

Transforming Airbnb Hosting

Maximizing Revenue and Host Excellence through Predictive Analytics



Aishwarya Ajaykumar



Samarth Bansal



Siddhant Treasure



Yi-Chun Huang



Executive summary

Objective:

Our objective was to understand how Superhost status influences a host's performance and revenue on Airbnb. We delved into the nuances of what it means to be a Superhost and how it differentiates from being a regular host.

Key Questions Addressed:

Value of Superhost Status: We explore the significance of attaining Superhost status on Airbnb, assessing its benefits in terms of guest satisfaction, increased bookings, and enhanced visibility.

Financial Implications: The analysis quantifies the financial benefits associated with becoming a Superhost, including potential increases in revenue and market competitiveness.

Findings:

Superhost Advantages: Our findings are quite revealing. Superhost status is not just a title; it's a significant driver of success. Superhosts consistently outperform regular hosts in terms of revenue, guest satisfaction, and booking rates.

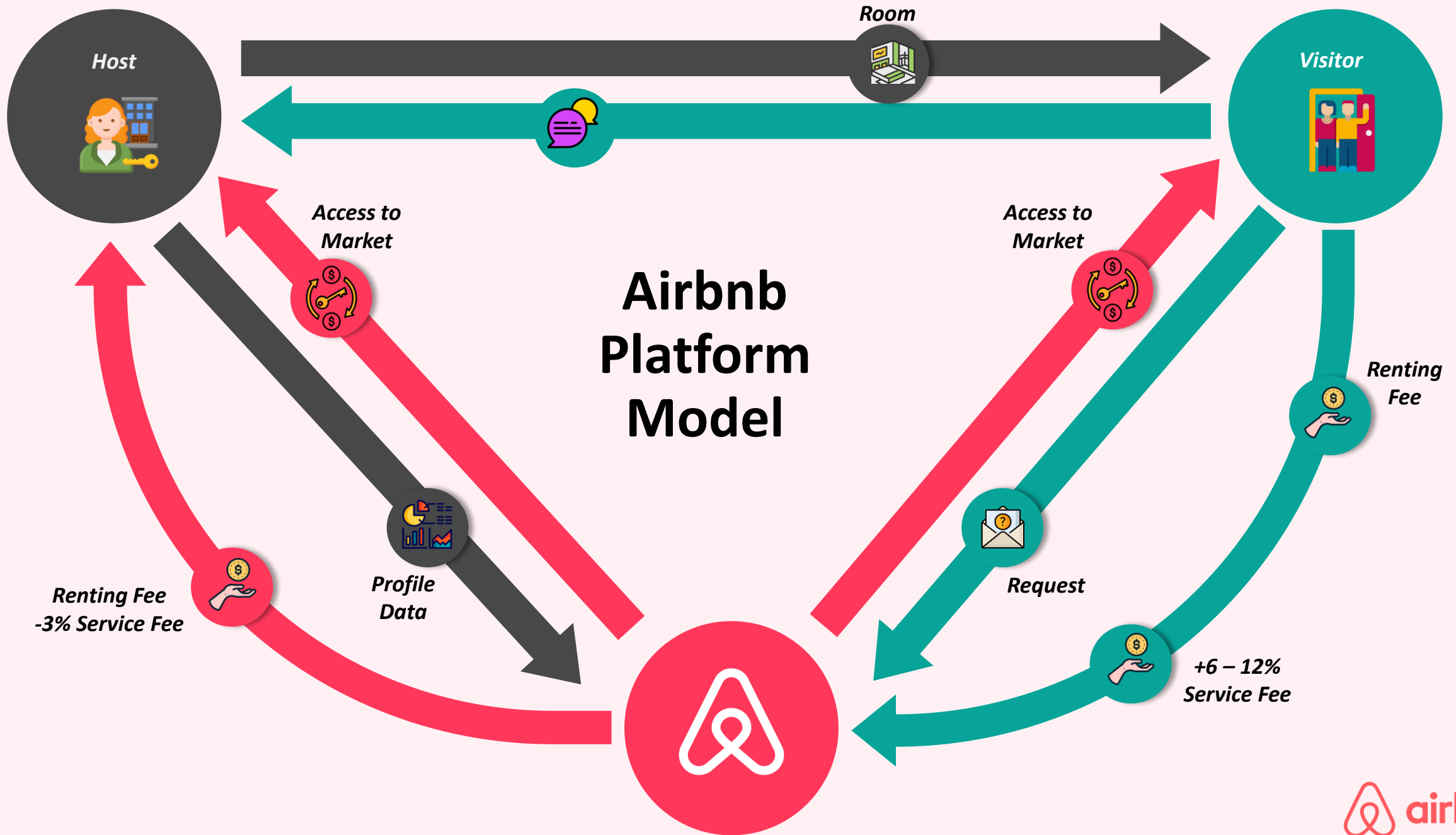
Revenue Impact: Quantitatively, Superhosts see a notable increase in earnings. This lift in revenue varies based on several factors, including location, property type, and the strategies employed by the host.

