**Investing in Austin**

**Problem Statement:**

Short-term rental properties have emerged as a popular investment avenue for those seeking reliable and substantial returns. Even though these properties are booming around the US, what specific factors or characteristics make Austin, TX appealing to investors? We aim to find the significant factors that drive rental prices, keeping a unit booked, and positive reviews and ratings. As a secondary analysis, we will be segmenting the Austin properties into groups to gain a better understanding of which characteristics of a property in Austin make it a success.

**Data Utilization**

The data set that was used for this analysis contained over 6,000 Airbnb rental units in the Austin area. For each property, excluding null values, there is a range of variables including nightly price, cleaning fee, accommodations, review scores, etc. Although having this much data is great, it certainly does not come clean. We identified four different methods to clean the data; Removing columns with high null percentages, replacing null values with substantial substitutes, transforming text columns into binary “Yes” or “No” (To show if it was included by the host), and then finally removing the remaining null values in the data set. After cleaning, there were 3,351 out of the original 6,000, which is more than acceptable for a real-world data set.

**Analytical Techniques**

Throughout the analysis, we used three different modeling techniques to extract the best possible outcomes. Each modeling technique has its strengths and weaknesses, so we used that to our advantage. The three modeling techniques used were linear regression, classification tree (logistic regression), and K-means clustering. The purpose of using linear regression was to be able to find the important factors that had a significant impact on the nightly rental price. Since we are dealing with mostly numeric features, linear regression was a perfect choice. The use of classification trees became more useful when dealing with binary variables. Determining what influences a unit to be booked was answered using this method. Classification trees are also an effective way to define which variables are significant to the target variable, using a variable importance table. Lastly, clustering is used to group the data into groups, without the reliance on a target variable. This is a form of unsupervised learning, which makes it very hard to test the results. In this case, we were able to find similar groups of properties that have higher rental prices, occupancy rates, and ratings with this method.

**What are the key factors driving the nightly rental price for units in the Austin area?**

The factors driving rental prices of Airbnb units in Austin were determined by a linear regression model. The model was put through many iterations, including a stepwise feature selection for better results ([See Figure 1](#dauphrtncu2o)). The R-squared value for this model was 60%, which is a moderate, but acceptable score. In total, there were 19 significant variables that had an impact on the nightly rental price of a unit. However, all 19 of these factors do not have a large impact. If a property has more bathrooms, bedrooms, or beds then the nightly rental price will be higher. On the other hand, if a property has availability, experiences listed, private or shared room types, and a tent property type, the property rental price will be lower. The most important takeaway is more rooms and bathrooms in the house causing higher nightly prices. But how can we be so sure these large high-priced houses are successful properties?

**What are the key factors driving the probability of keeping a unit booked in the Austin area?**

The original data set did not contain a variable for the probability of a property being booked. However, there were a set of variables, names availability 30,60, and 90 days. To calculate the probability of a unit being booked, we assumed that if there was less than 40% occupancy in the next 90 days, the unit would be “booked”. When running a classification tree against this variable, we were able to find out which factors had the most impact on a unit being occupied ([See Figure 2](#mqcq7p6zy1ho)). The number of reviews, house property type, and host response rate are just 3 of many important factors. However, the most important variable that influences a unit being booked is the number of reviews ([See Figure 3](#a9i9qfxsrwsx)).

**How important are guest reviews and ratings? How could they be improved?**

Guest reviews are the pinnacle of Airbnb properties. Having a great set of reviews ensures perceived confidence in you from the guests. Using the variable importance table from the previous classification tree, we found that the number of reviews has a significant impact on the probability of a unit being booked ([See Figure 5](#gapju1fvcfx5)). We know that the sheer number of reviews matters, but what about the quality of the rating? Of properties with a perfect review scores rating we found that price matters most. People care about a good deal and are willing to advocate for it! As an additive to price, these homes are typically smaller in size and come without the burden of additional fees. If better reviews are being targeted, start by taking away the hassling fees and prompting customers to give a review.

