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

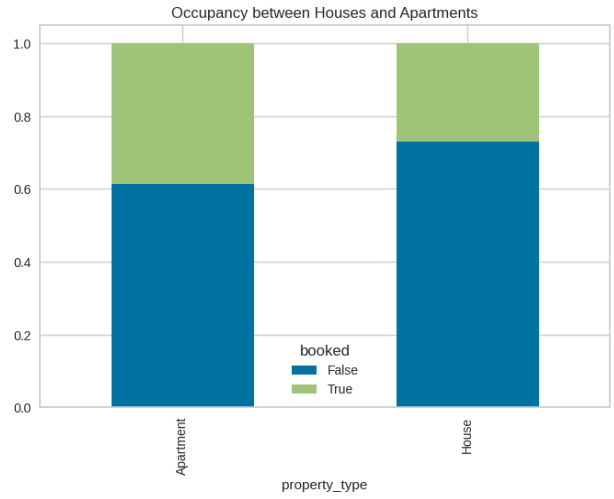
The driving factor behind revenue in the Airbnb industry is keeping a property occupied. To determine if a unit is booked, first, we expressed the available nights for the next 90 days as a percentage. If the availability is less than 40%, the unit is considered 'booked', otherwise it is 'not booked'

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



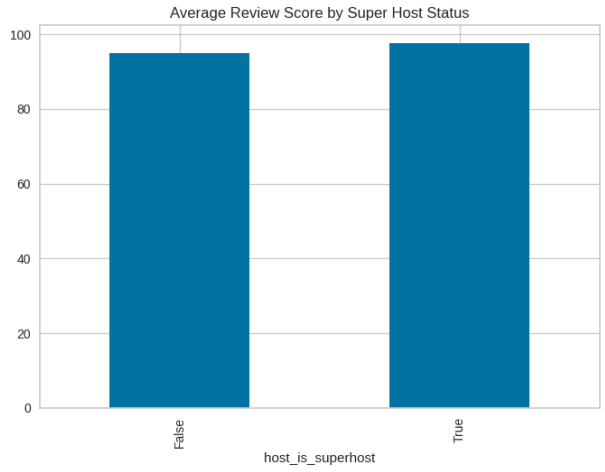
IMPORTANCE OF REVIEWS AND RATINGS

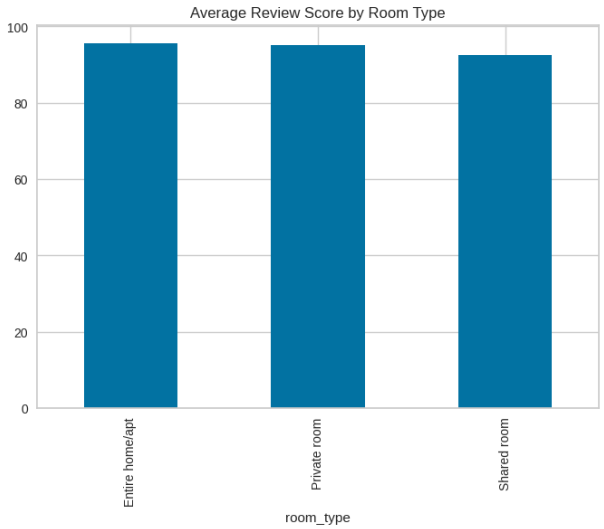
An extensive analysis was done using linear correlation and comparative column charts to discover if there are any features of a listing that may improve customer reviews. Do listings with higher reviews have a higher occupancy rate? Is there an ideal number of bedrooms or bathrooms for a unit to have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

- 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

Popular and Affordable

- 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

Single Room Travel

- 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

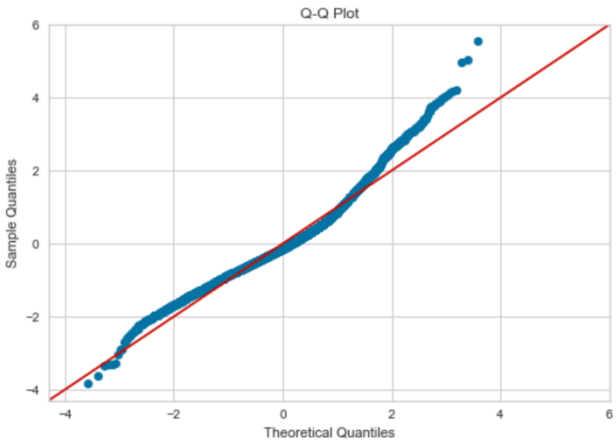
LINEAR REGRESSION - SUMMARY & RESIDUAL PLOTS

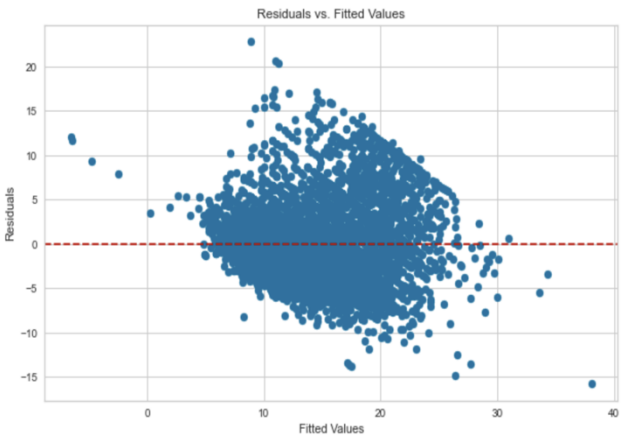
OLS Regression Results

Dep. Variable:	price	R-squared:	0.547
Model:	OLS	Adj. R-squared:	0.546
Method:	Least Squares	F-statistic:	427.1
Date:	Sat, 09 Dec 2023	Prob (F-statistic):	0.00
Time:	21:09:28	Log-Likelihood:	-16042.
No. Observations:	5668	AIC:	3.212e+04
Df Residuals:	5651	BIC:	3.223e+04
Df Model:	16		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	9.1972	0.272	33.822	0.000	8.664	9.730
accommodates	0.3322	0.039	8.587	0.000	0.256	0.408
bathrooms	1.8475	0.120	15.391	0.000	1.612	2.083
bedrooms	1.3754	0.096	14.286	0.000	1.187	1.564
host_identity_verified	-1.2794	0.122	-10.524	0.000	-1.518	-1.041
mostly_filled	-0.9382	0.130	-7.202	0.000	-1.194	-0.683
room_type_Private room	-3.6403	0.141	-25.729	0.000	-3.918	-3.363
guests_included	-0.0869	0.041	-2.113	0.035	-0.168	-0.006
number_of_reviews	-0.0249	0.002	-10.504	0.000	-0.029	-0.020
extra_people	0.0103	0.002	4.624	0.000	0.006	0.015
room_type_Shared room	-5.3595	0.385	-13.914	0.000	-6.115	-4.604
availability_30	0.0883	0.007	11.794	0.000	0.074	0.103
booked	1.7478	0.200	8.751	0.000	1.356	2.139
property_type_Tent	-13.0875	1.275	-10.264	0.000	-15.587	-10.588
cancellation_policy_moderate	-0.6444	0.131	-4.905	0.000	-0.902	-0.387
property_type_Townhouse	-1.1866	0.548	-2.165	0.030	-2.261	-0.112
host_is_superhost	-0.6997	0.169	-4.139	0.000	-1.031	-0.368

Omnibus:	712.878	Durbin-Watson:	1.804
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1259.701
Skew:	0.835	Prob(JB):	2.88e-274
Kurtosis:	4.595	Cond. No.	810.





DECISION TREE - VARIABLE IMPORTANCE