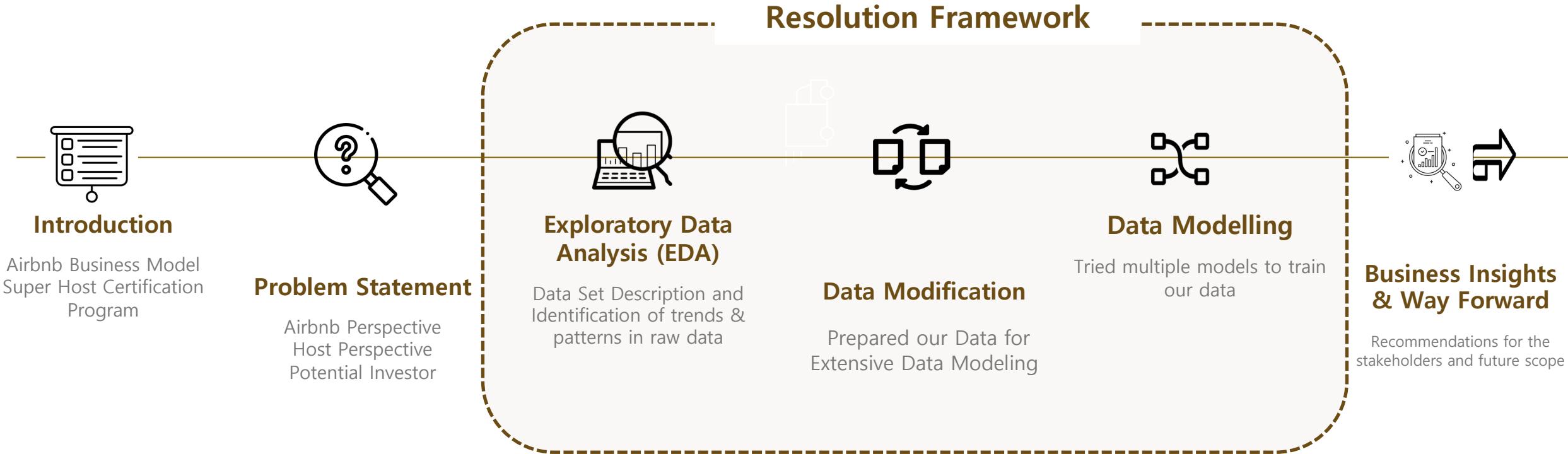




MGMT 683- AIRBNB | DATA-DRIVEN STRATEGIES FOR AIRBNB ECOSYSTEM IN CHICAGO
GROUP 27: AKANKSHA SINGH, DEVARSHI SHARMA, HARSHRAJ JADEJA, HUMMARAH SHAHZAD

AGENDA



Project Objective

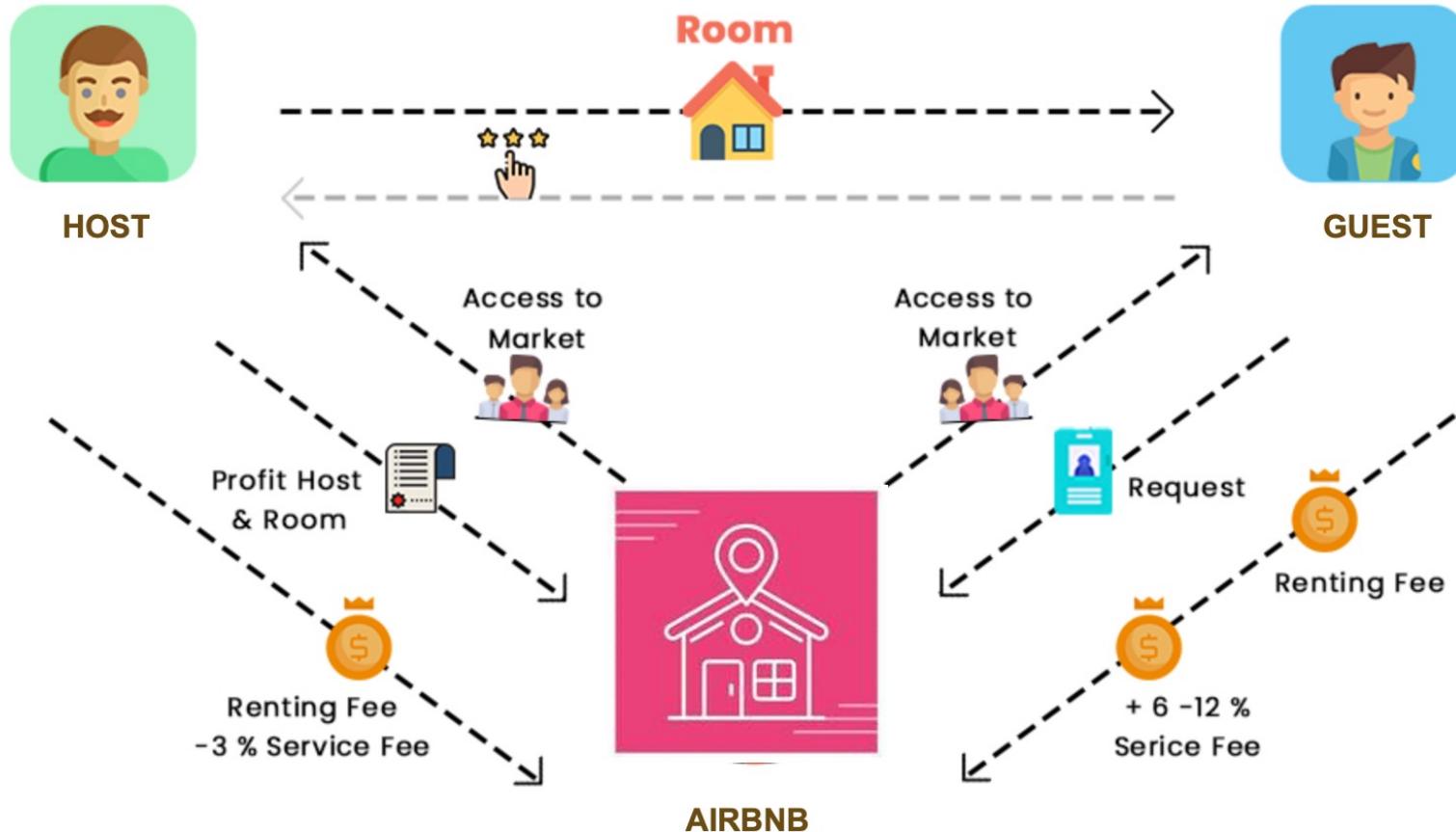
UTILIZE GIVEN DATA SETS ON AIRBNB TO SOLVE REAL TIME PROBLEMS USING ROBUST MODELING TECHNIQUES



INTRODUCTION

AIRBNB BUSINESS MODEL

Insight into Stakeholders within a Platform-Based Airbnb Business Operations.



SUPER HOST CERTIFICATION PROGRAM

The Superhost certification program yields mutual benefits across the entire spectrum of stakeholders within the Airbnb ecosystem.



Certification Program

Airbnb to provide guests with a clear and visible platform-initiated indicator of host quality

Criterion:

- Hosted at least 10 trips
- Maintained a 90% response rate when responding to guests' requests
- Completed all confirmed reservations without cancellation
- Received a 5-star review at least 80% of the time



Airbnb



Host



Guest

- **Enhanced Reputation-** Highlights reliable hosts, boosting platform credibility
 - **Improved Guest Satisfaction-** Attracts guests seeking quality stays, raising overall satisfaction.
 - **Competitive Edge-** Sets high standards, encouraging hosts to maintain quality
-
- **Increased Bookings-** Attracts more guests due to trust, potentially increasing earnings.
 - **Higher Earnings-** With more bookings and increased trust, Superhosts can often command higher rates for their listings, leading to potentially higher earnings
 - **Recognition & Support-** Access to better customer service, resources, and promotions.
-
- **Assurance of Quality-** Visible badge indicates host commitment to excellence.
 - **Trust & Confidence-** Reliable hosts with positive reviews ensure a better stay.
 - **Better Communication-** Responsive hosts address guest inquiries promptly, enhancing the experience.



