jxn628 / **dsc-phase-5-project**   Public

★ 1 star   0 forks

⭐ Starred ▼   👁 Unwatch ▼

<> Code   ⊙ Issues   ⑂ Pull requests   ⊙ Actions   ⊞ Projects   📖 Wiki   ⚠ Security   📈

⑂ main ▼   ⋯

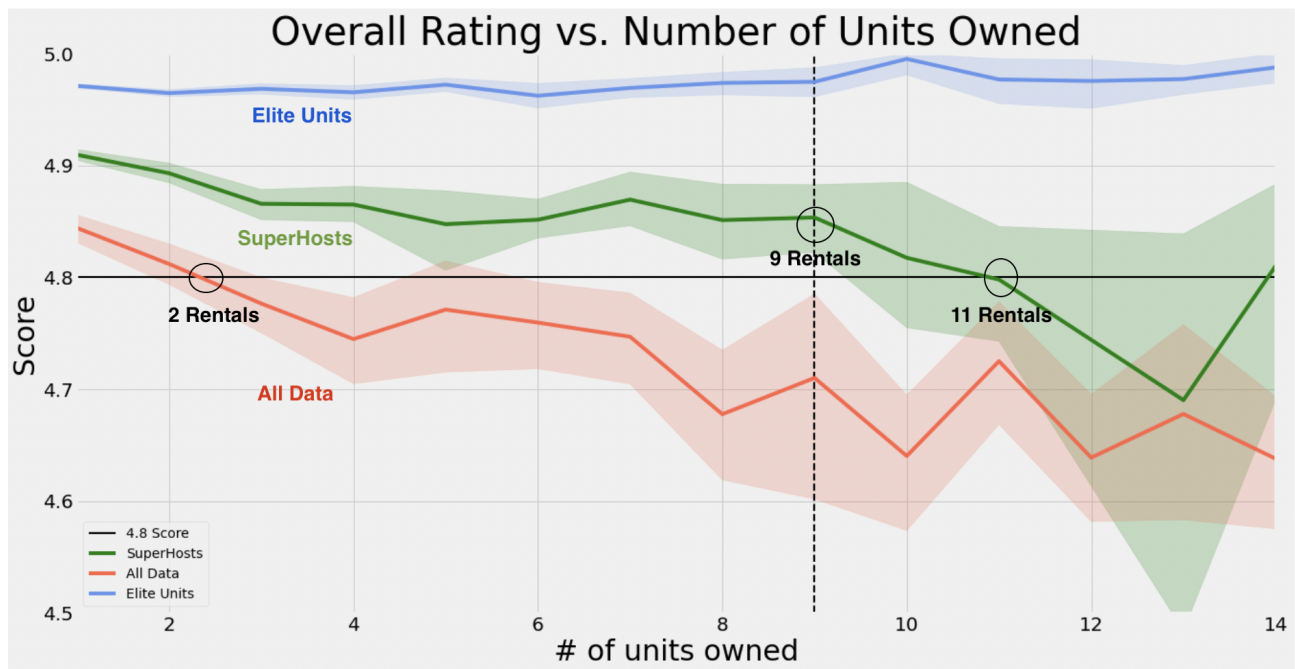jxn628 fixed link ⋯   3 minutes ago   🕐 63

View code

≡ README.md   ✏

Using Predictive Modeling to Get More 5-Star AirBnb Reviews

Airbnb has revolutionized the hotel and vacation industries by allowing people to easily rent their own houses out to other travelers. However, it can sometimes be hard for Airbnb Hosts to be as successful as they would like to be. The most common question from Hosts are "What can I do to get more 5-star reviews and a 5.0 Overall Rating?" This project uses data for the city of San Diego from InsideAirbnb.com to create a model which can predict whether a Rental Unit should get a 4.9-5.0 Overall Rating with 83% Precision.

Oceanside Property Management is a property management company located in San Diego California. Their main business is managing rental properties. However, they have recently noticed that a lot of Airbnb hosts have been reaching out to them for guidance. These hosts are mostly uninterested in having OPM manage their rentals, however they wany some help in increasing their success as Airbnb hosts.

There have been so many Airbnb hosts reaching out that OPM has decided that this can be a good side-business for them. They plan to officially add Airbnb consulting as a service that they provide. In their initial research they found that the top questions that potential clients who wish to utlize this service are:

- **1) "What can I do to get more 5 star ratings?"**
- **2) "Can you help me reach Superhost status? (or maintain Superhost status)**

These questions are understandable because Airbnb puts a huge focus on getting 5 star overall ratings. They also highly publicize the benefits of getting (and maintaining) Superhost status.

