EXECUTIVE SUMMARY

Unlocking Growth Opportunities: Leveraging Text Analysis for Airbnb Listing Optimization

Sara Kuqja

Master's in Business Analytics, MBAN1

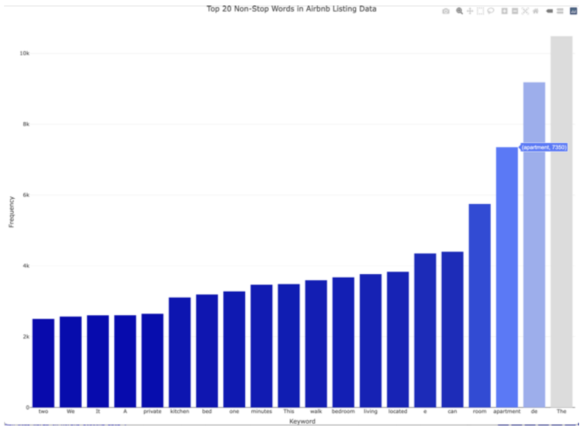
Boston, MA 202

INTRODUCTION:

This analysis aims to extract actionable insights from Airbnb listing data to inform business strategies and decision-making. The dataset comprises text descriptions of listings, including names, summaries, spaces, and descriptions. By leveraging text analysis techniques, we uncover valuable insights to enhance listing attractiveness, pricing strategies, and customer satisfaction.

1. DATA PREPARATION AND EXPLORATION:

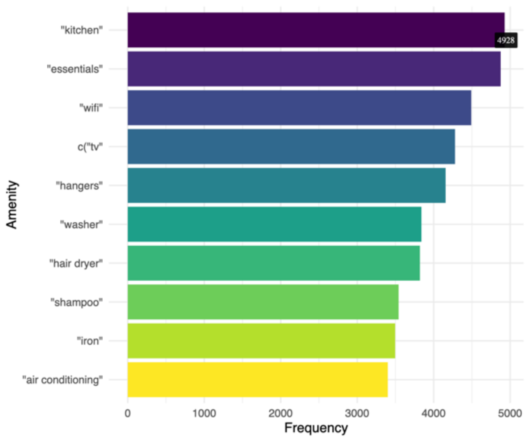
We combined multiple text columns to create a comprehensive text corpus for analysis. Common keywords were identified to understand the prevalent themes across listings. While the initial top words included generic terms such as 'The,' 'de,' and 'can,' further analysis revealed insightful keywords like 'apartment,' 'room,' and 'located.' These keywords provide valuable insights into the unique selling points of Airbnb accommodations, emphasizing features like 'cozy,' 'downtown,' and 'spacious' that attract guests.



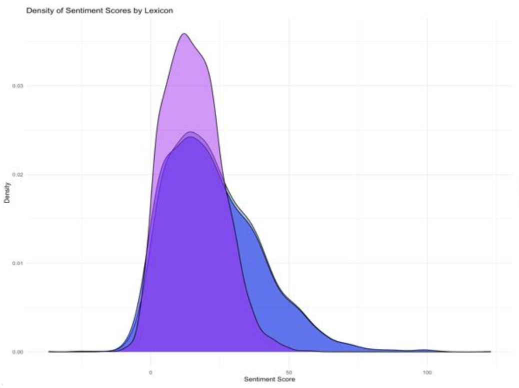
2. AMENITIES ANALYSIS:

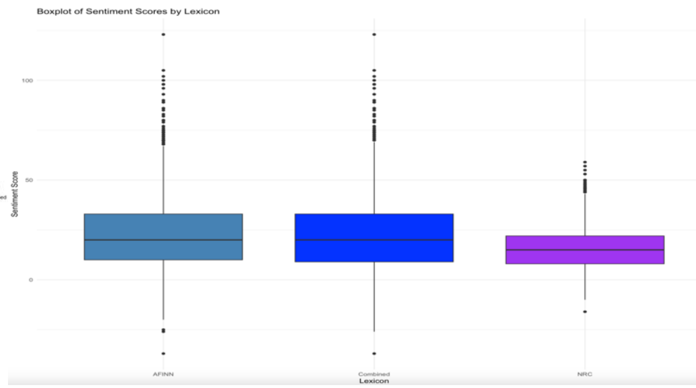
An analysis of amenities frequency revealed the most common amenities mentioned in listings.

