**AirBNB Reviews Sentiment Analysis**

**Problem Statement**

In the competitive landscape of hospitality and accommodation services, understanding customer sentiment through reviews is crucial for maintaining and enhancing service quality. Airbnb, a leading platform connecting hosts and guests worldwide, faces the challenge of managing and improving customer satisfaction continuously.

Consumer reviews in the Airbnb marketplace are one of the key attributes to measure the quality of services and the main determinant of consumer rentals decisions. Such feedback can impact both a new and repeated consumer's choice decision. Airbnb faces the challenge of effectively understanding and utilizing customer feedback to enhance service quality and maintain competitive advantage. Despite receiving a substantial volume of reviews, the company struggles to systematically identify and prioritize the most critical themes and sentiments expressed by its customers.

Some of the challenges including:

1. **Volume and Variability**: Managing a large volume of reviews across different languages and cultural contexts presents challenges in consistently extracting meaningful insights from customer feedback.
2. **Data Integration and Analysis**: Integrating sentiment analysis tools and techniques into Airbnb's existing data infrastructure is complex and requires robust analytics capabilities to derive actionable insights effectively.
3. **Actionable Insights**: Despite collecting extensive feedback, Airbnb faces difficulties in translating sentimental analysis results into actionable strategies that drive tangible improvements in service quality and customer satisfaction.

**Objectives**

The Natural Language Processing (NLP) Sentiment Analysis method is used in this project to predict and categorize the sentiments of each review in the dataset. The goals of this project are to:

* **Develop automated processes** or tools to categorize and prioritize themes and topics from customer reviews, enabling Airbnb to handle large data while still maintaining reasonable costs.
* **Implement a standardized sentiment scoring system** to quantitatively measure and track changes in customer satisfaction over time, facilitating data-driven decision-making and targeted service enhancements.

**Data Source**

The data set used in this project is obtained from Kaggle website, and this dataset can be used for NLP use cases, e.g. Exploratory Data Analysis (EDA), Text summarization, Sentiment Analysis, Intent Analysis and many more. The link to obtain the dataset is:

<https://www.kaggle.com/datasets/muhammadahmedansari/airbnb-dataset>

The dataset comprises 342,904 rows distributed across 6 columns.

**Metrics**

Accuracy in sentiment analysis is crucial because it directly impacts the reliability and usefulness of the insights derived from the analysis. Here’s an elaboration on why accuracy is so important in this context:

1. **Decision Making and Actionability**: Organizations often use sentiment analysis to make critical decisions about their products, services, or brand reputation. If the sentiment analysis is inaccurate, these decisions could be misguided, leading to poor strategic choices. For instance, a falsely positive sentiment analysis might lead a company to believe their product is well-received when it's not, impacting their marketing or product development strategies.
2. **Customer Insights**: Understanding customer sentiment helps businesses tailor their offerings to better meet customer needs and expectations. Accurate sentiment analysis ensures that these insights reflect the true feelings and opinions of customers, enabling businesses to make informed adjustments and improvements. For example, accurate analysis might reveal specific pain points that need addressing or highlight features that customers love.
3. **Reputation Management**: Brands monitor sentiment to manage their online reputation. Accurate sentiment analysis allows them to promptly address negative sentiment or capitalize on positive sentiment. Misinterpreting sentiment could result in inadequate responses or missed opportunities to engage with customers, potentially damaging brand perception.
4. **Market Research**: Sentiment analysis is integral to market research, providing insights into consumer preferences, trends, and competitor analysis. Accurate sentiment analysis ensures that market researchers can identify emerging trends or shifts in consumer sentiment accurately, which is vital for staying competitive and innovative.
5. **Customer Satisfaction and Loyalty**: Accurate sentiment analysis helps businesses gauge customer satisfaction levels. By accurately identifying positive and negative sentiment, businesses can take proactive measures to enhance customer experience, improve satisfaction, and foster loyalty. Inaccurate analysis could lead to missed opportunities to address customer concerns or recognize loyal customers.

