AirBnb Analysis

AI for Business Final Project

Group 8 : Shin Yu Wang, Elsa Chiang, Tzu Yun Huang

Agenda

1. **Background**
2. **Dataset Overview**
3. **Features Exploration**
4. **Occupancy Rate Prediction Model**
5. **Summary**

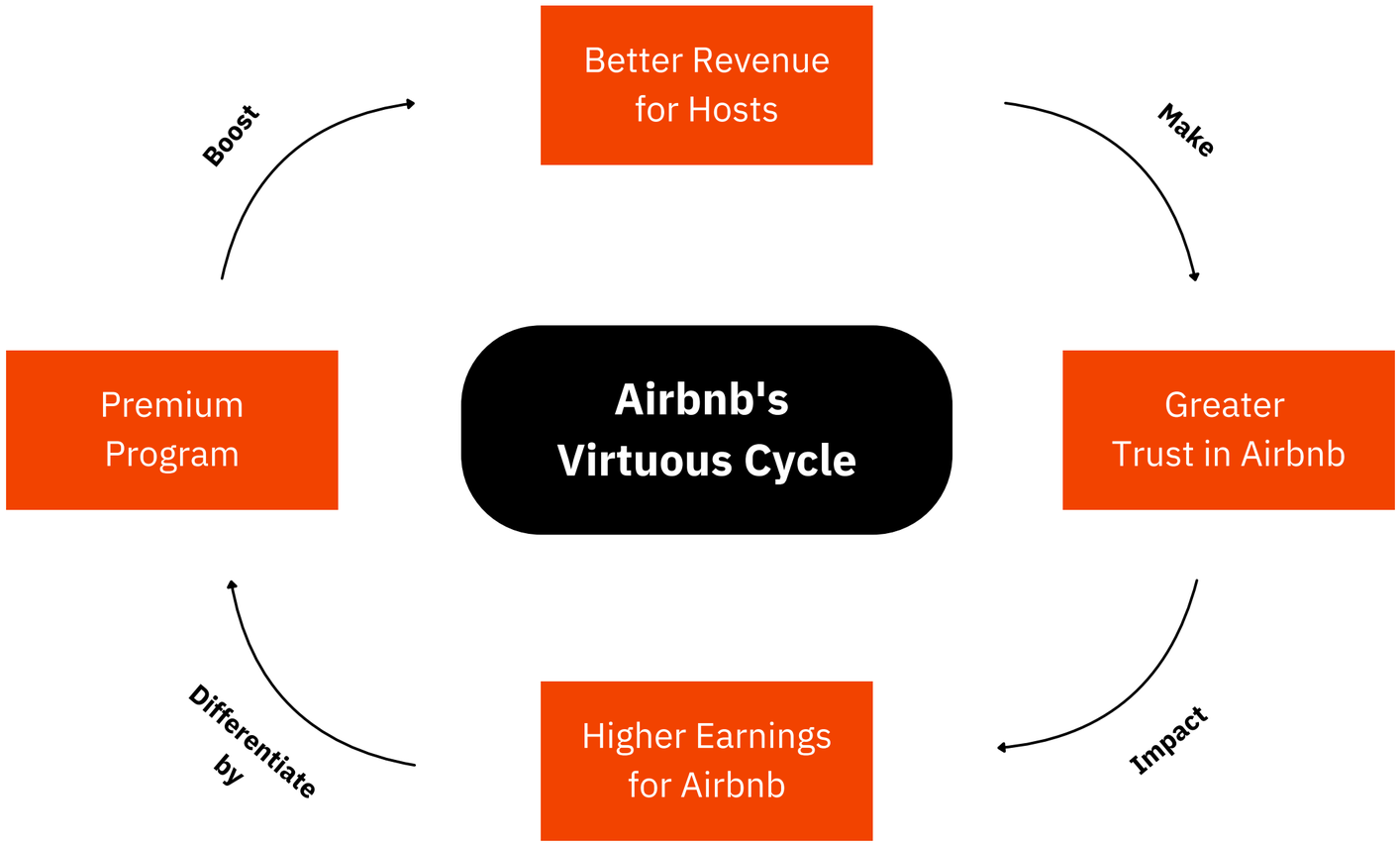
Background

Business Problem: Driving Revenue Growth for Airbnb and Clients

- Need to stand out in a competitive marketplace.
- Airbnb's Revenue Growth is tied to Clients' Success.
- Introducing a Premium Program

