**ANALYZING THE IMPACT OF AMENITIES ON AIRBNB REVIEW SCORES AND GUEST SENTIMENT ON LISITNGS.**

**Team members:**

**Pavan Kola**

**Abhiram Karthik Meduri**

**Abstract:**

In the evolving landscape of short-term rentals, understanding what drives guest satisfaction and pricing on platforms like Airbnb is critical for hosts aiming to remain competitive. This study explores the relationship between listing amenities and guest sentiment, using both structured data (such as price, amenities, and review scores) and unstructured text (guest comments). Drawing from a dataset of over 201901 Airbnb listings, we analyzed how different amenities impact pricing and review scores through linear and ridge regression models. Our findings show that while amenities like kitchens, air conditioning, and self check-in contribute to higher prices, others such as Wi-Fi can unexpectedly lower guest ratings, possibly due to unmet expectations. Sentiment analysis was performed using VADER and RoBERTa models. Rather than predicting outcomes from sentiment, we focused on identifying the most extreme guest experiences and common keywords associated with highly positive and negative reviews. Comments mentioning clean spaces, great locations, and ease of access were closely tied to satisfaction, while complaints centered around cleanliness issues, bathrooms, and unclear check-in processes. This combined approach reveals the relevant and emotional elements that influence listing success and offers actionable guidance for Airbnb hosts.

**Introduction:**

Airbnb has dramatically transformed the way people travel, offering a decentralized marketplace for short-term lodging. With millions of listings available worldwide, competition among hosts is fierce. For hosts and property managers, distinguishing their listings in a crowded market is essential for securing bookings and ensuring positive guest experiences. While prior research has established that certain structural features influence pricing and satisfaction, fewer studies have integrated textual reviews to understand their combined effect.

Our study aims to bridge these two domains structured data and unstructured feedback by investigating the extent to which amenities affect listing price and guest satisfaction, and by analyzing guest comments to uncover the emotional themes that characterize high and low satisfaction stays. Through a dual approach involving statistical modeling and sentiment analysis, we attempt to answer not just what features matter to guests, but also why they matter. In doing so, we hope to equip hosts with actionable insights that allow them to improve the guest experience and optimize their pricing strategies in a data-driven and context-aware manner.

Our research focuses on two central questions:

1. Which amenities are most strongly associated with increased prices and high guest ratings?
2. What patterns can be observed in guest sentiment, and how do they reflect the positive or negative experiences expressed in Airbnb reviews?

Answering these questions can inform data-driven decisions for optimizing listings. Our approach blends traditional statistical modeling with natural language processing, examining both the amenities and guest perceptions components of listing success. Through this lens, we aim to provide a more comprehensive understanding of what makes a successful Airbnb listing.

**Literature Review:**

The success of short-term rental platforms like Airbnb has prompted extensive academic inquiry into the determinants of listing performance, particularly regarding pricing and guest satisfaction. Much of the early research focused on structured data such as listing characteristics, location, and host attributes to understand pricing dynamics and consumer decision-making.

Wang and Nicolau (2017) examined price determinants across 33 Airbnb markets and found that amenities, property type, and neighborhood characteristics significantly influenced listing prices. Their regression analysis revealed that listings offering conveniences like Wi-Fi and free parking typically commanded higher prices, reflecting guests’ willingness to pay for comfort and accessibility.

Tussyadiah and Pesonen (2016) approached the subject from a consumer behavior angle, highlighting how service-related features such as host responsiveness and property cleanliness affected booking decisions. Their findings emphasized that experiential factors were just as important as listing attributes in driving guest satisfaction and repeat bookings.

Gibbs, Guttentag, Gretzel, Morton, and Goodwill (2018) contributed to the growing literature on host-guest dynamics by studying the relationship between host professionalism and guest ratings. They concluded that professionalized listings (e.g., those with consistent branding, high-quality images, and prompt communication) received more favorable reviews. However, the study also acknowledged that subjective guest experiences often expressed through written reviews played a crucial but underexplored role in shaping outcomes.

More recently, research has turned toward text mining and sentiment analysis to extract insights from guest reviews. Mostafa (2013) introduced sentiment analysis in hospitality reviews using lexicon-based techniques and demonstrated that customer emotions directly influenced their satisfaction scores. Tools like VADER (Hutto & Gilbert, 2014) have since gained popularity due to their effectiveness in processing social media-style text with emoticons, slangs, and punctuation-based emphasis.

In the era of deep learning, transformer-based models like RoBERTa have significantly improved sentiment classification accuracy. Unlike rule-based tools, RoBERTa captures context and nuance, making it particularly useful for interpreting sarcasm, subtle dissatisfaction, or mixed emotions in Airbnb reviews.