Oceanside Property Management has decided that the main focus of their service will be helping clients get more 5 star reviews. Therefore they have tasked me with providing the following:

- **1)A model** that will predict whether a specific rental unit should get a 5 Star Overall score based on other available information.
- **2)An industry analysis of AirBnb in San Diego.** Specifically looking for any insight that they can give to their clients that will give them a leg up on people who don't use their consulting service.

They also want me to answer the following questions:

- 1)Is there a significant advantage to being a Superhost? (is it worth all the effort to get this status and maintain it?)
- 2)How do we determine whether a Host "should" be getting 5 Star reviews?
- 3)What factors are most important in determining a 5 Star Overall Rating? (what aspects should they most focus on)

And finally, they want to know where their consulting service can make the most impact, so they know which features to market and/or which hosts to market to.

About AirBnb

Information from: https://listwithclever.com/research/airbnb-vs-hotels-study/ accessed 7/7/22

- Airbnb is becoming the preferred choice of vacationers — **60% of travelers who use both Airbnb and hotels prefer Airbnb over comparable hotels when going on vacation.**
- 68% of business travelers prefer staying in hotels when traveling for work, and they're more likely to have a negative experience at an Airbnb

Importance of 5 Star Reviews

AirBnb focuses on exceeding customer expectations, which is why they strictly **require that hosts maintain a near perfect rating in order to remain on the service.**

Importance of Superhost

- information from https://www.airbnb.com/d/superhost. Accessed 6/16/22

Advantages:

- Superhost badge to stand out among other hosts.
- Customers can filter search results to show only superhosts.

Requirements:

- Minimum 4.8 overall rating.
- 10 stays over the last year.
- < 1% Cancellation Rate.
- At least 90% Response Rate.

- Reassessed every 3 months.

Problems with the AirBnb Review Scale

The review data is incredibly skewed because Airbnb requires such a high rating. **Even though there is a 5 point scale, Anything lower than a 4.8 is seen as "bad".**

- So while this is technically a 5 point scale (as a reviewer can give 1 - 5 stars, with no partial stars allowed), getting a 4.0 average could result in being de-listed from the service!

The major problem with this review system is that **airbnb guests often assume that airbnb's review scale functions similarly to a hotel review scale, which also uses 5 stars**, with 3 considered average, 4 above average, and 5 star being the best possible experience.

from https://medium.com/@campbellandia/how-to-avoid-the-dreaded-4-star-review-a-guide-for-airbnb-hosts-cdf482d083fe

- *The problem stems from the fundamental difference in what most people think a 5-star rating system is, and what AirBnB's system actually is. The vast majority of people think that a 4-star review is perfectly appropriate; Their stay was good, they enjoyed themselves, but your place wasn't the Vanderbilt Suite at the Plaza. What they don't understand is that if a listing gets too many 4-star reviews the AirBnB platform begins to send warnings to hosts that their listing will be removed.*