Strategic Recommendations: To help hosts achieve and benefit from Superhost status, we've developed a set of strategic recommendations. These include focusing on enhancing guest experiences, adopting dynamic pricing strategies, and conducting thorough market analysis.

Conclusion:

In conclusion, our study highlights the Superhost status as a key factor in Airbnb hosting success. It's not just about earning a badge; it's about adopting strategies that lead to better guest experiences, higher earnings, and a more sustainable hosting model.





Problem Statement

A thorough analysis utilizing predictive models to project revenue for Airbnb superhost and host listings, offering actionable insights that highlight the benefits of achieving superhost status and delineating the pathway to becoming a superhost in the dynamic Airbnb marketplace

Growing Airbnb Market

Understanding Airbnb's revenue and superhost dynamics is key to host competitiveness and profit

Host Differentiation

Differentiating superhost experiences, guiding hosts on where to enhance and invest

Adapting to Consumer Preferences

Aligning host offerings with changing trends, boosting guest satisfaction and revenue

Technology Integration in Hospitality

Highlights the need for tech and data analytics in improving personal and small hospitality businesses

A map of Chicago, Illinois, with numerous red location pins scattered across the city, indicating the locations of Airbnb listings. The pins are concentrated in the downtown area and surrounding suburbs.

13689

Unique Airbnb Host IDs

725

Unique Census Tracts

25845

Unique Airbnb Property IDs

74

Unique Property Type

4

Unique Listing Types

56

Unique Zipcodes

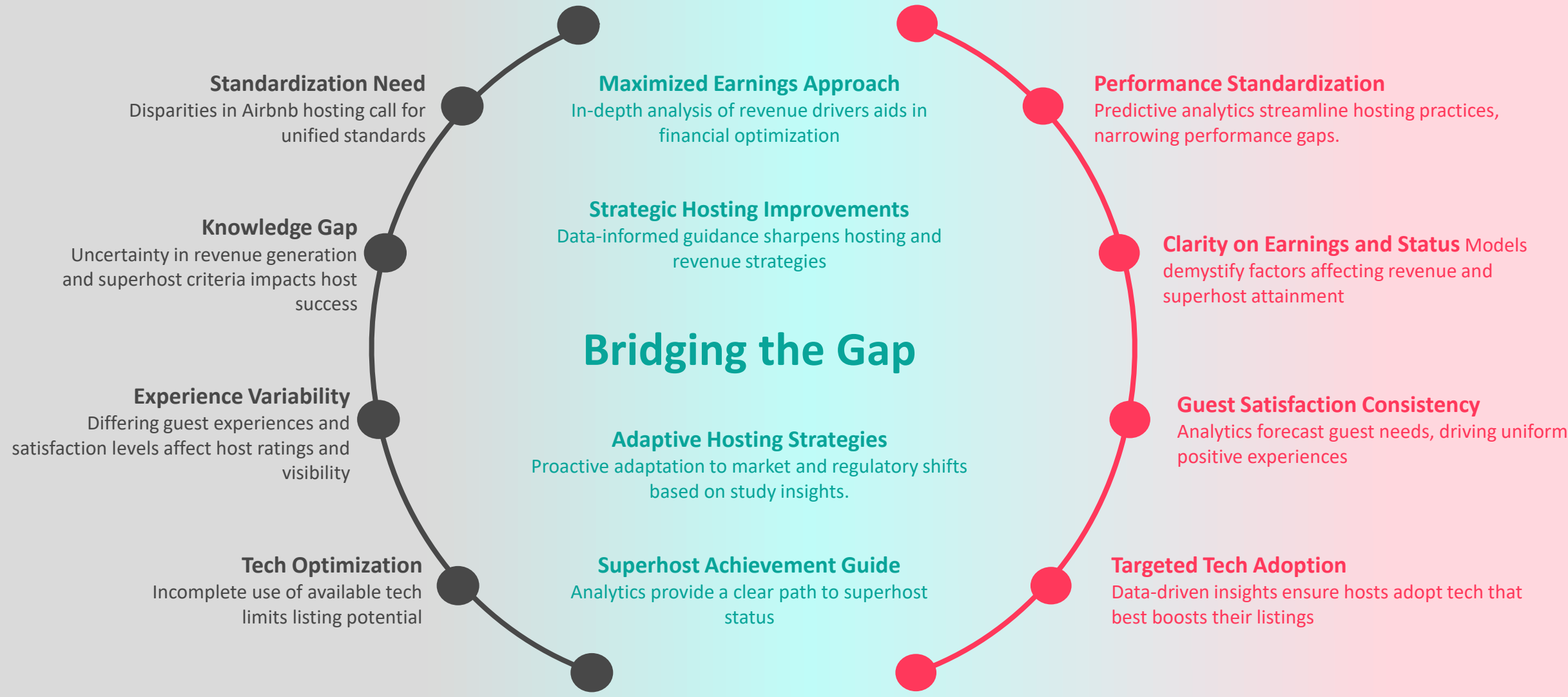
518639119

Total Pop. In Census Tracts

271025271

Total Rev. of Census tracts

Current State vs Future State



Decoding the Dataset: Core Elements and Statistics

Thematic breakdown

Host and Property Profiles

Identifiers, property types, and listing specifics

Financial and Booking Analysis

Revenue, nightly rates, booked days, and future reservations

Performance and Quality Indicators

Superhost status, ratings, reviews, and guest satisfaction metrics

Geographic and Demographic Insights

Location specifics and area demographics

Market Trends and Seasonality

Seasonal impacts on bookings, market dynamics, and competitive data

