Comprehensive NLP Analysis of Airbnb Reviews for Improved Decision-Making

George Washington University

Final Term Project - Individual Final Report

Author

Vishal Bakshi
Department of Data Science
George Washington University
Vishal.bakshi@gwu.edu

Abstract

This project explores methods to enhance sentiment and aspect-based analysis of Airbnb guest reviews using advanced Natural Language Processing (NLP) techniques. By employing transfer learning, aspect extraction, and summarization, we aim to improve the efficiency of feedback interpretation. Our approach utilizes transformer-based models, including BERT and T5, for sentiment classification and summarization, alongside Word2Vec and KMeans clustering for aspect-based sentiment analysis (ABSA). The study evaluates the influence of aspect granularity, dataset diversity, and pre-trained models on overall performance. Results are expected to demonstrate that combining transformer-based models with aspect-specific features enhances predictive accuracy and provides actionable insights for hosts. This project contributes to the field of text analysis by offering a framework to automate and refine the interpretation of large-scale user-generated content for short-term rental platforms.

2 Introduction

The rapid growth of platforms like Airbnb has generated vast user reviews offering insights into guest satisfaction, property conditions, and host performance. However, analyzing these reviews manually is inefficient. This report leverages Natural Language Processing (NLP) techniques to automate insights extraction through (1) sentiment analysis, (2) aspect-based sentiment analysis (ABSA), and (3) summarization. These methods aim to help hosts, guests, and platform administrators improve guest experiences and address recurring issues effectively.

3 Project Goals

This project comprehensively analyzes NYC Airbnb reviews (2021) using advanced NLP techniques. Specifically, we:

- **Preprocess and Sentiment Analysis (Phase 1):** Prepare and fine-tune a BERT model to classify reviews as positive or negative.
- **Aspect-Based Sentiment Analysis (Phase 2):** Identify and cluster key aspects (e.g., location, amenities) and map sentiments to them.

- **Summarization (Phase 3):** Generate concise summaries of multiple reviews using a transformer-based model (T5), thereby aiding quick decision-making.

4 My Role and Shared Work:

- While my teammate focused on preprocessing, initial sentiment labeling, and fine-tuning a BERT model for sentiment classification
- My primary contributions were:
  - Experimenting with aspect extraction methods, finalizing Word2Vec + KMeans clustering for ABSA.
  - Creating visualizations like bigrams, word clouds, and cumulative sentiment plots to analyze sentiments and aspects.
  - Designing and implementing the summarization phase using the T5 model, including chunking strategies and prompt engineering.
- We worked together on preprocessing strategies, aspect cluster naming, and ensuring a smooth data flow from sentiment analysis to aspect-level insights.

5 Description of My Individual Work and Methodology

5.1 Aspect-Based Sentiment Analysis (ABSA) Exploration

Before finalizing the ABSA pipeline, I experimented with spaCy's dependency parsing and named entity recognition for aspect identification. However, this approach was set aside due to its complexity and suboptimal clustering results.

5.2 Approach (Word2Vec + KMeans)

- **Word2Vec Embeddings**
  After Kusum completed thorough preprocessing and tokenization of reviews, I utilized the cleaned tokens to train a Word2Vec model. This provided semantic vector representations of words, enabling meaningful clustering.

- **Noun Extraction and Clustering**
  By focusing on nouns (POS tagging), we extracted potential aspect terms. I then applied KMeans clustering to group semantically similar nouns into distinct aspect categories (e.g., "Location and Transportation," "Value and Pricing"). Through iterative refinement and manual inspection, I assigned interpretable aspect names to each cluster.

5.3 Rationale (Word2Vec + KMeans)

- The Word2Vec + KMeans approach provided a flexible, data-driven method for identifying aspects directly from reviews, eliminating the need for domain-specific lexicons.
- Manual mapping of clustered terms to aspect names resulted in meaningful and human-interpretable categories

6 Description of My Individual Work and Methodology

6.1 Aspect-Based Sentiment Analysis (ABSA) Exploration

While my teammate worked on generating initial sentiment distribution bar charts and yearly trend lines, I concentrated on developing more detailed visualizations

6.2 Top Bi-grams in Positive Reviews

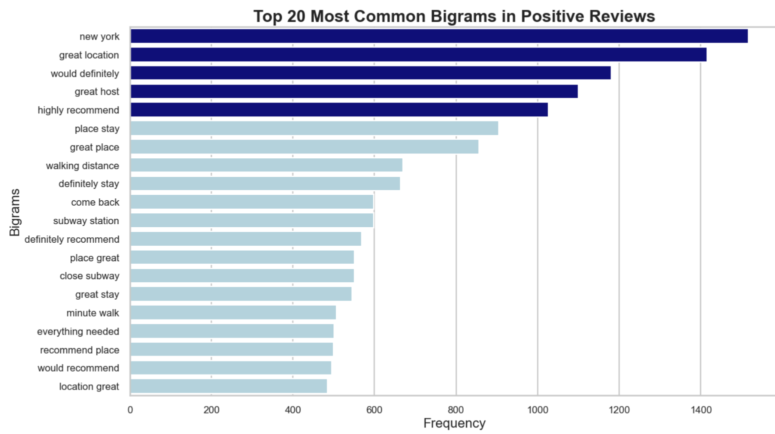

Fig 3. Top 20 Most Common Bi-grams in Positive Reviews

Common bigrams such as "great location," "great host," and "highly recommend" frequently appear in positive reviews.

