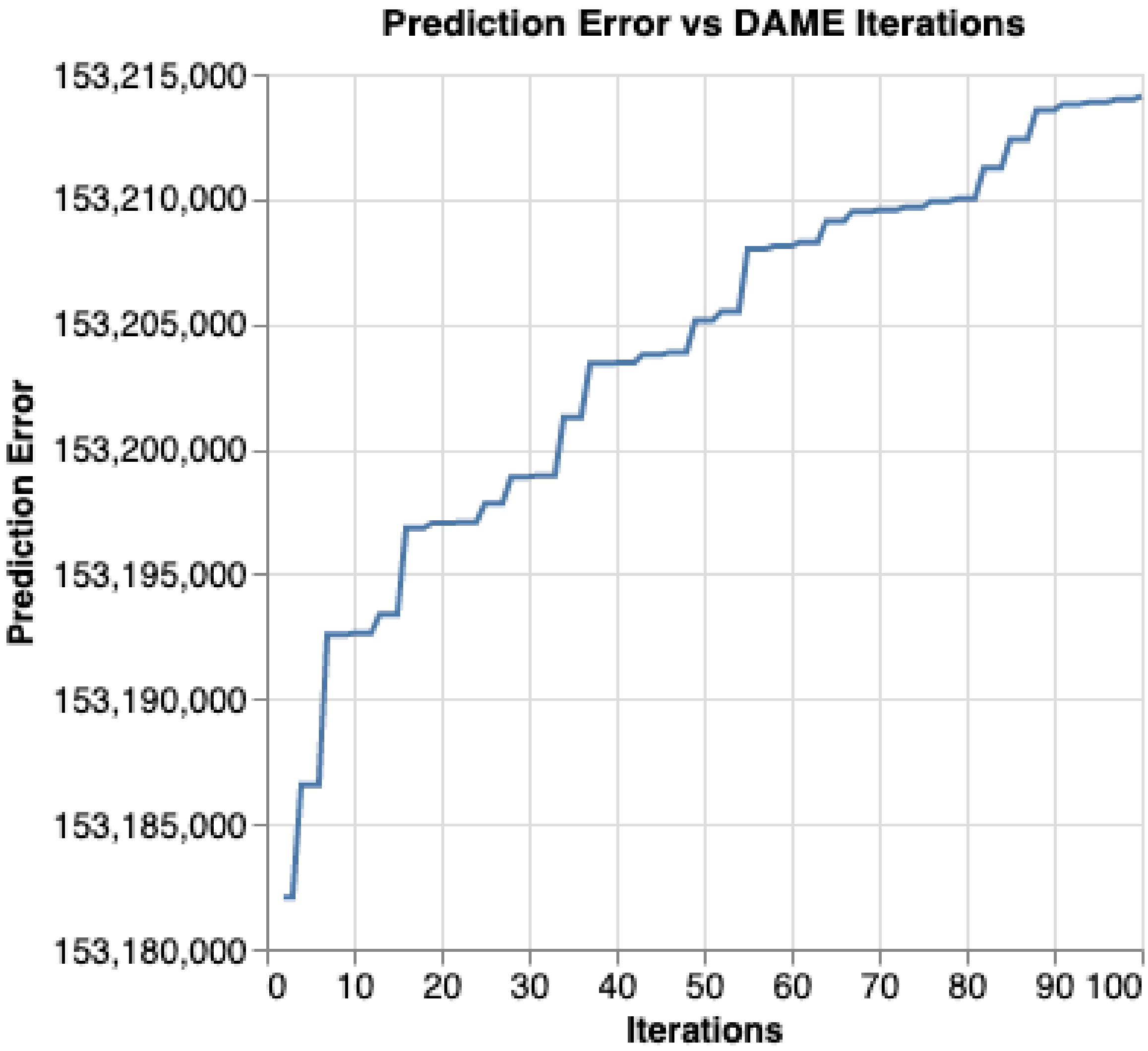
Poppick, S. (2015). "Airbnb Says Renting Your Place Is Like Getting a Big Raise." Retrieved 04/03/2022, from https://money.com/airbnb-raise-income-report/.