Operational Details

Listing features, operational aspects, and amenities

Seasonal Revenue Shifts

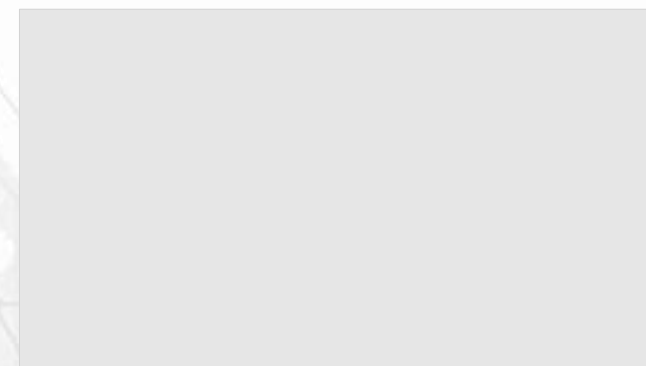
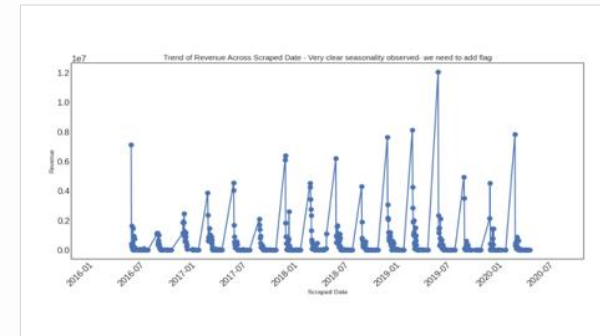
There are key revenue bursts in summer and Christmas

Geographical Insights :

Chicago Downtown zip codes generates more revenue than other areas

Top Property Revenue Insights:

Apartments, Condominiums, and Houses emerge as top earners



Data Cleaning

Missing Values

Address missing data by imputing with median values for numerical variables and modes for categorical variables

Special Characters Checks

Screen the dataset for special characters that may indicate data entry errors or encoding issues, ensuring textual data integrity

Duplicates Removal

Check performed to ensure elimination of duplicate entries to prevent data redundancy

Outliers Detection

Identify statistical anomalies through outlier detection methods

Standardization

Normalize data formats and conventions across the dataset to ensure consistency



Data Preprocessing

Encoding Categorical Variables

Leverage LightGBM's advanced handling of categorical variables.

Data Splitting

Only Census Tracts with 1 Million Revenue:

To focus the analysis on high-value areas

Dimensionality Reduction

Prune the dataset by removing features that do not contribute to the model's predictive power

Time Series Considerations

Panel Data with Seasonality: Incorporate dummy variables to represent seasonal effects and include lagged columns to capture time-dependent patterns within the panel data structure.

Feature Engineering

Temporal Features: Derive temporal attributes from date-time information, capturing essential dynamics like seasonality and trends.

Aggregating Features: Consolidate data points over the `superhost_period_all` to create summarized metrics that reflect longer-term patterns and host performance.

The Regression model

Revenue Prediction for Superhost Listings

OLS Regression Results						
=====						
Dep. Variable:	revenue	R-squared:	0.714			
Model:	OLS	Adj. R-squared:	0.714			
Method:	Least Squares	F-statistic:	4890.			
Date:	Fri, 08 Dec 2023	Prob (F-statistic):	0.00			
Time:	19:45:42	Log-Likelihood:	-7.6263e+05			
No. Observations:	84151	AIC:	1.525e+06			
Df Residuals:	84107	BIC:	1.526e+06			
Df Model:	43					
Covariance Type:	nonrobust					
=====						
	coef	std. err	t	P> t	[0.025	0.975]

