Anthony X. Ayala & Jennifer Ontiveros-Olivas
Dr. Tonya Balan
BAN 6025 Predictive Analytics & Data Mining
December 13, 2023

Final Project: Airbnb investing in Austin, Texas

Business Problem:

The business challenge focuses on enhancing investment returns in Austin, Texas's short-term rental market, particularly for Airbnb properties. The investor seeks insights into factors that affect nightly rental rates, the likelihood of consistent bookings, and the impact of guest reviews and ratings. This involves analyzing data from approximately 6000 Airbnb properties in Austin to uncover these critical elements. The analysis will pinpoint what attracts renters, what factors influence their ratings, and what leads to higher occupancy rates. Moreover, the task includes segmenting the Airbnb properties into various groups, profiling each, and identifying which groups command the highest nightly rates, receive the most favorable guest reviews, and achieve the highest booking rates. This thorough assessment provides the investor with essential information for informed decision-making in this sector.

Overview of Data:

The dataset contained detailed information about Airbnb listings in Austin, Texas, with a structure of 5835 rows and 54 columns. It includes a variety of descriptive variables for each listed property, such as location, neighborhood, summary, notes, size, and the number of bedrooms and bathrooms. Additionally, the dataset offers insights into the hosts, including their names, response times, response rates, and whether their identity is verified. It also provides specific information regarding the availability of each Airbnb property, detailing the number of days it was available over periods of 30, 60, 90, and 365 days.

Summary of the analytical techniques that were applied:

In our analysis of the dataset containing Airbnb listings, we tackled missing values through imputation techniques such as mode, median, and mean. Non-essential variables like listing URL, notes, and summary were omitted. We conducted term frequency analysis on textual data to assess sentiment trends. We used a linear regression model with 'Price' as the dependent variable to identify price influencers. Initially yielding an R-squared value of 0.458, the model

was refined using stepwise feature selection and log transformation, enhancing the R-squared to 0.627. This process pinpointed key variables affecting rental prices.

Decision tree algorithms were employed to predict booking likelihood, focusing on accuracy and simplicity. These models provided significant insights into booking patterns and probabilities. Additionally, we applied K-Means Clustering to segment the properties into distinct groups, aiming to profile each based on rates, guest ratings, and booking frequencies. This multifaceted approach offered a comprehensive view of factors driving rental pricing and booking trends in the Airbnb market.

Influential Factors on Rental Price:

- Bedrooms and Bathrooms: Properties with more bedrooms and bathrooms command higher nightly rates. This is evidenced by the positive and statistically significant coefficients (Bedrooms: 0.1190, Bathrooms: 0.2288) with a p-value < 0.05.
- Room Type: The type of room significantly impacts pricing. The negative coefficient (-0.5903) indicates that certain room types are priced lower than our model's baseline room type.
- Accommodation Capacity: Properties that accommodate more people are likely to have higher prices, as shown by the positive coefficient (0.0453).

Impact of Guest Reviews on Pricing:

- Location Review Scores: Higher location ratings positively correlate with increased prices, as the coefficient (0.0958) indicates. To enhance ratings, hosts can focus on improving factors like accurate location descriptions and amenities related to the location.

What factors determine if an Airbnb is "booked" or not?:

Our decision tree analysis identifies two main scenarios predicting the booking of Airbnb units, primarily based on their availability in the next 365 days, with a threshold of less than 309 days. The differentiation between the scenarios lies in the short-term availability over the next 60 days.

1. Scenario one: Units available for less than seven days in the next 60 days tend to be booked, regardless of whether their nightly rate is below or above $73.50.
2. Scenario two: Units with availability exceeding seven days in the next 60 days are also likely to be booked despite having fewer or more than five reviews.

The refined decision tree model concurs that lower availability in the next 60 or 365 days increases the likelihood of booking. However, it adds that the type of cancellation policy, the maximum occupancy, and the number of reviews also influence the booking probability. In summary, while short-term and medium-term availability are key determinants of a unit's booking status, other factors like cancellation policies and reviews also play a role, although to a lesser extent.