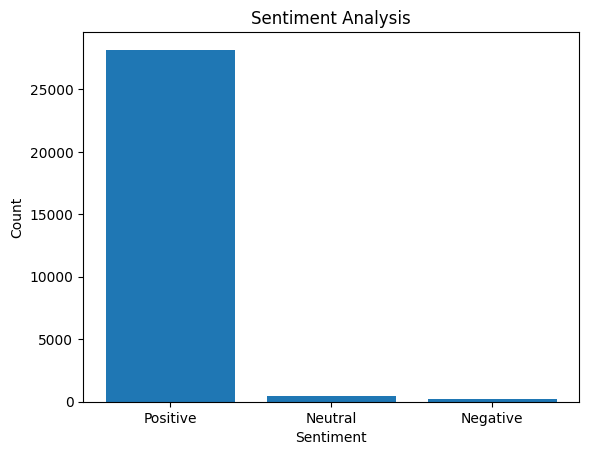
**Findings**

Upon conducting sentiment analysis on the dataset, the following results were obtained, as shown in Table 1.

|  |  |
| --- | --- |
| **Sentiments** | **Reviews Instances** |
| Positive | 28172 |
| Neutral | 418 |
| Negative | 222 |

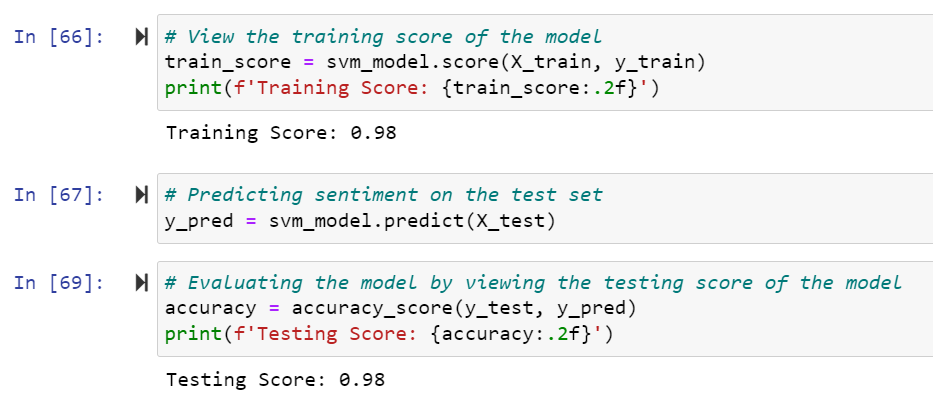
**Table 1**: Reviews Instances of Each Sentiment

These findings reveal a predominance of positive sentiment within the dataset, with a significant number of instances categorized as positive. The presence of neutral and negative sentiments, although less frequent, indicates a spectrum of emotional responses within the analyzed data. This distribution suggests that while most sentiments expressed are positive, there are also instances of neutrality and negativity that contribute to a more nuanced understanding of the overall sentiment landscape. Further analysis and contextual exploration may be beneficial to uncover underlying patterns and implications within the dataset. Figure 1 shows the sentiment analysis result in a visualization method.



**Figure 1**: Sentiment Analysis Results

To verify the accuracy of the model, we train the model and test it against the training score. The testing score is shown in Figure 2.

**Figure 2**: The Accuracy of the NLP Model

It can be concluded that the accuracy of the model is satisfactory.

**Risk/Limitations/Assumptions after Findings**

I have encountered challenges translating the non-English text in the dataset, as my current computer hardware limitations would require approximately 7 to 8 hours to translate the text in each of the two columns needing translation. Consequently, the total time for translating both columns would amount to 14 to 16 hours.

Therefore, for this project, I have decided to exclude the non-English text from the data analysis. However, I have provided Python code for the text translation and commented it out. If your computer has sufficient hardware resources, you may attempt to execute it.

Please disable the text removal portion of the Python code if you choose to use the translation method.

**References**

*Tutorial: Quickstart — TextBlob 0.18.0.post0 documentation*. (n.d.-b). <https://textblob.readthedocs.io/en/dev/quickstart.html>

Raza, M. R., Hussain, W., & Varol, A. (2022a). Performance analysis of deep approaches on airbnb sentiment reviews. *2022 10th International Symposium on Digital Forensics and Security (ISDFS)*. <https://doi.org/10.1109/isdfs55398.2022.9800816>