numCancel_pastYear	-18.9215	7.408	-2.554	0.011	-33.441	-4.402
prop_5_StarReviews_pastYear	-6.9856	7.540	-0.926	0.354	-21.764	7.793
Bedrooms	153.2522	12.618	12.145	0.000	128.521	177.983
Bathrooms	14.6158	8.951	1.633	0.102	-2.928	32.159
Max Guests	69.8653	12.538	5.572	0.000	45.290	94.440
Cleaning Fee (USD)	122.8330	9.938	12.360	0.000	103.354	142.312
Minimum Stay	-14.9348	7.695	-1.941	0.052	-30.017	0.147
Number of Photos	41.0041	8.164	5.022	0.000	25.002	57.006
Nightly Rate	-19.1674	12.580	-1.524	0.128	-43.823	5.488
Number of Reviews	-42.6129	7.761	-5.491	0.000	-57.825	-27.401
booked_days	2124.5019	10.274	206.775	0.000	2104.364	2144.640
tract_total_pop	-14.1193	9.365	-1.508	0.132	-32.474	4.235
tract_white_perc	43.5018	10.711	4.061	0.000	22.508	64.495
tract_black_perc	-2.1517	10.405	-0.207	0.836	-22.545	18.242
tract_asian_perc	-18.0451	7.979	-2.262	0.024	-33.684	-2.407
tract_superhosts	-146.7384	11.706	-12.535	0.000	-169.682	-123.795
revenue_period_city	-10.4891	7.512	-1.396	0.163	-25.212	4.233
revenue_period_tract	260.9832	12.239	21.324	0.000	236.995	284.971
booked_days_avePrice	2180.0384	11.778	185.102	0.000	2156.955	2203.122
available_days_aveListedPrice	70.7922	16.214	4.366	0.000	39.013	102.571
prev_available_days_aveListedPrice	-140.2719	18.966	-7.396	0.000	-177.445	-103.099
occupancy_rate	-2.1347	10.677	-0.200	0.842	-23.061	18.791
prev_booked_days_avePrice	-401.3552	13.661	-29.379	0.000	-428.131	-374.580
prev_available_days	-55.8364	10.793	-5.173	0.000	-76.991	-34.682
prev_revenue	937.1233	13.347	70.214	0.000	910.964	963.283
prev_booked_days	-469.7892	10.841	-43.335	0.000	-491.037	-448.541
available_days	-22.8885	12.033	-1.902	0.057	-46.472	0.695

Predicts the revenue for listings managed by Airbnb superhosts to understand the revenue potential of properties under the management of superhosts

Regression Model

Positive Drivers

Booked_days

Booked_days_avePrice

Cleaning Fee(USD)

Bedrooms

Prev_revenue

Negative Drivers

Prev_booked_days

Minimum_Stay

prev_available_days_aveListed
Price

Number of Reviews

prop_5_StarReviews_pastYear

Host Behaviour Shifts

Let’s look at the history of Hosts vs Superhosts dynamics



4.8+ overall rating

Superhosts have a 4.8 or higher average overall rating based on reviews from their Airbnb guests in the past year. Guests know they can expect outstanding hospitality from these hosts.



10+ stays

Superhosts have completed at least 10 stays in the past year or 100 nights over at least 3 completed stays. Your guests can feel confident staying with an experienced host.



