



## Leveraging Superhosts

Airbnb Chicago

SUSTAINABILITY AND  
PROFITABILITY

MGMT 683 - TEAM 13





# Team



Santosh Pawan  
Kurada



Sathwik  
Kanukuntla



Soham Agarwal



Rohit Mittal



Varun  
Annapareddy

# Executive Summary

**Objective:** This project aims to enhance Airbnb listing performance by analyzing the dynamics of achieving Super host status. Utilizing advanced machine learning models and a rich panel dataset, we provide insights to improve guest experiences and boost overall platform success.

**Outcome:** Our comprehensive analysis and strategic recommendations aim to significantly improve host performance and guest satisfaction, reinforcing Airbnb's position in the market.

## Key Focus Areas

**Superhost Status Dynamics:** We explored factors influencing Super host status across property types, its business benefits, and identified critical metrics like guest ratings, cancellation rates, and response times.

**Performance Metrics Correlation:** Our investigation revealed a strong link between Super host status and key performance indicators such as booking rates, occupancy, revenue, and guest satisfaction.

**Improving Overall Ratings:** We assessed factors impacting host ratings, emphasizing the importance of consistent quality, responsiveness, and property features.

**Market Trends and Seasonality:** The study examined market dynamics and seasonal trends, offering strategic insights for pricing and marketing.

**Actionable Recommendations:** We provided strategies for hosts to enhance listing appeal and service quality, along with guidance for Airbnb to support its hosting community.

**Super Product Integration:** Integrated machine learning insights with the GPT API, facilitating real-time, data-driven decision-making for hosts and Airbnb management.



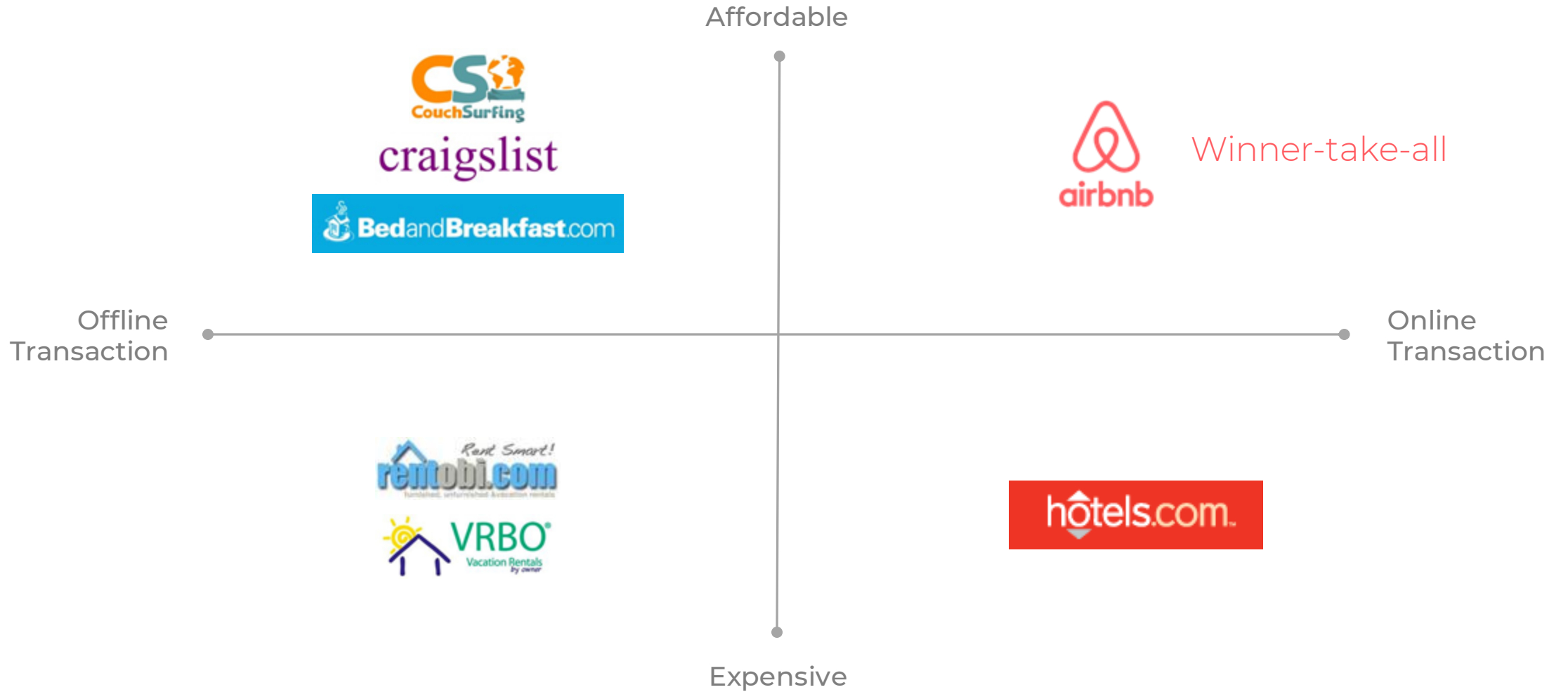
The  
Pinnacle of  
Popularity



A Verb Brand Status



# Competition





# Airbnb Superhosts Chicago

10k+

superhosts  
as of 2018

97.2

property rating  
as of 2018

## Who are they?

Top-rated, most experienced hosts on Airbnb, committed to providing you with outstanding hospitality

