Aspect-Based Sentiment Analysis of Airbnb Reviews:

Insights from the Melbourne Dataset

Yunzhen WU

1. Introduction and Problem Statement

The emergence of peer-to-peer accommodation platforms such as Airbnb has revolutionized the hospitality industry by enabling individuals to offer unique lodging experiences. While this model provides travelers with more personalized stays, the rapid growth of listings and reviews has made it increasingly difficult for hosts and platform operators to extract actionable insights from unstructured guest feedback. Traditional numerical ratings often lack the depth needed to capture specific aspects of satisfaction or dissatisfaction expressed in textual reviews.

To address this gap, this project applies aspect-based sentiment analysis (ABSA) to Airbnb guest reviews from Melbourne, Australia, leveraging a publicly available dataset from Kaggle. The dataset includes extensive records of Airbnb activity in Melbourne—a city known for its high Airbnb density—including listing information, calendar availability, and guest reviews. This study focuses exclusively on its textual reviews, which offer a rich source for extracting both sentiments and topics of concern.

Our approach employs a structured natural language processing (NLP) pipeline that includes data preprocessing, aspect extraction using three methods—rule-based keyword matching, Latent Dirichlet Allocation (LDA), and BERTopic—and sentiment analysis using both VADER (a lexicon-based tool) and BERT (a transformer-based classifier). We further conduct aspect-sentiment cross-analyses to explore the relationships between key topics and emotional tone, enabling the identification of specific areas for service improvement.

This project builds upon and extends prior research in sentiment analysis and opinion mining. For example, Tussyadiah and Zach (2016) highlighted experience-based attributes through content analysis of hospitality reviews, while Zhu et al. (2022) demonstrated the advantages of contextual embeddings in ABSA for product reviews. Poria et al. (2016) proposed multi-level sentiment analysis frameworks, providing useful foundations for modeling aspect-specific opinions. Drawing from these studies, our project integrates interpretable techniques with modern language models to strike a balance between transparency and accuracy. This hybrid approach supports Airbnb hosts and platform operators in identifying service strengths and potential areas for improvement.

2. Research Design and Modeling Methods

2.1 Data Preparation and Preprocessing

The dataset used is the *reviews_dec18.csv* file from the Kaggle Airbnb Melbourne collection, containing over 480,000 reviews. The project focuses exclusively on the textual reviews and review IDs in order to streamline the analysis. The preprocessing pipeline includes:

- Normalization: Texts were cleaned by removing punctuation, HTML tags, and special characters, while also applying lowercasing and repeated character normalization.
- Tokenization: Each review was split into tokens using NLTK's *word_tokenize*.
- Lemmatization: Words were reduced to their root forms using WordNet's lemmatizer.
- Stopword Removal: Both standard English stopwords and a custom set (e.g., "would", "much", "very", "airbnb", "melbourne",) were removed to focus on meaningful tokens.
- Corpus Inspection: Word clouds and top-N word frequency charts were generated to provide an initial understanding of the corpus.

This preprocessing workflow produced a final cleaned and tokenized corpus (*final_docs*), which serves as the input for the aspect extraction and sentiment analysis tasks.

2.2 Aspect Extraction Methods

Three complementary approaches were implemented to capture aspects from guest reviews:

- Rule-Based Keyword Matching: 9 aspect categories (location, cleanliness, communication, amenities, comfort, value, checkin, recommendation, host) were manually predefined, and then relevant tokens were assigned to each category. This method provides transparency and high interpretability.
- Latent Dirichlet Allocation (LDA): Unsupervised topic modeling was applied using Gensim's LdaModel, trained with 5 topics and 10 passes. The *num_topics* parameter was selected based on topic coherence and interpretability, with each document assigned a dominant topic.
- BERTopic: BERTopic was used for advanced topic modeling. Due to memory constraints and the large scale of dataset (~480k reviews), a stratified random sample of 10,000 reviews was used. The model captured nuanced topics across the review corpus, linking them back to individual reviews via *bertopic_topic* and *bertopic_probability*.

Each aspect extraction method produces an individual column in the merged dataframe *aspect_df*, allowing for comparative analysis and flexibility in downstream tasks.

2.3 Sentiment Analysis Methods

Two techniques were employed to assess the emotional tone of each review:

- VADER (Valence Aware Dictionary and sEntiment Reasoner): Applied directly to the cleaned review texts, providing rule-based polarity scores (positive, negative, neutral) and enabling quick insights across the dataset.
- BERT-Based Sentiment Analysis: Leveraging Hugging Face's pipeline with the distilbert-base-uncased-finetuned-sst-2-english model, reviews were passed through the

model with a batch size of 16 for efficiency. The output labels (“positive”/“negative”) and confidence scores were mapped to the standardized sentiment categories.

2.4 Aspect-Sentiment Cross-Analysis

The final step involved merging aspect extraction outputs with sentiment predictions. By linking aspect categories (from rule-based, LDA, BERTopic) to sentiment scores (from VADER and BERT), the project explored which aspects are most associated with positive or negative sentiments. This cross-analysis provides rich, interpretable insights.

3. Results

3.1 Exploratory Data Analysis

The EDA revealed distinct patterns in the Airbnb guest reviews. The word cloud (Figure 1) highlights prominent terms such as great, location, place, host, and clean, indicating reviewer focus on property quality, convenience, and cleanliness. This is supported by the top 20 word frequency chart (Figure 2).

3.2 Aspect Extraction Results

(1) Rule-Based Aspect Extraction

The rule-based method captured frequent and meaningful aspects such as location, host, communication, cleanliness, and comfort. These align with typical Airbnb evaluation points and are distributed as shown in Figure 3.

(2) LDA Topic Modeling

LDA generated five interpretable topics. These include themes like public transport and location (Topic 0: “close”, “tram”, “walk”, “station”, etc.), hospitality and home comfort (Topic 1: “home”, “recommend”, “back”, etc.), room conditions (Topic 2: “bed”, “bedroom”, “kitchen”, “bathroom”, etc.), communication and service quality (Topic 3: “communication”, “easy”, “host”, etc.), and aesthetics (Topic 4: “nice”, “clean”, “lovely”, etc.). The distribution of review counts across these topics is shown in Figure 4, where Topic 0, 1 and 3 dominates.

(3) BERTopic Modeling

BERTopic produced over 100 fine-grained topics. The intertopic distance map (Figure 6) shows clear separation among top topics, and the topic frequency distribution (Figure 7) highlights semantic diversity. The top topics shown in Figure 5 include urban location (Topic 0), transport access (Topic 1), cleanliness and tidiness (Topic 3), dining and nightlife (Topic 4), sleeping comfort (Topic 5), communication (Topic 6), local attractions (Topic 7), amenities (Topic 8), booking issues (Topic 10), value for money (Topic 11), and family-friendly homes (Topic 12). However, certain topics reflect noisy or incoherent content (Topic 2), while others exhibit divergent sentiment patterns, warranting deeper investigation.