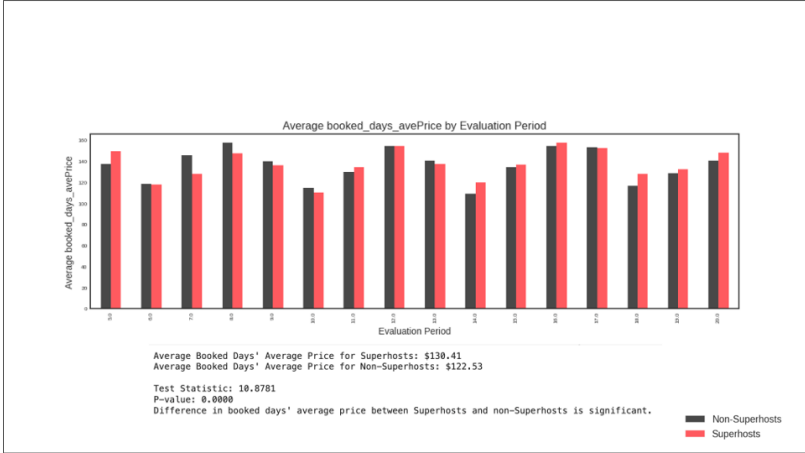
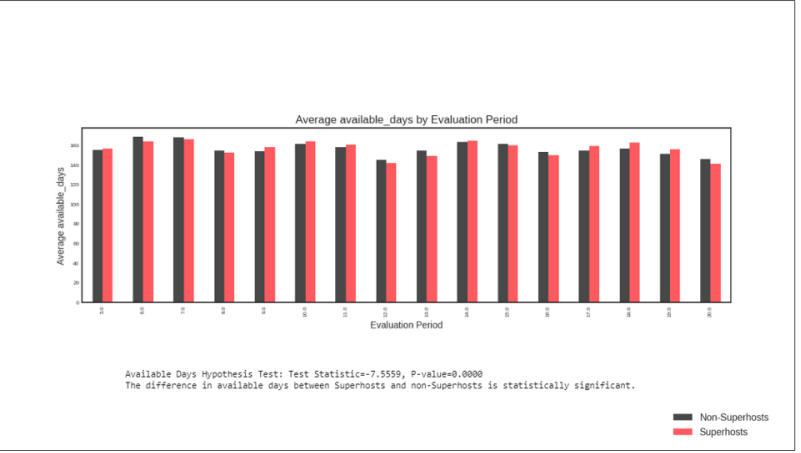
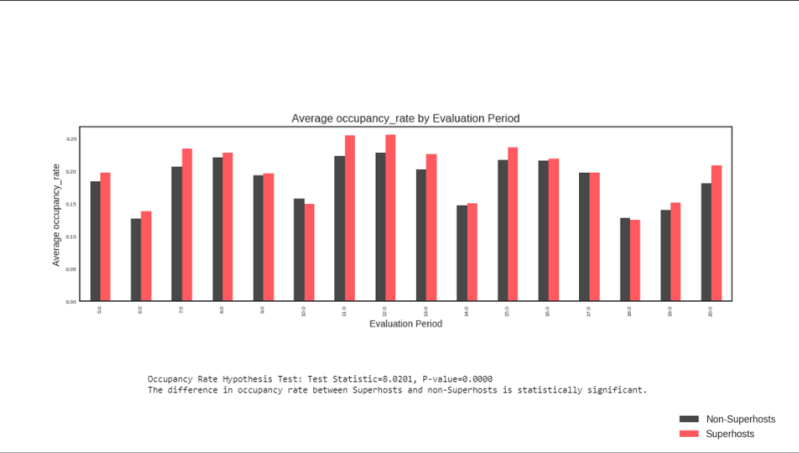
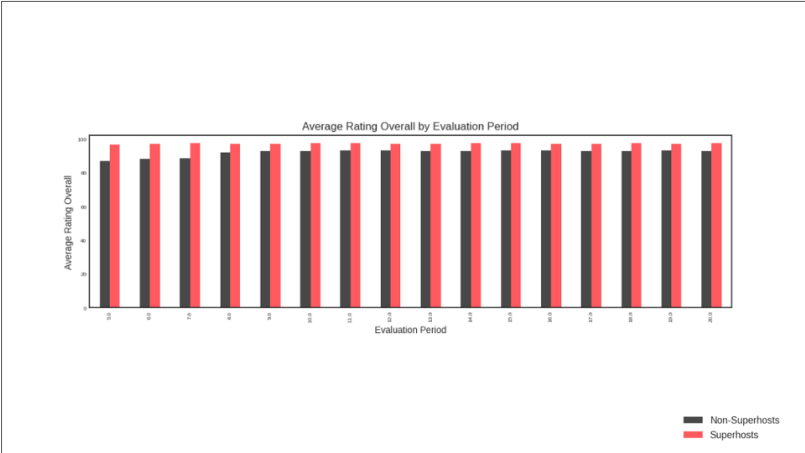
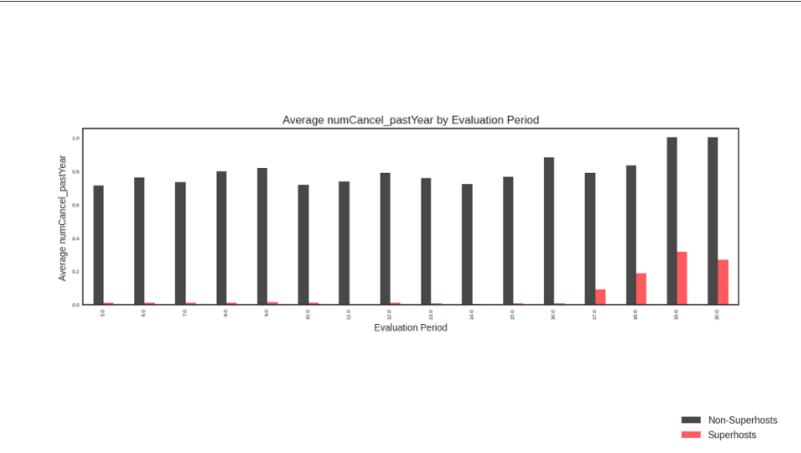
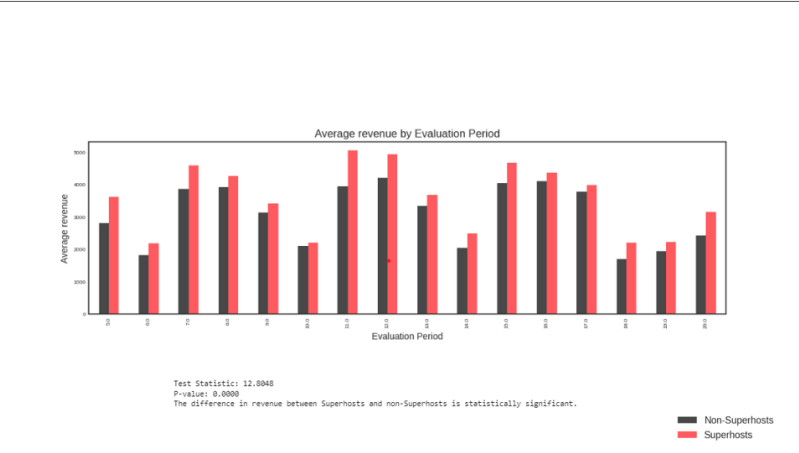
<1% cancellation rate

Superhosts cancel less than 1% of the time, not including extenuating circumstances. This means 0 cancellations for hosts with fewer than 100 reservations in a year. Rare cancellations mean peace of mind for guests.