Segmentation and Profiles of Airbnb Properties:

- Cluster 0: High host response rate (79.95%), accommodates up to five people, one bathroom, but has the lowest review and cleanliness scores despite the most listings.
- Cluster 1: Most frequently booked, suitable for up to four people, one bathroom, highest review scores, and great value for money.
- Cluster 2: Largest capacity (5 beds, accommodates eight), highest average price ($669.57) with a no-refund policy, listed approximately 27 times.
- Cluster 3: Most reviewed (14 on average), high review scores across categories, available for bookings throughout the year, and the most affordable on average at $178.01

Market Insights:

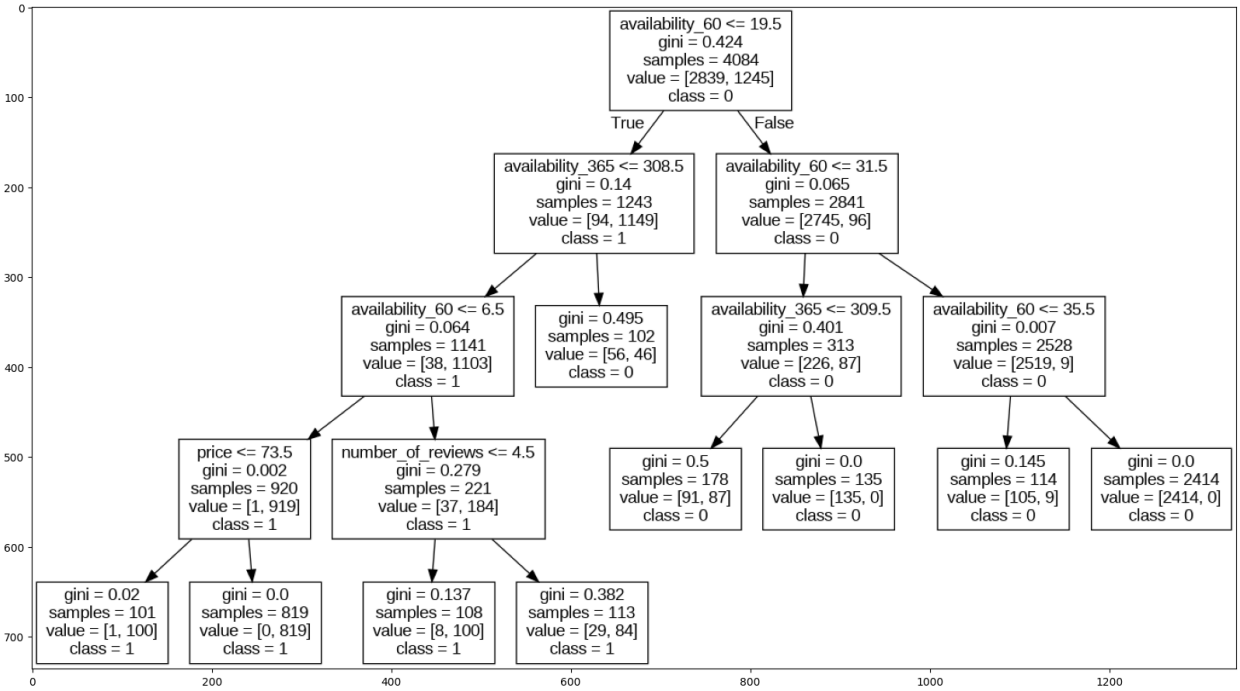
Properties offering more accommodations and bathrooms (Cluster 2) are more likely to charge higher rates. High review scores, particularly in value, as seen in Clusters 2, 3, and 4, can influence availability for the next month or two and next year.

Conclusion/Recommendations:

This analysis offers key insights into the Austin Airbnb market, which is crucial for informed property investment decisions. We recommend that investors consider not just the pricing and features of a property but also emphasize its availability and the diversity of review scores. Availability has emerged as a crucial driver for booking success, complemented by review metrics and cancellation policies. Austin's Airbnb properties are particularly appealing due to their proximity to downtown (figure four), enhancing their allure amid the vibrant local scene. For investors, a deep understanding of these elements is crucial to successfully navigating Austin's dynamic short-term rental market.