The Problem
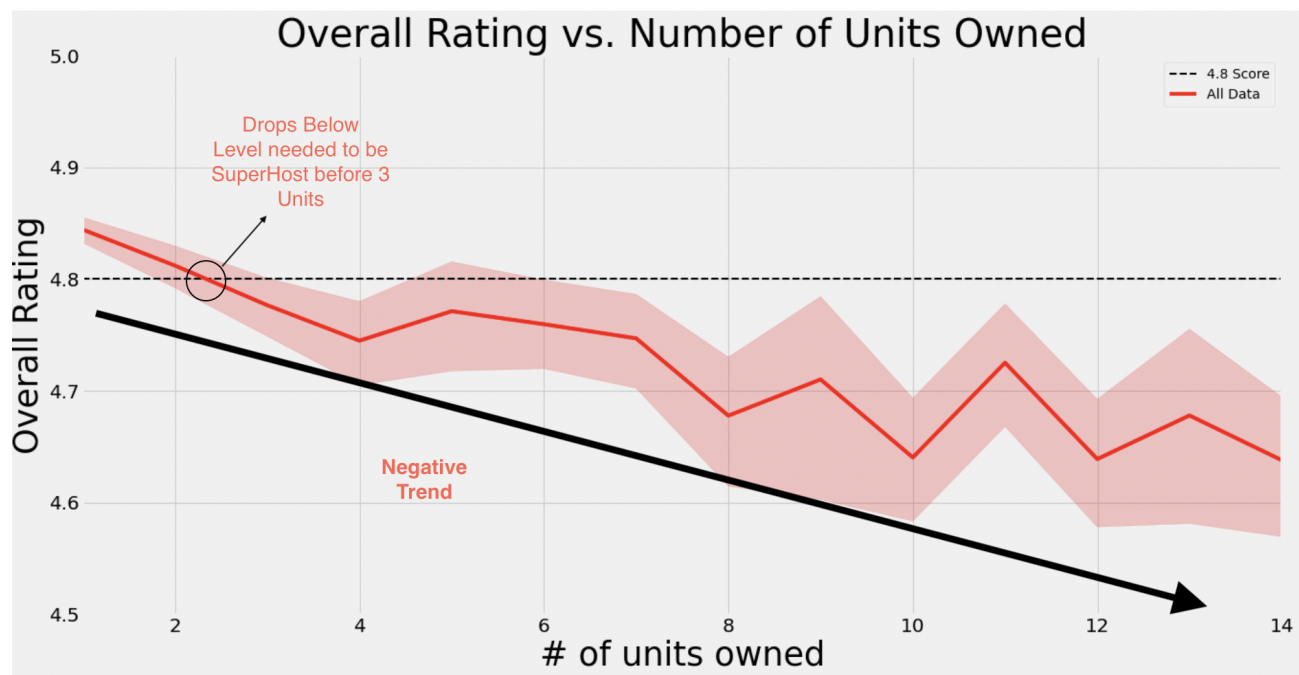
Target: Elite Units

- Elite Units: Any Unit with a 4.9-5.0 Overall Rating.
- - 4.9 is still an incredibly high score, and is above thresholds for success (4.8 rating, etc), so it is well worth capturing units with a 4.9 Rating as high performers as well.

## Choosing Model Evaluation Metrics

- My goal is to predict whether a person will get a 4.9-5.0 Airbnb Overall rating.
- Which is worse?
  - Model predicts that a unit is an Elite Unit, but they actually aren't? (more false Positives)
  - Model predicts that someone is NOT an Elite Unit but they actually are? (more false negatives)

Decision

- I want false Positives to be as low as possible.
- If my model says that a property is an Elite Unit, I want it to be true.
- If it misses some of the Elite units in the process, that is fine.
- **Therefore, I am most concerned with Precision, balanced out by F1 score.**

Final Model Evaluation: Decision Tree 2

- **Precision:** This Model correctly picks whether a rental will have an overall AirBnb rating between 4.9-5.0, 83% of the time.
    - This is 33% better than random guessing.
    - The Final Model is also a improvement over the baseline model. (about 7% better)
- **F1 Score:** While other models had slightly better F1 Scores, Decision Tree 2's F1 Score is only slightly worse. The F1 Score indicates that Precision is reasonably balanced with Recall, so I don't need to worry about this being an unbalanced and un-usable model. Therefore I'm fine choosing a model with a lower F1 in order to get more precsision.
- **ROC AUC Score:** Shows the True Positive Rate vs. the False Postive Rate. Some models had slightly higher scores than my Final Model, but again, it was very slight.
- **Cross Validation Score:** This model performs fairly well on data that it was not trained on and is comporable to the Cross Validation Scores of the other models.