**How do property profiles differ?**

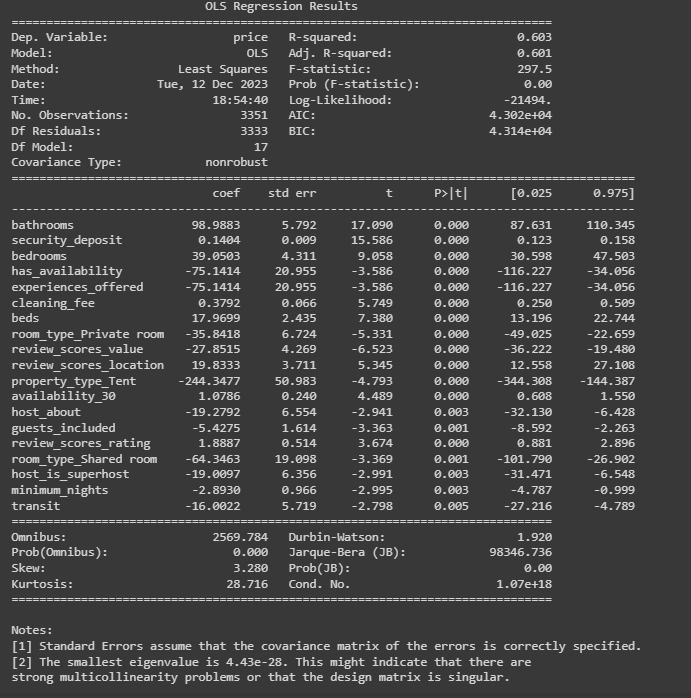
By clustering the properties into different groups, we can find similarities within those property groups without defining a target variable. In our research, we split the properties into two different groups, and the defining characteristics were price, size, and commonality. Cluster 0, otherwise known as “Elite Estates”, contains properties that are not sought after, as their price is too high for most customers. These properties are large, containing multiple bedrooms, bathrooms, beds, etc. On top of the nightly price being expensive, these properties tag on the additional fees for cleaning and security. Cluster 1, known as “Budget Bungalows”, consists of non-traditional properties (non-home) that have lower prices. These units are booked very often in a 90-day window due to their low pricing. No additional fees are attached to these properties. You get what you pay for, and people love that ([See Figures 6, 7, and 8](#ru2mnpbvscnz))