6.3 Top Bi-grams in Negative Reviews

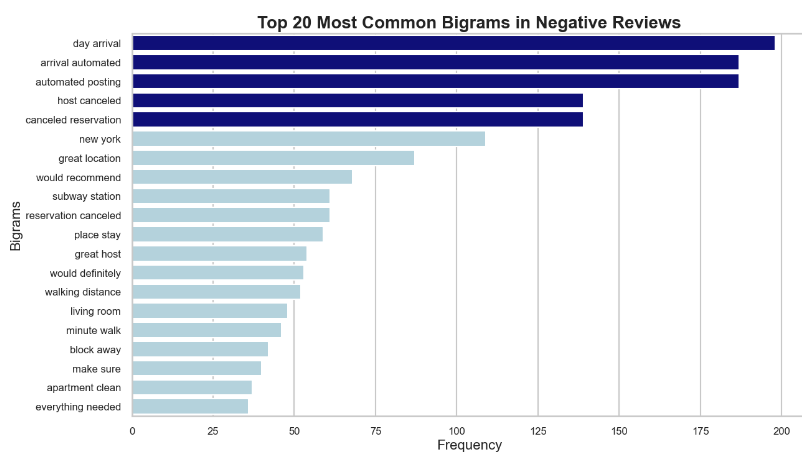

Fig 4. Top 20 Most Common Bi-grams in Positive Reviews

Frequently occurring negative bigrams, such as "host cancelled" and "canceled reservation," highlight dissatisfaction primarily related to booking issues and host reliability.

Observation: The bigram analyses offer actionable insights by enabling hosts to reinforce aspects frequently associated with positive sentiment and address recurring issues emphasized in negative feedback.

6.4 Word Clouds

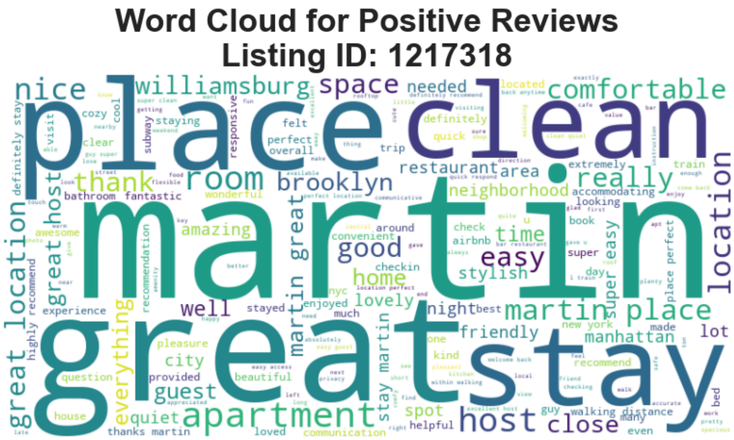

Fig 5. Word Cloud for Positive Reviews

I created word clouds from the top positive reviews to highlight the most frequently mentioned words. Terms like "great," "clean," and "comfortable" prominently reflect positive sentiments and provide a quick overview of what guests value most.

Observation: Word clouds, though straightforward, are intuitive tools that allow stakeholders to quickly identify key themes. They complement bigram analysis and sentiment metrics by visually summarizing the most salient terms.

6.5 Cumulative Evolution of an Aspect Over Time

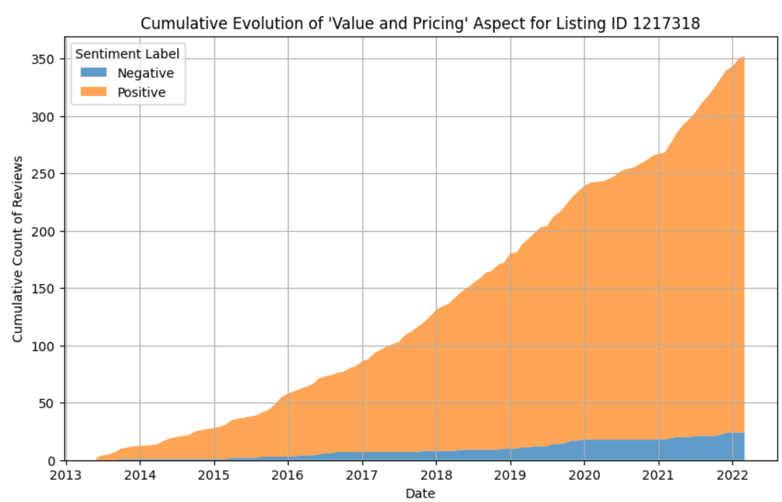

Fig 8. Stacked Area Chart of Cumulative Positive and Negative Reviews Over Time for "Value and Pricing"

A stacked area plot illustrates the cumulative growth of positive and negative sentiments for the "Value and Pricing" aspect. The chart shows that positive sentiment steadily increases over time, while negative sentiment grows at a slower and more stable rate.