90% response rate

Superhosts respond to 90% of new messages within 24 hours. When guests ask you questions, they know that a quick response is only a message away.

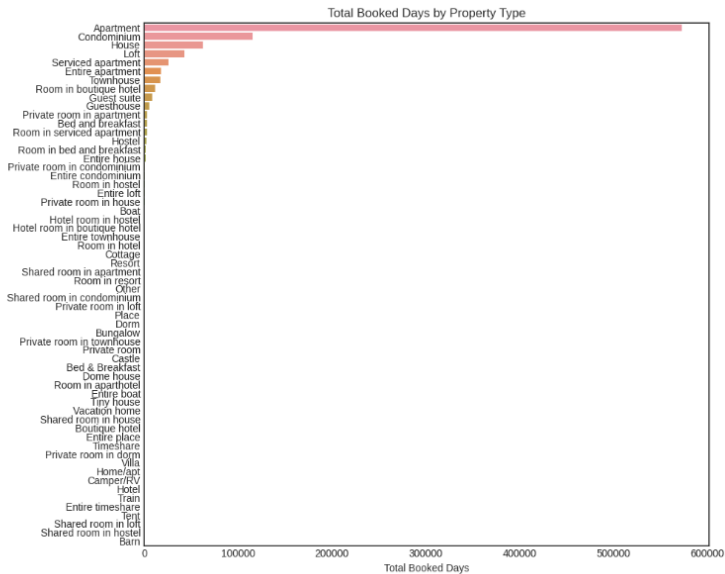
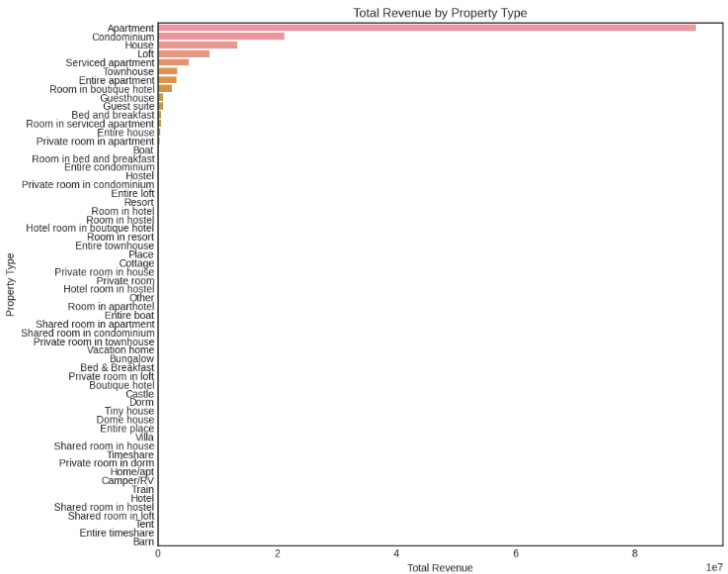


Modeling Metrics: Model Design and Strengths

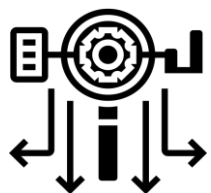
Revenue Prediction for Superhost and generic host Listings

MAPE on the Validation Set for Superhost Model: 0.033388698405795836
MSE for Superhost Model: 224325.26450103635
R-squared for Superhost Model: 0.9895288358892635
Feature Importance for Superhost Model:

	feature	importance
5	booked_days	901
8	booked_days_avePrice	767
13	Cleaning Fee (USD)	148
9	available_days_aveListedPrice	144
14	occupancy_rate	103
2	Airbnb Host ID	93
16	Nightly Rate	92
10	prev_available_days_aveListedPrice	87
18	Bedrooms	81
7	prev_booked_days_avePrice	66
21	prev_revenue	60
22	Number of Reviews	56
15	prev_occupancy_rate	55
0	index	47
17	prev_Nightly Rate	42
12	prev_available_days	41
19	Bathrooms	39
11	available_days	37
20	Number of Photos	36
3	Airbnb Property ID	33
1	superhost_period_all	23
23	prev_Number of Reviews	23
6	prev_booked_days	18
4	Superhost	8



Predicts the revenue for listings managed by Airbnb hosts to understand the revenue potential of properties under the management of superhosts