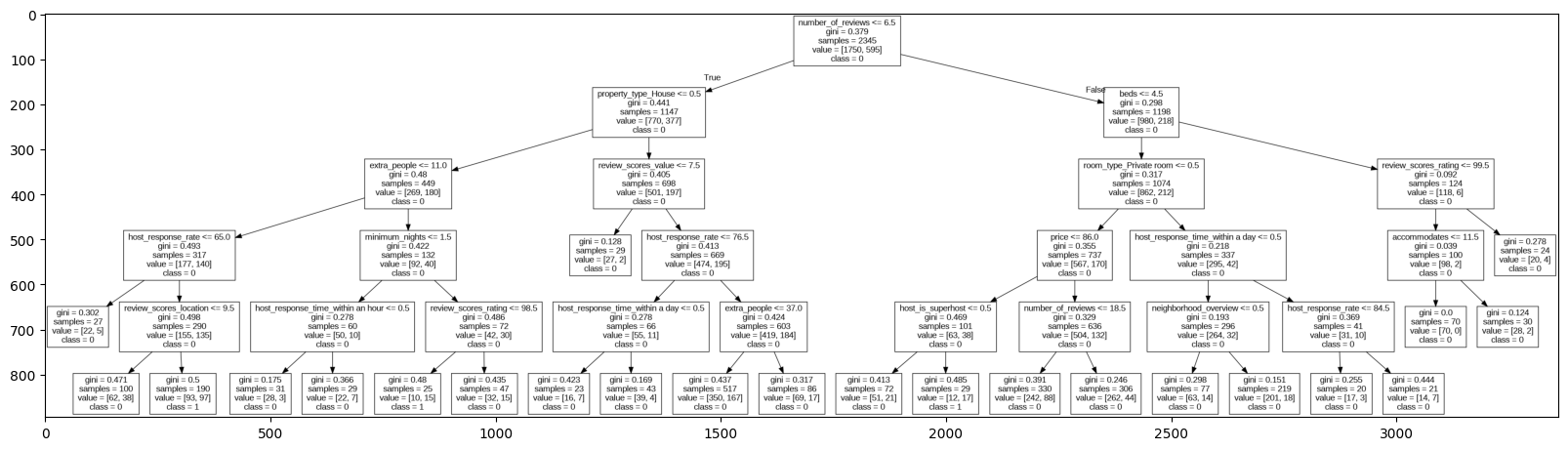
**Recommendations**

The ultimate goal for our investment group is to invest in properties that will provide substantial long-term returns when listed on Airbnb. Looking at multiple different target variables across a large Airbnb dataset, we were able to define which properties were successful. Your investment group should invest in properties that are a part of the “Elite Estates” cluster. These properties are typically owned by other investment groups just like us. Once these properties have been acquired and listed on Airbnb, the nightly price should be lowered. Due to the price being a large factor in the “Elite Estates”, dropping the price would allow customers to be more interested in. If it's possible, cleaning fees and security deposits should be avoided as well. As we know, once these properties are acquired, it is essential to have a high number of reviews. Encourage your guests to leave a review with some level of incentive. The “Elite Estates” collection in Austin, TX stands out for its commitment to offering value. By strategically pricing these properties are lower nightly rates, decreasing additional fees, and providing customers with the incentive to leave reviews, we ensure that these properties will provide unmatched value to the company, and more importantly, an exceptional experience for every guest.