Appendix

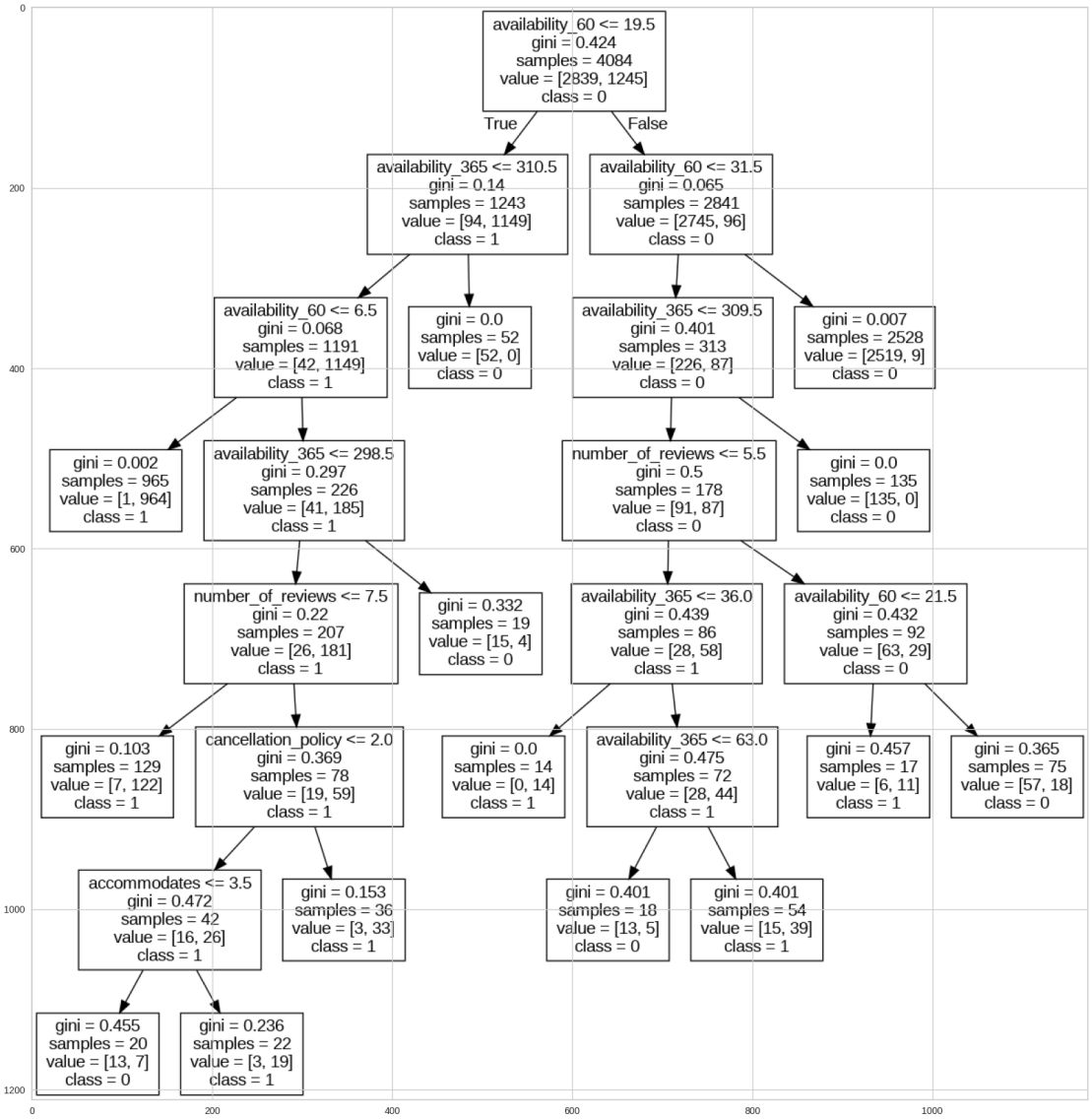
1. Decision Tree with Model Metrics (Simpler Model)




```
 -- train set --
Accuracy : 0.9559
Precision: 0.9667
Recall.  : 0.8859

 -- test set --
Accuracy : 0.9589
Precision: 0.9743
Recall.  : 0.8931
```