## Superhost Criteria

10+

stays

4.8+

overall rating

<1%

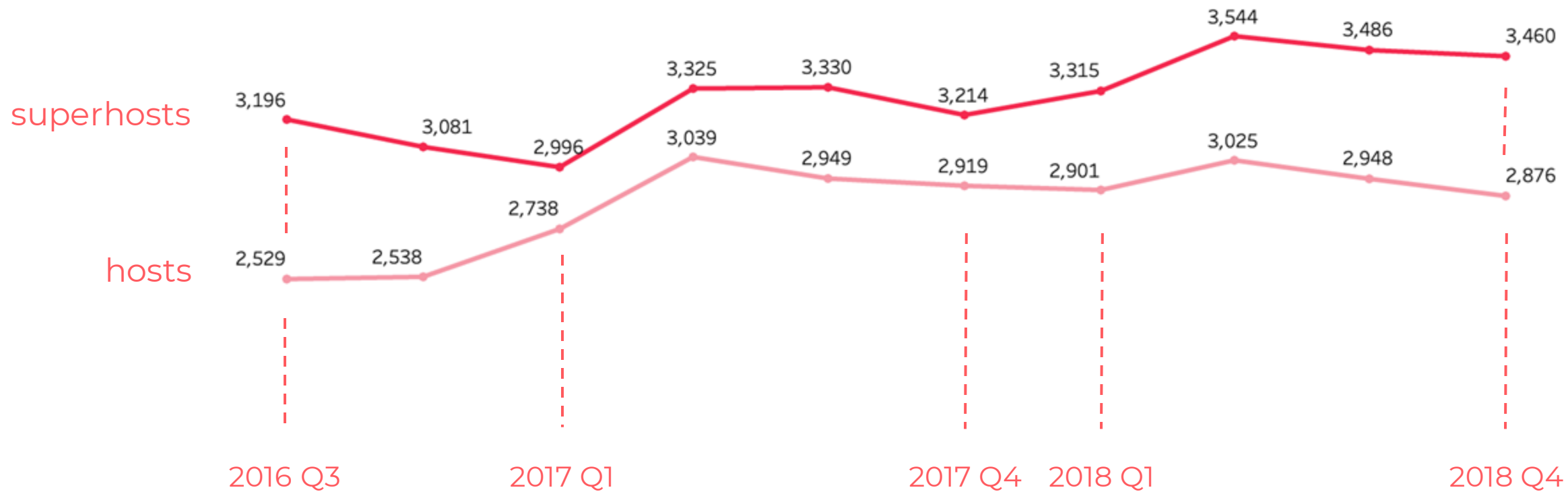
cancellation  
rate

90%

response rate



# Revenue Comparison





Let's begin by breaking our analysis of Airbnb's **superhost program** into **two parts**



What are the **numbers** telling us



What are **numbers** **not** telling us





....to lay it down



What are the  
**numbers** telling us



Am I ever going to be a superhost in this lifetime?

Am I going to lose my superhost status?

3 out of 4 superhost requirements are in my control, but how do I increase my average rating?



What are **numbers not**  
telling us



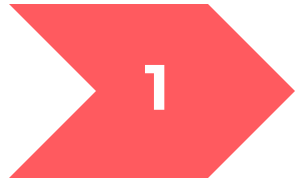
I want to be the CEO of Airbnb one day what more can I do to leverage the superhost program and maximise profitability?



What are the **numbers**  
telling us



# Hosting to Superhosting



## Exploratory Data Analysis

Class Imbalances

Identifying Missing Values, distributions, data types & data statistics

Superhost patterns - Geospatial pattern, Revenue trends



## Data Pre Processing

Manual removal of similar columns

Feature Transformation

Imputing missing values

Removing outliers

Encoding Categorical values



## Feature Selection

Gradient Boosting

Removal of highly correlated variables



## Modelling

Logit Regression

Polynomial Regression

Backward Regression

Adaptive Lasso



## Icing on the cake

State of the art business insights generator to integrate in the Airbnb app



# Model 1

Predicting superhost status

## Problem

Am I ever going to be a superhost in this lifetime?

## Gradient Boosting

Prediction

97.3%

Accuracy

96.9%

Precision

96.2%

Recall

## Logistic Regression

Interpretation

95.1%

Accuracy

86.7%

Precision

86.7%

Recall

## Top 15 features from boosting affecting your chances

1. rating\_ave\_pastYear
2. numCancel\_pastYear
3. num\_5\_star\_Rev\_pastYear
4. superhost\_change\_gain\_superhost
5. numReviews\_pastYear
6. prev\_rating\_ave\_pastYear
7. prev\_num\_5\_star\_Rev\_pastYear
8. prop\_5\_StarReviews\_pastYear
9. prev\_prop\_5\_StarReviews\_pastYear
10. prev\_numCancel\_pastYear
11. hostResponseAverage\_pastYear
12. tract\_superhosts\_ratio
13. prev\_numReviews\_pastYear
14. tract\_prev\_superhosts\_ratio
15. Number of Reviews

Target: Predicting superhost status

## Features analyzed

-2.225%

cancellations past  
year

1.027%

superhost : host in  
a tract

0.44%

avg. rating for  
past year in the  
previous period

2.59%

avg. rating past  
year

0.05%

responses in the  
past year

0.301%

reserved days in  
the past year



# Business Implications

Quality of  
Service and  
Guest  
Satisfaction

Market  
Positioning and  
Competitive  
Environment

Operational  
Efficiency and  
Strategic  
Management



## Model 2

Superhost loss classifier

## Problem

Am I going to lose my  
superhost status?

## Gradient Boosting

Prediction

97.8%

Accuracy

86.4%

Precision

## Logistic Regression

Interpretation

95.8%

Accuracy

## Top 15 features from boosting affecting your chances

1. prev\_numCancel\_pastYear
2. prev\_numReviews\_pastYear
3. prev\_rating\_ave\_pastYear
4. hostResponseNumber\_pastYear
5. superhost\_period\_all
6. revenue
7. prev\_hostResponseAverage\_pastYear
8. numCancel\_pastYear
9. tract\_superhosts\_ratio
10. prev\_Number of Reviews
11. numReviews\_pastYear
12. scrapes\_in\_period
13. tract\_superhosts
14. prev\_hostResponseNumber\_pastYear
15. rating\_ave\_pastYear

Target: Predicting loss of superhost status

Features analyzed

0.756%

Cancellations in  
past 1 year

-0.073%

Annual Revenue

-0.575%

Number of Reviews  
past year

0.174%

Number of Super  
hosts in the tract





# Business Implications

Ensure clear communication and guest engagement to avoid cancellations

High Revenue may help the host retain super host status

Focus on guest satisfaction for more reviews



# Model 3

Predicting Average  
Rating

## Problem

3 out of 4 superhost requirements are in my control, but how do I increase my average rating?

**Linear Regression**  
Prediction & Interpretation

**61.5%**  
R<sup>2</sup> Score

## Top 15 features from affecting your chances

1. Neighborhood\_River North
2. rating\_ave\_pastYear
3. Neighborhood\_Armour Square
4. prop\_5\_StarReviews\_pastYear
5. prev\_prop\_5\_StarReviews\_pastYear
6. Neighborhood\_Auburn Gresham
7. Neighborhood\_Washington Heights
8. prev\_rating\_ave\_pastYear
9. Neighborhood\_Calumet Heights
10. Property Type\_Room in a parthotel
11. Neighborhood\_Back of the Yards
12. Neighborhood\_Chatham
13. Zipcode\_60621
14. Neighborhood\_Belmont Cragin
15. prev\_Rating Overall

Target: Predicting Host Rating

Features analyzed

-12.328%

Neighborhood in River North

-1.603%

Property Type - an apartment turned hotel.

