****

**Austin Airbnb Market**

6025 Predictive Analytics Final Project

Hewei Shen Macon Barker

# Business Problem

Our client has tasked us with analyzing Airbnb listings in hopes of entering the Austin market for short-term rentals. The ultimate goal is to recommend the features that may improve guest reviews and occupancy, as well as see what factors into pricing.

# Data Overview

The provided data set contained details about nearly 6000 Airbnb listings in the Austin area. There were 5 main categories:

* **Host information** - Is the host a super host? What is their response time?
* **House information** - descriptions, number of bedrooms, house type, number accommodated, etc.
* **Availability** - number of days property available over the next 30, 60, 90, and 365 days
* **Prices** - nightly rate, cleaning cost, security deposit, cost per additional guest
* **Ratings** - overall score, accuracy of listing, cleanliness of the property, ease of check-in, etc.

# Methodology

Below is a brief description of the techniques used to help answer our questions of interest.

* **Correlation and Comparative Charts** - help us see relationships between different variables
* **Multiple linear regression** - uses linear relationships to help predict a target variable
* **Decision Tree** - a classification method used to make a binary by determining which features are most important and grouping based on similar attributes
* **K-Means Clustering** - a method used to group similar data points. Clusters can then be further analyzed in different ways

# Findings

## Nightly Pricing

Rental price is an important factor for investment profit in Airbnb properties. We built a linear regression model that helps predict nightly rental rate for any given Airbnb property and detect the most influential metrics to the rental rate. Our regression model can explain 54.7% of the variability in nightly rental rate.

**Key takeaways**

* Houses have almost all the luxury options (over $2,000 nightly rental rate) than other property types, while Villas have the highest average price of $393
* Room type is one of the most important factors for rental rate. Properties of entire home/apt have a significantly higher price than private rooms or shared rooms
* The capacity of the property is the main determinant of rental rate, specifically the maximum number of people who can stay in the house, the number of bathrooms and bedrooms
* The significance of host information is ambiguous. Identified host and superhost status harm the rental price according to the model

## Keeping a Unit Booked

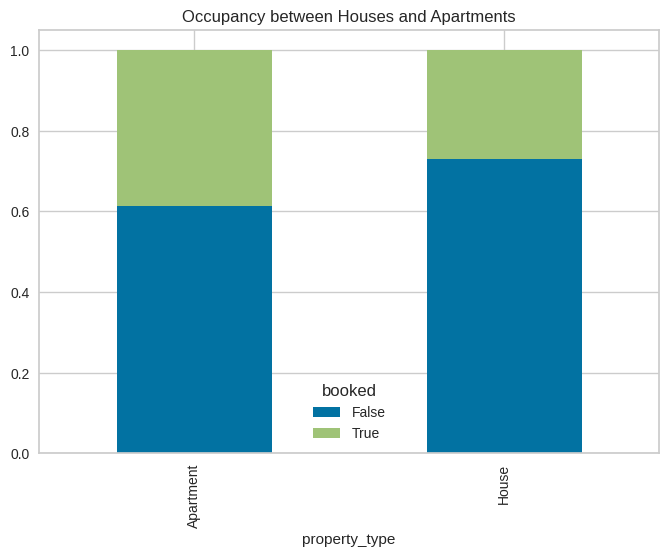
The driving factor behind revenue in the Airbnb industry is keeping a property occupied. To determine if a unit is booked, first, we expressed the available nights for the next 90 days as a percentage. If the availability is less than 40%, the unit is considered ‘booked’, otherwise it is ‘not booked’

To determine the key factors to keeping a property occupied, we used the Decision Tree method. This technique will let us know specific features to focus on and can also help us predict if a given property will be ‘booked’ over the next 90 days. Our model had an accuracy of 67.33%, which is the percentage of the time that the correct status was predicted. The precision was 45.50%, which tells us the percentage of properties predicted as booked, that were actually booked.

After analyzing the output, we believe that the following features will assist in keeping a property booked

**Most important factors**

* **Price** - while important, price ranges vary depending on other features offered
* **Cleaning Fee** - listings with higher fees have low occupancy rates
* **Security Deposit** - requiring a security deposit may hurt occupancy rates
* **Property Type** - apartments are more likely to be booked compared to houses
* **Extra person cost** - additional charges for extra guests may lower the booked rate
* **Room type** - entire house/apartment tends to have higher booked rates