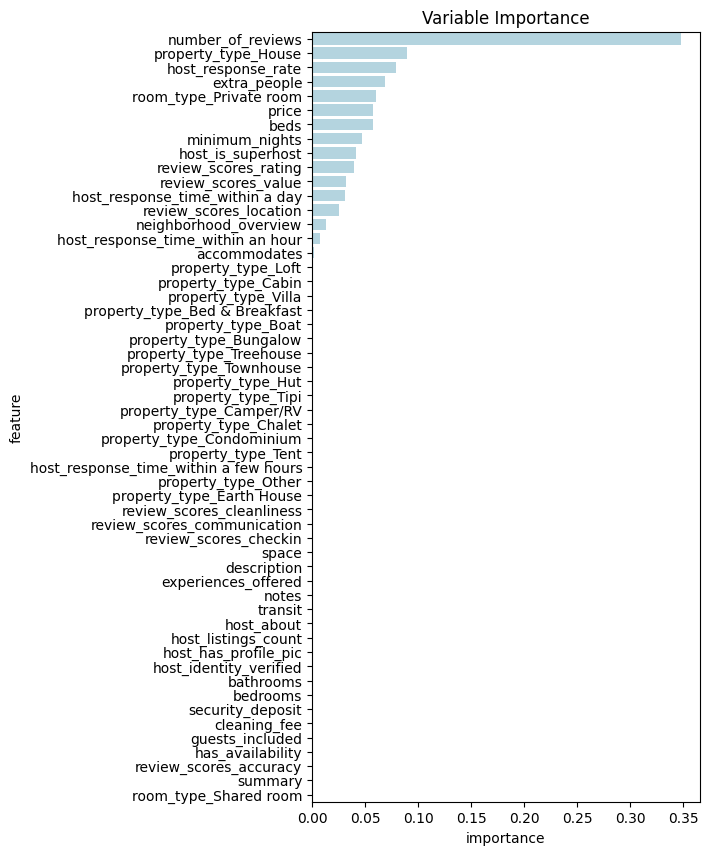
Appendix

**Price Regression Results (Figure 1)**

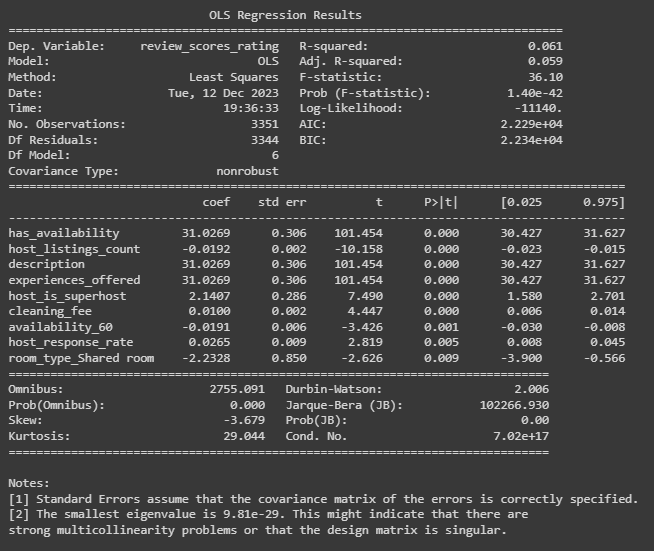
****

**Booked Classification Tree (Figure 2)  
**

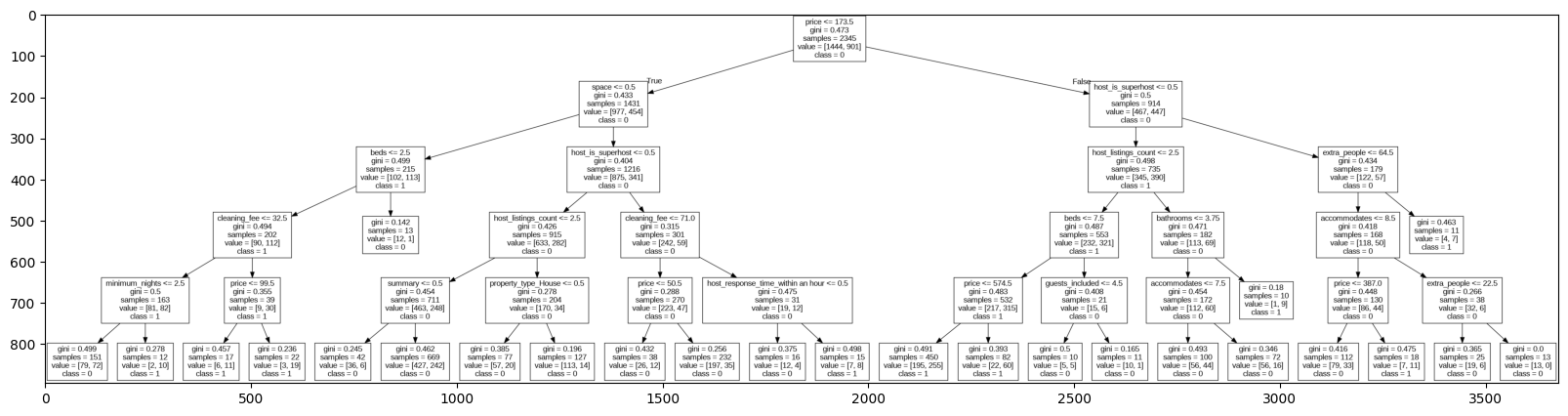
**Variable Importance for Booked (Figure 3)**

****

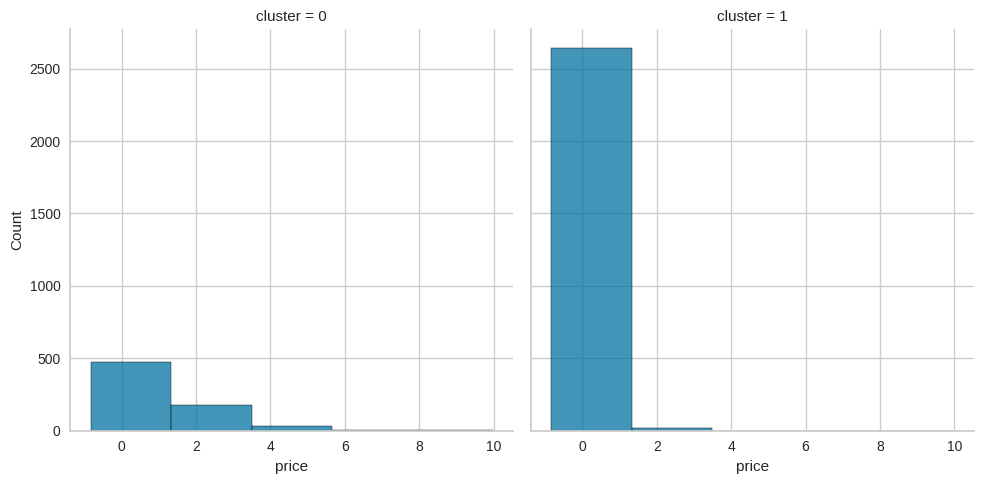
**Guest Reviews Linear Regression Results (Figure 4, Not Used, Poor R^2)**