PROBLEM IDENTIFICATION

PROBLEM STATEMENT

Implemented a comprehensive, multi-faceted strategy to pinpoint real-time challenges experienced by all stakeholders in a platform-based business model



- Re-evaluation of the criteria for a Superhost status with additional parameters
- Potential new Airbnb investments by investors



- Predict Host or Superhost status based on Various Factors
- Significance of Superhost status using DiD approach
- Better understanding of the factors that influence occupancy and drive business



- Competition measure by Herfindahl index
- Optimizing Return on Investments by neighborhood for new investments

Superhost Evaluation Criterion Re-evaluation:
Robustness of Current evaluation matrix

Superhost Status Prediction:
Host status in the upcoming period

Occupancy Rate Prediction based on Status:
What to expect in terms of occupancy with respect to Status

Investment Decision:
As a potential host where to invest to add Airbnb listing



SUPERHOST STATUS? DIFFERENCE IT WOULD MAKE? OCCUPANCY RATE?



Problems Identification

WHERE TO INVEST?



DATA SET INFORMATION

Implemented a comprehensive, multi-faceted strategy to pinpoint real-time challenges experienced by all stakeholders in a platform-based business model



No. of Rows & Columns

121K Rows & 111 Columns



Average Booked Days

26 Days Booked



Time Period

16 periods



Average Occupancy Rate

19% Occupancy



Total No. of Hosts

13.6 K Hosts



Average Revenue

\$3,301



Super Hosts Status

~4,600 Hosts



Average Rating Overall

94%



EXPLORATORY DATA ANALYSIS

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis to shape the path for problem solving approach | Host, Superhost Differences



Descriptive Statistics

Variable	Superhost	Mean	StDev	Minimum	Median	Maximum
rating_ave_pastYear	0	4.6490	0.3054	1.0000	4.6800	5.0000
	1	4.8854	0.0814	4.0000	4.8889	5.0000
numReviews_pastYear	0	106.49	220.18	1.00	30.00	2650.00
	1	100.68	207.87	1.00	43.00	2305.00
numCancel_pastYear	0	0.8009	1.9975	0.0000	0.0000	53.0000
	1	0.0810	0.7383	0.00000	0.00000	16.0000
numReservedDays_pastYear	0	682	1955	0.0	208	115926
	1	594.0	1307.7	0.00	236.0	17317.0
hostResponseAverage_pastYear	0	96.392	7.676	0.0000	99.558	100.000
	1	99.078	2.273	41.250	100.000	100.000
available_days	0	163.49	64.96	1.00	182.00	245.00
	1	160.70	58.18	1.00	174.00	245.00
available_days_aveListedPrice	0	129.29	151.43	10.00	94.56	8191.13
	1	129.99	130.38	10.87	98.60	7109.81
booked_days	0	24.570	19.792	1.000	20.000	212.000
	1	26.791	19.674	1.000	24.000	220.000
booked_days_avePrice	0	132.14	132.70	10.00	95.47	3002.45
	1	136.53	130.03	10.00	100.20	3318.50
Bedrooms	0	1.5747	1.0876	0.00000	1.0000	12.0000
	1	1.6720	1.0763	0.00000	1.0000	12.0000

Key Takeaway

Super Hosts Generally have

Higher:

- Rating average past year
- Host response average past year
- Booked days
- Average price booked days
- Bedrooms

Lower:

- Number of Reviews past year
- Number of Reserved Days past year
- Available Days past year

Similar:

- Average Listed Price for Available Days

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis to shape the path for problem solving approach | Host, Superhost Differences



Descriptive Statistics

Variable	Superhost	Mean	StDev	Minimum	Median	Maximum
Bathrooms	0	1.3516	0.8430	0.00000	1.0000	25.0000
	1	1.3422	0.6486	0.00000	1.0000	25.0000
Max Guests	0	4.2441	2.9443	1.0000	4.0000	16.0000
	1	4.3871	2.9061	1.0000	4.0000	16.0000
Cleaning Fee (USD)	0	63.875	52.916	5.000	50.000	510.000
	1	67.505	52.230	5.000	55.000	550.000
Minimum Stay	0	3.695	11.569	1.000	2.000	365.000
	1	3.843	13.040	1.000	2.000	365.000
Number of Photos	0	17.732	12.652	0.0000	14.000	161.000
	1	22.511	14.079	0.0000	19.000	133.000
Instantbook Enabled	0	0.49279	0.49995	0.00000	0.00000	1.00000
	1	0.49729	0.50000	0.00000	0.00000	1.00000
Nightly Rate	0	156.87	146.87	10.00	114.00	1949.00
	1	150.04	132.85	10.00	112.00	1890.00
Number of Reviews	0	36.684	49.153	0.000	18.000	604.000
	1	61.905	65.605	0.000	40.000	621.000
Rating Overall	0	92.985	7.874	0.0000	94.000	100.000
	1	97.331	3.163	0.0000	98.000	100.000
revenue	0	3060	4431	10.0	1780	128125
	1	3524	4634	15.0	2250	120838
occupancy_rate	0	0.18771	0.18221	0.00408	0.14070	1.00000
	1	0.19783	0.17478	0.00408	0.15698	1.00000

Key Takeaway

Super Hosts Generally have

Higher:

- Max Guests
- Cleaning Fee
- Minimum Stay
- Number of Photos
- Number of Reviews
- Overall Rating
- Revenue
- Occupancy Rate

Lower:

- Number of Bathrooms
- Nightly Rate

Similar:

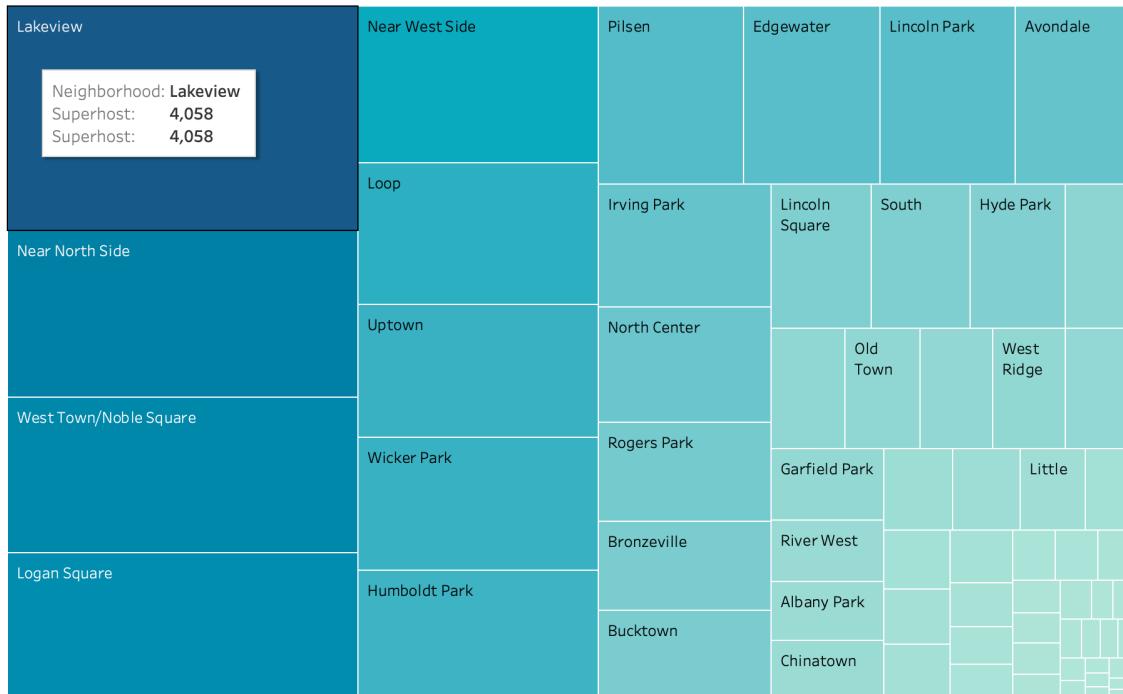
- Proportion of Instant Booking Enabled

EXPLORATORY DATA ANALYSIS



Superhosts Count for a Neighborhood does not determine the Revenue of that Neighborhood