6.255%

Avg rating of the past year

3.425%

previous proportion of 5 Star Reviews in the past year

3.190%

Being in the Neighborhood Auburn Gresham

2.296%

Being in the neighborhood Washington Height

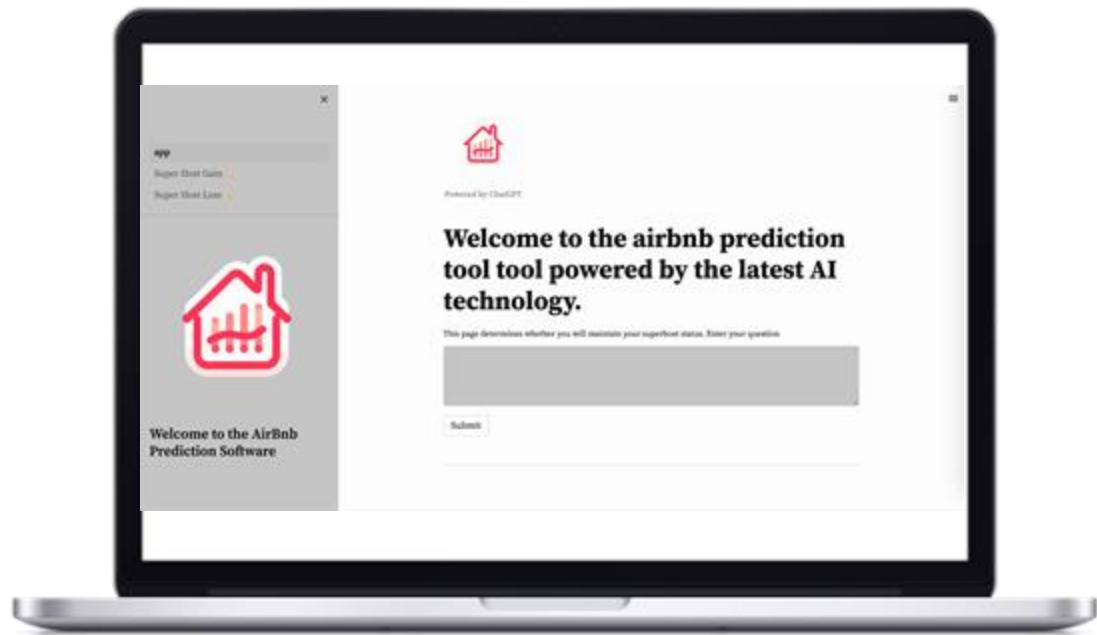


# Business Implications

Location  
Influence and  
Neighborhood  
Characteristics

Property Type  
and Guest  
Expectations

Quality of  
Service and Past  
Performance



## Our “Super” Product

As a host, imagine you could know on how to become a superhost or where you're going wrong!

Our website is a cutting-edge platform designed for Airbnb hosts, integrating three specialized machine learning models with the GPT API for enhanced business insights. Additionally, our site employs GPT's advanced AI capabilities to answer business-related questions, providing hosts with a comprehensive, data-driven approach to optimize their Airbnb operations.



What are the **numbers not**  
telling us



Trade off between the motivation to become a Superhost and Exclusive Status of a Superhost

## Dynamic Metrics to become a superhost based on regional performance

Optimize the Superhost Program by creating custom metrics so that hosts never miss an opportunity and Airbnb never misses an extra buck from the superhost premiums

Meet your Superhost

📍 Little Rock, Arkansas

👤 3 years of hosting experience

🏠 Myranda's house is rated ★4.96



# Why Dynamic Criteria for Superhosts?

## Motivation

There are several psychological and cognitive factors that contribute to the decrease in motivation when faced with an overly difficult task:

1. Perceived competence
2. Goal discrepancy
3. Extrinsic vs. intrinsic motivation

## Exclusivity

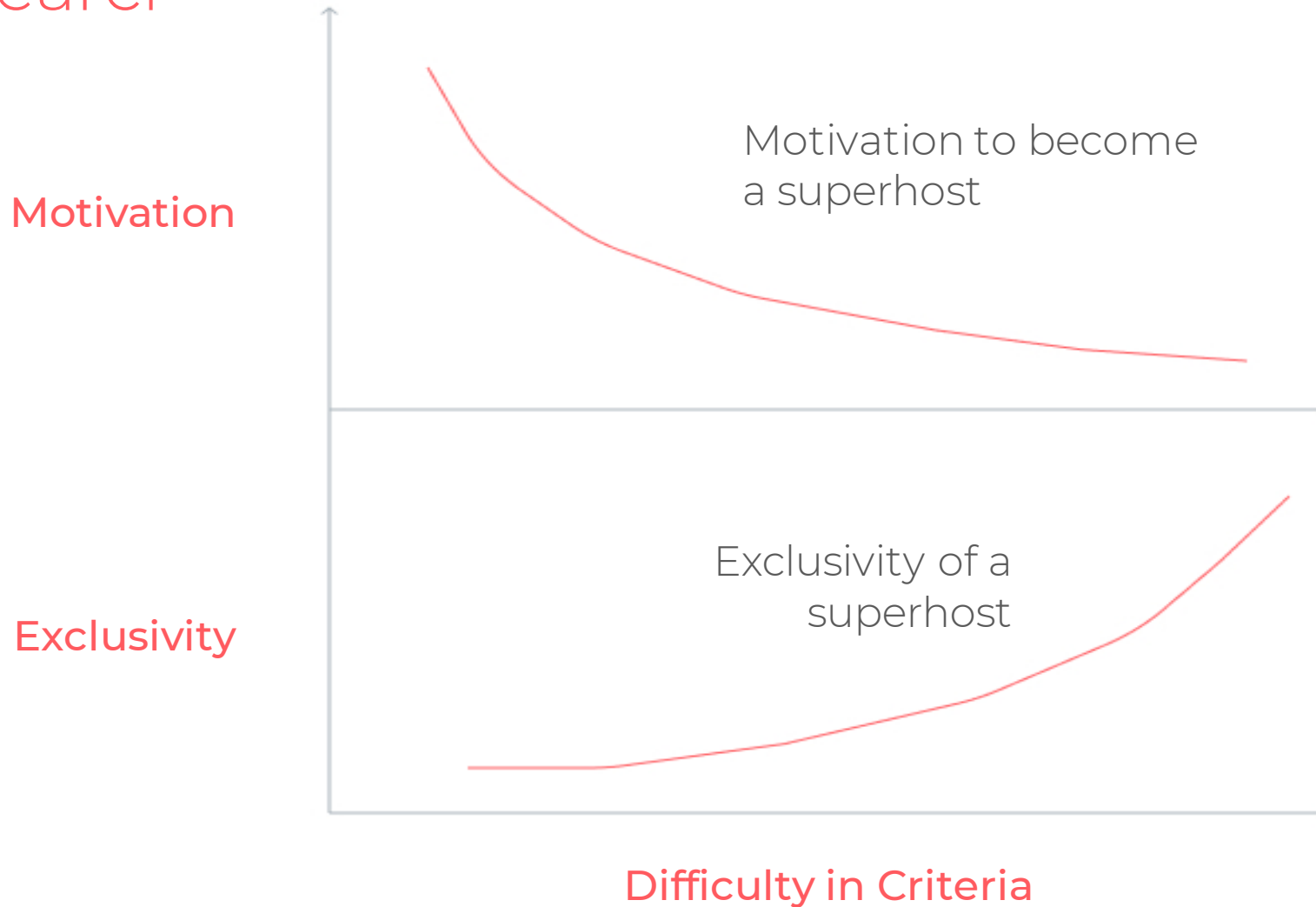
An exclusive membership or exclusive status can lose its exclusivity in many ways:

1. Overgrowth
2. Accessibility
3. Loss of differentiation





# Representation Based on Qualitative Research





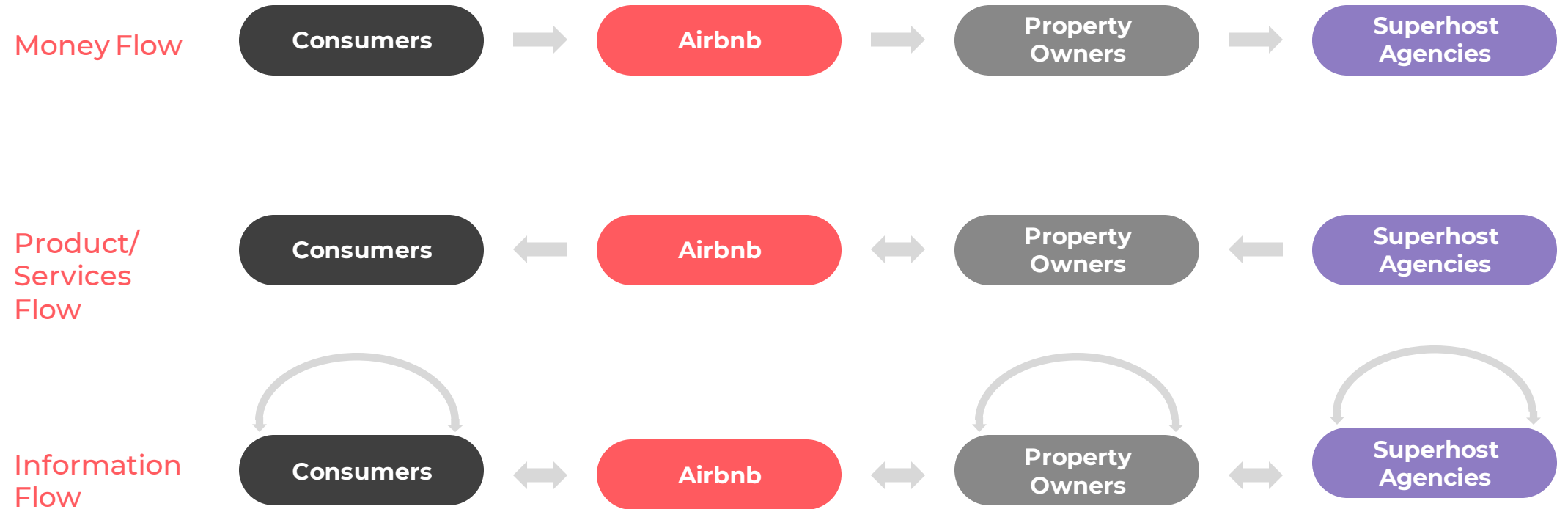
Leveraging a hidden player in the value chain

## Airbnb Certified Property Management Agencies

Increase your odds of getting perks for a property being managed by a superhost!



# Airbnb Value Chain

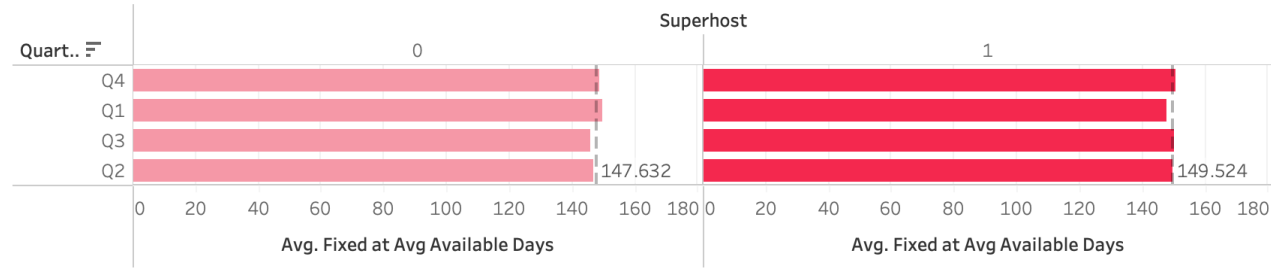




Thank you  
Happy Hosting!