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 Management **47**: 413-421.

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 Research **43**(6): 919-929.

Zervas, G., D. Proserpio and J. W. Byers (2017). "The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry." Journal of Marketing Research **54**(5): 687-705.

APPENDIX



Appendix 1: *Initial DAME diagnostics showing the first increase in prediction error between 0 and 10 iterations.*

Our first regression after matching was a linear regression with annual revenue as our dependent variable and the categorical variable of whether or not a host is considered a superhost as our predictor:

$$annual\ hotel\ revenue\ \sim C(host\_is\_superhost)$$

With the following output:

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	4.407e+04	4259.398	10.346	0.000	3.57e+04	5.24e+04
C(host_is_superhost)[T.1]	8493.5169	5822.852	1.459	0.145	-2922.021	1.99e+04

The second regression was a linear regression that used annual revenue as the dependent variable and both the superhost categorical variable and the regional variable of state (California vs Florida):

$$annual\ hotel\ revenue\ \sim C(host\_is\_superhost) + C(state)$$

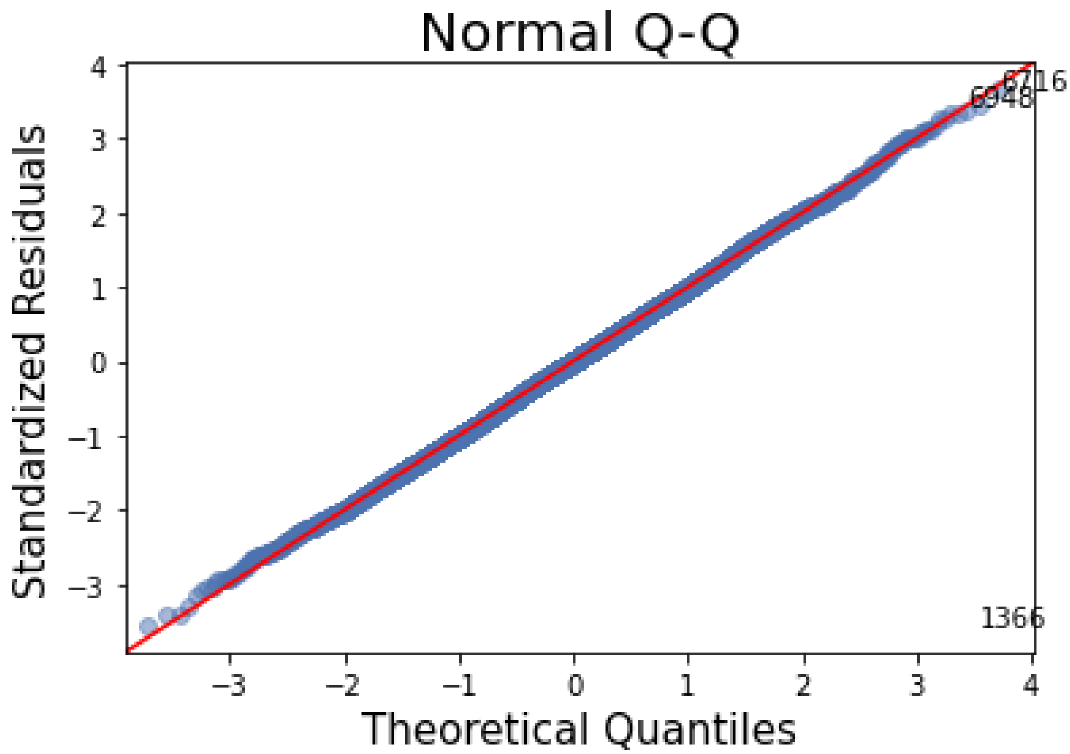
With the following output:

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	3.686e+04	4548.860	8.102	0.000	2.79e+04	4.58e+04
C(host_is_superhost)[T.1]	8493.5169	5811.151	1.462	0.144	-2899.081	1.99e+04
C(state)[T.1]	3.032e+04	6806.971	4.454	0.000	1.7e+04	4.37e+04

The results of our initial regression using data that was not log transformed yielded the following output:

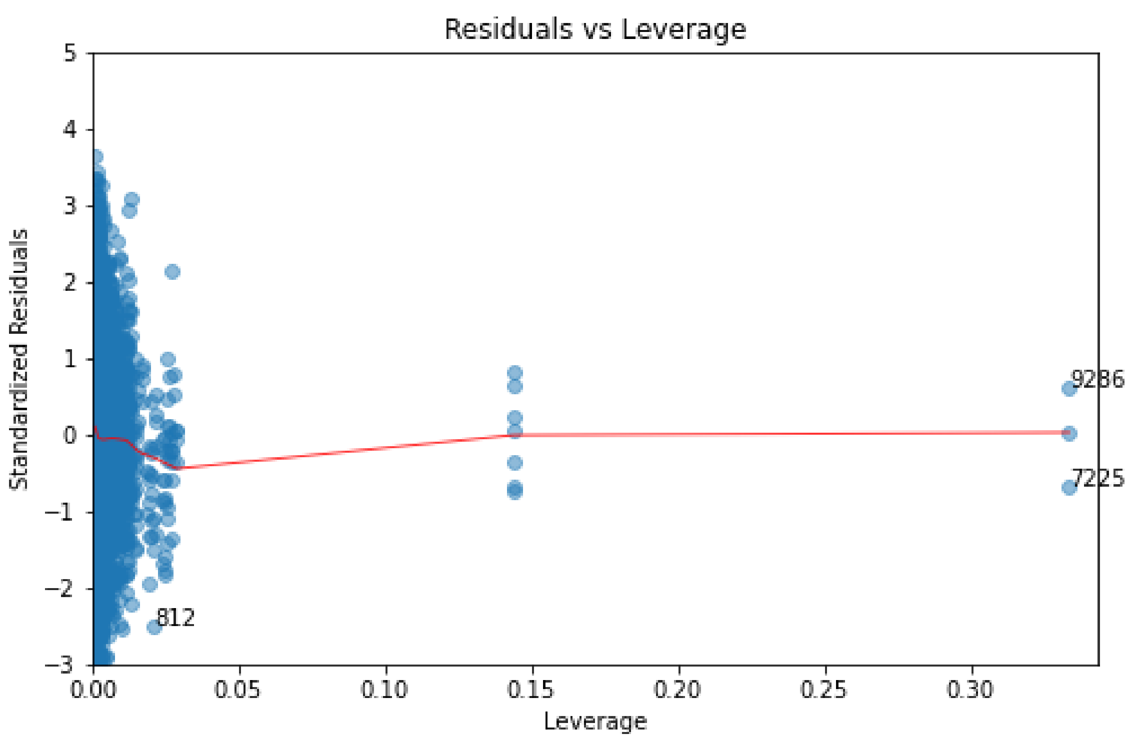
```
                            OLS Regression Results
==============================================================================
Dep. Variable:     annual_listing_revenue   R-squared:                   0.061
Model:                                OLS   Adj. R-squared:              0.059
Method:                     Least Squares   F-statistic:                 24.59
Date:                    Sun, 24 Apr 2022   Prob (F-statistic):      2.00e-118
Time:                            22:26:03   Log-Likelihood:        -1.0858e+05
No. Observations:                   10223   AIC:                     2.172e+05
Df Residuals:                       10195   BIC:                     2.174e+05
Df Model:                              27
Covariance Type:                nonrobust
==============================================================================================================
                                              coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------------------
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
==============================================================================
Omnibus:                    15981.562   Durbin-Watson:                   1.827
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         22883714.193
Skew:                           9.574   Prob(JB):                         0.00
Kurtosis:                     233.990   Cond. No.                     2.25e+15
==============================================================================
```

We ran diagnostics on this data and realized that the assumption of normality for linear modelling were violated, thus we performed a log transformation and re-ran our diagnostics.