****

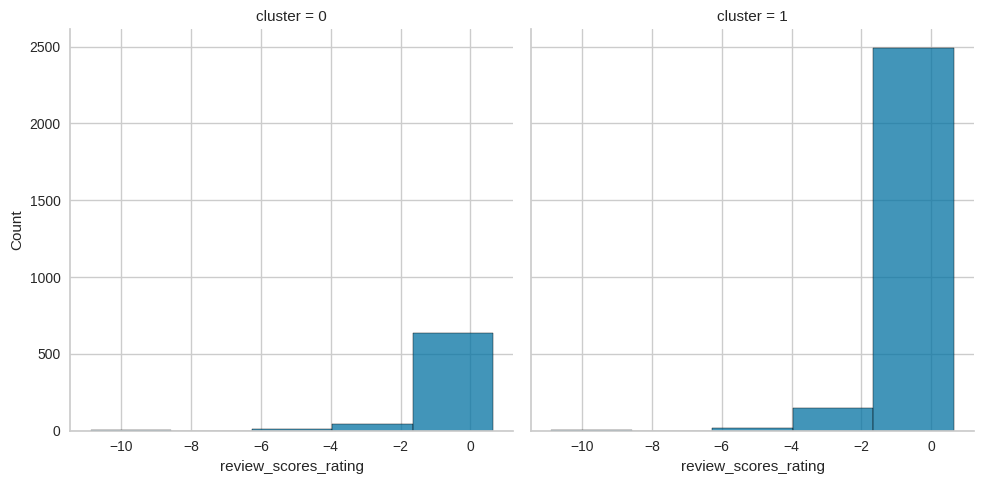
**Perfect Ratings Classification Tree (Figure 5)**

****

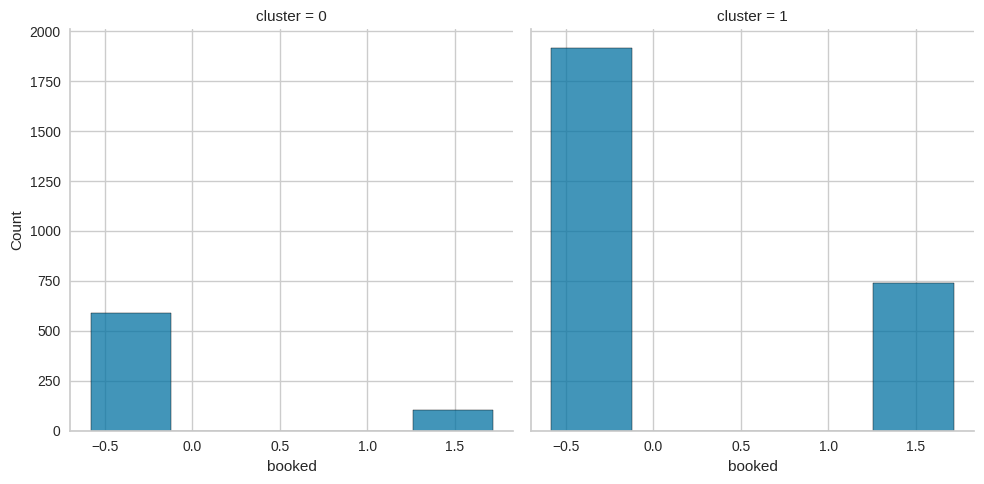
**Cluster Analysis - Price (Figure 6)**

****

**Cluster Analysis - review\_scores\_rating (Figure 7)**

****

**Cluster Analysis - booked (Figure 8)**

****