3.3 Sentiment Analysis Results

(1) VADER Sentiment Analysis

VADER analysis reveals that over 90% of reviews are positive (Figure 8). Neutral and negative sentiments are minimal. Aspect-level results (Figure 9) show high positivity across rule-based themes. LDA topic-level analysis (Figure 10) reflects the same pattern, though Topic 0 has a slightly higher negative proportion. For BERTopic (Figure 11), most topics remain positive, but Topic 2 and 10 show high neutrality, approaching 100%, while Topic 14 has elevated negativity (~17%).

(2) BERT Sentiment Analysis

BERT results mirror VADER's trends (Figure 12), with high positive classifications and strong model confidence (Figure 13). Rule-based aspects under BERT (Figure 14) show high positivity, but with check-in showing slightly more negative reviews. LDA topic sentiment (Figure 15) points to Topic 2 as relatively negative. BERTopic aspect-level sentiment (Figure 16) indicates that most topics are positively perceived, but Topic 10 (reservation issues) and Topic 14 (cleaning fees) again stand out for having higher non-positive sentiment.

4. Analysis and Interpretations

The combined findings of lexical analysis, topic modeling, and sentiment classification reveal consistent themes in Airbnb guest feedback, while highlighting specific topics that diverge in sentiment or semantic clarity.

Guest emphasis on location, cleanliness, and host interaction appears repeatedly across EDA, rule-based aspects, and topic models. This convergence supports the reliability of these factors as core to positive guest experience. The strong lexical alignment between frequent terms and extracted aspects confirms that most guests emphasize practical and interpersonal attributes of the stay.

The topic modeling results provide more nuanced interpretations. LDA produced five coherent and broad topics, particularly around transport, hospitality, and service communication. In contrast, BERTopic offers greater specificity, revealing over 100 semantically distinct clusters. While this provides more nuanced insights (e.g., dining/nightlife, booking issues, amenities), it also introduces noise, such as Topic 2, which contains sparse or ambiguous tokens likely resulting from tokenization errors or duplicate content.

Sentiment analysis via both VADER and BERT confirms a strong overall positivity bias in the dataset. Over 90% of reviews were classified as positive, and the rule-based aspects and LDA topics also largely followed this trend. However, subtle differences emerge in BERTopic sentiment patterns. Specifically, Topics 2 and 10 showed extremely high neutrality under VADER, likely due to vague or auto-generated phrases and booking-related issues, while Topic 14 had a slightly higher proportion of negative sentiment under VADER, which may be related to complaints about cleanliness or extra fees. These outlier topics may reveal friction points in the Airbnb experience that are difficult to detect with frequency-based methods alone.

The BERT sentiment results further confirm these findings while providing more detailed model confidence information. Most topics remain positive, but topics about booking cancellations and cleanliness-related fees stand out again, with an increase in negative sentiment. Though such topics represent a minority, they significantly influence overall perception due to the emotional salience of booking failures or additional charges.

Overall, the results illustrate the strengths of a hybrid approach combining traditional NLP (rule-based, LDA) and neural modeling (BERTopic, BERT). While general satisfaction is high, some topics repeatedly labeled with neutral or negative sentiment suggest that Airbnb and its hosts should pay closer attention to topics involving cleanliness policies, booking reliability, and perceived value.

5. Conclusions

This project applied a structured ABSA pipeline to analyze Airbnb guest reviews in Melbourne. By combining rule-based and unsupervised topic modeling with both lexicon-based and deep learning sentiment tools, we identified core satisfaction drivers such as location, cleanliness, and host interaction. While most reviews were overwhelmingly positive, specific topics—like booking issues and cleaning fees—consistently showed higher levels of neutrality or negativity.

For hosts, this highlights the importance of transparent communication, especially around booking procedures and extra charges. For the platform, fine-grained, interpretable analysis of textual sentiment insights—beyond star ratings—can better identify and address specific service gaps to improve quality monitoring and guest retention.

Overall, this approach offers a scalable framework for mining user feedback in peer-to-peer marketplaces. Future work may expand this framework to incorporate multimodal inputs (e.g., listing descriptions, images) or temporal dynamics to track shifts in guest expectations and satisfaction over time.

References:

Poria, S., Cambria, E., Hazarika, D., & Vij, P. (2017). A deeper look into sarcastic tweets using deep convolutional neural networks. *Proceedings of the 26th International Conference on World Wide Web Companion*, 583–588.

Tussyadiah, I. P., & Zach, F. (2016). Identifying salient attributes of peer-to-peer accommodation experience. *Journal of Travel & Tourism Marketing*, 34(5), 636–652. https://doi.org/10.1080/10548408.2016.1209153

Zhu, L., Xu, M., Bao, Y., Xu, Y., & Kong, X. (2022). Deep learning for aspect-based sentiment analysis: A review. *PeerJ Computer Science*, 8, e1044. https://doi.org/10.7717/peerj-cs.1044