Neighborhood wise Superhosts Count



Highest no. of Superhosts in Lakeview followed by Near North Side & West town Noble Square

Neighborhood wise Revenue



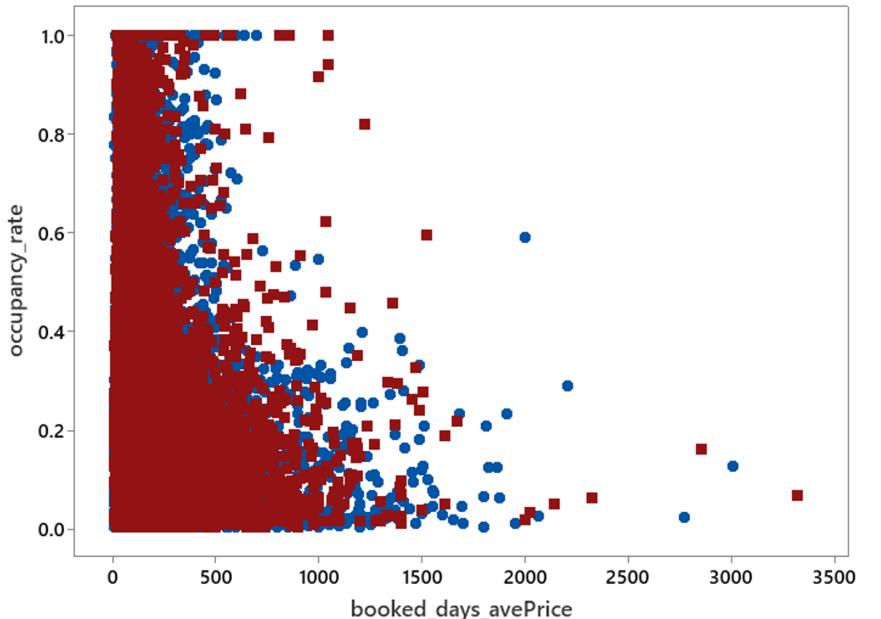
Highest Revenue at Near North Side followed by Old Town, Loop & Lincoln Park

EXPLORATORY DATA ANALYSIS

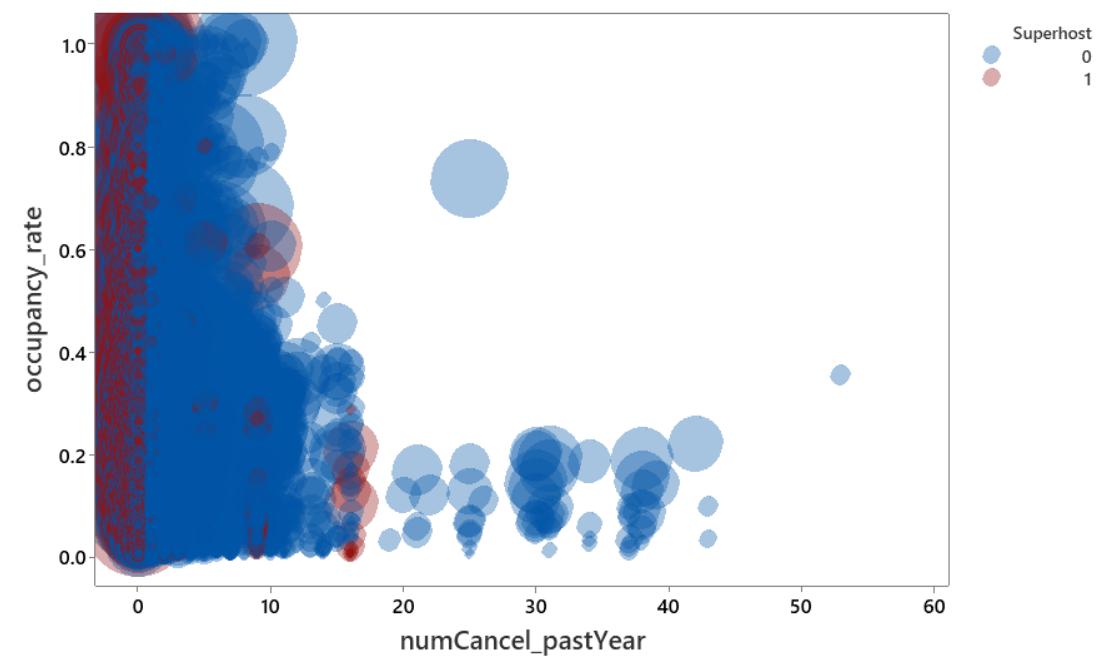
Occupancy rate with average price of booked days and number of cancellations past year



Scatterplot of Occupancy Rate vs Booked days Avg. Price



Bubble Plot of Occupancy Rate vs numCancel_pastYear



Increase in the booked_days_avePrice decreases occupancy rate for both hosts and super hosts, Highest occupancy rate for the booked days average price range \$500- \$1500

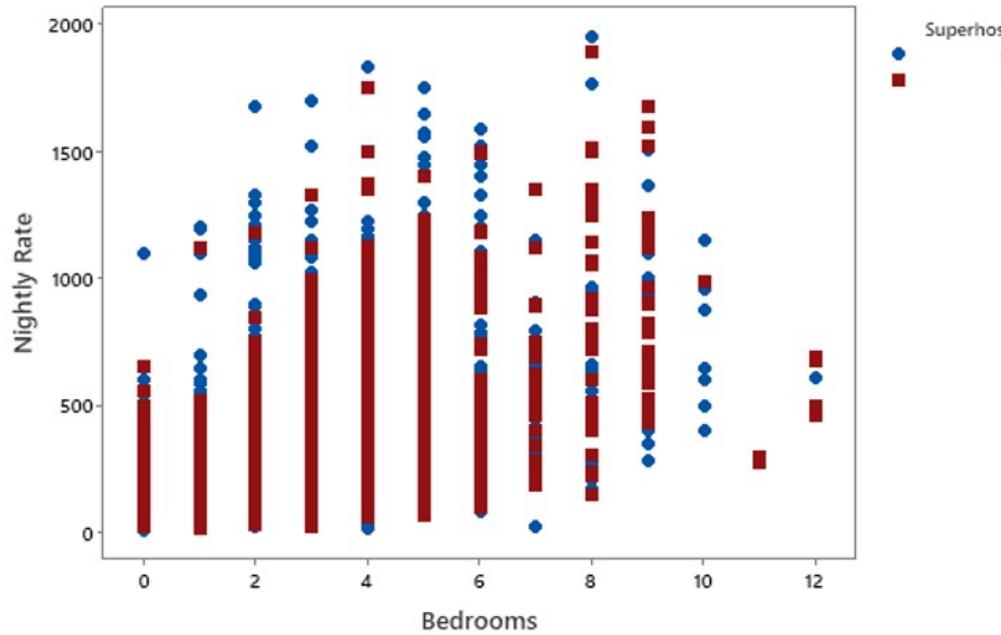
As numCancel_pastYear increases occupancy rate decreases and so does the revenue for both Hosts & Superhosts . Bubble size here is as per the revenue of the host.

