

# Executive Summary

## I. Overview of Business Problem

LJ Investment Group is considering investing in property in Austin, Texas with the goal of listing the property for rent on Airbnb. Airbnb rose to popularity for its home feel, bigger spaces, and lower pricing in comparison to hotels. For investment groups, short term rental properties have become a sought-out investment strategy for generating passive income. Due to this sought-out opportunity, the market has become saturated and complex. It can be difficult to know what features consumers look for, what factors highly influence customer ratings, what information to highlight on the listing, and how to gauge the potential occupancy rate for a property.

## II. Overview of Austin Airbnb Data

The analytics team at LJ Investment Group collected data from multiple neighborhoods in Austin, Texas. The data collected centered around rental property features such as number of bathrooms and bedrooms, amenities, square feet, and neighborhood features such as transit, neighborhood description, as well as property owner features such as response time and rate, how many listings they have, if they are a super host, and when they started hosting rental properties. The team also collected data on scores given by customers. The analytics team took time to clean the data before conducting exploratory analysis. The team followed the statistical rule for dropping columns which are missing more than 20% of the data. They also determined multiple columns would not add value to the analysis and/or add too much complexity to the models. The analytics team followed all industry recommended cleaning.

## III. Summary of Analytical Techniques

Three analytical techniques, in addition to a forecast, were used to answer four key questions about the short-term rental market in Austin. The team produced two linear regressions, two decision trees, and a k-means clustering. The team chose one of each of the linear regressions and decisions trees which most closely produced a good, accurate model. The linear regression defined key factors for nightly pricing. Nightly pricing was a significant target variable because every potential customer looks at the pricing before booking. To assist LJ Investment Group with setting their potential future rental property's nightly price, the linear regression provided insight into what variables are highly correlated with nightly pricing. Correlation determines the strength and direction of the relationship between a variable and the target variable. Therefore, the closer the correlation is to one, the stronger the relationship is. The decision tree reflected key factors for keeping a unit booked, as well, how important are guest reviews. Decision trees are binary, classification models which determine the characteristics of the end consumer. Lastly, the k-means clustering defined the customer profile of each segment. Therefore, the investment group will know which target market (consumer segment), the group wants to market to.

## IV. Key Findings

- a. What are the key factors driving the nightly rental price for Airbnb's in Austin?

Through exploratory analysis and the creation of several linear regressions found the highest correlation to nightly prices was bathrooms. It had a correlation score of 59.48% rounded. Following right behind was bedrooms with a score of 54.12% rounded. As seen in Figure 1, located in the appendix, there were five moderate-to-highly correlated variables. The correlated variables were bathrooms, bedrooms, accommodates, beds, and cleaning fee. Each of these variables would provide a positive increase to the

nightly price. These variables do not surprise the analytics team because, socially, it is understood that when more is offered to a consumer such as one more bathroom, bedroom, and bed, the higher the price the consumer must pay. Along with these highly correlated variables, there was a plethora of either very little correlated or little-to-moderately negative correlated variables. It is important to note the scores of accuracy, communication, check in, and value are all negatively correlated. The number of reviews variable was the highest negatively correlated variable. The analytics team is confident that LJ Investment Group should consider finding investment properties which contain the highly correlated variables. These variables will assist in pricing the rental property, as well as marketing to a wider verity of consumers.

- b. What are the key factors driving the probability of keeping a rental booked in the Austin area?

Looking at Figure 2, the decision tree breaks down what characteristics/key factors customers look for when they book a short-term rental property. The decision tree produced three sets of different key characteristics which will keep a short-term rental booked. In order to determine if a rental property will stay booked, the analytics team used the 90-day availability column with a less than 40% availability. When looking at Figure 2, zero means the property did not stay booked and one means the property did stay booked. The first set of characters is the host response time is not available, the host didn't have rental properties since 2015, the property type is not a house, and the number of reviews was greater than 1.5. The second set of characters are the host response time is available and have had a rental property since 2015, as well the neighborhood is not provided and the review score for check in is less than 9.915. The third set follows extremely closely to the second set but does continue a little further. To stay booked, the host response time is provided, the host started in 2015, the neighborhood is not provided, the review score for check in is lower than 9.915, the price is higher than \$177.5, and the cancellation policy is not strict. All of these characteristics provide insight into what consumers look for when they are researching short-term rentals on Airbnb. The key factors, as seen in Figure 3, are host response time not provided, host since 2015, and number of reviews. When challenges arrive, it is important to customers that they are able to reach out and get a response from their host. Also, the experience a host has and the amount of reviews provide assurance to customers.

- c. How important are guest reviews and ratings?

Guest reviews and ratings are important to consider but they do not hold the upmost importance. As stated earlier in the report with the support of Figure 1, the correlation of scores to nightly prices is low. The overall ratings from customers have a correlation of 2.5% rounded. This holds very little influence and strength in determining prices. As well, seen in Figure 2, the score of check in held importance. Potential consumers will always look reviews and there is a potential for the negative reviews to sway the consumers away. Through this analysis, it has been disproven that guest reviews are highly influential to future consumers booking a rental property.

- d. What are the customer profiles of each segment/cluster?

Cluster 0: Cluster zero does care about how many listings a host does have. They want to see a host have many properties because they know a host can take of their property. The customers in this cluster also want properties which accommodate four or more guests, one or more bathrooms, one and half bedrooms with two beds and priced around \$191. They also have the second highest cleaning fee of \$43 with a want of one or more guests and the guests can have nine extra people. They want a minimum of two nights with a 90-day availability of 61. They have the lowest desire to consider all of the reviews.