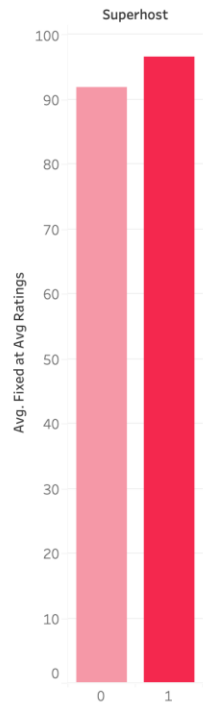
# Superhost VS Host Analysis

Avg Available Dates



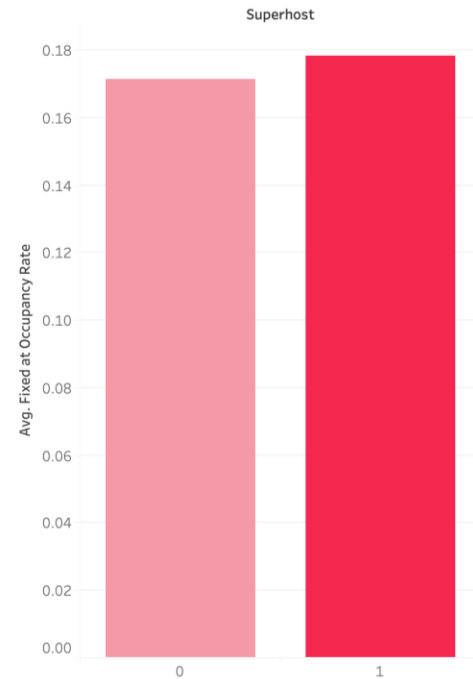
0 – Host  
1 - Superhost

Ratings for Host vs Superhost

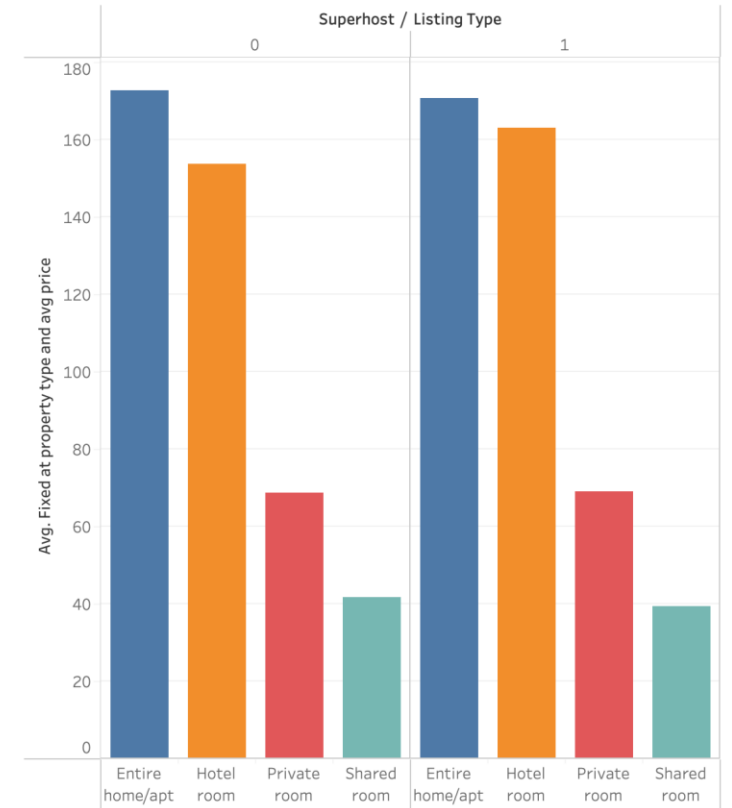


Occupancy Rate

(can say small differences, but these small margins can sometimes just be the difference to have that extra share of revenue you can enjoy as a superhost)