EXPLORATORY DATA ANALYSIS

Change in Occupancy Rate & Nightly Rate with respect to the number of bedrooms for Hosts & Superhosts

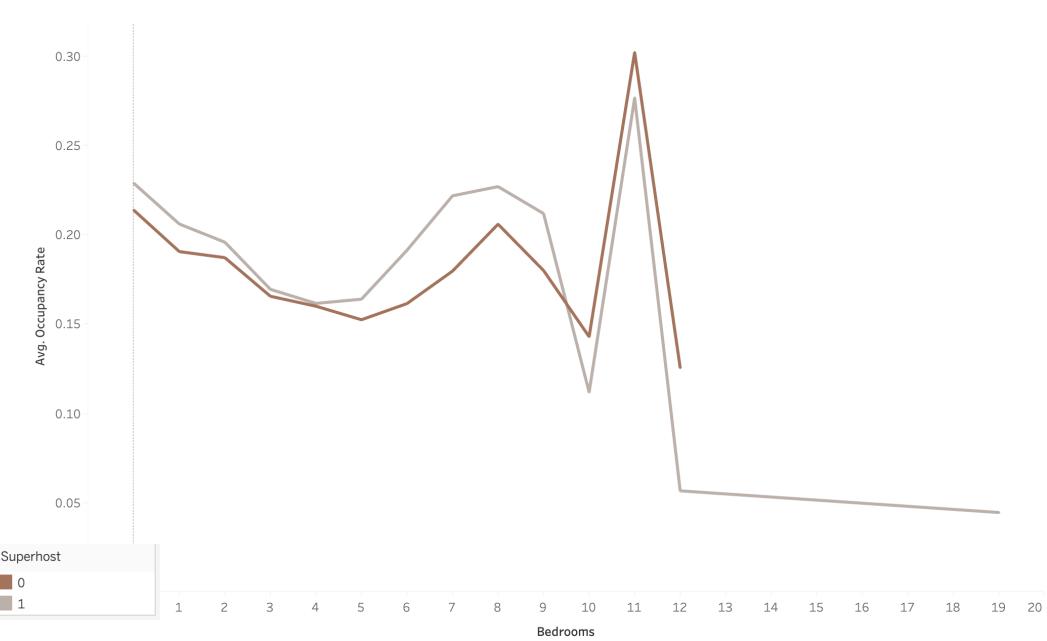


Scatterplot of Nightly Rate vs Bedrooms



Increase in the number of bedrooms increases nightly rate for both hosts & Superhosts. For higher number of bedrooms there are more Superhosts than hosts.

Avg. Occupancy Rate vs No. of Bedrooms



Occupancy rate for Superhosts is generally higher than host across all periods. Occupancy Rate is maximum for the number of bedrooms at 11 for both hosts & Superhosts beyond which it starts decreasing drastically (this could be due to the decrease in demand as well).

EXPLORATORY DATA ANALYSIS

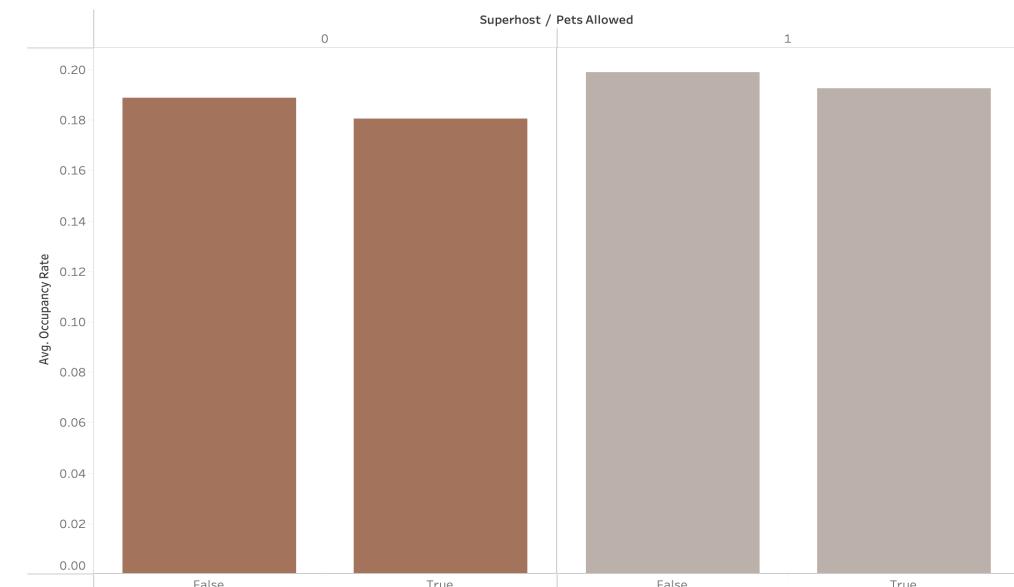
Higher Average no. of guests hosted by Superhosts vs Hosts; Occupancy rate higher for both Hosts & Superhosts for Pets not Allowed



Average no. of Guests for Hosts & Superhosts

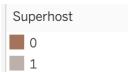


Occupancy Rate for Hosts & Superhosts for Pets Allowed



Increase in the booked_days_avePrice decreases occupancy rate for both hosts and super hosts, Highest occupancy rate for the booked days average price range \$500- \$1500

Increase in the booked_days_avePrice decreases occupancy rate for both hosts and super hosts, Highest occupancy rate for the booked days average price range \$500- \$1500