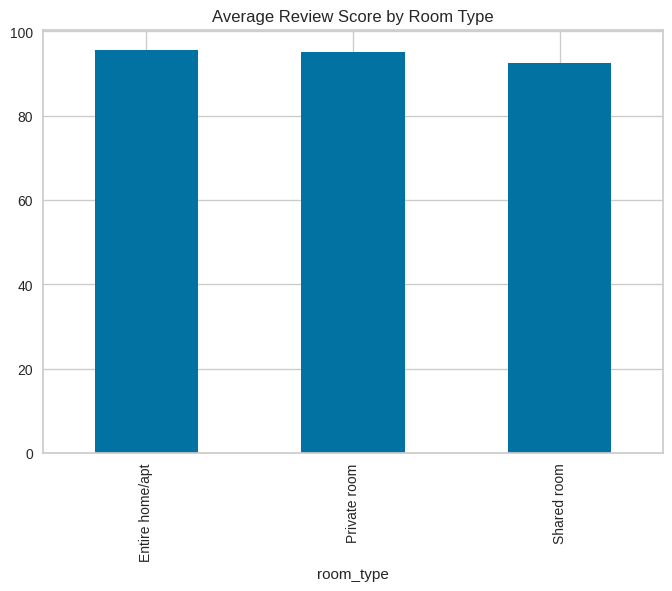
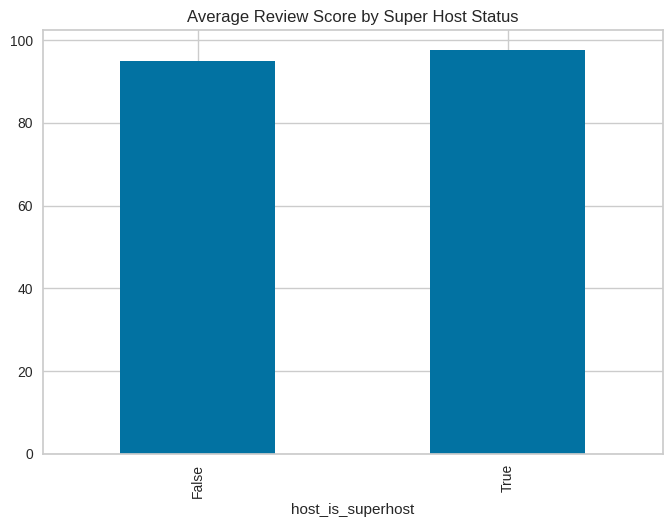
## Importance of Reviews and Ratings

An extensive analysis was done using linear correlation and comparative column charts to discover if there are any features of a listing that may improve customer reviews. Do listings with higher reviews have a higher occupancy rate? Is there an ideal number of bedrooms or bathrooms for a unit to have to achieve higher reviews?

While there was not a clear leading feature that drove review scores, we believe that focusing on the following may slightly improve guest reviews.

**Super Host Status**

Listings from super hosts had an average review score of 97.5 while non-superhosts had an average of 94.9. Some things to focus on to achieve super host status are responsiveness to guests and not canceling any reservations.



**Room Type**

There were three different room types in the data set: ‘Entire home/apt’, ‘Private room’, or ‘Shared room’. Listings with a shared room had the lowest average review score at 92.6, 3 points below the other 2 room types. This makes sense as customers sharing a room with another person may be more likely to have a negative experience.

## Unit Segmentation

We clustered the Austin Airbnb properties into three groups and for each group, we provided a profile.

|  |  |  |
| --- | --- | --- |
| **Large and Luxurious** | **Popular and Affordable** | **Single Room Travel** |
| * Properties with a median $500 rental rate and $63 per person * Mostly houses * Have more bedrooms and beds, and can accommodate 8 guests on average * Host has high number of listings on Airbnb * All the hosts have profile pictures and most have verified identity * Significantly high-security deposit and cleaning fee * Strict cancellation policy | * Properties with a median $165 rental rate and $43 per person * Extremely popular properties that are mostly booked within 90 days * More flexible or moderate cancellation policy * Have a high accuracy of the listing score | * Properties with a median $125 rental rate and $49 per person * Have nearly half of the room type as private room or shared room * Have a lot of reviews * Least popular properties that are mostly available within 90 days * More flexible or moderate cancellation policy |