Appendix

Figure 1. Word Cloud of Airbnb Reviews

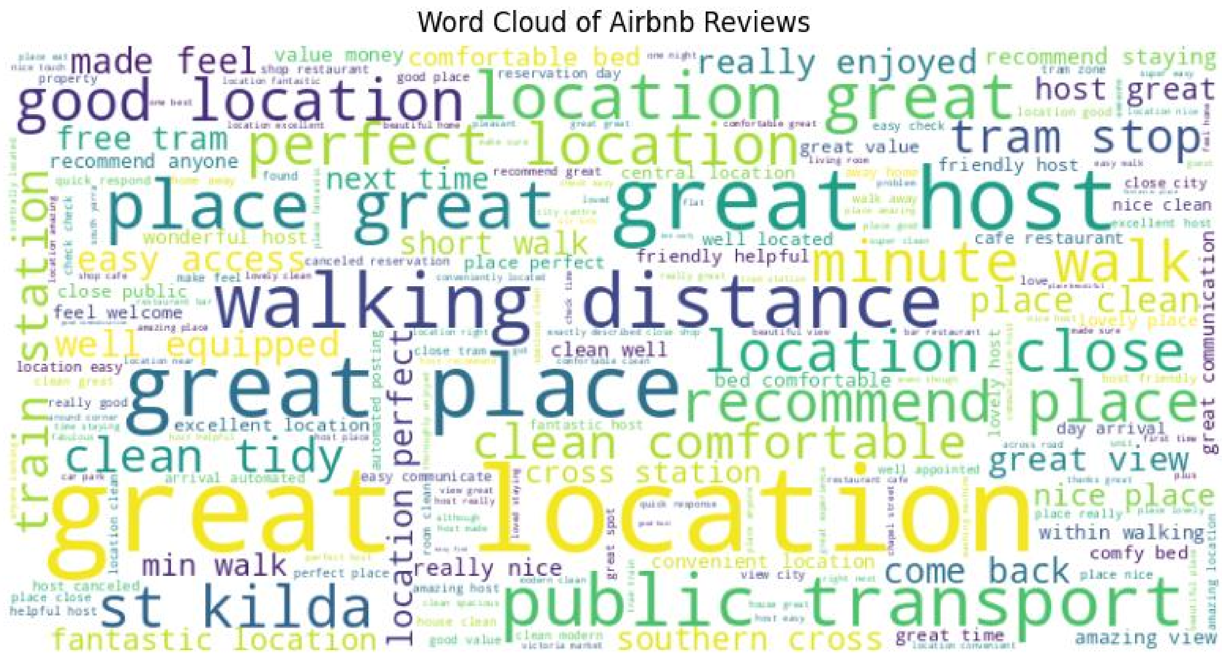

Figure 2. Top 20 Frequent Words in Airbnb Reviews

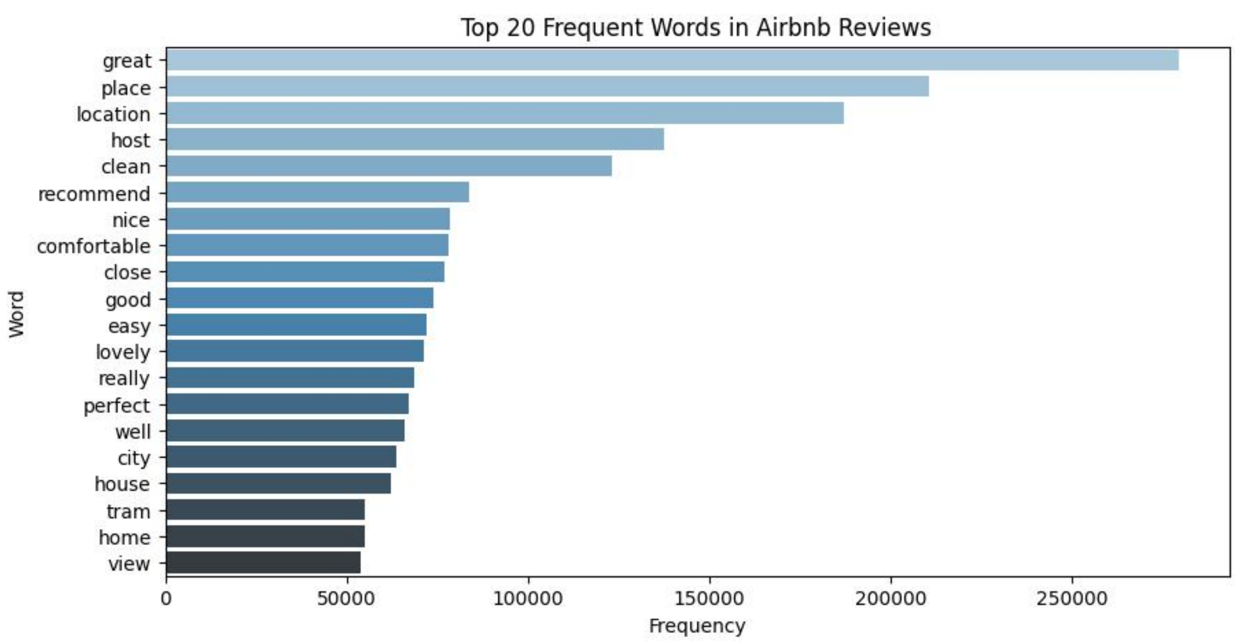

Figure 3. Top Aspects from Rule-Based Extraction

