BAN 6025 Final Project

Airbnb in Austin, TX

Aadya Choudhary and Lambert Li

Short term rental properties have become a popular investment strategy for many individuals interested in generating a passive income stream. However, the dynamics of the short term rental market are complex, and it can be difficult to know what features tend to draw in potential renters, what factors have the most influence on ratings, what information to highlight on the listing, and how to gauge the potential occupancy rate for a property.

You work for a real estate investor who is considering an investment property in Austin, TX with the goal of listing the property for rent on Airbnb. You have been provided with a data set that contains information on almost 6000 Airbnb rental units in Austin. For each property, you have the following information:

Variable Name	Variable Description
id	Unique property identifier
listing_URL	Direct link to property listing on Airbnb
name	Property name as listed on Airbnb
summary	Property description as listed on Airbnb
space	Description of the living spaces available in the property
experience	Airbnb experiences available with the property
neighborhood	Description of the neighborhood where property is located
notes	Additional comments about the property
transit	Description of transportation options accessible from the property
host_id	Unique host identifier
host_name	Host name
host_since	Date when host listed first rental property on Airbnb
host_location	Geographic area where host is actually located
host_about	Information shared by the host about themselves
host_response_time	How quickly you can expect a response if you contact the host
host_response_rate	% of time host responds to questions or inquiries
host_is_superhost	t = superhost, f = not superhost
host_listings_count	Total number of listings host has on Airbnb
host_has_profile_pic	Listing includes photo of host
host_identity_verified	Host has been verified by Airbnb
neighbourhood	Austin neighbourhood where property is located
city	City where property is located
property_type	Type of property (house, apartment, camper, etc)
room_type	Type of room available (entire house, private room, etc)

accommodations	Max number of people who can stay in the house
bathrooms	Number of bathrooms
bedrooms	Number of bedrooms
beds	Number of beds available
bed_type	Type of bed available (real bed, futon, etc)
amenities	Amenities available with the property
square_feet	Square footage (size) of the property
price	Nightly rental rate (in USD)
weekly_price	Rental rate for an entire week (in USD)
security_deposit	Refundable security deposit required to rent property (in USD)
cleaning_fee	Cleaning fee assessed at checkout (in USD)
guests_included	Number of guests included in rental prices (in USD)
extra_people	Additional charge for guests above the number included (in USD)
minimum_nights	Required minimum number of nights to rent
has_availability	Has at least one open night within the next year = t
availability_30	Number of days the property is available out of the next 30
availability_60	Number of days the property is available out of the next 60
availability_90	Number of days the property is available out of the next 90
availability_365	Number of days the property is available out of the next 365
number_of_reviews	Total number of reviews that have been left for the property
review_scores_rating	Overall property rating (1 to 100 with 100 being the best)
review_scores_accuracy	Accuracy of listing (1 to 10 with 10 being the best)
review_scores_cleanliness	Cleanliness of the property (1 to 10 with 10 being the best)
review_scores_checkin	Ease of check in to the property (1 to 10 with 10 being the best)
review_scores_communication	Communication with host (1 to 10 with 10 being the best)
review_scores_location	Location of property (1 to 10 with 10 being the best)
review_scores_value	Property was worth the price (1 to 10 with 10 being the best)
instant_bookable	Available for instant booking without approval from host
cancellation_policty	Cancellation policy (flexible, moderate, strict, etc)

The client has several questions that he would like to have answered based on the data that has been provided:

- **What are the key factors driving the probability of keeping a unit booked in the Austin area? To answer this question, you can classify a unit as "booked" if the 90-day availability for the property is less than 40%.**

1. What analytical technique(s) did you choose to address this scenario?

Ans. For the analysis of the key factors driving the probability of keeping a unit booked a classification tree was chosen.

The technique was used to predict whether a unit would be 'booked' based on its attributes.