Appendix 2*: Q-Q-plot to evaluate the assumption of normality for log-transformed data.*

There were no relevant leverage points in our data:



After the log transformation was performed, we received the following results on our regression:

OLS Regression Results

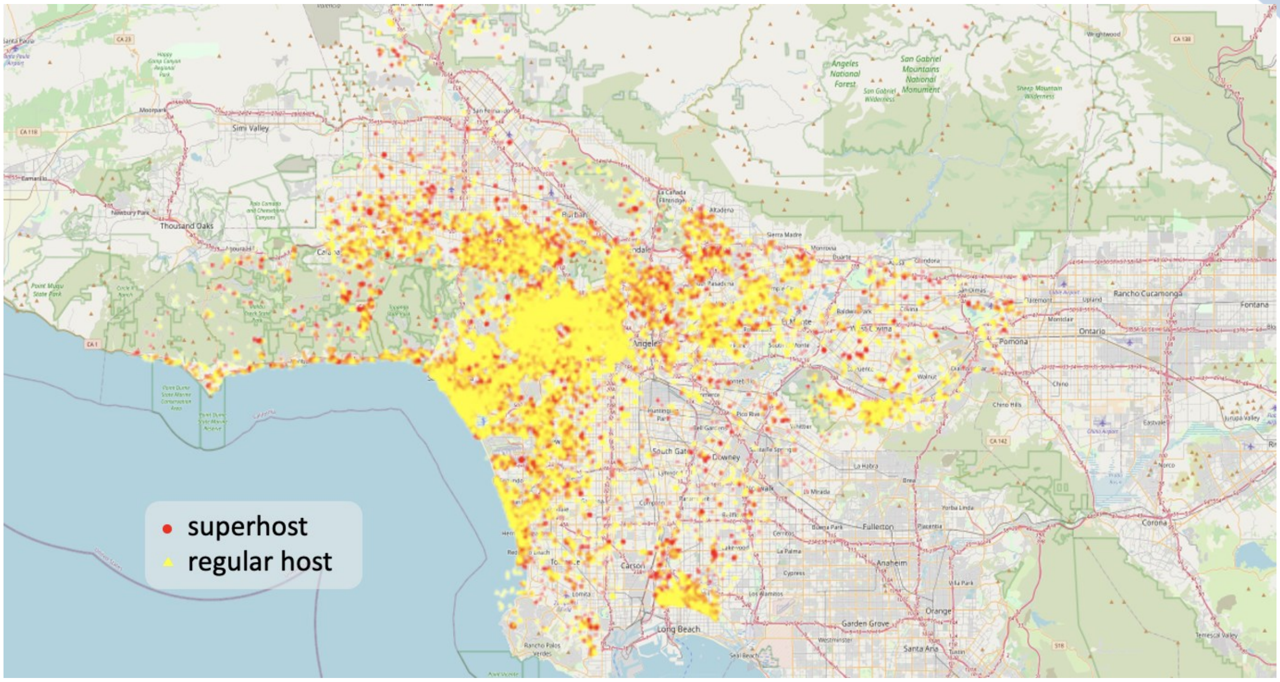
Dep. Variable:	log_revenue	R-squared:	0.432
Model:	OLS	Adj. R-squared:	0.430
Method:	Least Squares	F-statistic:	286.7
Date:	Sun, 24 Apr 2022	Prob (F-statistic):	0.00
Time:	22:26:15	Log-Likelihood:	-20355.
No. Observations:	10223	AIC:	4.077e+04
Df Residuals:	10195	BIC:	4.097e+04
Df Model:	27		
Covariance Type:	nonrobust		