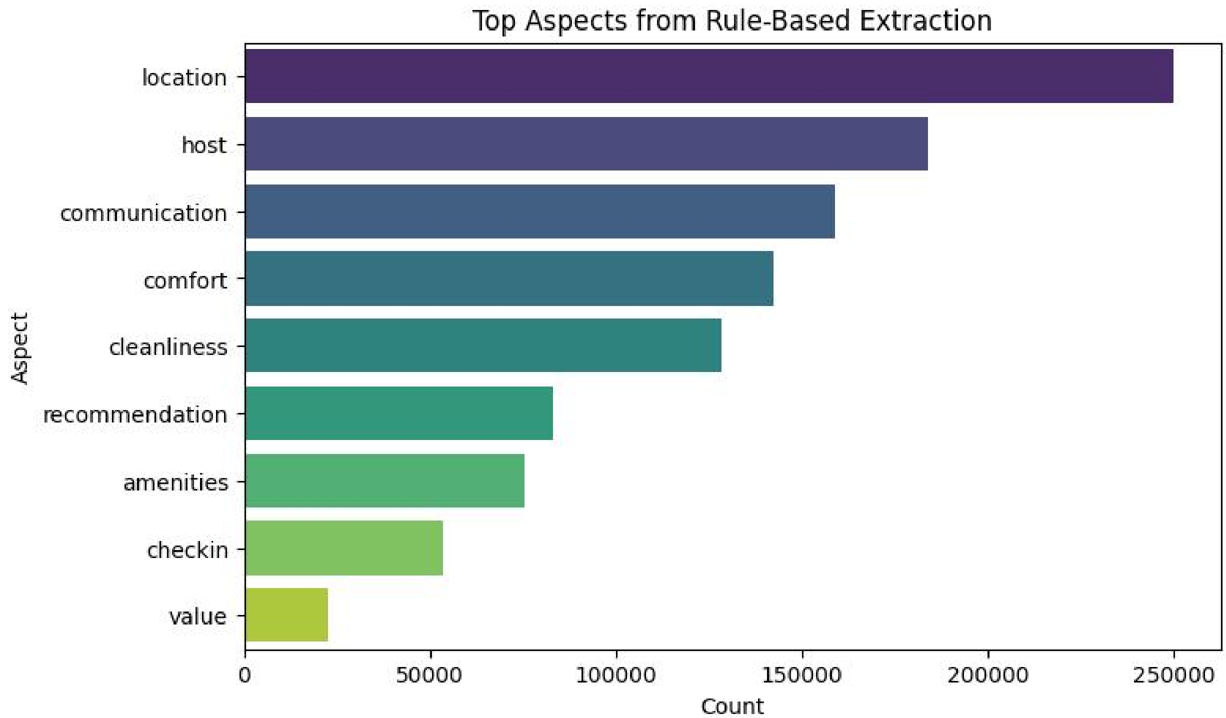

Figure 4. Top LDA Topics in Airbnb Reviews

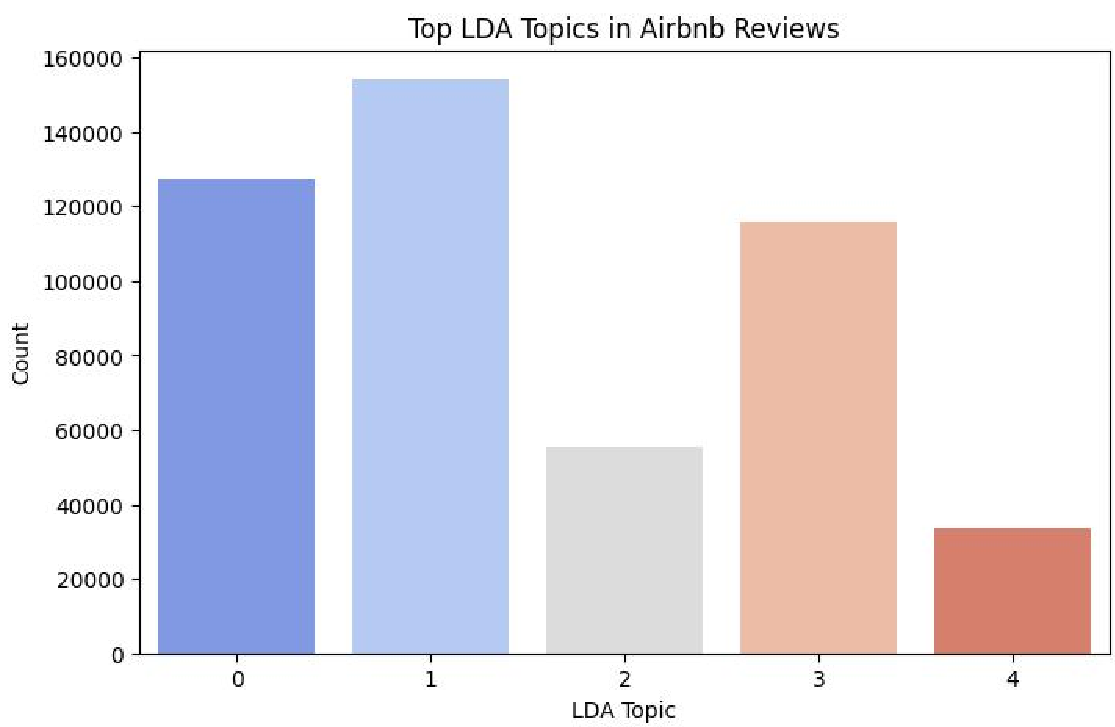

Figure 5. BERTopic Word Scores

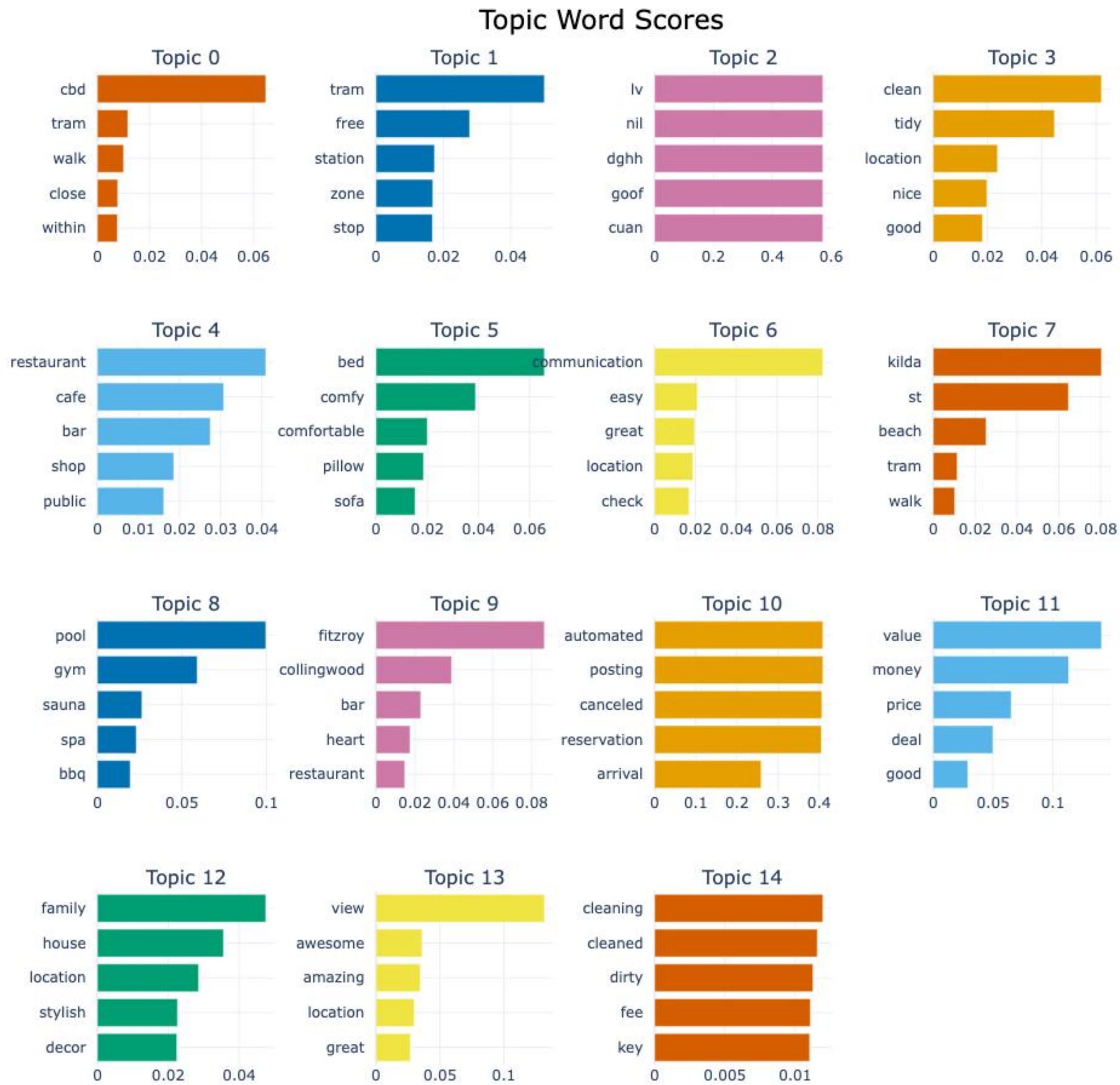

Figure 6. BERTopic Intertopic Distance Map