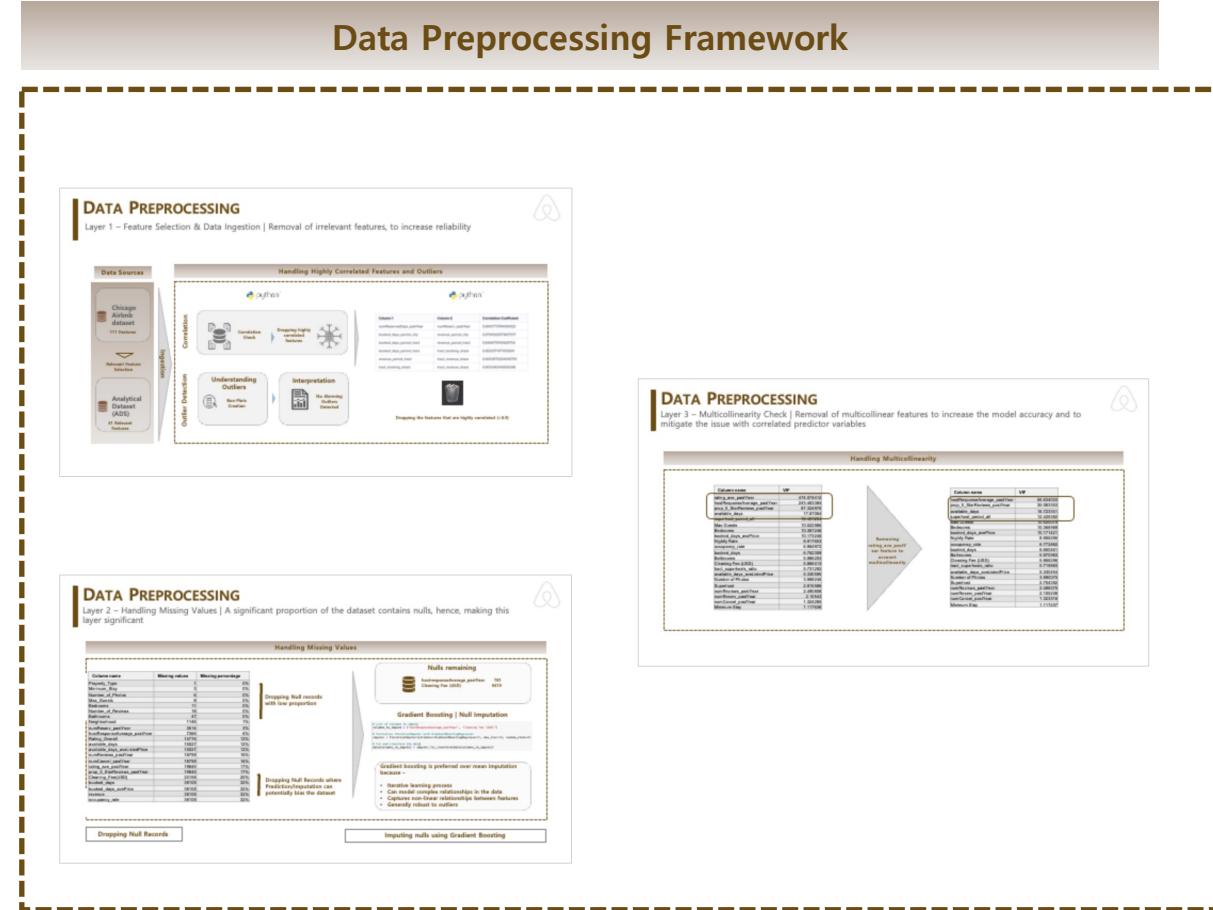
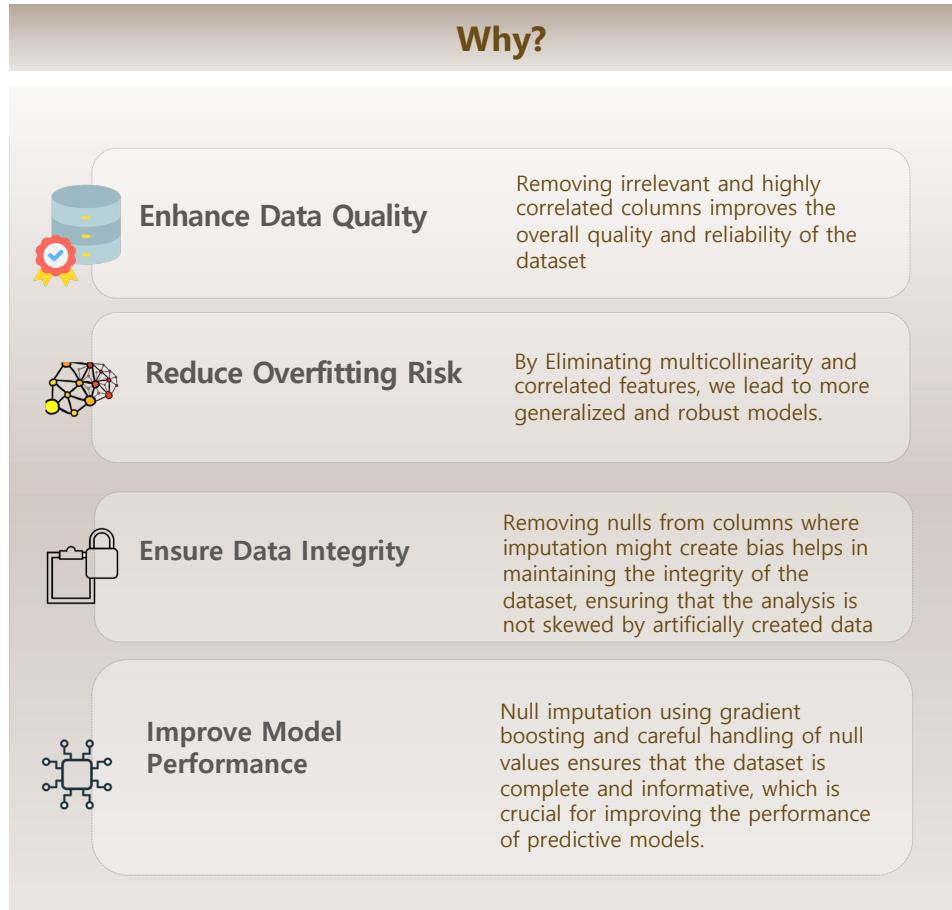




DATA MODIFICATION

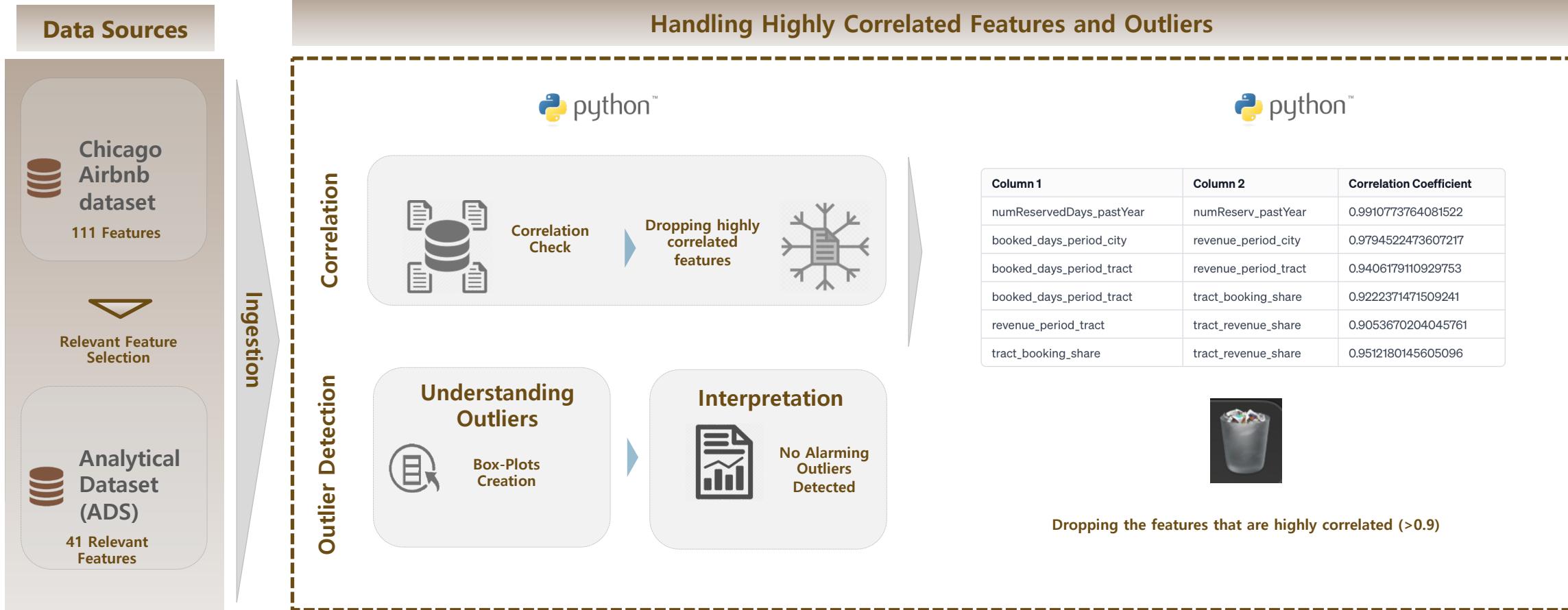
DATA PREPROCESSING

Data preprocessing enhances the quality and reliability of the data, paving the way for more accurate and meaningful insights



DATA PREPROCESSING

Layer 1 – Feature Selection & Data Ingestion | Removal of irrelevant features, to increase reliability



DATA PREPROCESSING

Layer 2 – Handling Missing Values | A significant proportion of the dataset contains nulls, hence, making this layer significant



Handling Missing Values

Column name	Missing values	Missing percentage
Property_Type	1	0%
Minimum_Stay	3	0%
Number_of_Photos	6	0%
Max_Guests	8	0%
Bedrooms	11	0%
Number_of_Reviews	18	0%
Bathrooms	47	0%
Neighborhood	1146	1%
numReserv_pastYear	3514	3%
hostResponseAverage_pastYear	7366	6%
Rating_Overall	14776	12%
available_days	15027	12%
available_days_aveListedPrice	15027	12%
numReviews_pastYear	18798	16%
numCancel_pastYear	18798	16%
rating_ave_pastYear	19845	17%
prop_5_StarReviews_pastYear	19845	17%
Cleaning_Fee(USD)	24158	20%
booked_days	38108	32%
booked_days_avePrice	38108	32%
revenue	38108	32%
occupancy_rate	38108	32%