Cluster 1: Cluster 1 contains the highest means for a large amount of the columns/variables. Cluster 1 holds interest in a high host listing count at 30.19, which also accommodates the most guests at 7.56. As well, this cluster wants 2.31 bathrooms, 3.16 bedrooms, 4.06 beds at a price of \$583.88. This segment commands the highest nightly rates. This segment also looks for the most guests included at 3.21 with extra people at 23.231. Following similarly to cluster 0, cluster 1 also wants a minimum of 2 nights per stay.

Cluster 2: Cluster 2 does not need the identity of the host to be provided or for them to be a super host. Also, they do not need a response from the host at any time or to know when the host started renting out properties. This segment does look for value in review scores around 9.26. This segment looks highly to reviews to know if they should boo for not. They have a booked rate of 0.200. They also have the lowest 90-day availability of 43.93. This cluster does have the second highest nightly price of \$229.93.

Cluster 3: Cluster 3 follows similarly to cluster 2 and cluster 4. This segment has the lowest nightly price at \$181.94. Cluster 3 wants a low cleaning fee of 25.15, listings count of 2.69 which includes 1.17 bathrooms, 1.22 bedrooms, 1.153 beds and accommodates 3.21 guests. One interesting characteristic of this segment is they want to ability to have 10 extra people for a minimum of 2 nights. Cluster 3 had a booked rate of 0.203 and a 90-day availability of 54.74. This segment also looks highly to review scores and wants to see high scores in their respective ranges.

Cluster 4: Cluster 4 has the highest booked rate at 0.224 and the highest customer rating at 96.498. This segment had the second lowest 90-day availability with the highest number of reviews at 14.77. The price for cluster 4 who wants to accommodate 4 guests, 1 bath, 1 bedroom with two beds is \$212.68. This price falls in the middle for all of clusters. This segment also wants a relatively high cleaning fee of \$34.82. This segment is ideal for any investment group because the group can accommodate both the middle class and the upper middle class.

Overall, none of the clusters cared to have a defined neighborhood or city, except for being in Austin. All but cluster 2 cared to have their host identity verified, their host responded within an hour and a few hours. All clusters also cared to have a house as their property type and a real bed as their bed type.

## **V. Recommendations**

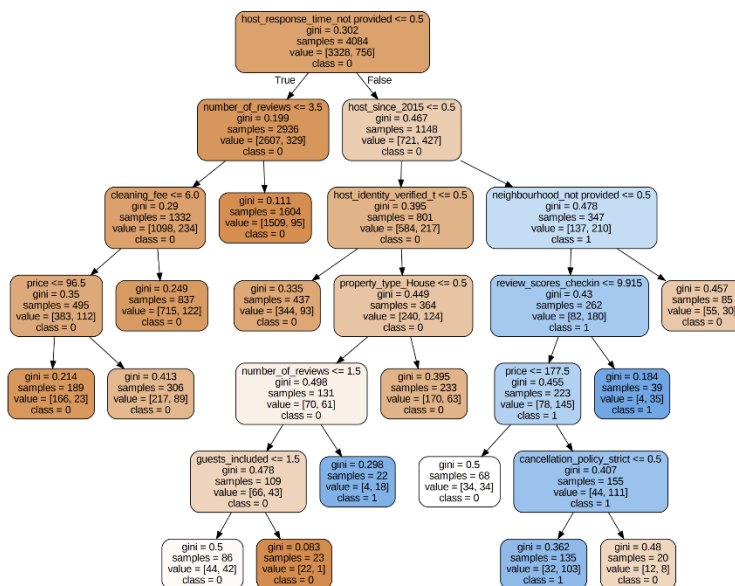
LJ Investment Group should look for property in the city of Austin. The neighborhood is not as significant because it was not important to any of the clusters. The property should be of a decent size with at least 2 bathrooms, 3 bedrooms, and 4 beds. The analytics team recommends focusing on marketing the potential future rental property to cluster 4. These customers land right in the middle allowing the company to gain profit from several different types of customers. Also, this group had the highest booked rate, making the analytics team confident that this cluster will keep the property booked the majority of the time. By looking for a property a little bigger than what cluster 4 needs, this will allow them to be able to have the 11 extra guests they want. The analytics team is confident, LJ Investment Group will be able to be highly successful in Austin, Texas looking for a property to do short-term rentals. Once the investment group finds a few properties to potentially invest in, the analytics team will collect more data on those specific neighborhoods, consumers, and amenities to ensure the success of the investment group.

## Appendix

### 1. Correlation Table

	correlation
price	1.000000
bathrooms	0.594774
bedrooms	0.541275
accommodates	0.476145
beds	0.456678
cleaning_fee	0.411318
guests_included	0.136695
extra_people	0.060012
availability_30	0.044624
review_scores_cleanliness	0.028781
review_scores_rating	0.025868
availability_60	0.016321
minimum_nights	0.015142
review_scores_location	0.014104
availability_90	0.004672
review_scores_accuracy	-0.007962
review_scores_communication	-0.009759
review_scores_checkin	-0.013676
review_scores_value	-0.042818
number_of_reviews	-0.136838

### 2. Decision Tree (attached in the submission)



### 3. Variable Importance Bar Graph