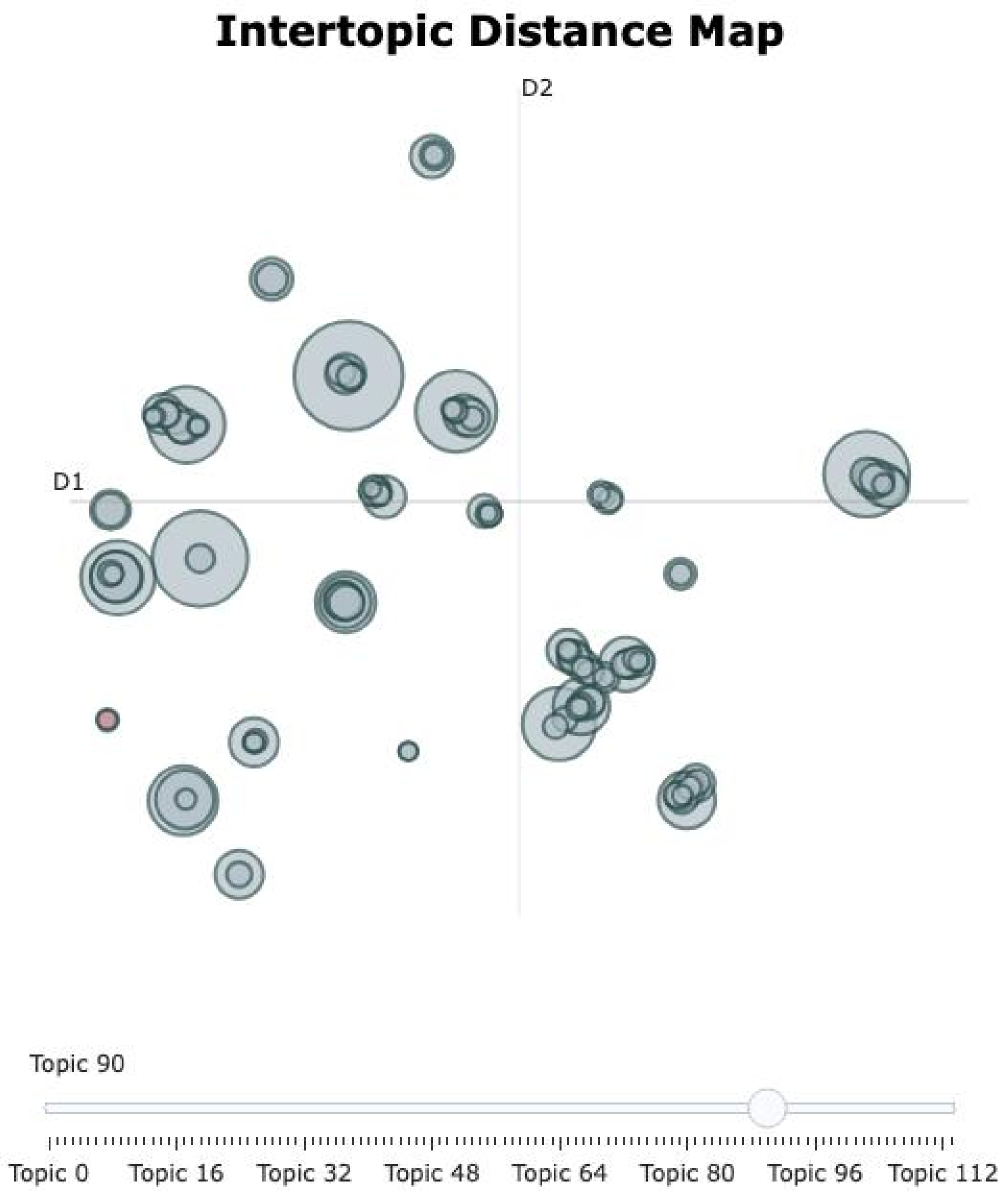

Figure 7. Top BERTopic Topics in Airbnb Reviews

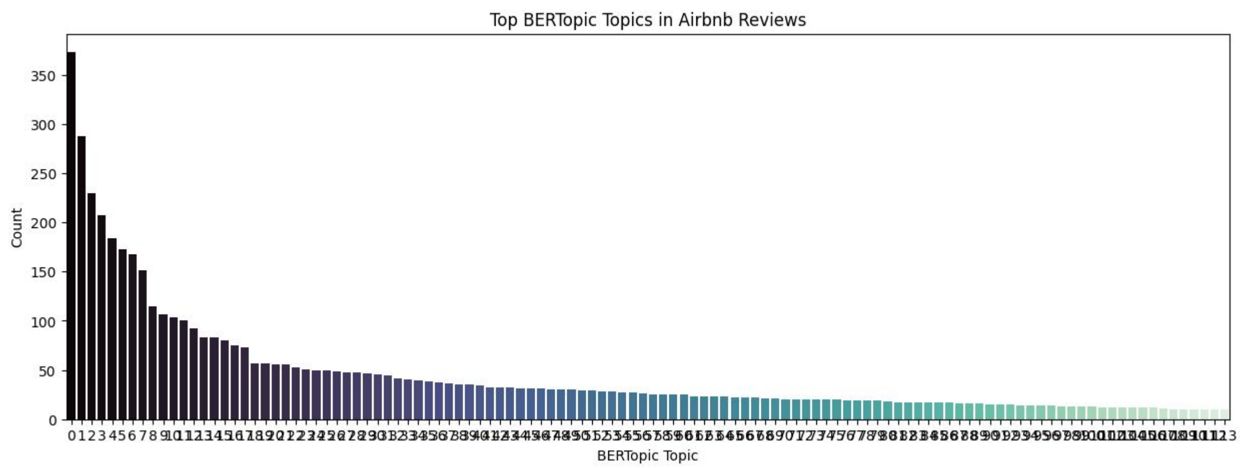

Figure 8. VADER Sentiment Distribution

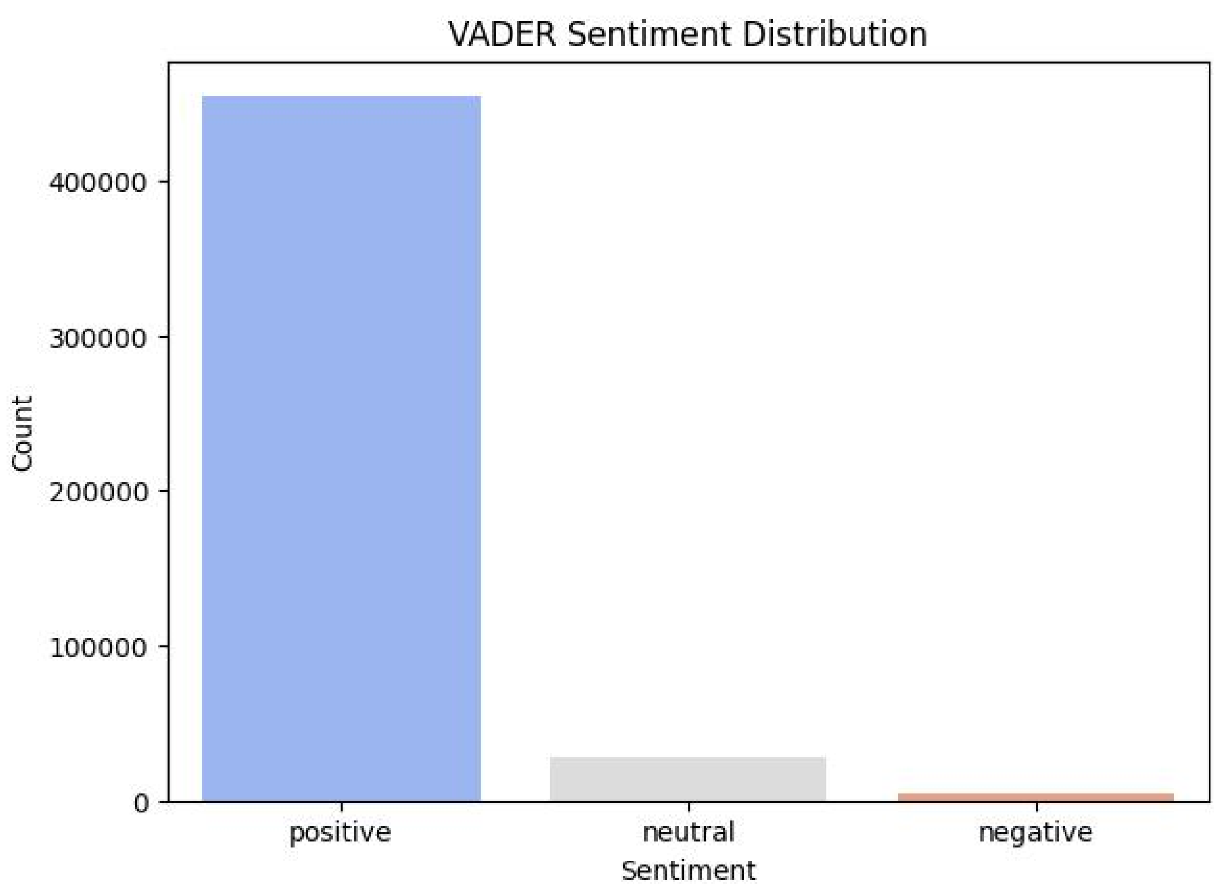