BACK TO AGENDA PAGE

Why We Increase Client Revenue:



airbnb

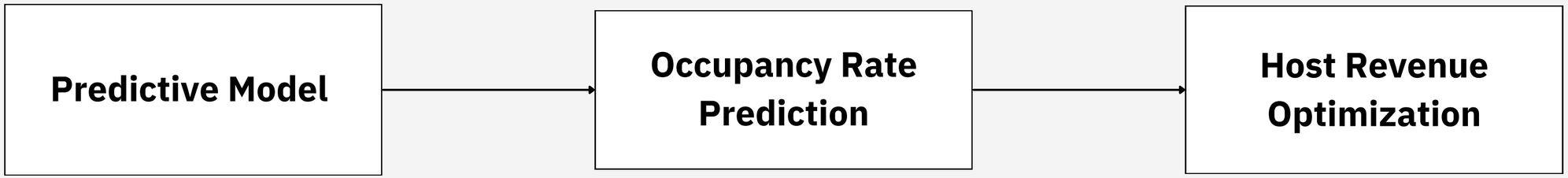
BACK TO AGENDA PAGE

Background

Revenue Formula in AirBnb's:

$$\text{Revenue} \uparrow\downarrow = \text{Nightly Rate} \times \underline{\text{Occupancy Rate}} \uparrow\downarrow$$

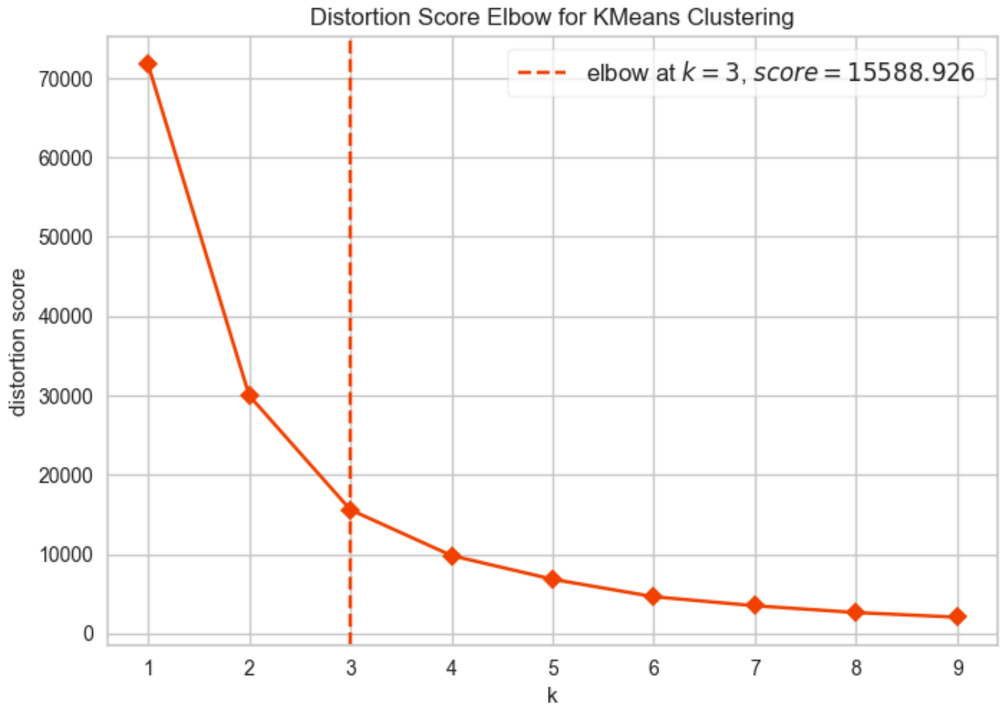
How We Increase Client Revenue:



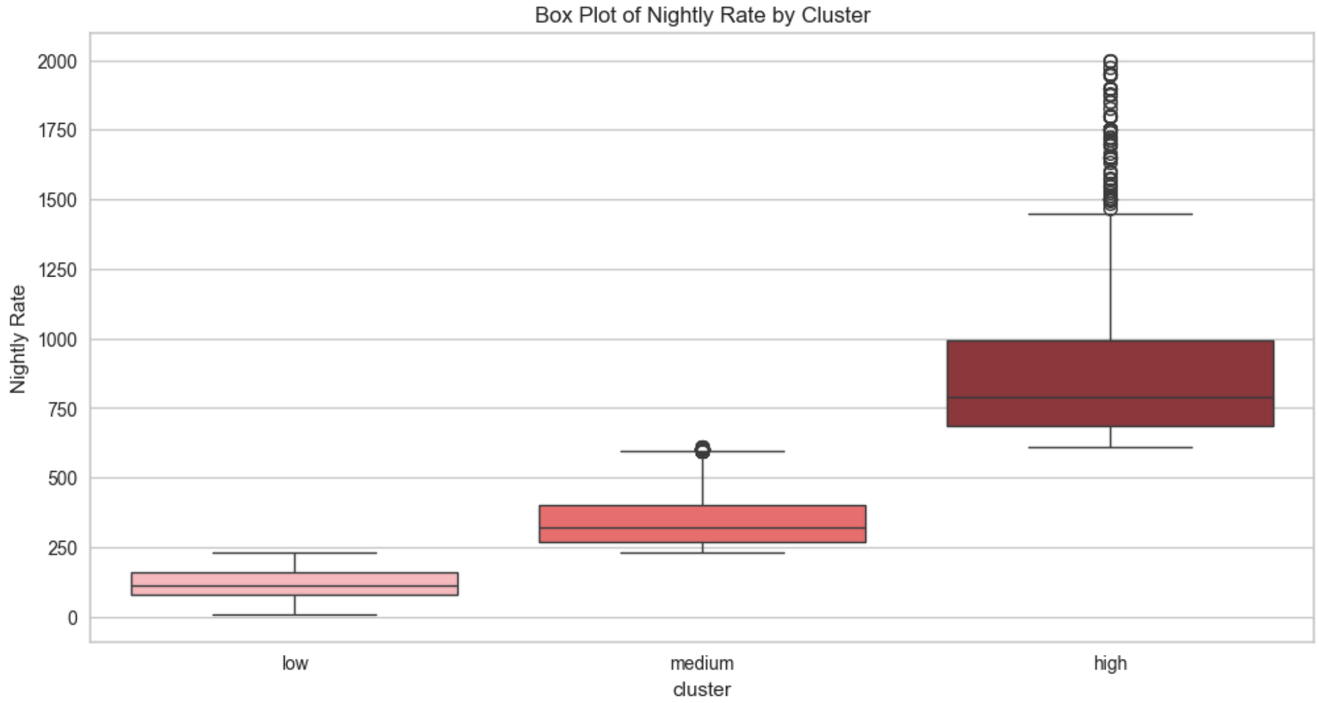
Dataset Overview

Clustering Airbnb Listings by Nightly Rates

Highlight the elbow point at k=3, showing the optimal cluster count.



Average nightly rate by three clusters, including low, medium, and high.