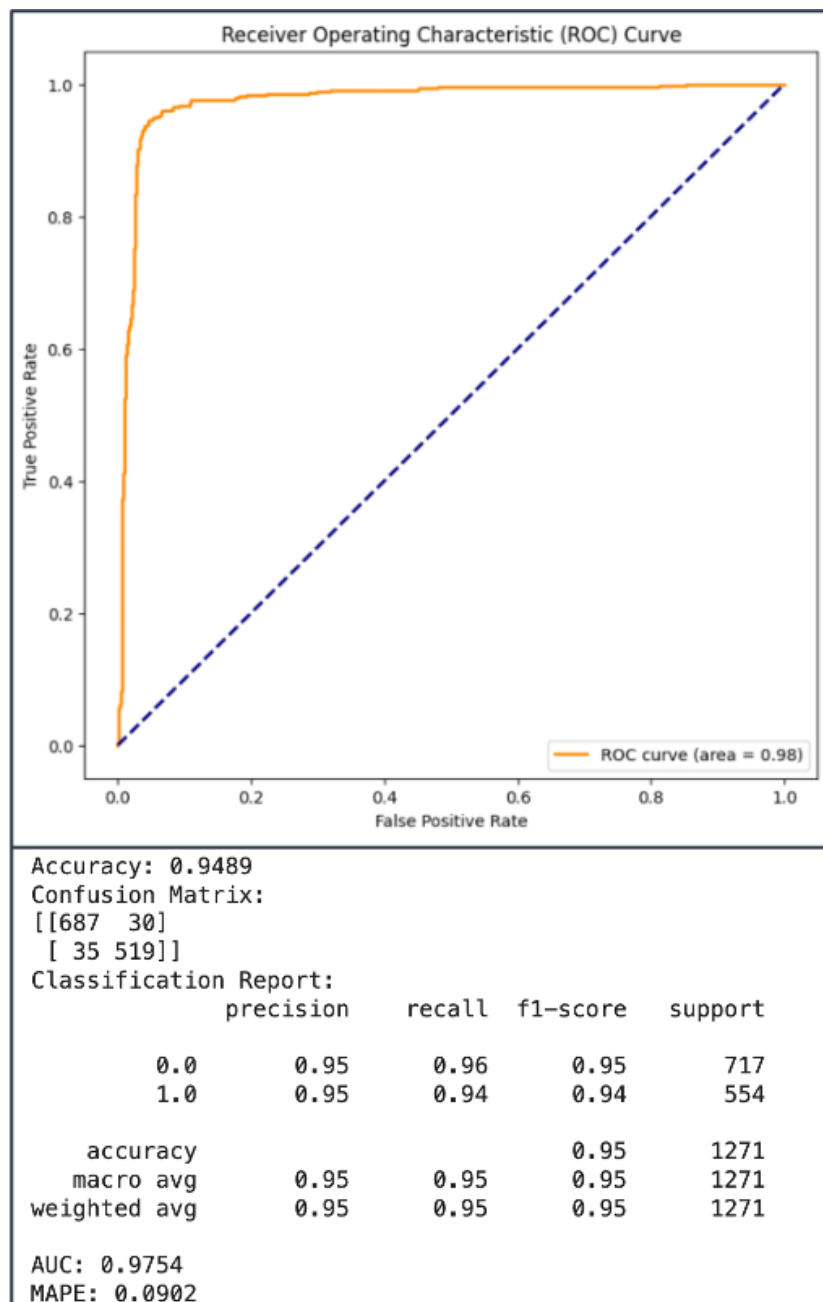
Superhost Status Prediction

A statistical model crafted to calculate the chances of an Airbnb host gaining or keeping superhost status, utilizing analysis of past performance indicators pertinent to Airbnb's superhost criteria.

Gradient boost Model 3

Top 20 Most Important Features:

	Feature	Importance
5	prev_rating_ave_pastYear	0.328803
0	rating_ave_pastYear	0.240714
3	num_5_star_Rev_pastYear	0.117364
6	prev_numReviews_pastYear	0.112582
7	prev_numCancel_pastYear	0.074841
1	numReviews_pastYear	0.036434
2	numCancel_pastYear	0.020898
25	Number of Reviews	0.010501
26	prev_Number of Reviews	0.009876
4	prop_5_StarReviews_pastYear	0.004606
48	tract_superhosts	0.004126
13	available_days	0.003688
30	occupancy_rate	0.003058
11	prev_numReservedDays_pastYear	0.002942
9	numReservedDays_pastYear	0.002796
37	tract_housing_units	0.002777
32	prev_occupancy_rate	0.002595
27	Rating Overall	0.001941
8	prev_num_5_star_Rev_pastYear	0.001721
50	tract_price_variance	0.001550



The superhost status prediction model will provide strategic value by identifying the key performance indicators that most significantly influence a host's probability of achieving superhost status. This model will serve as a tool for hosts to gauge their current performance against superhost standards and to pinpoint areas requiring improvement.