Figure 9. Rule-based Aspect vs VADER Sentiment

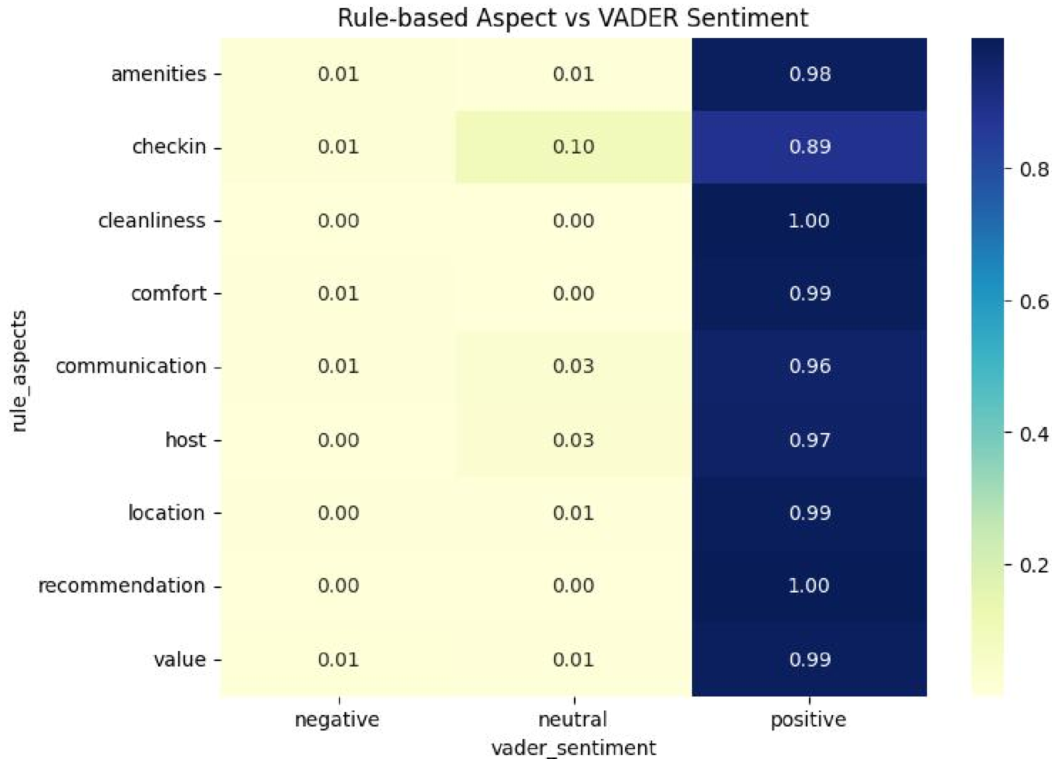

Figure 10. LDA Topic vs VADER Sentiment

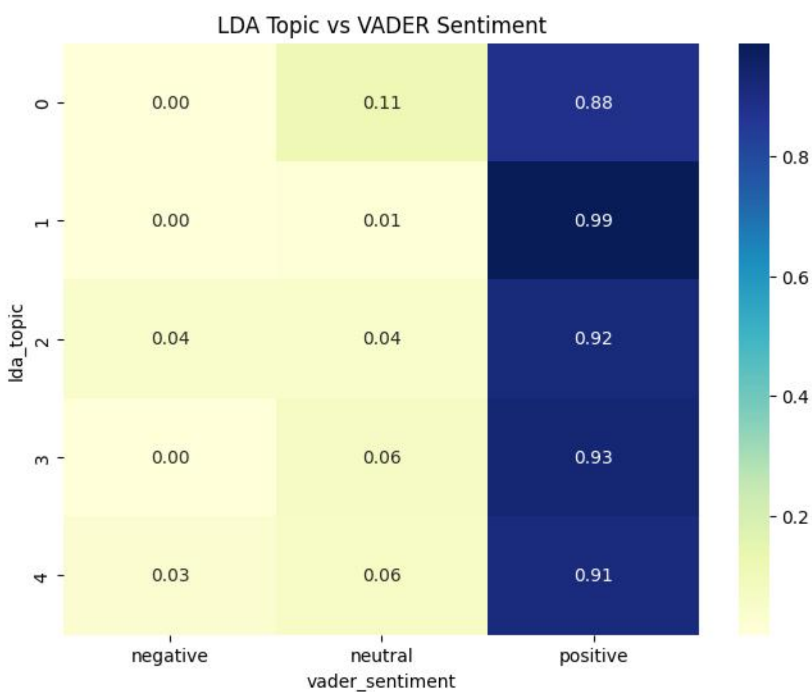

Figure 11. BERTopic Topic vs VADER Sentiment