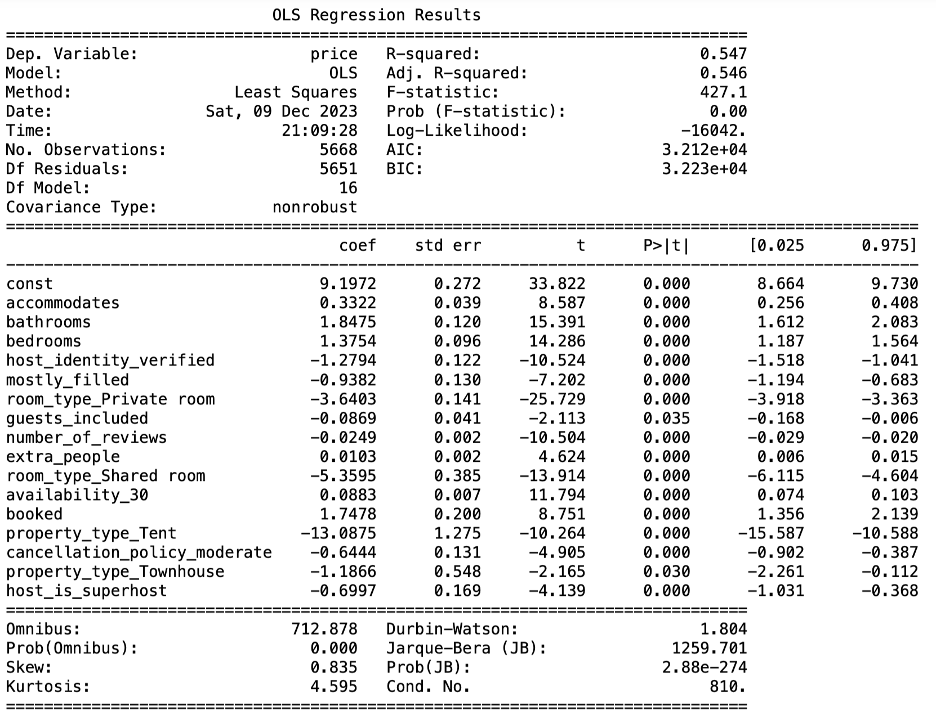
# Recommendations

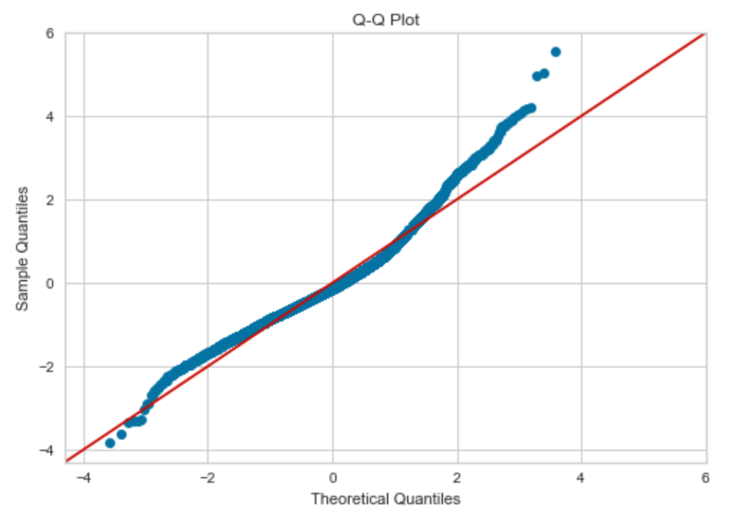
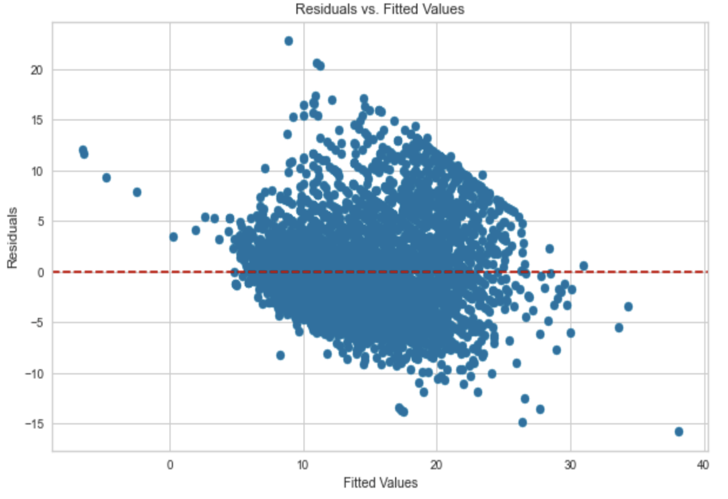
The real estate market is quite difficult to read and predict. However, based on the findings from our analysis, we recommend considering the following when deciding on a property and the features to include for a better chance of being successful in the Airbnb market.

* **Choose an apartment over a house.** While houses have more space available and tend to earn more revenue, apartments are more likely to stay booked.
* **Offer the entire home instead of just a room.** These listings have a much higher chance of staying booked and are likely to score higher reviews. Prices can be set higher as more space is available.
* **Keep cleaning fees and security deposits to a minimum.** High cleaning fees and security deposits may turn away potential customers.

# Appendix

## Linear regression - summary & residual plots





## Decision tree - Variable Importance