Despite these advancements, a notable gap remains: most studies treat structured and unstructured data in isolation. Few have explored how these two data types can complement each other to provide a fuller understanding of guest preferences. Additionally, while sentiment tools have been used to categorize guest comments, they are rarely integrated with pricing or satisfaction prediction models.

Our plan:

Our study fills this gap by jointly analyzing structured features (e.g., amenities, room types) and unstructured guest feedback (via sentiment analysis) to identify what truly drives price and satisfaction on Airbnb. We do not just measure the presence of certain amenities or count review stars; instead, we analyze how emotional tone in guest language interacts with listing characteristics to influence outcomes. By combining traditional regression with sentiment interpretation, this research offers a more holistic and actionable framework for hosts aiming to optimize their listings in a competitive market.

**Methodology:**

We used a dataset containing 201,901 Airbnb listings from Inside Airbnb. The dataset includes 81 columns covering numerical variables (e.g., price, review\_scores\_rating), categorical variables (e.g., room\_type, property\_type), and textual reviews (comments).

1. Target Variables:

* price: A continuous variable indicating the nightly rate for each listing.
* review\_scores\_rating: A continuous variable representing the guest satisfaction rating

1. Input variables:

* Amenities
* Guest comments

1. Data cleaning:

* Out of the 81 columns, we have observed that three columns doesn’t contain any values. So we deleted all these three columns. Now, there were 78 columns available.
* Next, we sort the missing values for each column and then imputed them with the statistical methods, that is if the variable is integer or float we have performed mean, if the variable is string we have applied mode and imputed those values.
* Other variables containing missing values are not useful for our analysis, so we deleted those columns as well.

1. To conduct our analysis, we employed a combination of **exploratory data analysis, natural language processing,** and **linear regression modeling.**

Linear regression model:

We are going to perform linear regression to find the relationship between amenities and its impact on review\_scores\_rating and prices, we are also going to check the multicollinearity and perform ridge regression in order to avoid this.

Sentiment analysis:

We **engineered features** from both structured data (amenities and listing details) and unstructured text (comments). For sentiment analysis, we used **VADER**, a rule-based sentiment engine, and **RoBERTa**, a transformer-based model capable of understanding contextual nuance. These tools helped us extract sentiment polarity and classify reviews into Positive, Neutral, or Negative categories.

Reasons why we used these methods for sentiment analysis:

VADER: It is particularly effective when dealing with short, informal text that includes punctuation, capitalization, slang, or emojis such as customer reviews or social media comments. It’s a rule-based system that assigns a compound score to each sentence, helping us quickly assess whether a review leans positive, negative, or neutral. VADER is ideal when speed and general sentiment orientation are more important than deep contextual nuance. Because it is lightweight and fast, it allowed us to efficiently process thousands of Airbnb comments

RoBERTa: It is a deep learning model that excels when understanding the full **context**of a sentence is critical. It reads the structure of the language more like a human would, making it especially good at detecting **subtle sentiment shifts, sarcasm**, or comments that blend praise with criticism. While it is more computationally intensive than VADER, RoBERTa provides greater accuracy in scenarios where emotional tone is embedded in complex or nuanced language.

In our project we used both to compare the results it provided, VADER code took very less time to execute but it took a lot of time for RoBERTa, we used gpu’s to run the code quickly.

Exploratory data analysis:

A graph of a price

AI-generated content may be incorrect.

This chart shows how listing prices are spread across the platform. Most properties are priced between $50 and $150, with the highest concentration around $100. There's a steep drop in the number of listings as the price increases, and only a small number of listings are priced above $400. This suggests Airbnb primarily attracts budget-conscious and mid-range travelers, with luxury listings being rare.

A graph with numbers and a line

AI-generated content may be incorrect.

Review scores are heavily skewed towards the higher end. Most guests give ratings between 4.5 and 5.0, with a noticeable peak around 4.8. This trend indicates that guest satisfaction is generally very high. The small number of lower ratings likely reflects specific negative experiences rather than widespread issues.

A graph of a property type

AI-generated content may be incorrect.

This boxplot highlights the variation in pricing across different property types. Entire homes, entire townhouse, and villas are among the most expensive, which makes sense because of privacy and space available in them. At the same time more unconventional options—like shared room in rental unit or tiny houses tend to be cheaper and show less price variability. Outliers (very expensive listings) are present across many categories but are most common among luxury property types.

A graph of a number of rooms

AI-generated content may be incorrect.

As expected, entire homes and apartments command the highest prices. Private rooms are more affordable and have a tighter distribution, indicating more pricing consistency. Hotel rooms fall somewhere in between. Shared rooms are the cheapest option, making them ideal for budget travelers. The outliers in each room type suggest that some listings charge unusually high prices, possibly due to location, seasonality, or luxury amenities.