The Financial Impact of Achieving Superhost Status

Revenue Analysis Overview

Focused on 72 census tracts with over \$1 million in revenue

Earning Potential for Hosts

20% of hosts in these tracts have the potential to increase earnings by an average of \$2,074.87 as superhosts

Underperformers' Revenue Increase

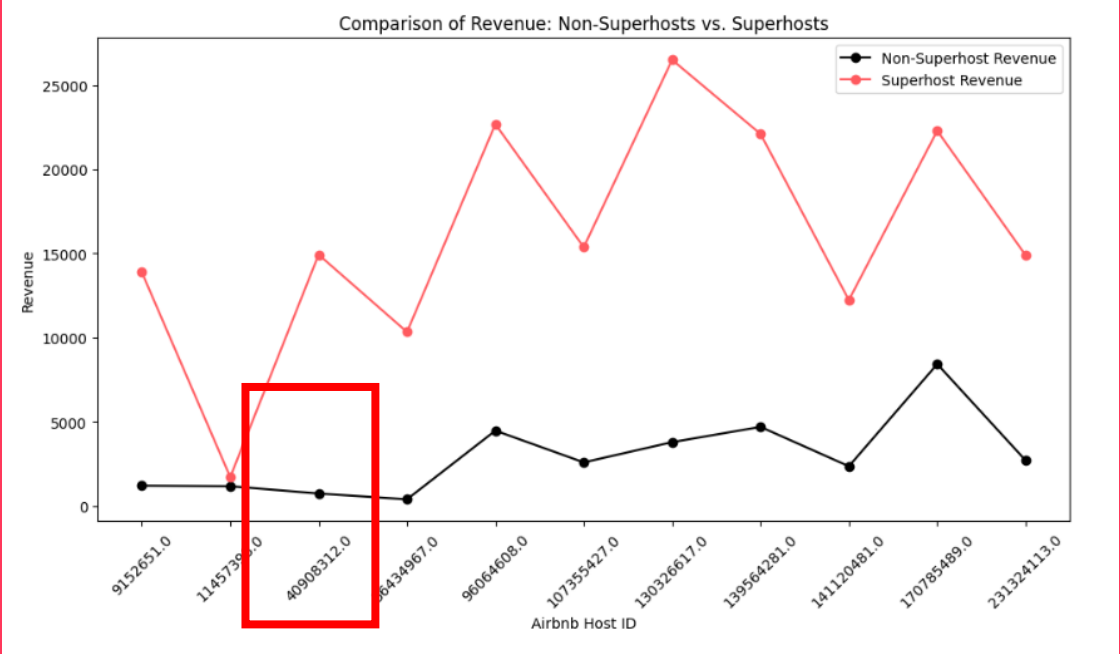
Superhosts earn at least \$500 (based on random sample) more than hosts at the lowest differential, showcasing consistent financial benefits

Chance of Advancement

Hosts with the lowest revenue lift have a 50% baseline chance of becoming superhosts, which improves with targeted enhancements

Strategic Financial Benefits

Visual data supports the financial upside of superhost status, encouraging hosts to make necessary improvements to advance



Probability of Airbnb Host ID 11457398 becoming a Superhost: 0.5078

Airbnb Host ID	Superhost	Predicted Probability
2153.0	0.0	0.013997
2613.0	1.0	0.000000
5775.0	1.0	0.000000
17928.0	1.0	0.000000
33004.0	0.0	0.000000
...
141528449.0	0.0	0.000000
44131859.0	1.0	0.000000
72165035.0	0.0	0.000000
21517759.0	0.0	0.000000
310181286.0	0.0	0.011693

Top 20 Most Important Features:		
	Feature	Importance
5	prev_rating_ave_pastYear	0.328803
0	rating_ave_pastYear	0.240714
3	num_5_star_Rev_pastYear	0.117364
6	prev_numReviews_pastYear	0.112582
7	prev_numCancel_pastYear	0.074841
1	numReviews_pastYear	0.036434
2	numCancel_pastYear	0.020898
25	Number of Reviews	0.010501
26	prev_Number of Reviews	0.009876
4	prop_5_StarReviews_pastYear	0.004606
48	tract_superhosts	0.004126
13	available_days	0.003688
30	occupancy_rate	0.003058
11	prev_numReservedDays_pastYear	0.002942
9	numReservedDays_pastYear	0.002796
37	tract_housing_units	0.002777
32	prev_occupancy_rate	0.002595
27	Rating Overall	0.001941
8	prev_num_5_star_Rev_pastYear	0.001721
50	tract_price_variance	0.001550

Recommendations



Easy Level: These are tasks that can be implemented quickly and with minimal effort:

Enhance Listing Quality:

Take and upload more high-quality photos of the property.

Update listing descriptions regularly with detailed and accurate information.

Encourage Guest Reviews:

Promptly ask guests for reviews after their stay.

Respond to all reviews, both positive and negative, in a professional manner.

Moderate Level: These tasks require more time and effort but are crucial for long-term success:

Guest Experience Enhancement:

Actively seek and apply guest feedback to improve the service.

Offer amenities or services that stand out (e.g., free Wi-Fi, welcome gifts).

Optimize Online Presence:

Use social media or local tourism websites to promote the listing.

Stay active on the Airbnb platform to maintain high visibility.

Hard Level: These are long-term strategies that involve significant effort or investment:

Property Improvements:

Consider renovations or redecorations to improve guest appeal and comfort.

Increase Bookings and Reserved Days:

Develop and implement strategies to increase the number of reservations and length of stays.

Offer discounts for longer bookings or off-season stays.

Minimize Cancellations:

Establish a reliable and flexible hosting plan to reduce the number of cancellations.

Thank You

Appendix