Dropping Null records with low proportion

Dropping Null Records where Prediction/Imputation can potentially bias the dataset

Nulls remaining

	hostresponseAverage_pastYear	781
	Cleaning Fee (USD)	9419

Gradient Boosting | Null Imputation

```
# List of columns to impute
columns_to_impute = ['hostResponseAverage_pastYear', 'Cleaning Fee (USD)']

# Initialize IterativeImputer with GradientBoostingRegressor
imputer = IterativeImputer(estimator=GradientBoostingRegressor(), max_iter=10, random_state=0)

# Fit and transform the data
data[columns_to_impute] = imputer.fit_transform(data[columns_to_impute])
```

Gradient boosting is preferred over mean imputation because –

- Iterative learning process
- Can model complex relationships in the data
- Captures non-linear relationships between features
- Generally robust to outliers

Dropping Null Records

Imputing nulls using Gradient Boosting

DATA PREPROCESSING

Layer 3 – Multicollinearity Check | Removal of multicollinear features to increase the model accuracy and to mitigate the issue with correlated predictor variables



Handling Multicollinearity

Column name	VIF
rating_ave_pastYear	474.074412
hostResponseAverage_pastYear	243.463384
prop_5_StarReviews_pastYear	87.324976
available_days	17.67064
superhost_period_all	12.431234
Max Guests	10.522986
Bedrooms	10.367246
booked_days_avePrice	10.173246
Nightly Rate	8.917653
occupancy_rate	6.964972
booked_days	6.762309
Bathrooms	5.986203
Cleaning Fee (USD)	5.866413
tract_superhosts_ratio	5.731292
available_days_aveListedPrice	5.330595
Number of Photos	3.998245
Superhost	2.815586
numReviews_pastYear	2.490808
numReserv_pastYear	2.10543
numCancel_pastYear	1.324285
Minimum Stay	1.117636

Removing
rating_ave_pastYear feature to
account
multicollinearity

Column name	VIF
hostResponseAverage_pastYear	56.634033
prop_5_StarReviews_pastYear	35.083353
available_days	16.723341
superhost_period_all	12.426382
Max Guests	10.520079
Bedrooms	10.366989
booked_days_avePrice	10.171421
Nightly Rate	8.906295
occupancy_rate	6.772882
booked_days	6.660461
Bathrooms	5.970983
Cleaning Fee (USD)	5.866296
tract_superhosts_ratio	5.719883
available_days_aveListedPrice	5.330454
Number of Photos	3.990075
Superhost	2.704392
numReviews_pastYear	2.488375
numReserv_pastYear	2.105238
numCancel_pastYear	1.323316
Minimum Stay	1.117237



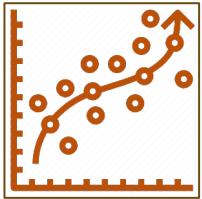
DATA MODELING

DATA MODELING TO PREDICT Host | SUPERHost STATUS

Aid Hosts – A Predictive Model Identifying Host-Manageable Factors Influencing Superhost Recognition



Modeling Approach



- Chicago dataset starting from **period 5-20**
- Training data:** Period 5-14
- Testing data:** Period 15-19

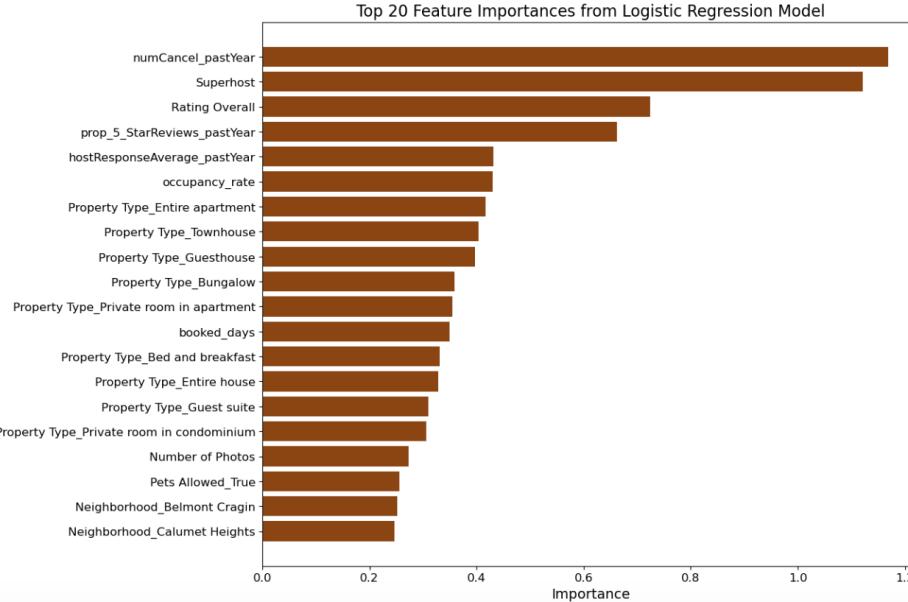
- Target variable:** Next period **Superhost status**
- Algorithm used: **Logistic Regression**
- Probabilities used for each class in **Difference in Difference Model**

Result

Predictions made on period 20 to identify potential Superhosts in period 21 and factors influencing it

Model Performance

	precision	recall	f1-score
0.0	0.88	0.79	0.84
1.0	0.80	0.89	0.84
accuracy	0.84		



Results & Insights



For Period 21, the predictions say that 79% of properties that don't allow pets will have a Superhost status



Lakeview neighbourhood is set to see the greatest number of Superhosts



Hosts of Studio apartments are most likely to gain/remain Superhosts in the next period



Having 10-20 photos of the property boost the chances of becoming a Superhost

DATA MODELING | PREDICT OCCUPANCY RATE FOR HOST STATUS

Aid Hosts and the Platform - Forecast the Occupancy Rate for the Upcoming Period to Assist Hosts in Effective Planning to Achieve or Sustain their superhost status



Modeling Approach



Dataset

- Chicago dataset starting from period 5-20
- Training data: Period 5-20 with predicted Next Period Superhost status as an additional parameter
- Testing data: Period 20

- Target variable: Occupancy rate
- Algorithm used: Random Forest

Result

Predicted occupancy rate for Period 21

Performance Evaluation

	MSE	MAPE
Validation	0.0004	1.89%
Test	0.001	3.80%

Alternate and more robust model

- We made an alternate more robust model based on rolling forecast model
- This resulted in an Improvement on MAPE

	MAPE
Validate	0.89%
Test	1.38%

Results & Insights

Increase in occupancy rate for hosts predicted to be Superhosts in period 21 by 8%

DATA MODELING | DIFFERENCE IN DIFFERENCES ANALYSIS



Aid Hosts - Becoming a super host is associated with an average increase of about 0.79% in occupancy rate compared to the control group (where the host stays with the Host status)

Treatment Group – The Hosts whose Predicted Status is Superhost with the associated Predicted Probability less than 0.7