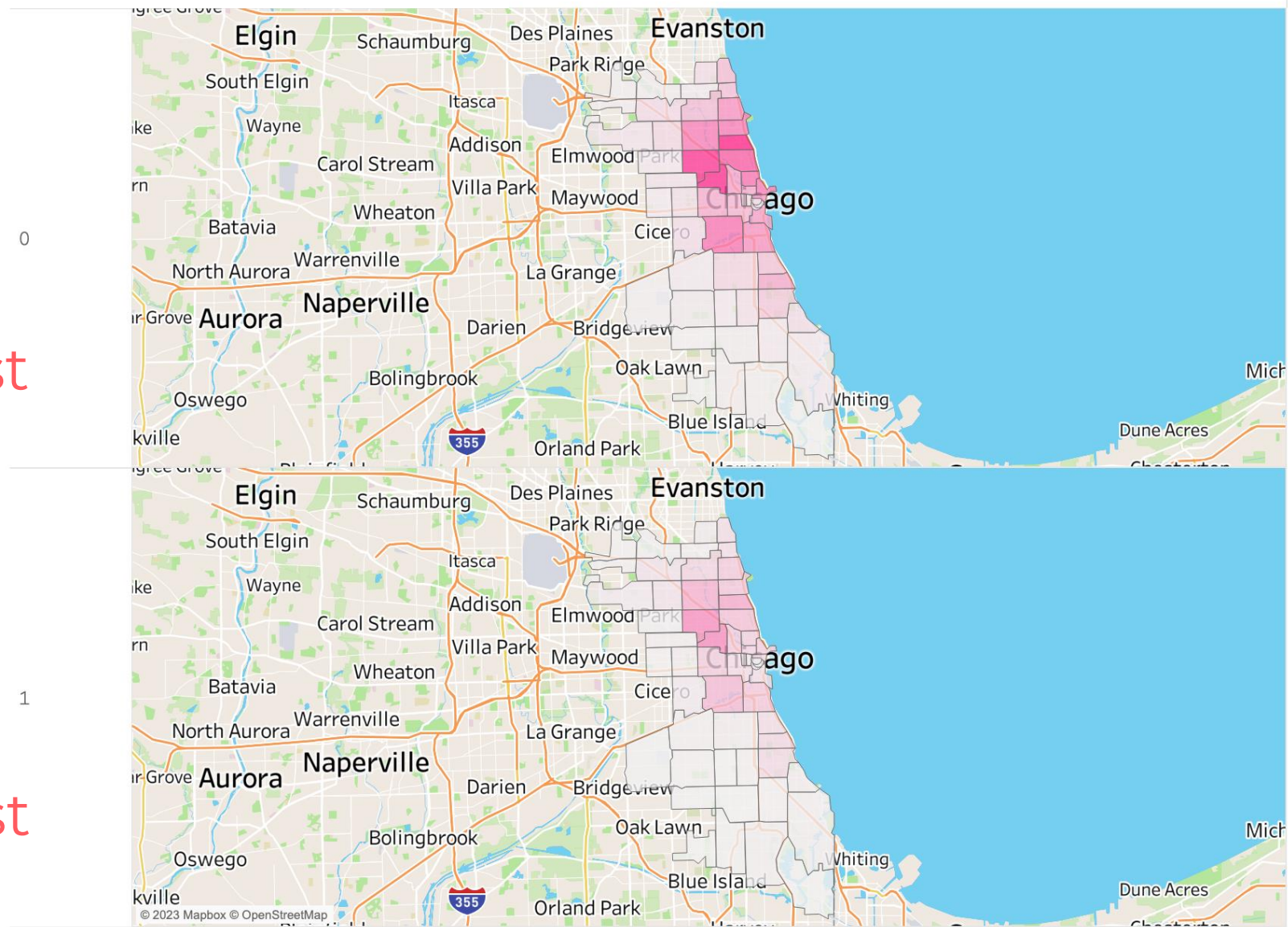
Property Type and Booking Prices



# Geospatial Spread of Hosts and Superhosts

Host

Superhost



# Model 1 Summary

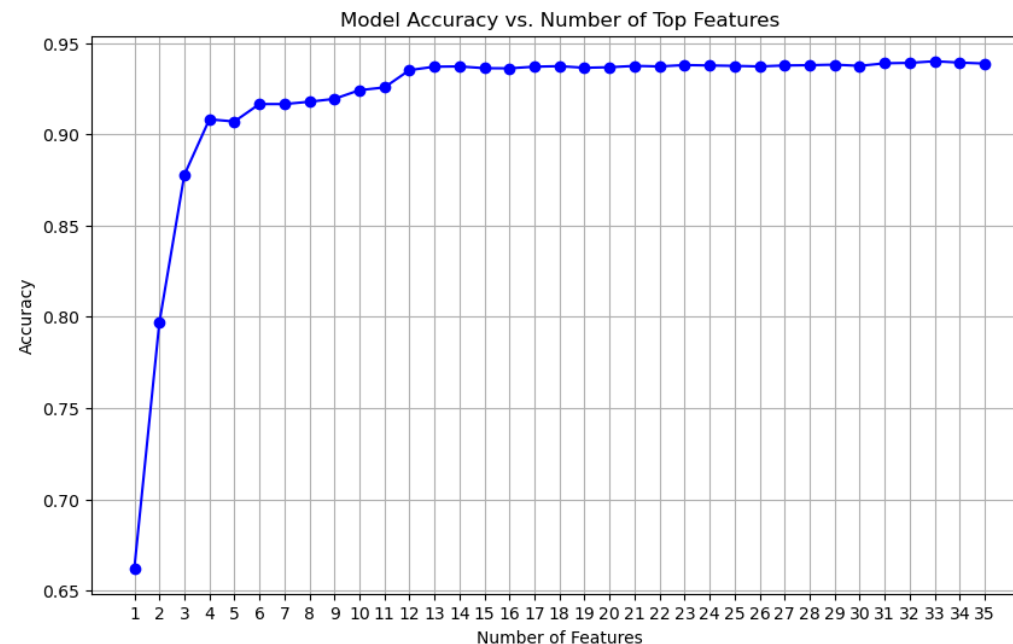
Optimization terminated successfully (Exit mode 0)  
Current function value: 0.24343313218582469  
Iterations: 94  
Function evaluations: 95  
Gradient evaluations: 94

## Logit Regression Results

```
=====
Dep. Variable:      Superhost      No. Observations:      55808
Model:              Logit          Df Residuals:          55788
Method:             MLE            Df Model:              19
Date:               Thu, 07 Dec 2023 Pseudo R-squ.:          0.6312
Time:               14:43:58       Log-Likelihood:        -13586.
converged:          True           LL-Null:               -36840.
Covariance Type:   nonrobust       LLR p-value:           0.000
=====
```

	coef	std err	z	P> z	[0.025	0.975]
const	-2.3016	0.027	-86.244	0.000	-2.354	-2.249
prev_rating_ave_pastYear	0.4386	0.038	11.536	0.000	0.364	0.513
tract_superhosts_ratio	1.0221	0.021	48.634	0.000	0.981	1.063
Minimum Stay	0.1321	0.017	7.948	0.000	0.100	0.165
numCancel_pastYear	-2.2497	0.042	-53.191	0.000	-2.333	-2.167
tract_total_pop	0.0050	0.016	0.308	0.758	-0.027	0.037
hostResponseAverage_pastYear	0.5592	0.038	14.534	0.000	0.484	0.635
hostResponseNumber_pastYear	0.0498	0.023	2.135	0.033	0.004	0.096
Period Category	-0.2382	0.018	-13.076	0.000	-0.274	-0.202
prev_num_5_star_Rev_pastYear	-0.1632	0.030	-5.505	0.000	-0.221	-0.105
rating_ave_pastYear	2.5570	0.049	52.688	0.000	2.462	2.652
numReservedDays_pastYear	0.3006	0.031	9.854	0.000	0.241	0.360
prev_Number of Reviews	0.1940	0.027	7.136	0.000	0.141	0.247
scrapes_in_period	-0.0577	0.029	-2.000	0.045	-0.114	-0.001
prev_scrapes_in_period	-0.1651	0.031	-5.300	0.000	-0.226	-0.104
prev_numCancel_pastYear	0.0041	0.031	0.131	0.896	-0.057	0.065
num_5_star_Rev_pastYear	2.2578	0.039	57.495	0.000	2.181	2.335
prev_hostResponseAverage_pastYear	-0.0477	0.032	-1.498	0.134	-0.110	0.015
Number of Reviews	0.0640	0.026	2.435	0.015	0.012	0.116
prev_numReserv_pastYear	-0.0187	0.022	-0.832	0.406	-0.063	0.025

## Elbow Graph for feature selection



- **K fold cross validation:**

Fold Accuracies: [0.9635321100917431, 0.9587155963302753, 0.9572247706422018, 0.9576834862385321, 0.9583715596330276, 0.9592889908256881, 0.9591743119266055, 0.9589449541284404]  
Mean Accuracy: 0.9591169724770643

# Model 2 Summary

## Logit Regression Results

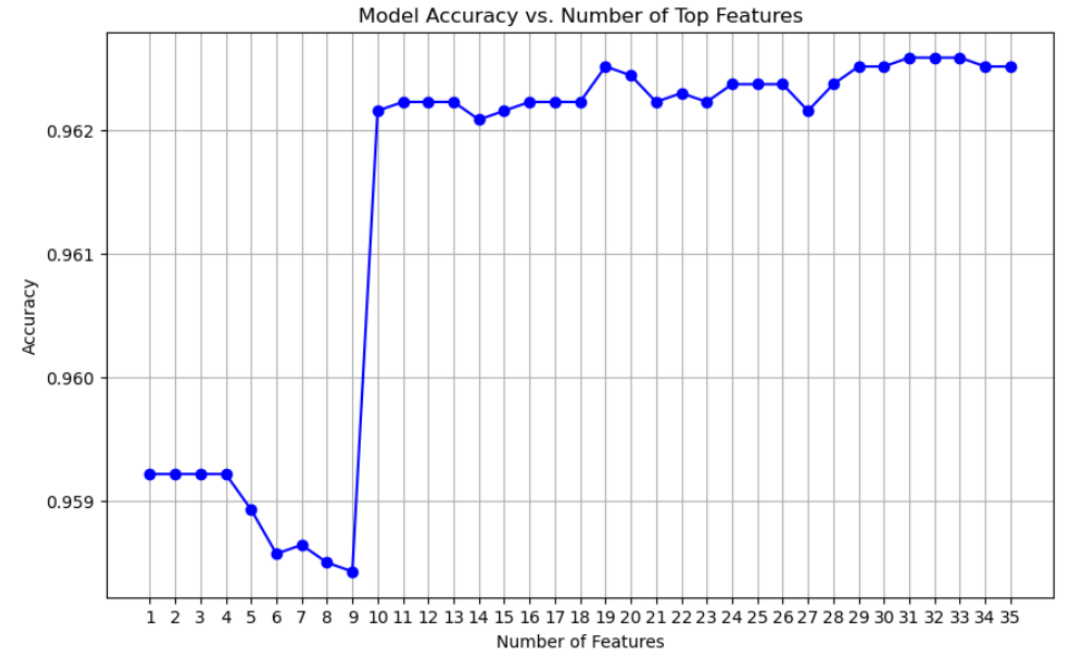
```

=====
Dep. Variable:    superhost_change_lose_superhost    No. Observations:    55808
Model:           Logit                               Df Residuals:        55788
Method:          MLE                                 Df Model:            19
Date:            Thu, 07 Dec 2023                    Pseudo R-squ.:       0.1823
Time:            14:58:09                            Log-Likelihood:      -7918.5
converged:       True                               LL-Null:             -9684.0
Covariance Type: nonrobust                          LLR p-value:         0.000
=====