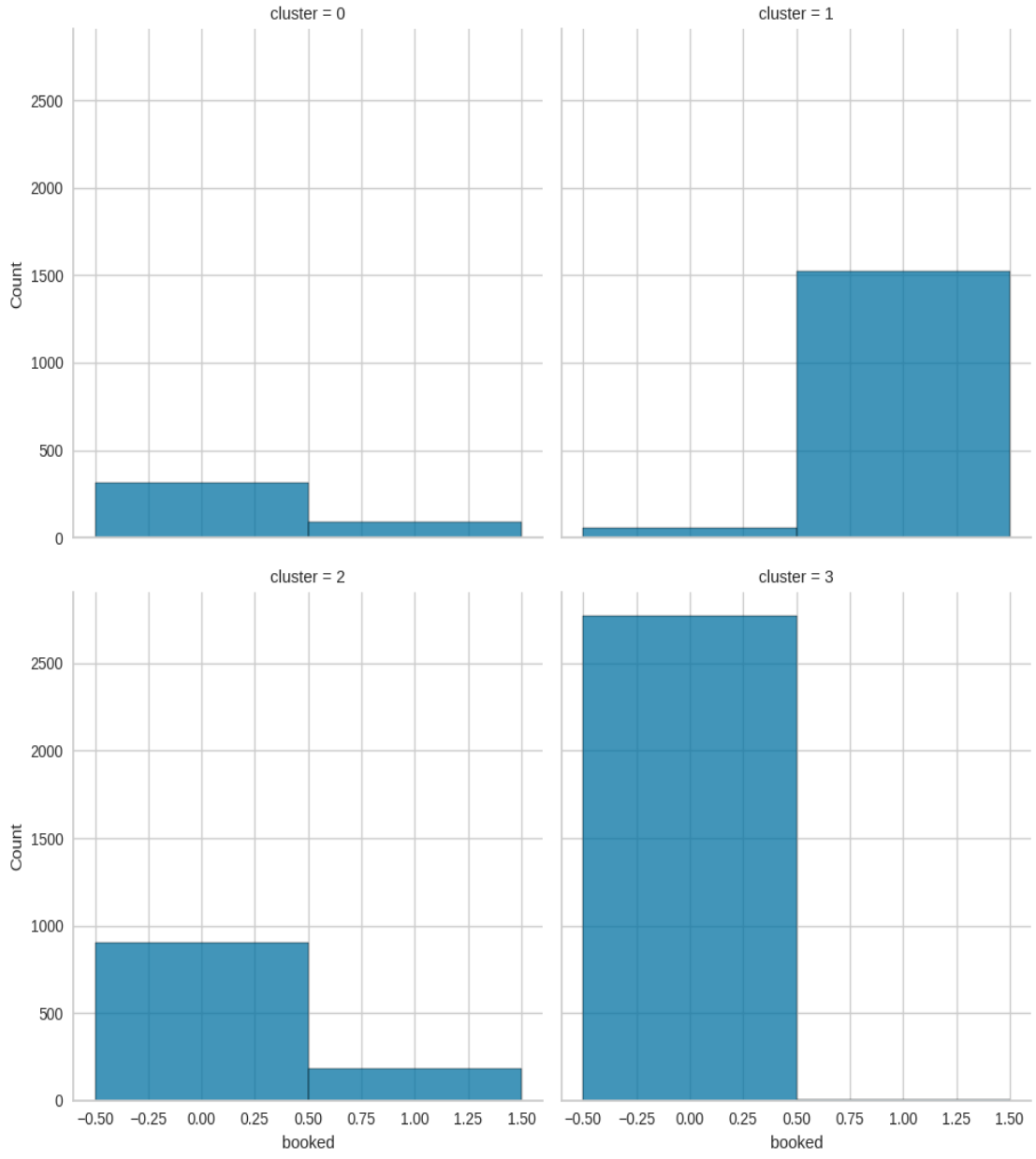
2. Tuned Decision Tree (Slightly more complex, but could be more realistic)




```
 -- train set --
Accuracy : 0.9809
Precision: 0.9717
Recall.  : 0.9655

 -- test set --
Accuracy : 0.9732
Precision: 0.9599
Recall.  : 0.9547
```

3. Clusters and Bookings



4. Word Cloud of Austin Airbnb Data