	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

Omnibus:	3.964	Durbin-Watson:	1.706
Prob(Omnibus):	0.138	Jarque-Bera (JB):	3.883
Skew:	0.030	Prob(JB):	0.144
Kurtosis:	2.925	Cond. No.	2.25e+15

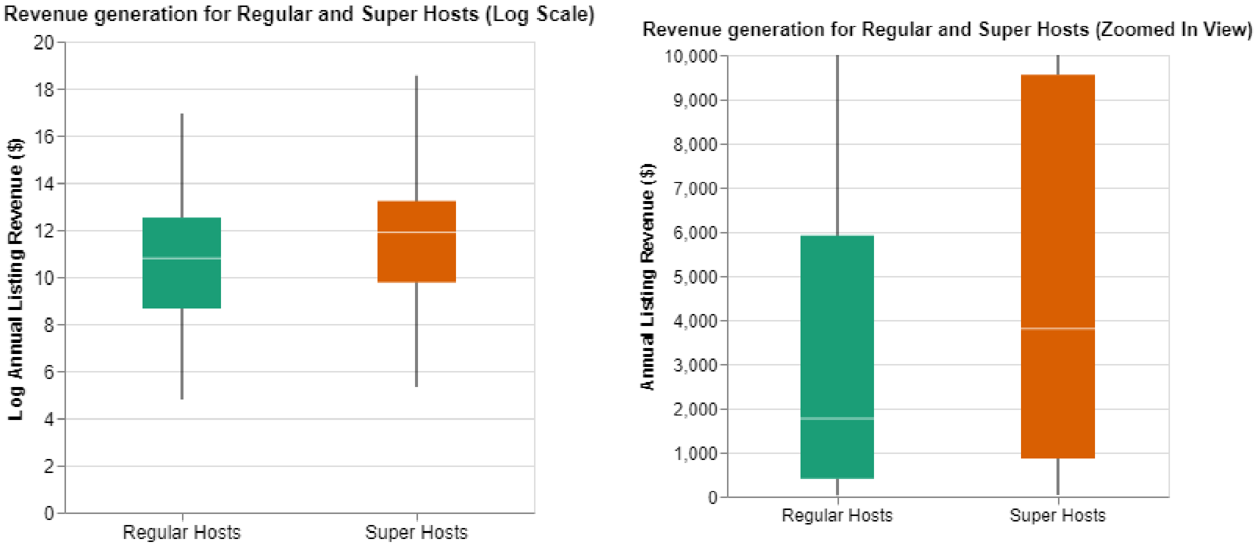
The final regression we ran was added to cluster our results by host:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            log_revenue   R-squared:                       0.432
Model:                            OLS   Adj. R-squared:                  0.430
Method:                 Least Squares   F-statistic:                 3.192e+14
Date:                Sun, 24 Apr 2022   Prob (F-statistic):           3.56e-08
Time:                        22:26:21   Log-Likelihood:                -20355.
No. Observations:               10223   AIC:                         4.077e+04
Df Residuals:                   10195   BIC:                         4.097e+04
Df Model:                          27
Covariance Type:              cluster
===========================================================================================================
                                          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