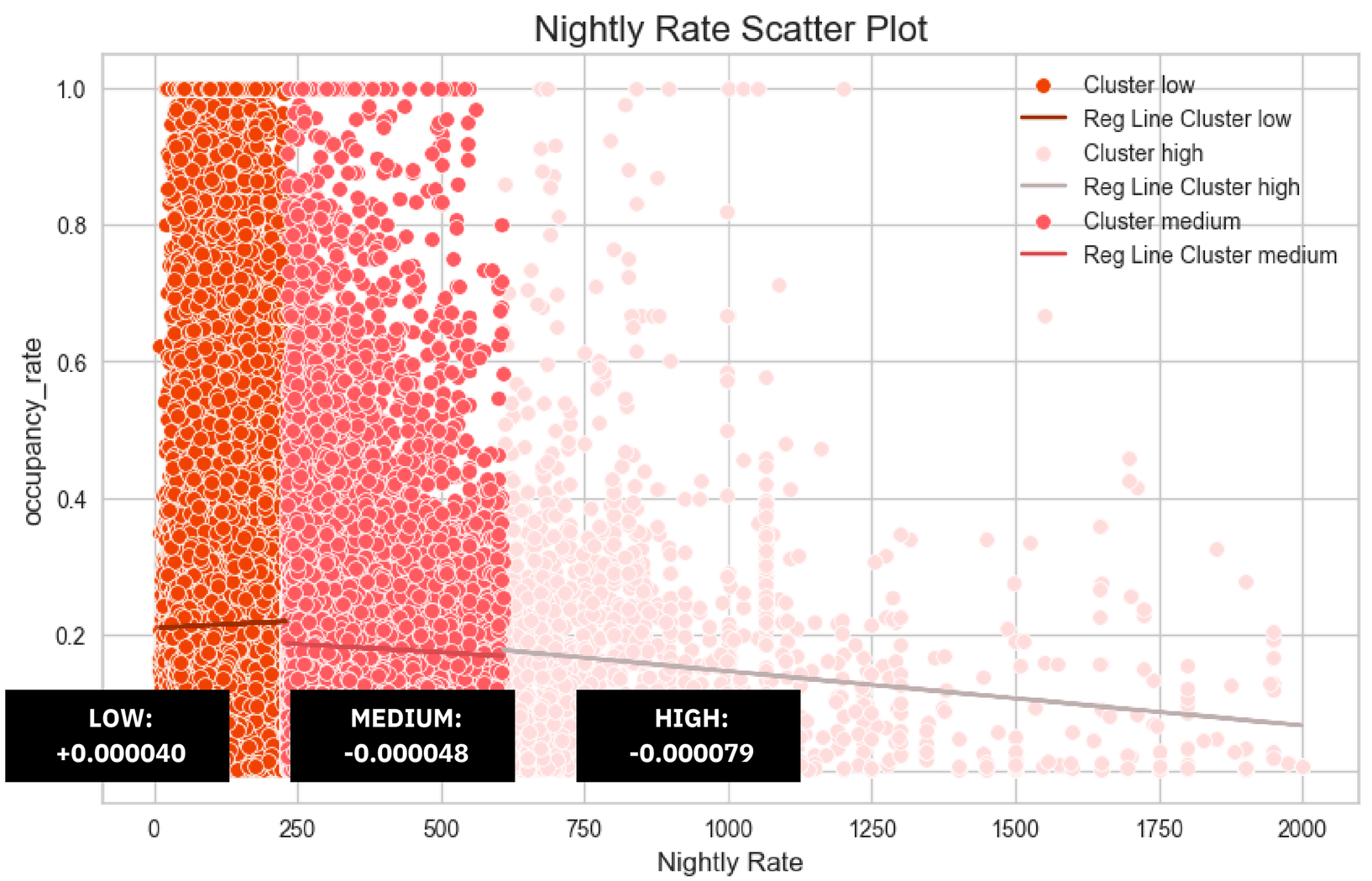
LOW: $119

MEDIUM: $342

HIGH: $873

Dataset Overview

Understanding Price Sensitivity in Clusters



BACK TO AGENDA PAGE

Features **Exploration**

Significant Variables to Occupancy Rate

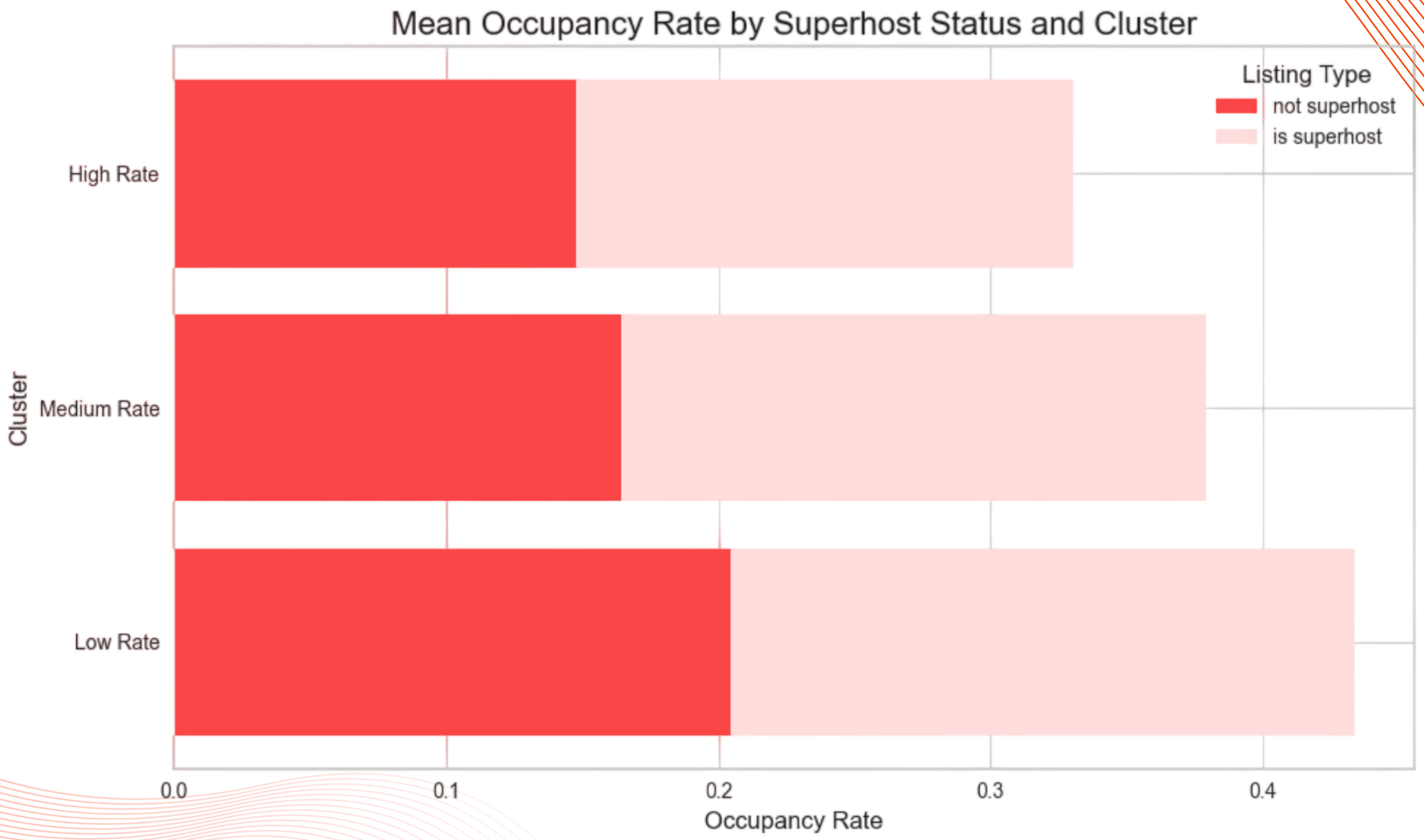
- Use linear regression to identify which variables significantly impact the occupancy rate

Rate Cluster	Low	Medium	High
Positive features	• booked_days 0.00681 • rating_ave_pastYear 0.00526 • Max Guests 0.00135 • rev_available_days 0.00037 • Number of Reviews 0.00019 • Rating Overall 0.00013	• tract_booking_share 2.10435 • tract_superhosts_ratio 0.08094 • booked_days 0.00602 • zip_black_nothispanic_percent 0.00117 • tract_prev_superhosts 0.00083 • hostResponseNumber 0.00011	• Pets Allowed 0.01554 • booked_days 0.00655 • numCancel_pastYear 0.00141 • prev_numReserv_pastYear 0.00014 • numReservedDays_pastYear 0.00004
Negative features	• Pets Allowed -0.00594 • available_days -0.00193 • Number of Photos -0.00026 • Number of Reviews -0.00021 • Cleaning Fee (USD) -0.00011 • booked_days_avePrice -0.00007	• Pets Allowed -0.00808 • available_days -0.00163 • tract_superhosts -0.00074 • Cleaning Fee (USD) -0.00007 • booked_days_period_tract -0.00002 • numReserv_pastYear -0.00001	• available_days -0.00155 • tract_unique_prices -0.00097 • prev_time_to_date_mean -0.00027 • numReserv_pastYear -0.00011 • numReservedDays_pastYear -0.00005 • revenue -0.000001 -0.00001