A graph of a review score

AI-generated content may be incorrect.

Shared and private rooms tend to receive very consistent and high review scores. Interestingly, hotel rooms show greater variation and sometimes lower ratings, possibly due to higher guest expectations or standardized service. Entire apartments still maintain strong ratings, but the range is wider, which may reflect inconsistencies in host service or property condition.

**Results and Discussion:**

After cleaning the dataset, we checked the descriptive statistics of some important target variables:

Price:

* Mean: $178.549
* Median: $122
* Range: $10 to $10,000
* Standard Deviation: 332.116

Review Scores Rating:

* Mean: 4.81635
* Median: 4.86
* Range: 1 to 5
* Standard Deviation: 0.17362

Most ratings were listed between 4.5 to 5, this is also shown in Exploratory data analysis this might indicate that there are many positive reviews of the Airbnb hotels.

**Amenities vs Review scores:**

Here the ridge regression is performed, because it is good for handling multicollinear variables, this yielding r-squared value of about 24.8%, although it’s low but since it is human related review data this can be considered good.

For this, we have preprocessed the data by labelling the amenities columns to its individuals by using multilabel binarizer and studying each amenity.

Key findings:

The key findings were determined by looking at the coefficients of the amenities.

Some amenities contributed to the positive review scores, while some contributed to negative scores because of the poor presentation and poor equipment and functioning.

 Amenities such as extra pillows and blankets, keyless entry, hair dryers, and exercise equipment are linked to higher guest ratings. These thoughtful, comfort-driven features enhance the overall stay.

 Surprisingly, some standard features like Wi-Fi, cribs, and coffee makers had small negative coefficients, which could suggest that when expectations are unmet, they harm satisfaction more than help.

**Amenities vs Price:**

Here, the price is the target variable, this tells us which amenities contribute to higher prices and which contribute to lower.

Air conditioning : 0.067

Self check-in : 0.157

Lock on bedroom door : -0.152

These are some of the results, from these we can say that:

 Self check-in is associated with higher prices. Guests often value the flexibility of being able to arrive on their own schedule without coordinating directly with the host, which can make a listing more appealing especially for travelers arriving late or staying just one night.

 Air conditioning also contributes to higher pricing. This isn’t surprising, especially in warmer climates where comfort is a priority. Guests are often willing to pay more for properties that ensure a pleasant indoor environment.

 On the other hand, listings with a lock on the bedroom door tend to be priced lower. This likely reflects the fact that these listings are often in shared homes, where the lock was mentioned and not available.

**SENTIMENT ANALYSIS**

VADER:

This classified the comments into :

**vader\_sentiment\_category**

**Very Positive 146719**

**Positive 43646**

**Neutral 8174**

**Negative 2247**

**Very Negative 1115**

After this we have found the top 5 and bottom 5 comments of customers and their respective listing id’s:

We analyze them and sort them according to best and worst by looking at the compound scores:

A graph with red and green bars

AI-generated content may be incorrect.

We also extracted the comments of these listing id’s.

After this, we thought to found the reasons behind these positive or negative reviews and found the most common words of these reviews, by analyzing them we can find the reason behind their performance:

A close-up of words

AI-generated content may be incorrect.

Most common words: Great location, Great place, downtown dallas, nice place, great host.

Key finding and suggestions:

Everything here in this word cloud is related to the location, this means if the location is in downtown Dallas and everything is within walking distance then the customer satisfaction is going to be higher. Also being a friendly host is also important.

A close up of words

AI-generated content may be incorrect.

This gives us what else to improve for better customer satisfaction:

Most common words: host, bathroom, broken, clean, place, somewhere

Key findings and suggestions:

* Most of the negative reviews contain about host so, the host should improve his behavior and stay friendly, may be he needed some training in management.
* Somewhere, place: The place also played a negative role, the locations are somewhere outside, so the location should be familiar and should be in main areas.
* Bathroom, dirty, clean, clean, towel : These reflect that if the properties are not maintained well then the customer satisfaction decreases. Everything even the minute amenities like towel effect the customer satisfaction.

RoBERTa:

This is a very complex analysis, the same analysis is performed here similar to VADER. But here the process is different, we used lemmatization for creating tokens and then loaded Roberta transformer model.