==============================================================================
Omnibus:                        3.964   Durbin-Watson:                   1.706
Prob(Omnibus):                  0.138   Jarque-Bera (JB):                3.883
Skew:                           0.030   Prob(JB):                        0.144
Kurtosis:                       2.925   Cond. No.                     2.25e+15
==============================================================================
```

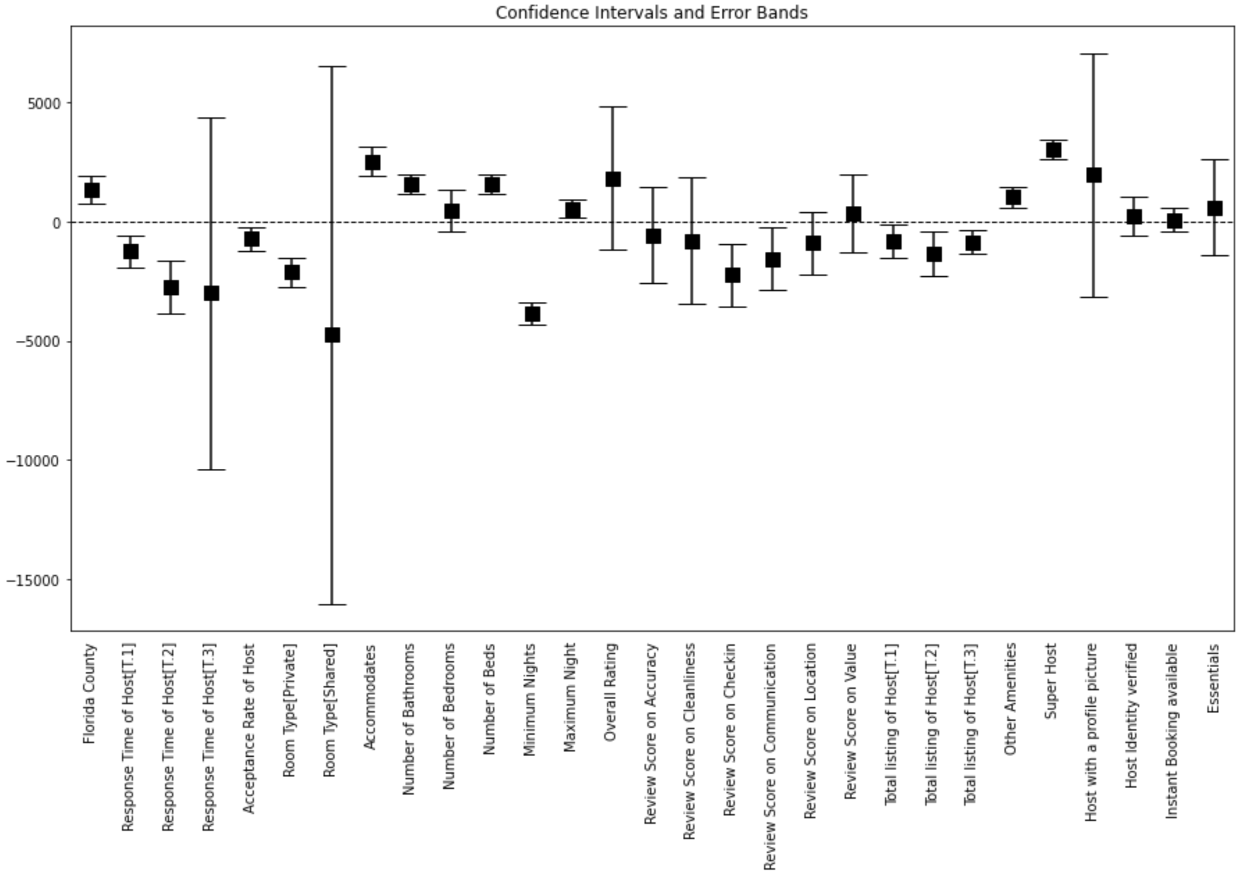


Appendix 3*: Shows the distribution of listings belonging to superhosts (red circles) and listings belonging to regular hosts (blue triangles) in LA County, California.*

Our fixed effects linear regression analysis clustered by host in the following box plots for the analysis of the average treatment effect (ATE):



***Appendix 4**: Log scale ATE on left and zoomed in ATE transformed back on right.*



Appendix 5: *Error Bars with 95% confidence interval as whiskers. If the whiskers cross the 0-line, the variable is not likely to be of statistical significance.*