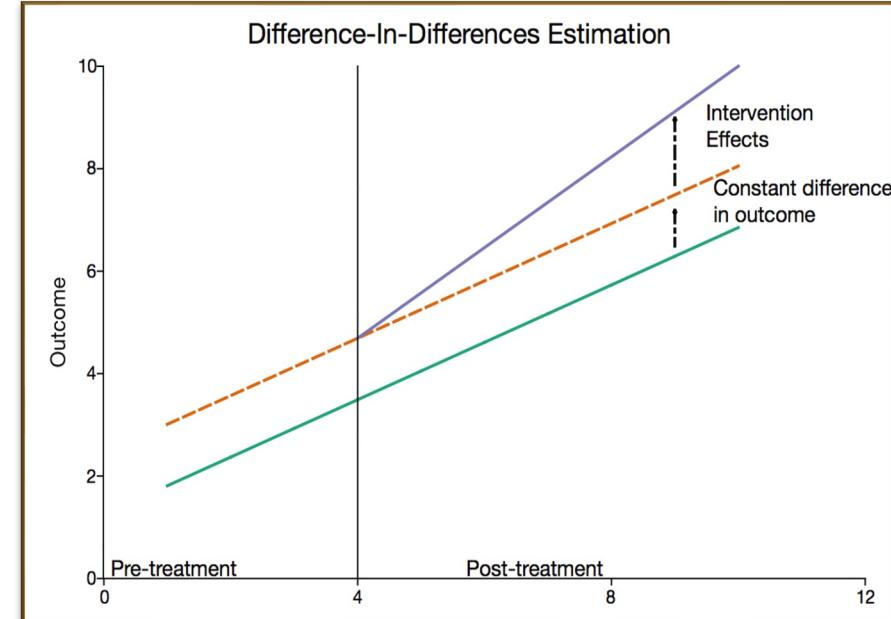
Control Group – The Hosts whose Predicted Status is Host with the associated Predicted Probability greater than 0.3

Treatment timeline:

Pre-Treatment – Period 19

Post-Treatment – Period 20

Average Occupancy Rate		
Group/Timeline	Pre-Treatment	Post-Treatment
Treatment Group	0.145238	0.182302
Control Group	0.135956	0.165151



The formula for calculating the Difference in Differences (DiD) is:

$$\text{DiD} = (\text{Treatment Group Post-Treatment} - \text{Treatment Group Pre-Treatment}) - (\text{Control Group Post-Treatment} - \text{Control Group Pre-Treatment})$$

Implemented a Difference in Differences approach to calculate the treatment effect of becoming a super host

DATA ANALYSIS TO IDENTIFY COMPETITION | HERFINDAHL INDEX

Aid Investors | A common measure of market concentration used to determine market competitiveness for potential investors in properties



$$HHI = s_1^2 + s_2^2 + s_3^2 + \dots + s_n^2$$

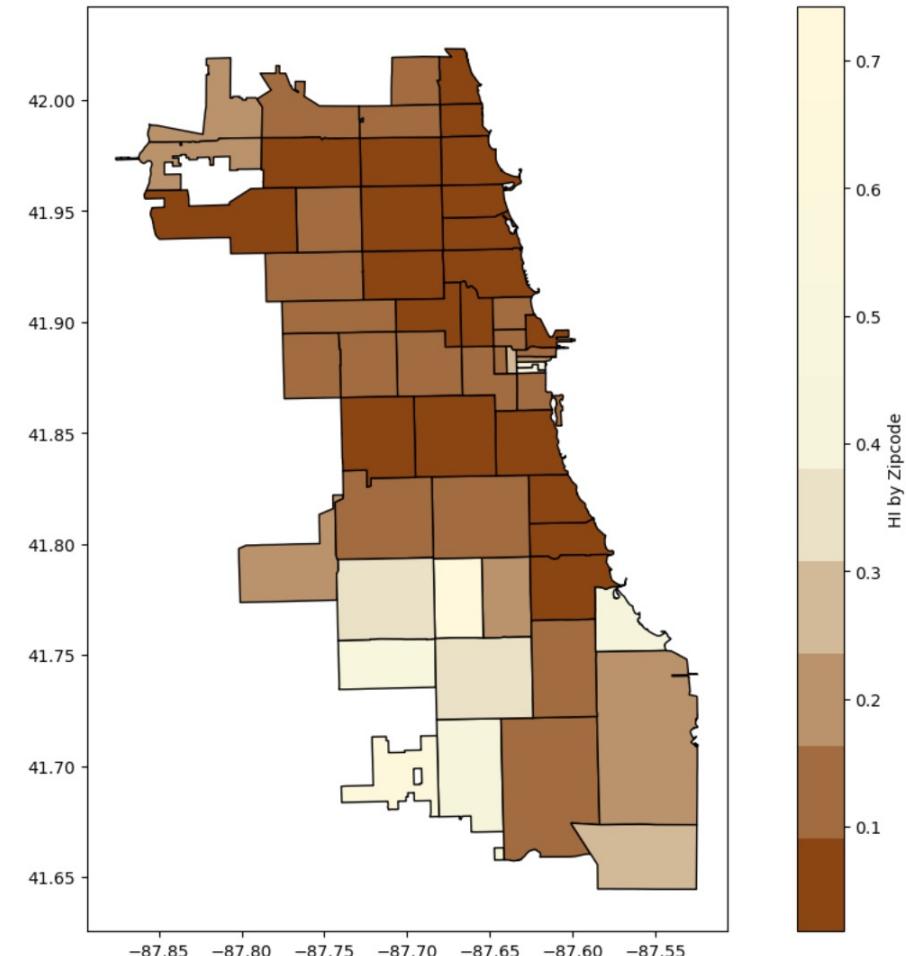
where: s_n = the market share percentage of property

- Higher the Herfindahl Index, Lower the competition
- Here revenue for each property is used to calculate the market share at a zip code level

Most competitive neighborhoods in Chicago

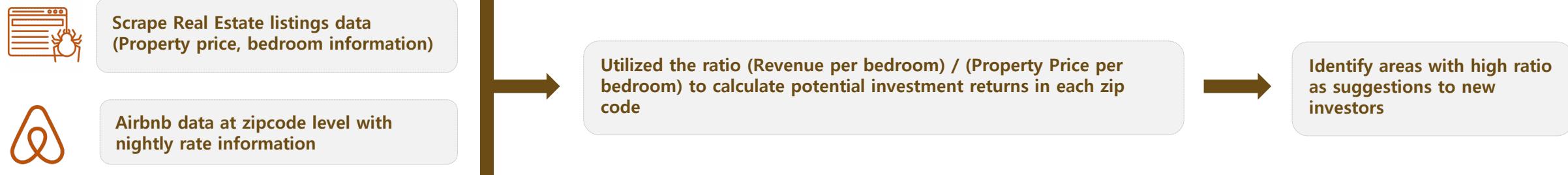
Zipcode	Herfindahl Index	Neighborhoods
60647	0.018	Logan square
60608	0.020	Pilsen
60622	0.029	West Park
60657	0.030	Lakeview
60614	0.031	Lincoln park
60618	0.034	Irving Park
60642	0.036	Goose Island
60653	0.037	Bronzeville
60660	0.041	West Edgewater
60611	0.048	Magnificent Mile
60625	0.049	Lincoln square
60637	0.049	West Woodlawn

Dark shades represent higher competition



DATA ANALYSIS TO IDENTIFY POTENTIAL INVESTMENTS

Aid Investors | Exploring Investment Opportunities in Airbnb Properties through Property Prices and Revenue Analysis to Identify Profitable Areas for New Investments



Potential ROI for most competitive neighborhoods in Chicago

Zipcode	Returns Ratio	Neighborhoods
60642	2.58	Goose Island
60660	5.44	West Edgewater
60625	3.34	Lincoln square
60637	7.09	West Woodlawn





BUSINESS INSIGHTS & WAY FORWARD

BUSINESS INSIGHTS & RECOMMENDATIONS

Comprehensive Business Insights to deal with real-time challenges experienced by all stakeholders in a platform-based business model



- **Superhost Promotion:** Airbnb should continue to promote the Superhost program, as it clearly influences quality standards and guest satisfaction (higher occupancy rate)
- **Support for New Hosts:** Provide resources or incentives for new hosts to become Superhosts, especially in areas with fewer Superhosts.
- **Neighborhood Highlighting:** Feature neighborhoods like Lakeview in marketing campaigns to showcase areas with high concentrations of Superhosts
- **Encourage Comprehensive Listings:** Educate hosts about the importance of thorough and attractive property listings, emphasizing the impact of having enough photos.