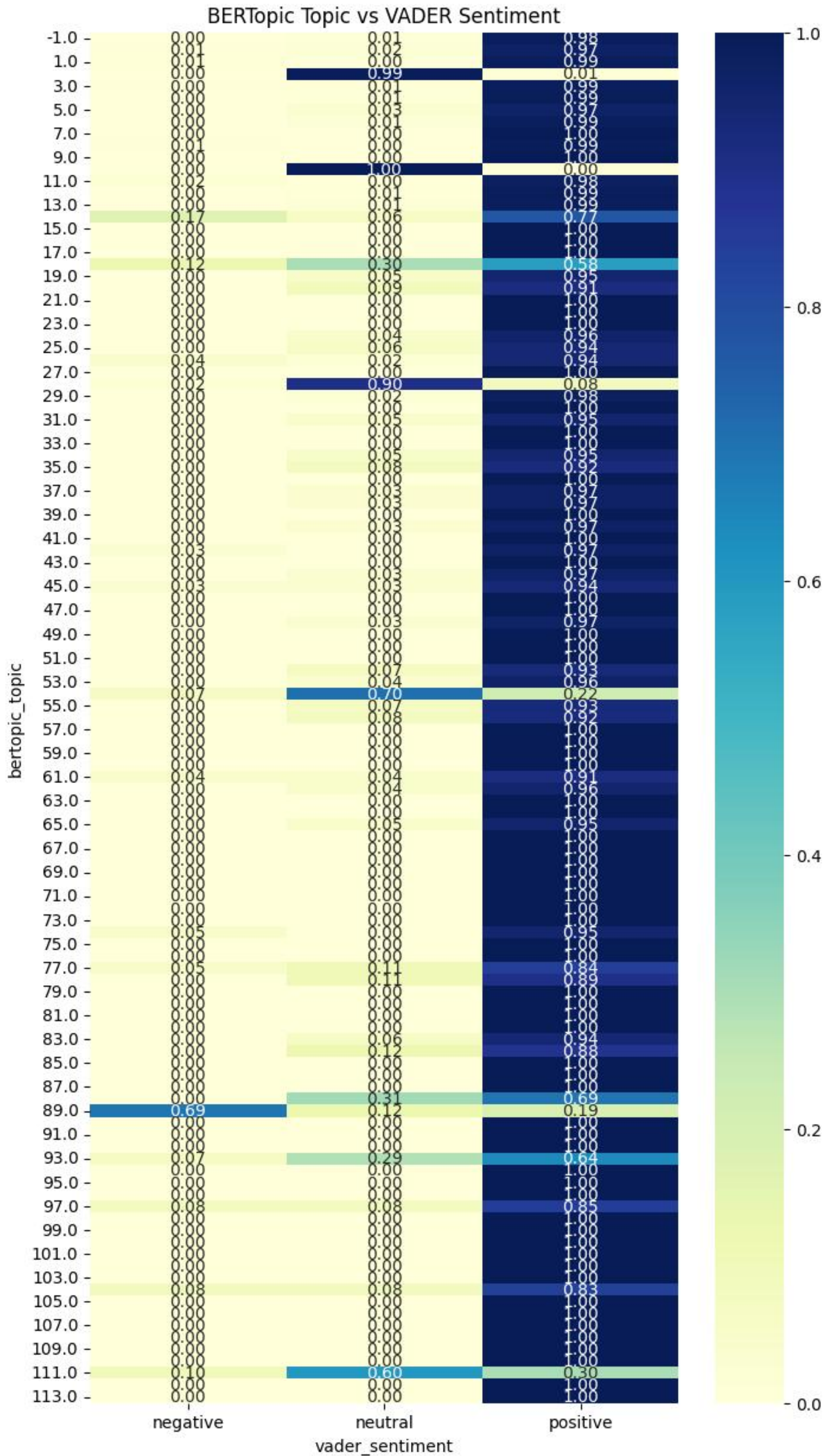

Figure 12. BERT Sentiment Distribution in Airbnb Reviews

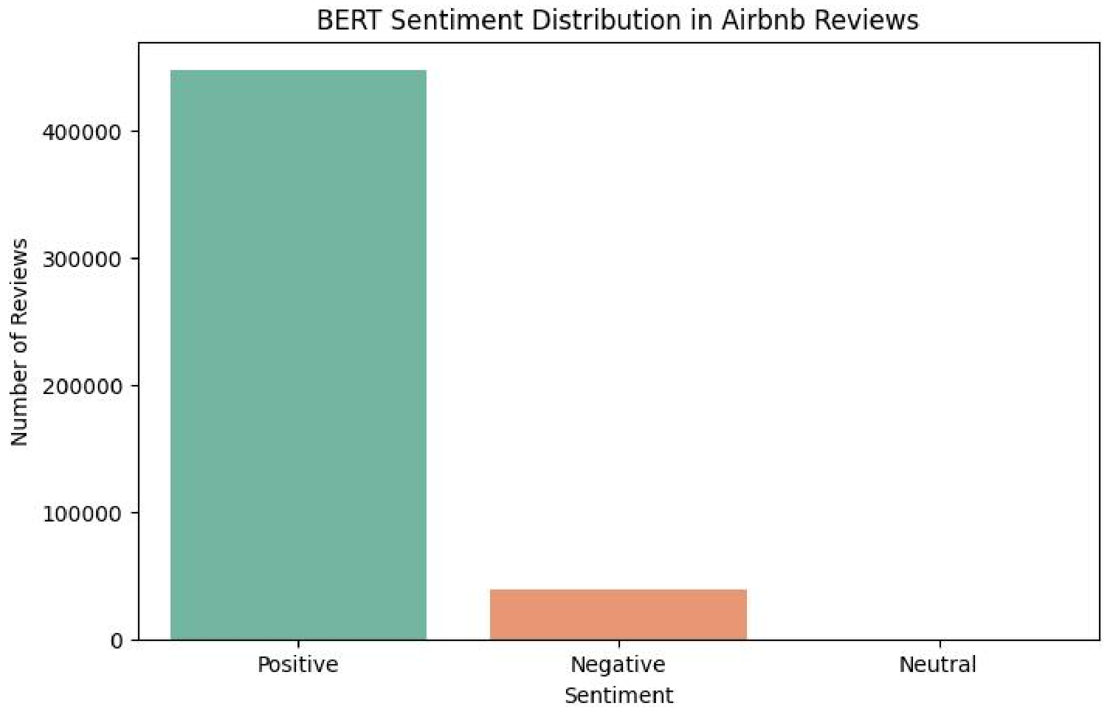

Figure 13. BERT Sentiment Confidence by Category

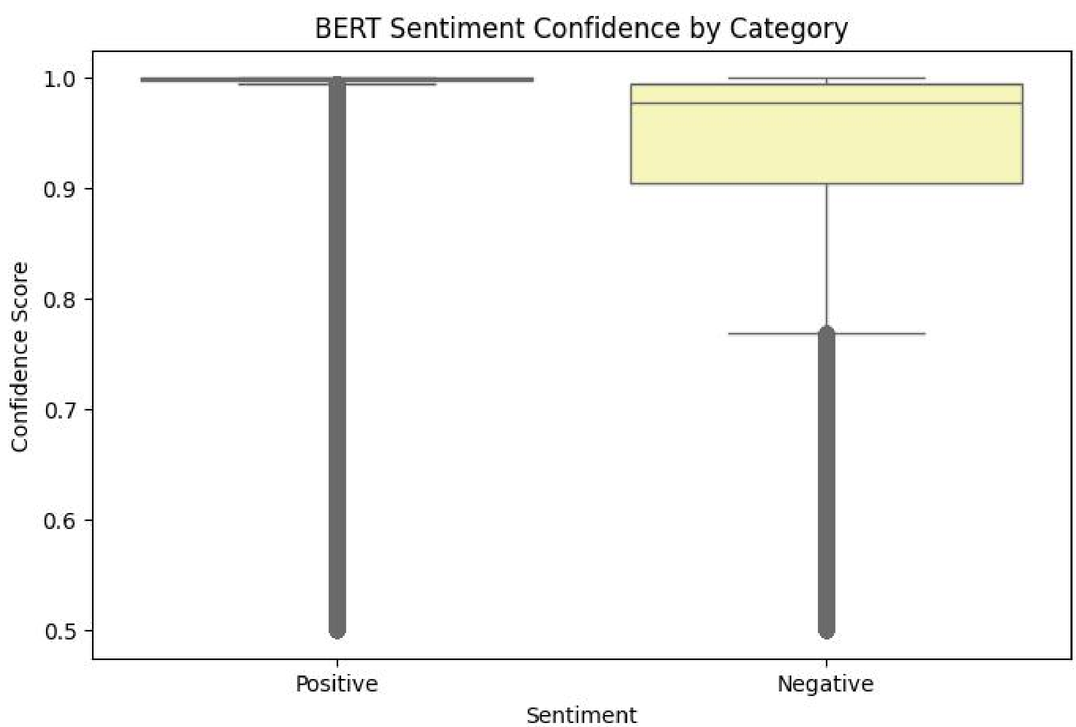