Final Model Confusion Matrix
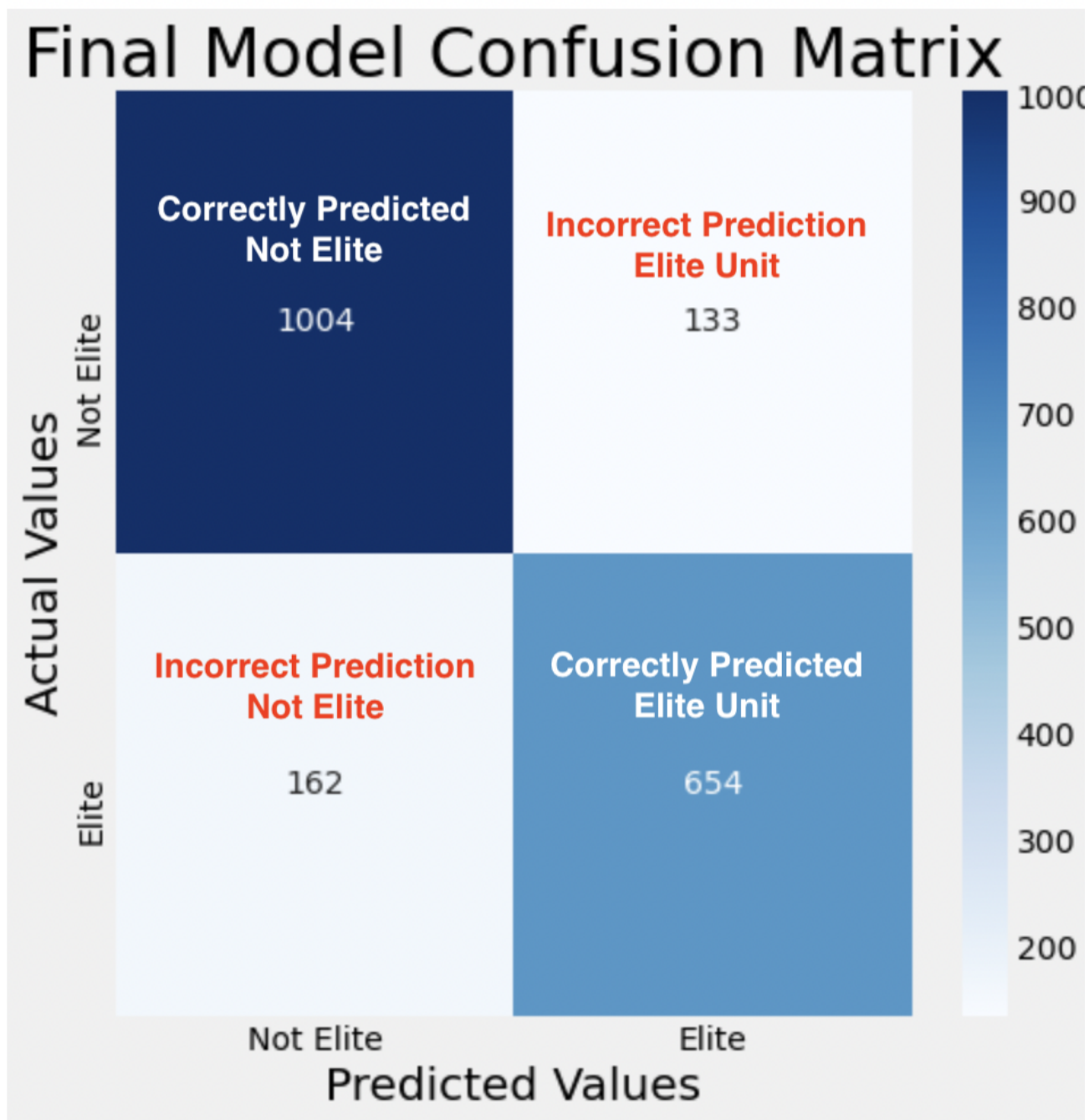
How to use This Model going forward:

- OPM can take the data from new clients and run the model to determine whether they are performing at 5-Star level or not.
- If they are, they should be able to obtain Superhost status and OPM can focus on helping them **maintain** everything that they are doing right.
- If they are not an Elite Unit, OPM can give them advice and help get them closer to a 5.0 Overall Rating.

Caveats:

- No model is perfect, and this one certainly isn't.
- This model relies on review scores from the 6 review categories. **If you don't have that data, the model does not perform reliably enough to be used.**
- That said, it can be reliably trusted as only 133 records from the test set of 1,953 were incorrectly labeled as being Elite Units when they were, in fact, not. (We aren't worried about the ones that were predicted to be not Elite incorrectly)

Top Features
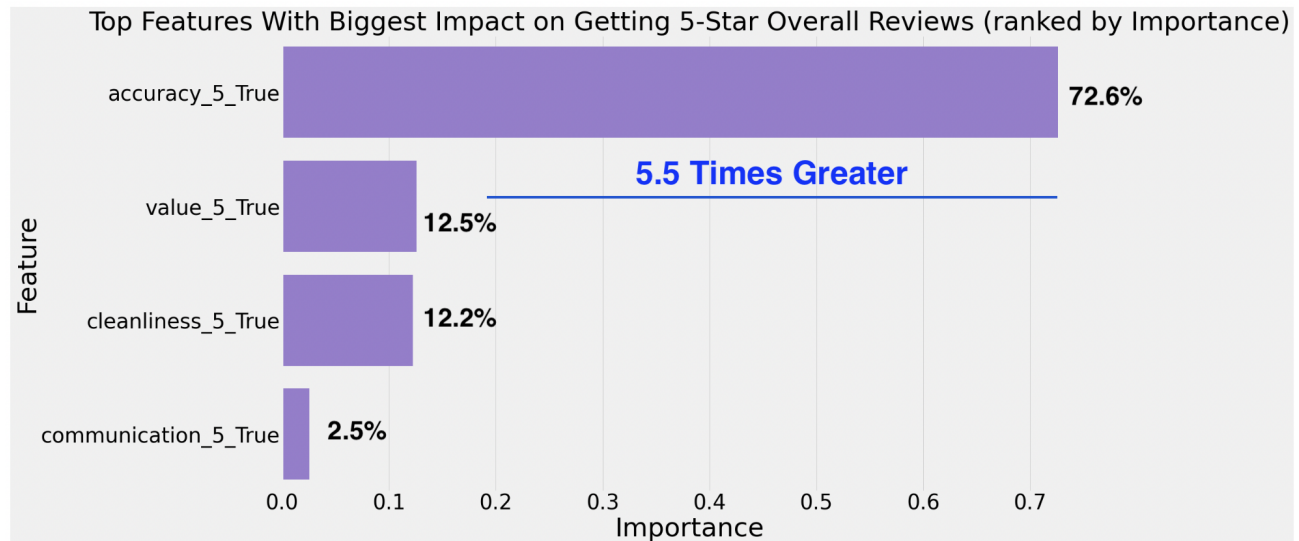
Analysis:

- **Accuracy is by far the most important feature**
- It is **5.5 Times more important** than the next features.
- **Value and Cleanliness are also important, but not nearly as much as Accuracy**
- **Communication** also has importance, but not nearly as much as the others.

Top Feature Accuracy Score

Accuracy is by far the most important feature in my model. Let's look at the relationship between Accuracy Score and Overall Rating.

Relationship Between Overall Rating and Accuracy Score
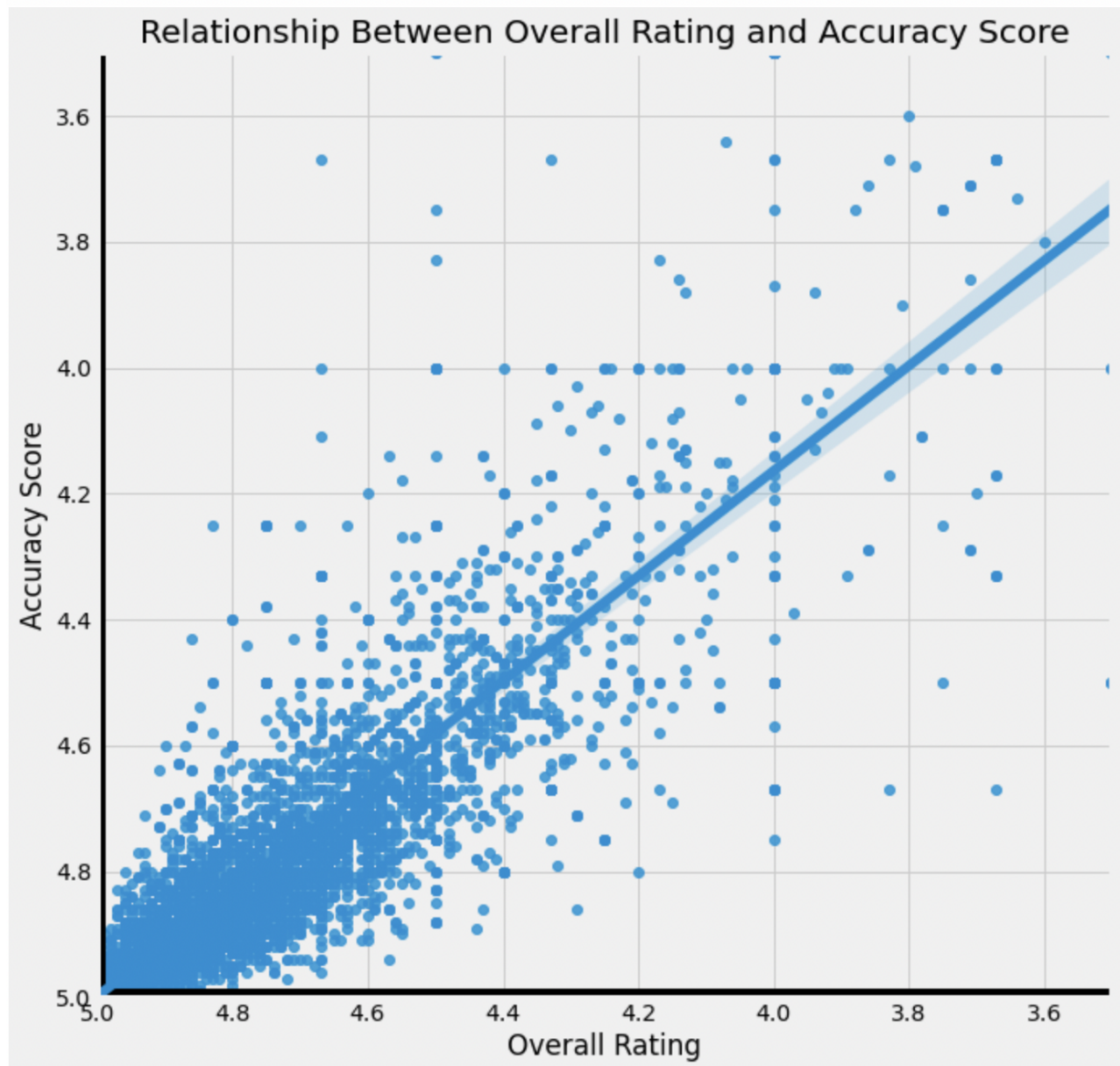
Analysis:

- There is a linear relationship between the two. Whatever the Accuracy Score is, the Overall Rating will likely be very similar as there is a nearly direct linear relationship.
- **Therefore, focusing on Accuracy is the best way to get 5 Star Reviews.**
- This matches what I found in my research. **The most important aspect of renting an AirBnb is that the listing is accurate, to ensure that Guest expectations are met.**
- Nearly all units that have an accuracy score of 4.9-5.0 also scored high in the other 5 review metrics.
- Nearly all units that did not have an accuracy score of 5 did not score highly on others as well.
- **Significantly more likely to be Elite Units and/or SuperHosts**

- More likely to have a 100% Response Rate
- **73% of units that scored 4.9-5.0 on accuracy were in our target 5-star range.**
- They are less likely to use the instant book feature, although 41% of units with 5.0 accuracy do each.

Questions Answered

Is there a significant advantage to being a Superhost? (is it worth all the effort to get this status and maintain it?)
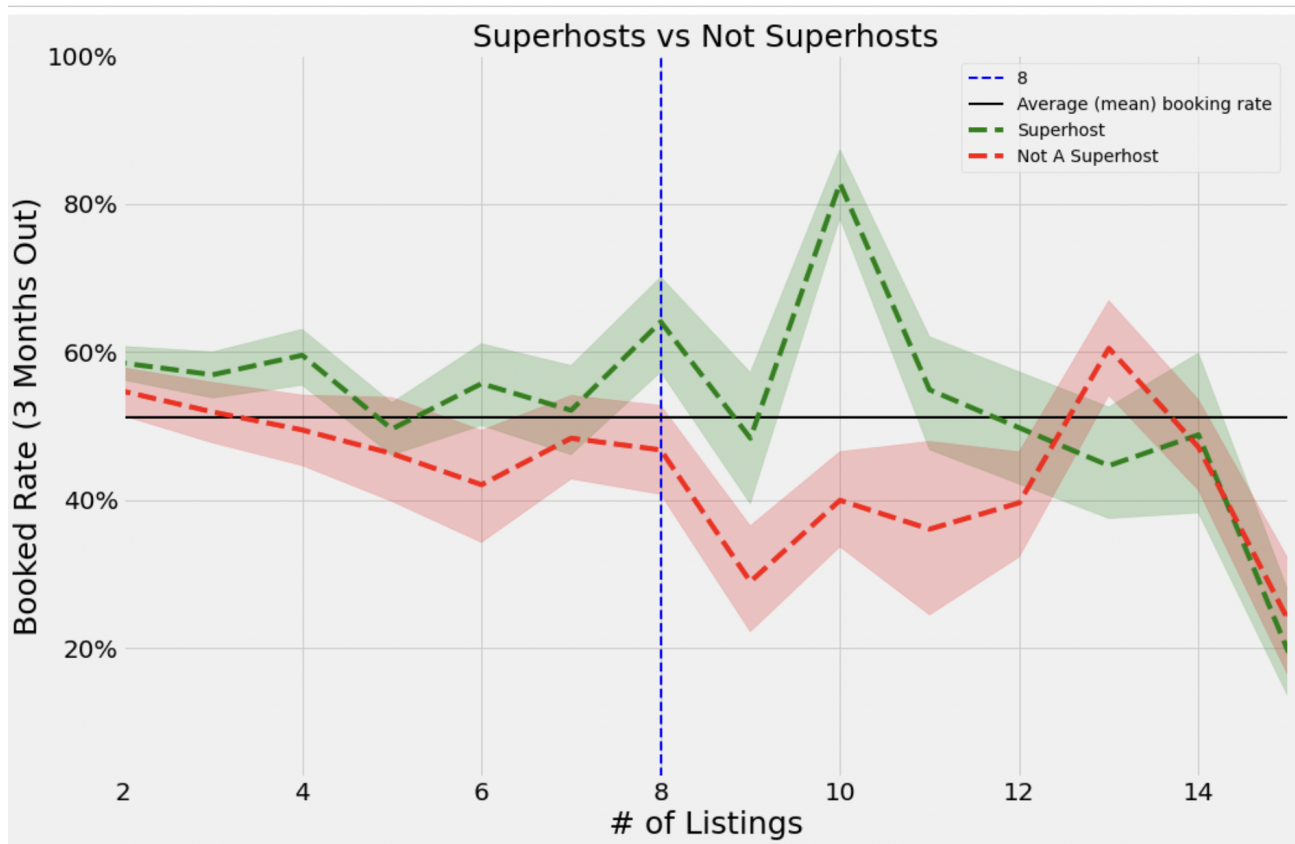
- **YES!**
- **Superhosts are 21% more likely to be Elite Units than non-superhosts.**
- Superhosts and the 4 Most Important Features:
- 81% of Superhosts have at least 4.9 Communication Score. (30% better than non-superhosts)
- 64% of Superhosts have at least 4.9 Accuracy Score. (26% better than non-superhosts)
- Superhosts have similar Value Scores to Non-Superhosts.
- Superhosts are better able to to handle higher numbers of listings.
- Most Superhosts can have 10 listings before it affects their 3 Month Booking Rate.