- Amenities like "kitchen," "wifi," and "TV" indicate guests prioritize essential facilities and modern conveniences.
- Features such as "essentials," "hair dryer," and "air conditioning" emphasize comfort and convenience, enhancing the appeal of your listings.
- Comparing amenities with competitors can highlight strengths and areas for improvement, ensuring your properties remain competitive.
- Providing sought-after amenities like "washer" and "iron" contributes to positive guest experiences and reviews.
- Highlighting popular amenities in your listings attracts guests seeking specific features during their stay.



3. SENTIMENT ANALYSIS:



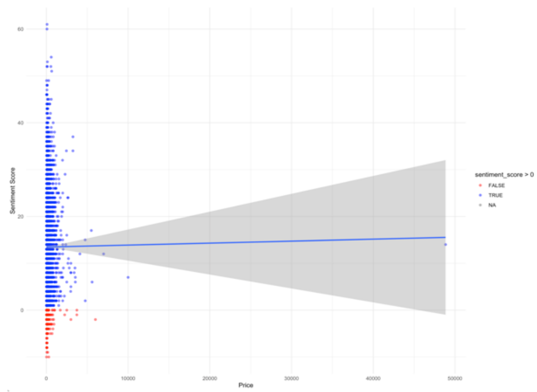


Distribution Consistency: AFINN and NRC show similar sentiment distributions, indicating alignment in sentiment intensity interpretation.

Balanced Sentiment: The "Combined" score falls between AFINN and NRC peaks, suggesting a balanced approach to sentiment analysis.

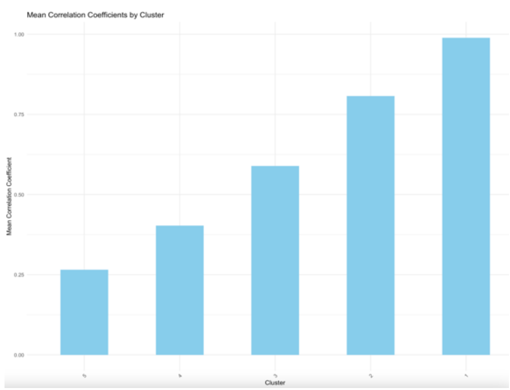
AFINN Sensitivity: AFINN exhibits wider sentiment variability and more extreme scores, potentially capturing diverse emotional expressions within Airbnb listings.

Furthermore, we explored the correlation between listing prices and sentiment scores, revealing how positive language could potentially influence a listing's pricing strategy.



Listings tend to have positive descriptions, but the sentiment does not strongly influence the price. This could mean that while hosts try to sound positive, travelers may not be willing to pay significantly more for a listing just because the description has a positive tone. Other factors likely play a more significant role in pricing. These might include location, amenities, size of the listing, and host reputation.

Performing hierarchical clustering based on correlation coefficients:



Note: for the tokens refer to the R- script aribnb_analysis.r

1. Cluster 5’s tokens like 'delay' and 'error' highlight critical issues that need immediate attention to mitigate negative impacts on performance.

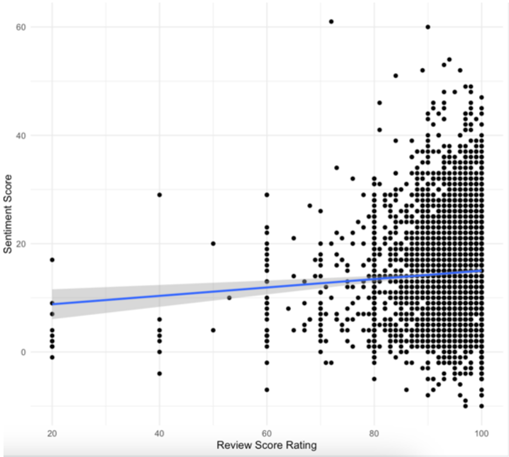
2. Tokens such as 'wait' and 'retry' in Cluster 4 suggest minor inefficiencies that could be streamlined to improve user experience.

