PROJECT DRAFT

Introduction:

The growth of the sharing economy has disrupted many traditional industries as it allows people to access the resources at lower cost and provides income to those who share their resources, and the travel industry is no exception. Airbnb, one of the most successful examples of the sharing economy, has rapidly expanded its presence worldwide since its founding in 2008. The platform allows individuals to rent out their homes, apartments, or rooms to travelers seeking a unique experience and an alternative to traditional hotel accommodations. As Airbnb continues to grow, it is important to understand how customers perceive their experience on the platform in order to improve the user experience and support the sustainable growth of the company.

Motivation for Study:

The objective of this study is to analyze customer reviews of Seattle Airbnb listings using sentiment analysis, with the aim of providing business owners with a rapid way to grasp the general opinions and feedback of their customers. By analyzing customer reviews, we can gain insights into what customers like and dislike about Airbnb listings in Seattle. This information can be used to identify areas for improvement and to provide actionable feedback to hosts. In addition, the study can provide insights into the overall sentiment of the Seattle Airbnb market, which can be useful for businesses and policymakers.

Relevant Theory:

Sentiment analysis is a technique where text data is analyzed to determine the emotional tone of the text. It involves the use of natural language processing (NLP) techniques to identify words and phrases that express positive or negative sentiment. Sentiment analysis has been widely used in marketing and customer experience management to gain insights into customer feedback and improve business processes. In the context of Airbnb, sentiment analysis can be used to identify areas for improvement and to support the sustainable growth of the platform.

Methods:

Data:

The Seattle Airbnb reviews dataset will be used as the main source of data for this project. The dataset contains over 400,000 rows and 6 columns, which include information on the listing ID, review ID, date of review, reviewer ID, reviewer name, and comments.Each row in the dataset represents a review for a particular listing, and includes information such as the listing ID, date of the review, reviewer ID, reviewer name, and comments or review given by customers about the listing.The comments field contains text reviews left by guests and is the primary focus of our analysis.Overall, the dataset seems to be focused on providing insight into the experiences of guests who have stayed at Airbnb listings in a particular location (Seattle). The data will be preprocessed by removing duplicates, handling missing values, and converting the text reviews into a machine-readable format.

Analysis Pipeline:

1. Data Collection: I collected the data for my sentiment analysis project from the Airbnb website. (http://insideairbnb.com/get-the-data)
2. Data Cleaning and Preprocessing: I cleaned and preprocessed the data to prepare it for analysis. This involved removing irrelevant data, removing stop words, stemming or lemmatizing words, removing punctuation and special characters, and converting all text to lowercase. Columns that i removed: id, reviewer_id, reviewer_name, comments. I removed the comments columns as i have added the clean comments column which is preprocessed. I have also removed the null values.
3. Text Analysis: I applied Natural Language Processing (NLP) techniques to extract useful information from the text data. This involved tokenization, part-of-speech (POS) tagging, and dependency parsing.
4. Sentiment Classification: I built a sentiment classification model using machine learning algorithms to classify each review as positive, negative, or neutral.
5. Evaluation: I evaluated the accuracy of our sentiment classification model using metrics such as precision, recall, and F1-score.
6. Visualization: I visualized the results of our sentiment analysis using a scatter plot, a line chart, and a stacked area chart. These visualizations helped me to gain insights into the sentiment of the Airbnb reviews.

7. Interpretation: I interpreted the results of the sentiment analysis and drew conclusions based on the insights gained from the analysis. I found that the majority of the reviews were positive and that the most common topics mentioned in the reviews were cleanliness, location, and amenities.

By following this analysis pipeline, I was able to successfully analyze the sentiment of the Airbnb reviews and gain insights into the customer experience. These insights can be used to improve the overall customer experience of our Airbnb listings by addressing common complaints or issues raised in negative reviews and promoting the positive aspects of the listings or services mentioned in positive reviews.

Result:

The sentiment analysis conducted on the dataset revealed some interesting insights into the sentiment of the comments. Firstly, it showed that the vast majority of the comments had a positive sentiment, with over 350,000 comments classified as positive. This is a promising indication of the overall positivity of the comments in the dataset.

The analysis also showed that the sentiment polarity distribution was clustered around a score of 0.25-0.75, with the highest frequency of datapoints occurring at a score of 0.50. This suggests that the sentiment expressed in the comments was largely moderate, with a slight tendency towards positivity.To further validate the model's performance, it was evaluated using a test set, which resulted in high precision, recall, and f1-score for all three categories, with an accuracy of 98%. This indicates that the model was effective in accurately predicting the sentiment of the comments.

In addition to this, various other analysis techniques were used, such as creating a heatmap, word clouds, scatter plots, and sentiment trends over time. These techniques provided valuable insights into the data, such as identifying the prevalence of negative sentiment in earlier years, and the relationship between review length and sentiment score.

Overall, the results of the sentiment analysis provided a comprehensive understanding of the sentiment expressed in the comments, which could be useful in various applications such as improving customer satisfaction or understanding public opinion.These findings can be useful for Airbnb hosts and managers to understand customer feedback and improve their services

Conclusion:

Implications of Results:

The results of this sentiment analysis can have several implications for Airbnb hosts and managers. Firstly, it can help them understand the sentiment of their customers and identify areas where improvements are needed. By focusing on the negative and neutral comments, they can take action to address the concerns of their customers and improve their overall experience. Secondly, it can help them identify the strengths of their services and leverage them to attract more customers. Positive reviews can be used as testimonials to showcase the quality of their services and attract new customers. Finally, it can help them monitor the sentiment trends over time and track the effectiveness of their efforts to improve customer satisfaction.

Limitations:

One of the limitations of this sentiment analysis is that it only analyzed the textual content of the reviews and did not take into account other factors such as the reviewer's demographics, location, or booking details. These factors could have an impact on the sentiment of the reviews and could provide more insights into customer behavior. Additionally, the analysis only focused on English reviews, which could limit the applicability of the results to non-English-speaking customers. Finally, the sentiment analysis was based on a pre-trained model and did not take into account the context or sarcasm in the reviews, which could lead to inaccurate results.

Next Steps:

To overcome the limitations of this sentiment analysis, future research can focus on analyzing other factors such as the reviewer's demographics, location, or booking details. This can provide more insights into customer behavior and help Airbnb hosts and managers tailor their services to meet the needs of different customer segments. Additionally, future research can focus on analyzing reviews in other languages to provide a more comprehensive picture of customer sentiment. Finally, future research can focus on developing more accurate

sentiment analysis models that take into account the context and sarcasm in the reviews. This can lead to more accurate results and provide more insights into customer sentiment.