- Most Superhosts are also able to have 10 listings before their Overall Rating Drops below 4.8.

Is there a significant advantage to having Elite Units?
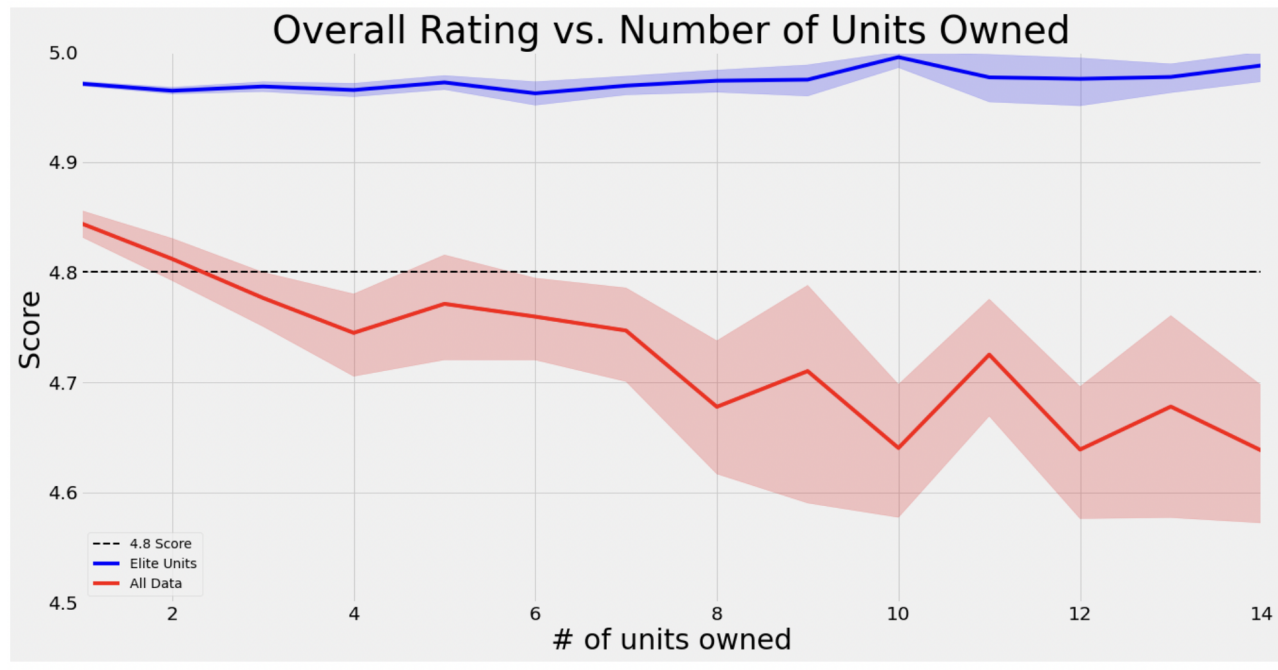
Analysis:

- While Elite Units perform slightly better in booking rate, **being an Elite unit is the best solution to the negative trend between Overall Rating and Number of Units.**
- **As long as you can keep your units performing at the highest levels, there is no limit on how many units you list.**
- The catch is of course, learning how many that you can manage and keep at that level. **This is an area where OPMs service will be invaluable!**
- Offer resources to help Hosts. (Preferrred cleaners, stagers, contractors for emergencies. Maybe even a dedicated customer service phone number)
- Most Notably, **Elite Units have the biggest increase in the Top Features: accuracy, cleanliness, value, and communication.**
- **This shows that the Target Feature does a good job of capturing the features that lead to more 5 Star Overall Reviews!**
- Elite Overall units score much higher in review metrics. This makes sense because they should have to score high in all of them to get a high overall score (even though it is a seperate metric in terms of AirBnb).
- They are also more likely to be a superhost, and more likely to have less than 5 listings.