```

	coef	std err	z	P> z	[0.025	0.975]
const	-4.4170	0.063	-70.196	0.000	-4.540	-4.294
prev_numCancel_pastYear	-2.5219	0.098	-25.609	0.000	-2.715	-2.329
prev_numReviews_pastYear	0.9206	0.055	16.671	0.000	0.812	1.029
prev_rating_ave_pastYear	1.1947	0.045	26.558	0.000	1.106	1.283
hostResponseNumber_pastYear	0.1681	0.050	3.384	0.001	0.071	0.265
superhost_period_all	0.1396	0.026	5.473	0.000	0.090	0.190
revenue	-0.0726	0.022	-3.375	0.001	-0.115	-0.030
prev_hostResponseAverage_pastYear	0.1579	0.030	5.333	0.000	0.100	0.216
numCancel_pastYear	0.7532	0.026	28.906	0.000	0.702	0.804
prev_numReserv_pastYear	0.0422	0.054	0.777	0.437	-0.064	0.149
prev_scrapes_in_period	-0.0295	0.045	-0.652	0.514	-0.118	0.059
Property Type_Serviced apartment	0.0270	0.021	1.283	0.199	-0.014	0.068
tract_superhosts_ratio	-0.6263	0.039	-16.077	0.000	-0.703	-0.550
prev_Number of Reviews	0.2072	0.036	5.688	0.000	0.136	0.279
numReviews_pastYear	-0.5765	0.046	-12.400	0.000	-0.668	-0.485
numReserv_pastYear	-0.0143	0.057	-0.252	0.801	-0.126	0.097
scrapes_in_period	0.1136	0.041	2.773	0.006	0.033	0.194
tract_superhosts	0.1737	0.034	5.072	0.000	0.107	0.241
prev_hostResponseNumber_pastYear	-0.1501	0.039	-3.860	0.000	-0.226	-0.074
rating_ave_pastYear	-0.7674	0.034	-22.579	0.000	-0.834	-0.701

## Elbow Graph for feature selection



## K fold cross validation:

Fold Accuracies: [0.9599770642201835, 0.959059633027523, 0.9576834862385321, 0.9596330275229358, 0.9576834862385321, 0.9616972477064221, 0.9569954128440367, 0.9547018348623854]  
 Mean Accuracy: 0.9584288990825688



# Model 3 Summary

OLS Regression Results

```

=====
Dep. Variable:      Rating Overall    R-squared:      0.632
Model:             OLS               Adj. R-squared: 0.631
Method:           Least Squares      F-statistic:    359.3
Date:             Thu, 07 Dec 2023   Prob (F-statistic): 0.00
Time:             21:20:06           Log-Likelihood: -1.3221e+05
No. Observations: 53540            AIC:            2.649e+05
Df Residuals:     53284            BIC:            2.672e+05
Df Model:         255
Covariance Type:  nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	0.1587	12.020	0.013	0.989	-23.400	23.717
superhost_period_all	0.1666	12.074	0.014	0.989	-23.498	23.832
scrapes_in_period	0.0004	0.001	0.425	0.670	-0.001	0.002
prev_scrapes_in_period	-0.0065	0.002	-4.283	0.000	-0.009	-0.004
Superhost	0.0509	0.040	1.266	0.205	-0.028	0.130
superhost_change_lose_superhost	-9.361e-09	4.28e-07	-0.022	0.983	-8.47e-07	8.29e-07
superhost_change_gain_superhost	-0.1614	0.055	-2.931	0.003	-0.269	-0.053
rating_ave_pastYear	6.2753	0.245	25.581	0.000	5.794	6.756
numReviews_pastYear	-0.0219	0.004	-5.441	0.000	-0.030	-0.014

Linear regression  
Target: Rating

## Linear regression with Backward Selection Target: Rating

OLS Regression Results

```

=====
Dep. Variable:      Rating Overall    R-squared:      0.631
Model:             OLS               Adj. R-squared: 0.629
Method:           Least Squares      F-statistic:    484.7
Date:             Thu, 07 Dec 2023   Prob (F-statistic): 0.00
Time:             20:54:20           Log-Likelihood: -1.3232e+05
No. Observations: 53540            AIC:            2.650e+05
Df Residuals:     53351            BIC:            2.667e+05
Df Model:         188
Covariance Type:  nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	34.6874	323.247	0.107	0.915	-598.880	668.255
superhost_period_all	40.5795	378.086	0.107	0.915	-700.472	781.631
Superhost	0.0407	0.040	1.016	0.309	-0.038	0.119
superhost_change_lose_superhost	-1.653e-05	0.000	-0.107	0.915	-0.000	0.000
superhost_change_gain_superhost	-0.1669	0.055	-3.031	0.002	-0.275	-0.059
rating_ave_pastYear	6.2550	0.245	25.554	0.000	5.775	6.735
numReviews_pastYear	-0.0220	0.004	-5.488	0.000	-0.030	-0.014
numCancel_pastYear	-0.0892	0.034	-2.596	0.009	-0.157	-0.022
num_5_star_Rev_pastYear	0.0313	0.005	6.135	0.000	0.021	0.041
prop_5_StarReviews_pastYear	-4.0764	0.354	-11.522	0.000	-4.770	-3.383
prev_rating_ave_pastYear	-2.2485	0.272	-8.262	0.000	-2.782	-1.715
prev_numReviews_pastYear	0.0135	0.004	3.111	0.002	0.005	0.022
prev_numCancel_pastYear	0.0748	0.039	1.930	0.054	-0.001	0.151
prev_num_5_star_Rev_pastYear	-0.0201	0.005	-3.661	0.000	-0.031	-0.009
prev_prop_5_StarReviews_pastYear	3.4248	0.381	8.993	0.000	2.678	4.171
numReservedDays_pastYear	0.0010	0.000	3.663	0.000	0.000	0.002