Features **Exploration**

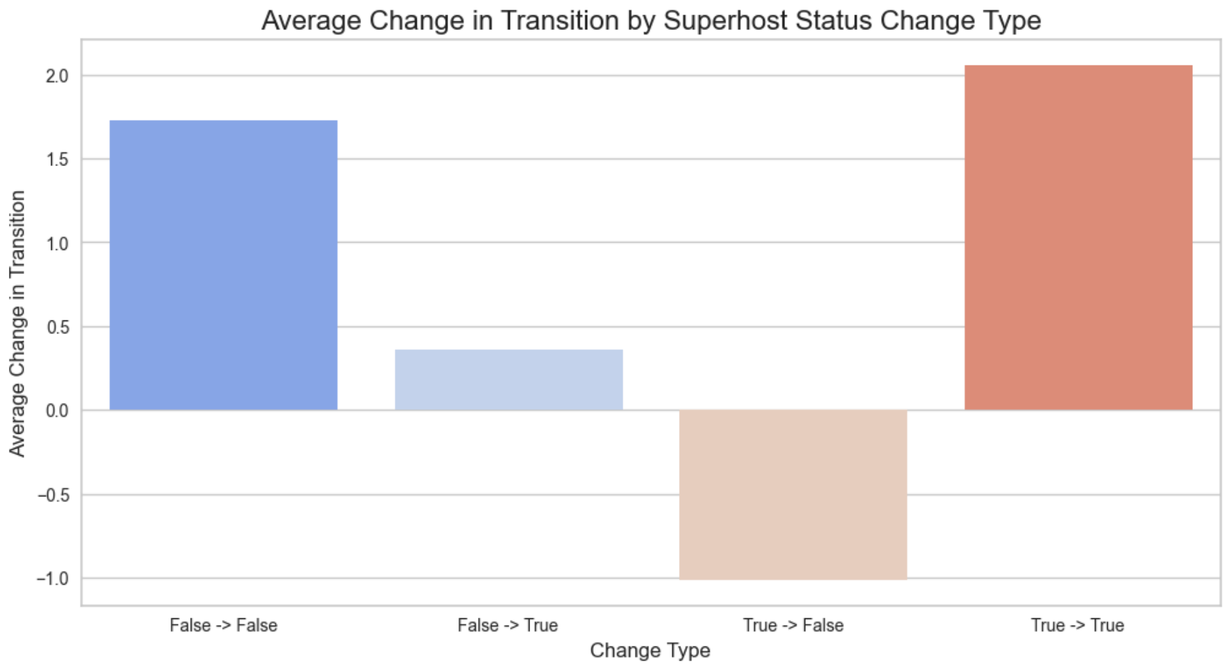
Variables That Impact Super Host

- Showing the average impact on occupancy rates of host with or without a superhost



Features **Exploration**

Superhost Transition



Examine how superhost status impacts occupancy rate

- Label
  - X: Superhost status changes
  - Y: Mean of changes in occupancy rate

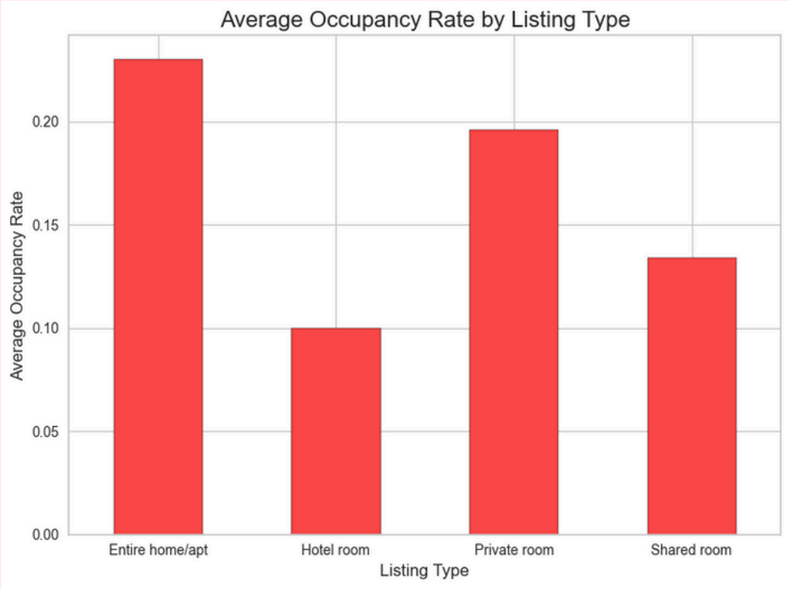
- Insight
  - Getting superhost status (False → True) leads to increased occupancy.
  - Losing superhost status (True → False) negatively impacts occupancy.

Features **Exploration**

Listing type

Low

a. **Entire home/apt**
Performs best and can be promoted more.
b. **Hotel room, Shared room**
Strategies need to be optimized to attract guests.

