

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_policy	Cancellation policy (flexible, moderate, strict, etc)

Now that you have an overview of the data and the business issues that need to be addressed, here is your assignment for the final project:

Deliverable:

Segment the Airbnb properties in the Austin area into different groups. Provide a profile of each segment. Which segments command the highest nightly rates? Which segments get the best guest ratings? Which segments have the highest percentage of “booked” units?

The luxury company-host properties have the highest nightly rate. The Best Family Experience has the best reviews and the highest percentage of booked units.

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

A clustering technique can best satisfy the need to segment different Airbnb properties into groups, so our study chose this method.

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

Benefits of using a cluster technique include:

- 1) The use of pattern recognition, which is a perfect match for this scenario that needs to categorize different features like price, review, and amenities.
- 2) Fit requirement of the task which is not to predict certain outcomes but make sense of different segments of data.
- 3) Create actionable insights and on-the-point recommendations by identifying groups that will create the highest nightly rates.
- 4) Flexibility and subjectivity where clustering profiles can leverage industry expertise to identify the most important features for targeted customers.

3. If you are using an unsupervised learning technique, how many clusters did you choose? Why did you choose this number?

I ended up choosing 7 clusters, which can seem excessive, but the decision came with running distortion score elbow for KMeans analysis multiple times while test try all the cases between 4 to 8 clusters. 7 clusters presented a different distribution of data point in the sense that it broke the largest cluster in 6-cluster case, which included 2268

airbnb listings, into 2. I believe this is a meaningful split and the insights will allow companies to gain nuanced understanding in the previously largest cluster. One potential reason that KMeans generated different variations of was due to having more than 100 variables, with more than 80 categorical variables being created by get_dummys method.

4. If you are using an unsupervised learning technique, please choose one cluster and include a short description of the cluster below.

Please find all cluster profile in the appendix section, as requested by the problem statement to provide a profile for each segment. The following description fits cluster 4, **the Luxury Company-Host:**

In this profile that includes 757 listings, the luxury company-host offer the priciest accommodations, average \$670 per night before cleaning and service fee, and have relatively low 90-day availability. From speculation, these properties are likely to be operated under corporate structures as opposed to individual host, as hosts of these properties have an average of 13 listings, a number that cannot be easily managed by a single person. This type of listing offers exclusive and premiere experiences, include one of the best amenities across the board. Due to its high-end nature, reviews for this type of listing are minimum.

The luxury company-host properties have the highest nightly rate.

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.
 - 1) Offering nightly rates in between \$160 - \$210 for properties that have accommodate 3 – 4 people will have a good “booked” rate in 90 days period.
 - 2) Investors should be cautious about offering properties that are middle-to-high end in price (above \$400 per night) because there is already a good number of these properties available.
 - 3) A high response rate from hosts will likely convert into receiving good reviews, so maintaining that communication will be key for properties to be received positively by existing and future customers.

Appendix for all cluster profiling:

cluster	
6	1368
5	1222
2	1178
0	998
4	757
1	173
3	124
Name: count, dtype: int64	

Cluster 0: Personal Accommodation

This profile includes 998 Airbnb listings in the Austin area, featuring highly responsive hosts who operate only a few properties by themselves. These listings often come in as private guestrooms, which are economic options for a single guest, and receive overwhelmingly positive reviews. Listings under this profile are predominantly houses, and the guests might be sharing living rooms with the hosts. This listing also offers the most lenient pet policies.

Cluster 1: Family Accommodation

This profile includes 173 Airbnb listings in the Austin Area, featuring hosts with more than 5 listings and have relatively fast response time. This type of property offers one of the best amenities within their accommodation and can host on average more than 4 people. These listings are second-to-the-most popular and are often booked ahead-of-time. The reviews for these types of properties are below average.

Cluster 2: Poorly Managed Inactive Listings

This profile includes 1178 listings that are relatively pricy operated by hosts that are worst in responsiveness. These hosts often have 1 – 2 listings under Airbnb and are just not very passionate about hosting. They offer accommodation for an average group size of 3 people and are rated the worst across the board. These listings can always be outcompeted by listings in other clusters, such as by family accommodation (cluster 1).

Cluster 3: Economic Accommodation

This profile includes 124 listings that offer almost exclusively shared room experiences, with one of the cheapest fares per night. This type of listing are the most

available in 30, 60, and 90 days period, likely due to most people disliking the shared living experience. The reviews for this type of listing is mixed, with each property having 58 reviews on average.

Cluster 4: Luxury Company-Host

This profile includes 757 listings. The luxury company-host offer the priciest accommodations, average \$670 per night before cleaning and service fee, and have relatively low 90-day availability. From speculation, these properties are likely to be operated under corporate structures as opposed to individual hosts, as hosts of these properties have an average of 13 listings, a number that cannot be easily managed by a single person. This type of listing offers exclusive and premiere experiences, including one of the best amenities across the board. Due to its high-end nature, reviews for this type of listing are minimum.

The luxury company-host properties have the highest nightly rate.

Cluster 5: Well-Rounded Company-Host

This profile includes 1222 listings, being one of the largest segments of Airbnb offers in the Austin area. This segment features hosts with high response rates and are most definitely corporates due to the large quantity of properties managed under each host, with an average of 37 listings. These properties have one of the best reviews, beaten slightly by cluster 6: the best family experience. This type of listing has the strictest cancellation policy, while offering one of the best amenities like washer and dryer.

Cluster 6: Best Family Experience

As the largest segment, Best Family Experience includes 1368 listings that on average accommodate 3 – 4 people. These listings offer below-average prices and are operated by hosts that manage under 5 assets. These listings are the most-in-demand, having only 50 days available on average in the 3-month period. These listings also have the best reviews, averaging 17 reviews with close-to-perfect individual ratings.

The Best Family Experience has the best reviews and the highest percentage of booked units.