**BAN 6025 Final Project**

**Airbnb in Austin, TX**

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 nightly rental price for units in the Austin area?
* 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%.
* How important are guest reviews and ratings? Is there anything the host could do to improve those ratings?
* 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?

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:

**Instructions:**

1. You must choose TWO of the business scenarios listed above. You will address those scenarios using the modeling techniques covered in BAN 6025.
2. While you may use always use any analytical techniques and any software that you have available to you, for this assignment, you must apply AT LEAST TWO of the techniques that you have learned in BAN 6025. The techniques that were covered in BAN 6025 include linear regression, forecasting, classification models (logistic regression and classification trees), and k-means clustering. **NOTE that for the purposes of this project, “classification models” is considered one technique.** So, this means that applying logistic regression and then applying classification trees does NOT count as satisfying the requirement of two techniques. Those are both variations of “classification models.”

**Deliverable:**

For each of the business scenarios that you chose in part (1) above, you should answer the following questions. These can be short answers and you can answer in bulleted list format or paragraph format. You should submit a document containing the answers to these questions for each of the business scenarios that you selected. (Note that you will answer 5 total questions for each scenario. Each question is worth 10 points for a total of 100 points for the submission).

You should also submit your python code and output as a separate file. Failure to do so will result in a 10 point deduction from your final grade.

1. What analytical technique(s) did you choose to address this scenario?
2. Why did you choose this technique? In other words, why was this technique a good match for this scenario?

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.

1. If you are using a supervised learning technique to answer this question, copy and paste the output from your final model below.
2. 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.”
3. If you are using an unsupervised learning technique, how many clusters did you choose? Why did you choose this number?
4. If you are using an unsupervised learning technique, please choose one cluster and include a short description of the cluster below.
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.