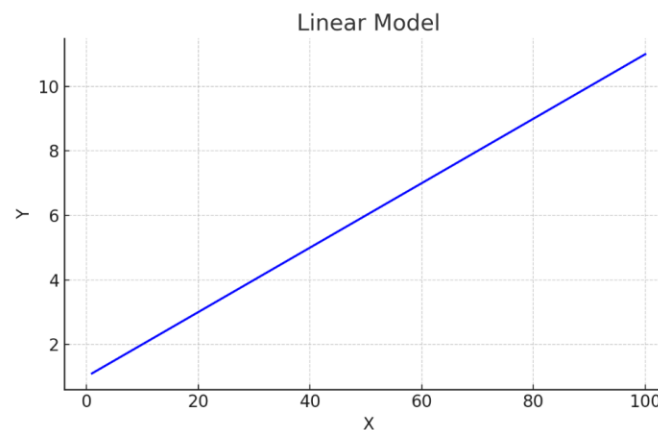


# Interpreting Coefficients – Regression Models

Generalized Graphs for Regression Models

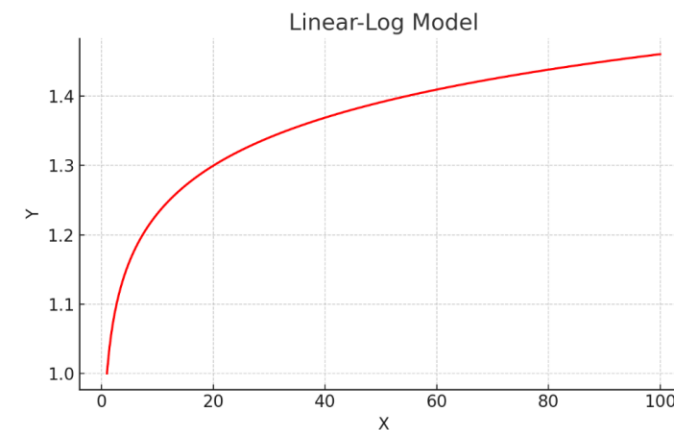
## Linear Regression (Model A)

- **Formula:**  $Y = b_0 + b_1 X$
- **Impact with  $d$  units:**  $b_1 \times d$  units increase in  $Y$ .



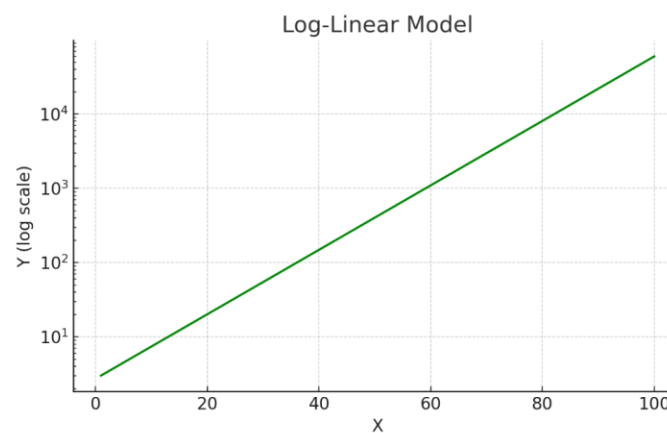
## Linear-Log Regression (Model B)

- **Formula:**  $Y = b_0 + b_1 \ln(X)$
- **Impact with  $d\%$  change:**  $\ln(1 + \frac{d}{100}) \times b_1$  units increase in  $Y$ .



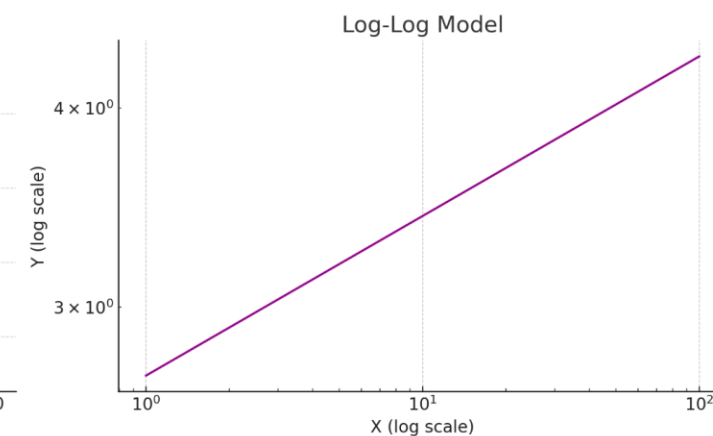
## Log-Linear Regression (Model C)

- **Formula:**  $\ln(Y) = b_0 + b_1 X$
- **Impact with  $d$  units:**  $(e^{b_1 \times d} - 1) \times 100\%$  increase in  $Y$ .



## Log-Log Regression (Model D)

- **Formula:**  $\ln(Y) = b_0 + b_1 \ln(X)$
- **Impact with  $d\%$  change:**  $(e^{b_1 \times \ln(1 + \frac{d}{100})} - 1) \times 100\%$  increase in  $Y$ .





# Model Links

**Superhost classifier:**

<https://drive.google.com/file/d/1jfziQv5-KHHcYnFBNj-kk1O01GXdQkvy/view?usp=sharing>

**Superhost loss predictor:**

<https://drive.google.com/file/d/19n6QO-nXUCa3CO-oSb5XdVWku33cuUyO/view?usp=sharing>

**Overall rating 1:**

<https://drive.google.com/file/d/1FntfRQDY8vegqRjQIzbu9vx3gceML7ep/view?usp=sharing>

**Overall rating 2:**

[https://drive.google.com/file/d/1A7S3dTpEtiKwY0s-Kv3HqM3o3026xFbY/view?usp=drive\\_link](https://drive.google.com/file/d/1A7S3dTpEtiKwY0s-Kv3HqM3o3026xFbY/view?usp=drive_link)

**Entire project link:**

<https://drive.google.com/drive/folders/1ghfZSWS389OFqudRu6aNp08G7A9AZi-?usp=sharing>

# References for recommendations

## MOTIVATION

- 1. Perceived competence:** Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman and Company. Schunk, D. H., & Pajares, F. (2002). Self-efficacy theory and conceptions of learning. In *Educational psychology: An introduction* (pp. 308-339). Routledge.
- 2. Goal discrepancy:** Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705. Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. *Advances in experimental social psychology*, 30(1), 1-46.
- 3. Extrinsic vs. intrinsic motivation:** Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press. Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.

## EXCLUSIVITY

- 1. Overgrowth:** Source: Schriver, K. (2012). *The secret lives of clubs: Inside the rituals and rivalries of the world's most exclusive memberships*. St. Martin's Press. (Chapter 3: The Perils of Popularity) Quote: "Once a club reaches a certain size, it becomes difficult to maintain the same level of intimacy and exclusivity. New members dilute the sense of belonging, and the club can start to feel more like a generic social gathering place."
- 2. Accessibility:** Source: Veblen, T. (1899). *The theory of the leisure class*. Dover Publications. (Chapter 4: Conspicuous Consumption) Quote: "As the goods or services associated with a particular status become more accessible, they lose their power to signal exclusivity. What was once a symbol of distinction for the wealthy elite becomes attainable by the masses, and its prestige diminishes."
- 3. Loss of differentiation:** Source: Thompson, C. J. (2005). *Strategies for competing in global markets*. Routledge. (Chapter 8: Branding and Differentiation) Quote: "In today's competitive landscape, exclusivity alone is not enough to sustain a competitive advantage. Consumers are constantly bombarded with choices, and they are likely to gravitate towards brands or experiences that offer unique value propositions beyond mere exclusivity."