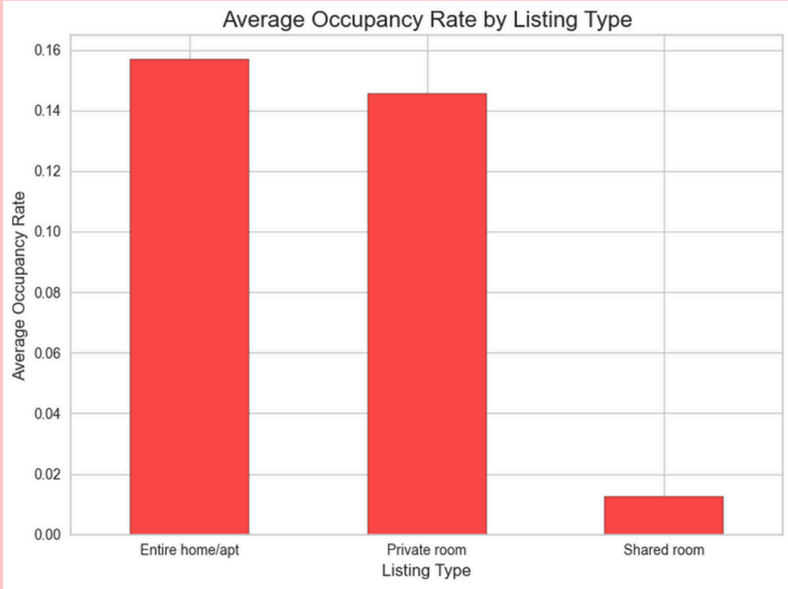


Medium

a. **Hotel room**
Best performing, first choice for mid-priced guests.
b. **Private room**
The weakest performance, needs to improve its attractiveness.

High

a. **Entire home/apt, Private room**
Performance is similar.
b. **Shared room**
almost no need and needs to be repositioned.



Features **Exploration**

Variable Selection

- **Dummy variables** that significantly influence the occupancy rate

Low
- host_is_superhost_in_period
- prev_host_is_superhost
- superhost_change
- Pets Allowed
- Instantbook Enabled

Medium
- host_is_superhost_in_period
- prev_host_is_superhost
- superhost_change

High
- host_is_superhost_in_period
- Pets Allowed
- Instantbook Enabled

- Identify significant **interaction terms**:
  a. Create interaction features based on correlations exceeding 0.8
  b. Feature selection using Lasso

Occupancy Rate **Prediction Model**

Gradient Boosting Tree

High predictive power

It provides better predictive accuracy.

Flexibility

It can be adapted for regression, classification, and ranking tasks.

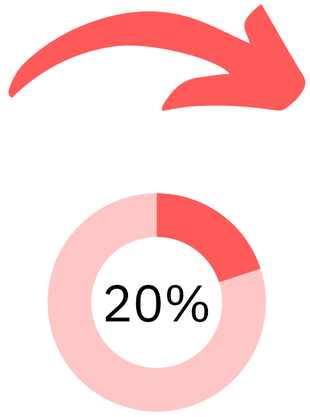
Suitability for moderate dataset size

It performs well on datasets of moderate size.

Model **Performance**

Gradient Boosting Tree

Input
All significant variables
• Low: 54 • Medium: 43 • High: 32



test dataset

Test Result	
Root Mean Squared Error	**R-squared**
• low: 0.0091 • Medium: 0.0093 • High: 0.0305	• Low: 0.9977 • Medium: 0.9968 • High: 0.9591

Premium Program

Benefits

Driving Revenue Growth for Customers

Occupancy Rate Prediction

Maximize hosts’ profits, regardless of their price segment.

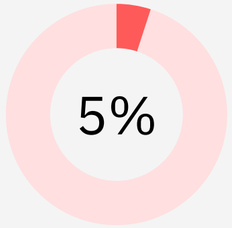
Customized Suggestions

Friendly for new hosts while also benefiting existing hosts by enhancing their service quality.

Premium Program

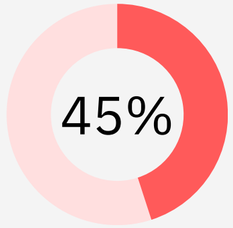
Customized suggestions

Assume you were a low-performing host, Eric:



Before Adjustment

Superhost = 0
Pets Allowed = False
available_days = 245
booked_days = 11



After Adjustment

Superhost = 1
Pets Allowed = True
available_days = 300
booked_days = 100