- They are less likely to have high Capacity, or use Instant Book feature, but the differences aren't major.
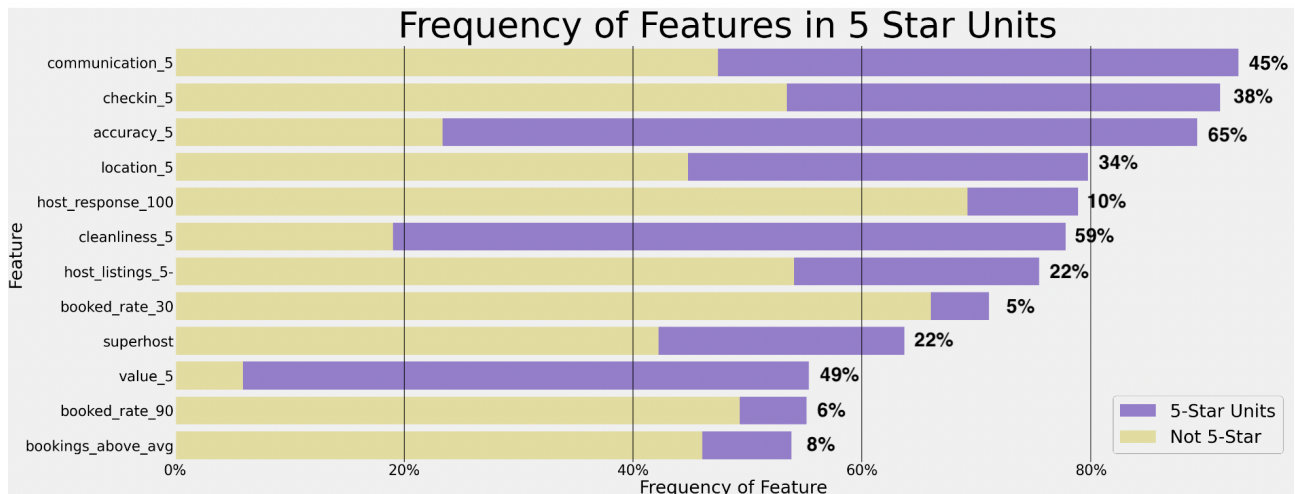
Conclusion

In my analysis of Airbnb rentals in San Diego California, I found that having a high overall rating (4.9-5.0), as well as having SuperHost status, were both beneficial to success on the platform.

- I also found that Accuracy was the biggest factor in getting a high overall rating, with a nearly 1 to 1 linear relationship.
- Other important features were Value, Cleanliness, & Communication.

Key Areas for OPM's AirBnb Host Consulting Service :

- **Accuracy:** By providing a listing service which assesses client's rental units and lists their units in such a way as to maximize the accuracy.
- **Bridging the Gap between Owner-Managed and OPM Managed:** OPM can provide a la carte services which help owners who wish to keep managing their own properties, but can't handle doing so at the highest quality levels. **This is also beneficial to OPM in creating a pipeline of potential fully managed units as hosts take on more properties that they can manage.** -- This can be structured in such a way to incentivize clients transitioning to OPMs full management service at certain thresholds (ie, 10 properties, etc).
- **Communication:** OPM can train hosts on what they can do to set expectations properly, and then exceed them with service (AirBnb's goal). This is done through how they communicate and how often they do it.
- **Provide Resources to Help Hosts Set Guest Expectations, and Then Exceed Them!**

- accurate listing

- explanation of airbnb's skewed review system.

- do this without being deceptive or cooercive.

- It doesn't matter if Hosts have all the metrics and analysis to **know** that their unit deserves 5-star reviews. Their fate is in the hands of the reviewers. If they really care about getting 5 star reviews (and they should since they are critical to success on AirBnb), they need to explain this to their guests.

Further Work

Use Natural Language Processing to analyze amenities.

- This DataSet includes **amenities, which would be very benefical to both the model and industry analysis.**

- However, they are all in string format and getting them into a useful format will be time intensive.

- Get them into a format where they can be one-hot encoded and fed into the model.

Increase the scope of this model.

- Incorporate data from the rest of California, and then the rest of the US.

Links

[Final Jupyter Notebook](#)

[Presentation](#)

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Languages

Languages

- **Jupyter Notebook** 100.0%