Observation: The consistent growth in positive sentiment suggests that guests increasingly recognize and appreciate the value offered by the listing. Meanwhile, the relatively low and stable growth of negative sentiment indicates that while some concerns exist, they are not escalating significantly. This trend reflects an overall improvement or sustained satisfaction in the value proposition, highlighting the importance of maintaining competitive pricing and addressing any recurring issues to enhance guest experiences further.to maintain and specific areas, such as resolving guest complaints, that require attention to improve the overall guest experience.

7 Summarization with a Transformer-Based Model (T5)

The final phase consolidated insights, sentiments, and aspects into concise, coherent summaries. While my teammate handled foundational data processing and sentiment classification, my contributions focused on the following

7.1 Model and Tokenizer Setup

I utilized the T5-large model, known for its strong abstractive summarization capabilities. By loading the model and tokenizer, I prepared the framework to transform multiple lengthy reviews into concise summaries.

7.2 Model and Tokenizer Setup

- **Chunking Strategy**
  Due to T5's input length limitations, I implemented a chunking method to divide long reviews into manageable segments, summarizing each individually.

- **Prompt Engineering**
  I experimented with prompts like "summarize:" and added context to emphasize key aspects such as host personality, location, and value. This ensured the model focused on attributes most relevant to guests.

7.3 Final Summary Generation

Intermediate summaries for each chunk were combined and re-summarized into a cohesive, human-readable narrative.

7.3.1 Sample Final Summary

> "The host is attentive, responsive, and maintains a clean, comfortable space. Guests frequently praise the location's convenience and safety and find the value fair relative to price. Most appreciate the easy access to transportation, indicating that the listing consistently meets or exceeds guest expectations."

Observation: This summarization significantly reduces the cognitive load for potential guests or hosts by providing a quick, high-level understanding of how a listing is perceived.

8 Results and Insights

8.1 Aspect-Level Findings

- Positive aspects such as "Accommodation Facilities and Aesthetics" and "Location and Transportation" emerged as key strengths.
- Smaller clusters highlighted recurring complaints, providing actionable insights for improvement.

8.2 Visual Insights

- Bigram and word cloud analyses validated the major themes guests appreciate or criticize.
- Temporal analyses, such as cumulative charts, revealed evolving guest perceptions, offering valuable insights for long-term planning.

8.3 Summarization Benefits:

- The final summaries condensed extensive reviews into concise paragraphs, enabling quicker and more informed decision-making.
- Hosts can easily identify top-praised attributes and address recurring issues without manually reading all reviews.

9 Conclusion and Future Directions

9.1 Summary of Contributions

My main contributions complemented the strong preprocessing and sentiment labeling foundation laid by my teammate.

- Explored multiple methods (spaCy and Word2Vec) for aspect extraction, ultimately selecting the Word2Vec + KMeans approach for its effective clustering.
- Developed various visualizations (bigrams, word clouds, heatmaps, radar charts, cumulative plots) to present insights in an intuitive manner.
- Implemented the final summarization phase using T5, refining prompts and chunking strategies to produce coherent and informative summaries.

9.2 Lessons Learned

- **Modularity:** Separating the pipeline into phases (sentiment, aspects, summarization) ensured flexibility and iterative improvement.
- **Customization:** Prompt engineering and chunking strategies can greatly enhance summarization quality.
- **Holistic Analysis:** Combining quantitative sentiment metrics with qualitative text analysis (bigrams, word clouds) and ABSA methods leads to richer insights.

9.3 Future Improvements

- Experimenting with other embedding methods or topic modeling techniques (e.g., BERTopic) for aspect discovery.
- Improving summarization prompts to emphasize specific user-defined criteria (e.g., focusing on safety for a listing in a busy neighborhood).

10 Code Provenance and Percentage from External Sources

10.1 Code Sources

- Adapted code from Hugging Face documentation for training and summarization.
- Used standard Python, NLTK, and Word2Vec functionalities as per official guides.

10.2 Approximate Percentage from Internet

- Similar to my teammate's calculation, a portion of the code for model loading, tokenization, and summarization logic (around 70 lines) were adapted from official documentation and examples.
- After customization and additional code (~200 lines), about 20% of the reused code remained close to the original structure.

References

- Hugging Face Transformers Documentation: https://huggingface.co/docs/transformers
- NLTK Documentation: https://www.nltk.org/
- Kaggle NYC Airbnb Reviews Dataset (2021)
- Contractions Library: https://pypi.org/project/contractions/
- Word2Vec (Mikolov et al., 2013)
- Scikit-learn KMeans: https://scikit-learn.org/