2. Why did you choose this technique? In other words, why was this technique a good match for this scenario?

Ans. A classification tree was chosen because:
a) It handles both numerical and categorical variables effectively.

b). It provides clear interpretability by visualizing decision paths, which is helpful for business stakeholders.

c). It identifies the most important features and ranks them that influence the booking probability.

If you are using a supervised learning technique, answer questions 3 and 4. If you are using an unsupervised learning technique, answer questions 5 and 6.

3. If you are using a supervised learning technique to answer this question, copy and paste the output from your final model below.

Ans.

Top 20 Variable Importance
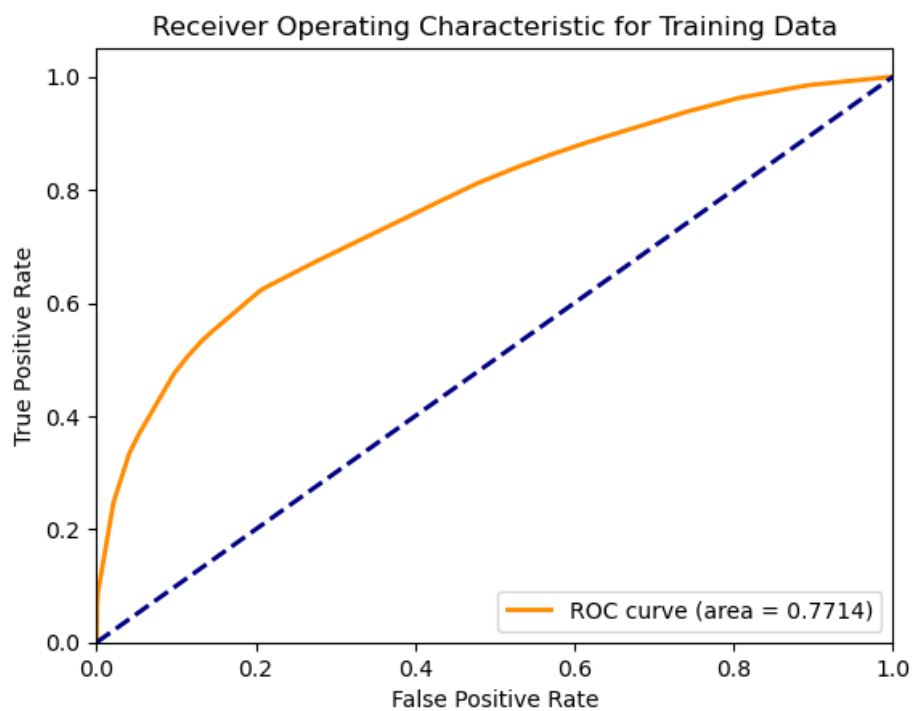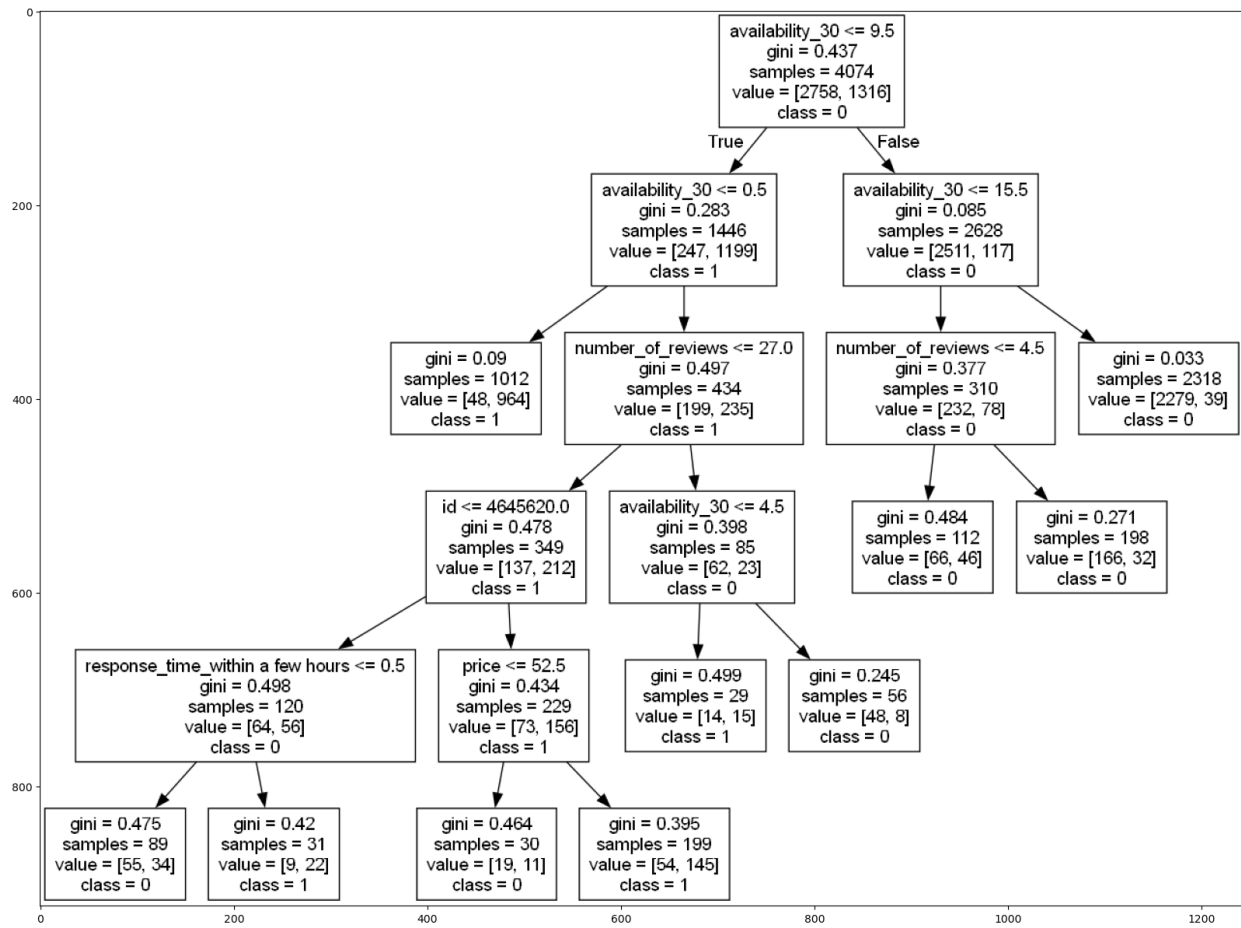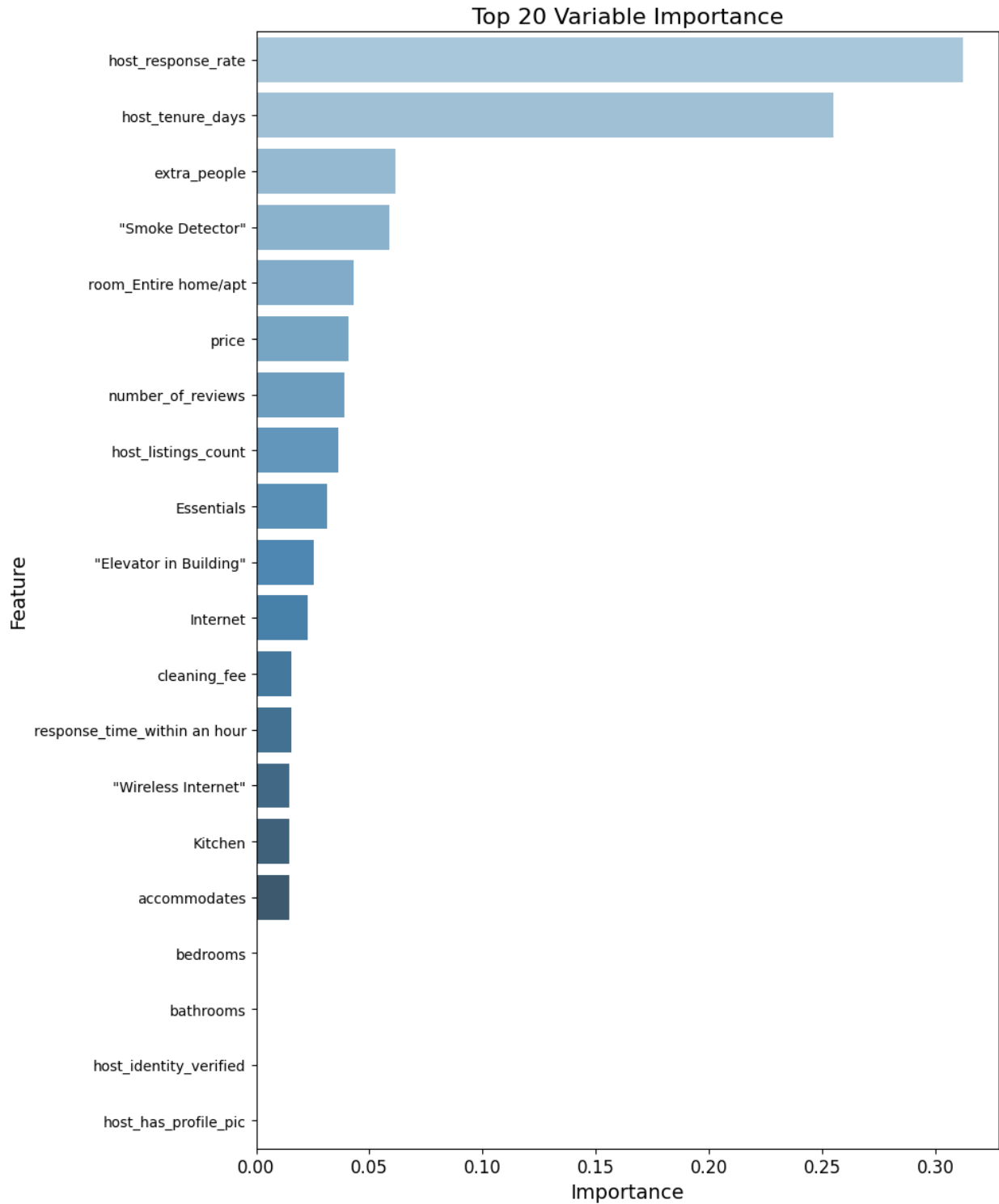
4. If you are using a supervised learning technique to answer this question, please explain how the model that you chose satisfies all the requirements of a "good model."