- **Location Focus:** Hosts in the Lakeview neighborhood should leverage their location's growing reputation for quality stays. Those outside Lakeview might consider strategies to compete or highlight unique aspects of their neighborhood.
- **Property Type:** Those with or considering investing in studio apartments should prioritize these for hosting, as they have a higher likelihood of achieving Superhost status.
- **Photo Strategy:** Hosts should aim to have 10-20 high-quality photos of their property. This not only boosts chances of becoming a Superhost but also enhances listing attractiveness.
- **Occupancy Rate Goals:** Understand that while becoming a Superhost is beneficial, the average increase in occupancy rate is modest (0.79%). Focus on other factors like pricing, amenities, and guest experience to increase occupancy.



- **Neighborhood Selection:** Guests looking for high-quality stays should consider the Lakeview neighborhood, as it's predicted to have a significant number of Superhosts.
- **Studio Apartments for Quality Assurance:** If you're a solo traveler or a couple, opting for studio apartments might offer a better chance of having a Superhost, potentially enhancing your stay experience.



- **Neighborhood Selection:** Investing in the following neighborhoods will maximize ROI despite the high competition if you maintain good service level

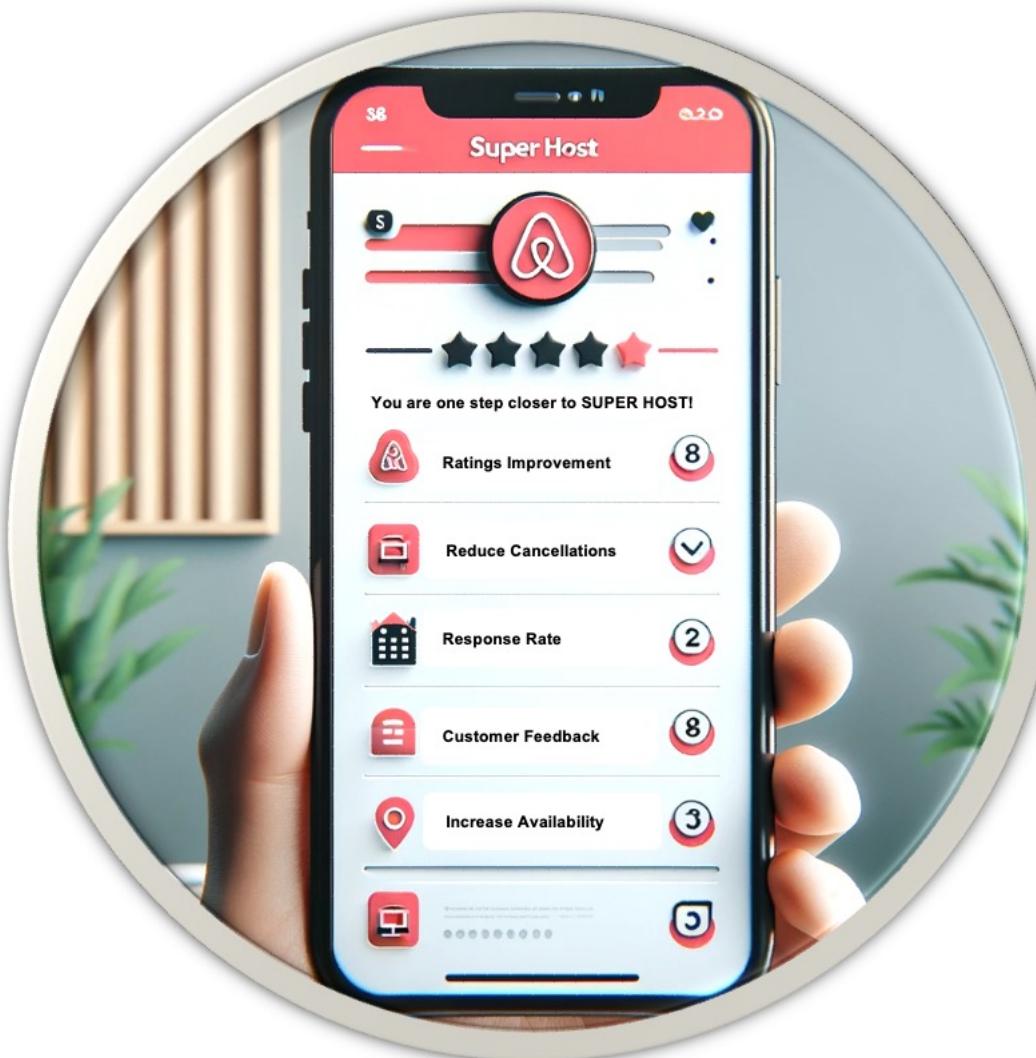
RECOMMENDATION – APP BASED TRACKER

Implement a tailored notification system to guide hosts with actionable insights & personalized advice, enhancing their journey towards becoming Superhosts



Benefits to Airbnb

- Enhanced Listing Quality
- Increased Customer Satisfaction
- Higher Retention Rates
- Positive Public Relations
- Brand Quality Commitment



Benefits to Hosts

- Clear path to improvement
- Personalized Recommendations
- Increased Booking Potential
- Real-Time Feedback
- Goal Setting & Tracking
- Enhanced Guest Experience



FUTURE SCOPE

KEY LEARNINGS & FUTURE SCOPE

Untapped Opportunities and Potential Avenues for Deeper Analysis in the future



Customer data – Availability of customer information would help solve problems for customers by customizing and tailoring their experience

Competition Analysis - Differentiate the brand through unique offerings (like Airbnb Experiences), focus on customer loyalty programs, and analyze market trends to stay competitive



Clean Data – The presence of nulls in the data leads to potential data loss as imputation does not always depict the exact picture



Sustainable and Responsible Tourism – Recommend ideas to Airbnb and Hosts for a sustainable and eco-friendly stay experience

Pricing and Revenue Management: Implement dynamic pricing algorithms based on demand and seasonality, offer competitive pricing strategies to the hosts





THANK YOU!



REFERENCES

REFERENCES

References to the Articles and Portals used for our analysis and research



How-to • Host

Tracking your Superhost status

Hello, superstar! If you have achieved Superhost status, it's important to make sure your account stays in good standing.

We'll let you know by email if you gained, retained, or lost Superhost status. Since assessments happen throughout the 12-month assessment period, we recommend you track how you are doing.

[Link to the Article](#)



House rentals in Chicago
Find and book unique houses on Airbnb

LOCATION
Chicago, IL

CHECK IN
Add Date

CHECK OUT
Add Date

Search

A large image shows a group of people sitting at a table on a rooftop deck overlooking a city skyline.

[Link to the Article](#)

Home > All topics > About hosting > Ways to host > Superhosts > How to become a Superhost

How-to

How to become a Superhost

Our top-performing Hosts are called Superhosts. Their listings feature a special badge that lets everyone know they're extra awesome.

In this article

- [How it works](#)
- [Requirements](#)

[Link to the Article](#)

EXPERIENCE

Airbnb [Add Topic +](#)

This Airbnb superhost was named Florida's most hospitable host with 205 perfect reviews



Brandon Girod
Pensacola News-Journal

Published 12:22 p.m. ET July 20, 2023

[Facebook](#) [Twitter](#) [Email](#) [More](#)



[Link to the Article](#)

Airbnb your place

AIRBNB PROPERTY ANALYSIS: FIND OUT IF YOU'LL TURN A PROFIT IN 10 EASY STEPS
Last updated October 17, 2023

Find the best places to invest

Enter an address, neighborhood, city or ZIP code [Start Analyzing](#)

Want to know how much your space is worth on Airbnb?

Avg Monthly Revenue	\$3474
Occupancy rate	55%
Market Avg Daily Rate	\$280

CATEGORIES

> Agent Resource Center

> Analysis

[Link to the Article](#)



APPENDIX