According to RoBERTa model:

**roberta\_sentiment\_category**

**Very Positive 179368**

**Positive 10834**

**Neutral 7333**

**Negative 2924**

**Very Negative 1442**

Most frequent words for positive review:

great: 111645

clean: 49695

location: 49072

nice: 34475

comfortable: 27279

perfect: 25870

beautiful: 20594

easy: 20217

space: 2021

* Words like “great”, “clean” and “location”  appeared most often in positive comments. These reflect the top priorities for many guests cleanliness, a convenient or desirable location, and an overall great experience.
* Other frequent terms such as “nice,” “comfortable,” “perfect,” and “beautiful” suggest that guests appreciated not just functionality, but also the aesthetic and emotional feel of the space.
* The word “easy” (e.g., easy check-in or easy communication) also appeared frequently, indicating that hassle-free interactions are an important part of a satisfying stay.
* Although “space” appears much less frequently than other words, its presence signals that guests value a well-designed or spacious environment.

Most frequent negative words:

clean: 962

door: 852

location: 844

time: 788

bed: 780

dirty: 773

bathroom: 734

check: 713

Key findings and analysis:

* Surprisingly, the word “clean” still appears prominently in negative comments (962 mentions), which suggests that when cleanliness is lacking, guests are quick to point it out.
* Words like “door,” “bed,” “bathroom,” and “dirty” reflect specific complaints about room condition and maintenance. These likely highlight issues that disrupted guest comfort or safety.
* The word “location” also appears in negative contexts (844 times), indicating that despite being a strength for many listings, poor or misleading location information can lead to dissatisfaction.
* “Time” and “check” likely refer to check-in/check-out issues—either delays or unclear instructions emphasizing the importance of smooth logistics.

**Overall findings and relevance to research questions:**

1.Key Insights from Structured Data:

* Air conditioning, kitchen, and self check-in significantly increased listing prices.
* Extra pillows, hair dryers, and keyless entry were linked to higher guest ratings.
* These results align with prior studies (e.g., Wang & Nicolau, 2017) that highlight the role of functional amenities in determining value.

2. Guest Expectations and Surprising Findings:

* Surprisingly, Wi-Fi, cribs, and coffee makers showed a negative impact on review scores.
* This suggests that when expected amenities underperform, they hurt guest satisfaction more than help.
* It reinforces that expectation management is just as important as the presence of a feature.

3. Sentiment Analysis Insights:

* Positive reviews included words like “great,” “clean,” “location,” and “comfortable.”
* Negative reviews focused on “dirty,” “bathroom,” “door,” and “check-in” issues.
* This mirrors earlier literature (e.g., Tussyadiah & Pesonen, 2016) that emphasized the emotional component of the guest experience.

4. Role of VADER Sentiment:

* The VADER compound score had a positive and statistically significant correlation with review scores.
* Listings with higher sentiment scores in comments also received higher ratings, confirming emotional tone as a useful indicator of guest satisfaction.

5. Contribution to Literature:

* While previous research often treated structured and unstructured data separately, this study integrates both.
* It builds on Mostafa (2013) by not just classifying sentiment but using it to contextualize guest expectations and reactions.

**Conclusion:**

This study examined the combined influence of amenities and guest sentiment on Airbnb listing performance, using both structured data and unstructured review text. Our analysis led to several key takeaways:

* Amenities Matter: Features like air conditioning, kitchen access, and self check-in consistently contributed to higher prices. Comfort-enhancing items such as extra pillows and smart locks were linked to better review scores.
* Expectations Shape Satisfaction: Surprisingly, amenities like Wi-Fi and cribs, which guests assume to be standard, were linked to lower ratings when expectations were not met.
* Sentiment Adds Depth: Sentiment analysis using VADER revealed that listings with more positive emotional language in reviews also received higher guest ratings. The most common positive words included “great,” “clean,” and “comfortable,” while negative experiences were often linked to cleanliness, check-in issues, and bathroom concerns.

Together, these findings underscore the importance of not just offering amenities, but maintaining them at a level that meets or exceeds guest expectations. Hosts can use these insights to adjust their offerings and presentation, potentially increasing revenue and guest satisfaction.

Recommendations for Future Research:

* Incorporate geographical data to analyze how sentiment and amenity importance vary by region or city.
* Apply time-series sentiment tracking to measure how reviews evolve over time with changes in service or amenities.
* Explore image-based features (e.g., listing photos) as another layer of guest perception and decision-making.

References:

* Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216–225.
* Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*. <https://arxiv.org/abs/1907.11692>
* Mostafa, M. M. (2013). More than words: Social networks’ text mining for consumer brand sentiments. *Expert Systems with Applications, 40*(10), 4241–4251. https://doi.org/10.1016/j.eswa.2013.01.019
* Tussyadiah, I. P., & Pesonen, J. (2016). Impacts of peer-to-peer accommodation use on travel patterns. *Journal of Travel Research, 55*(8), 1022–1040. <https://doi.org/10.1177/0047287515608505>
* Wang, D., & Nicolau, J. L. (2017). Price determinants of sharing economy-based accommodation rental: A study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management, 62*, 120–131. <https://doi.org/10.1016/j.ijhm.2016.12.007>