Ans. **Interpretability**: The decision tree clearly outlines the decision-making process through a series of splits based on feature values, making it easy to interpret.

Feature Importance: The variable importance chart provides insights into which features have the greatest impact on the booking probability, such as host response rate and number of reviews.

Performance:

- The ROC-AUC score of **0.7845** shows the model has good predictive capability while being simple enough to avoid overfitting.

- The tree depth and pruning parameters ensure the model generalizes well to unseen data.

Actionable Insights:

- Hosts can improve booking rates by focusing on the most important features, such as high response rates and encouraging guest reviews.

5. Please list at least two key business takeaways from your analysis that the real estate investor can use to help him address the chosen business scenario.

Ans.

Takeaway 1: Response Rate Is Crucial

- The host response rate is the most influential factor in determining whether a unit is booked. Hosts should prioritize improving their response times to inquiries.

Takeaway 2: Reviews Matter

- Number of reviews is one of the top predictors. Hosts with more reviews are more likely to keep their units booked, indicating the importance of guest feedback.

Takeaway 3: Room Type and Price

- Units categorized as Entire home/apt with competitive pricing (price) are more likely to be booked. Hosts should ensure their pricing aligns with market expectations.

Takeaway 4: Minimal Impact of Amenities

- Amenities like "Free Parking" and "Other pet(s)" have a minimal influence on bookings, suggesting hosts should focus on improving core aspects like response rate and reviews instead.