Figure 14. Rule-based Aspect – BERT Sentiment Distribution

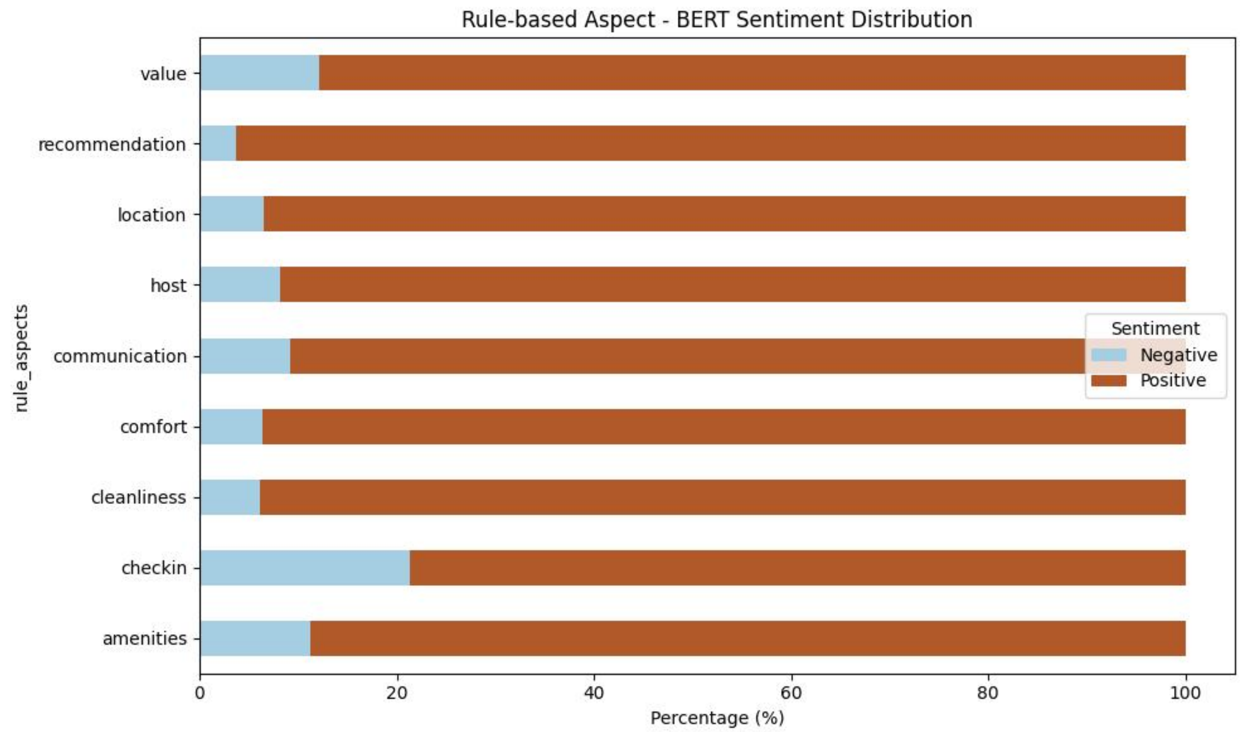

Figure 15. LDA Topic – BERT Sentiment Distribution

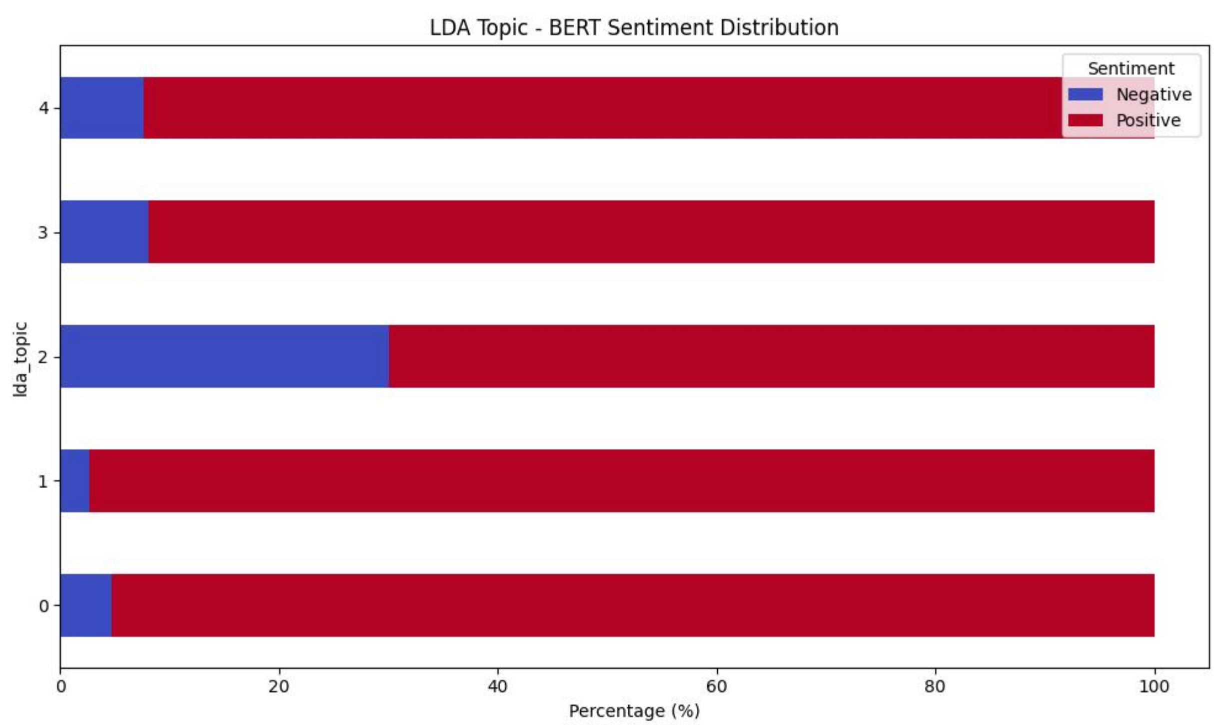

Figure 16. BERTopic Aspect – BERT Sentiment Distribution