3. The neutral tokens in Cluster 3, like 'data' and 'request', appear to have no significant effect and could be deprioritized in optimization efforts.

4. For Cluster 2, tokens like 'load' and 'connect' show a positive influence that should be reinforced to enhance system responsiveness.

5. The strong positive tokens in Cluster 1, such as 'success' and 'complete', are key performance drivers and should be the focus of best practices and resource allocation.

4. REVIEW ANALYSIS:



1. There is a positive association between sentiment scores and review scores, implying that more positive reviews are likely to result in higher overall ratings for listings.

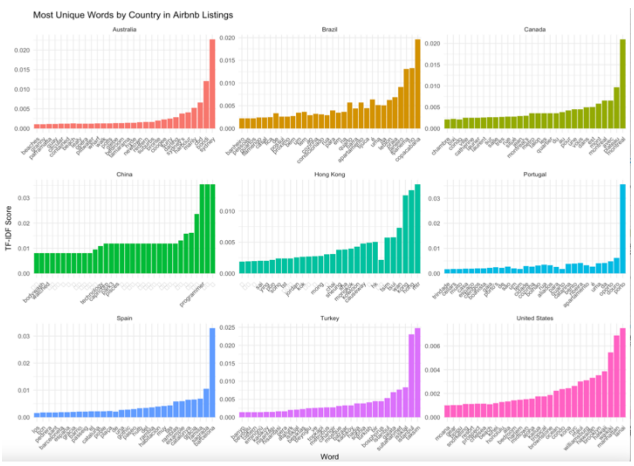
2. The data dispersion suggests variability in sentiment at lower review scores, which could indicate opportunities for targeted improvements in customer experience in these segments.

3. The dense clustering of data points at the higher end of the review scores suggests a ceiling effect, where despite high sentiment scores, the review scores do not proportionally increase, highlighting the need for differentiated strategies to further enhance the guest experience.

4. The breadth of sentiment scores at similar review score levels implies that guests weigh factors differently, underscoring the importance of personalized responses to guest feedback.

5. The visualization underlines the necessity of focusing on qualitative guest feedback to maintain high review scores, as there is not a one-to-one increase in review scores with sentiment scores.

5. GEOLOCATION



1. Unique words/phrases reveal cultural/linguistic differences (e.g., "brekkie" for Australia, "ar condicionado split" for Brazil).
2. They indicate popular property types/amenities in each country (e.g., waterfront for Canada).
3. High TF-IDF scores identify distinctive location-specific words/phrases for differentiating listings.
4. Incorporating these terms can improve visibility for local searches and SEO.
5. Localization and culturally-relevant messaging are important for resonating with audiences.
6. Analysis reveals potential new markets/segments with distinct word patterns.

For further insights into the distribution and importance of words in Airbnb listings' text data, refer to the TF-IDF framework highlighting significant words for each country and Zipf's Law demonstrating a statistical regularity in word frequency distribution in the r-script.

CONCLUSION:

Refering at the R-script we have build a machine learning model to predict whether Airbnb listings are booked or not based on textual information such as name, summary, and description.

Given the nature of the data here the target variable 'booked' is defined based on whether the listing has reviews.

The code aims to develop a predictive model that can classify Airbnb listings as either booked or not booked based on textual information, providing valuable insights for property owners and managers.

Future steps using the forecasting model:

Optimize Listings:

- Identify key features associated with bookings.
- Provide guidelines for hosts and implement a listing review process.

Improve Customer Experience:

- Enhance listing descriptions and offer customer service training.

Targeting and Positioning:

- Analyze market demand and tailor marketing campaigns.

Resource Allocation:

- Prioritize support for high-potential listings and invest in platform features.

Demand Forecasting:

- Implement dynamic pricing tools and plan for peak times.

Risk Management:

- Monitor and support low-performing listings and adjust cancellation policies.

Feedback Loop and A/B Testing:

- Collect feedback, conduct A/B testing, and refine the predictive model.