Sentiment Analysis

Business Understanding

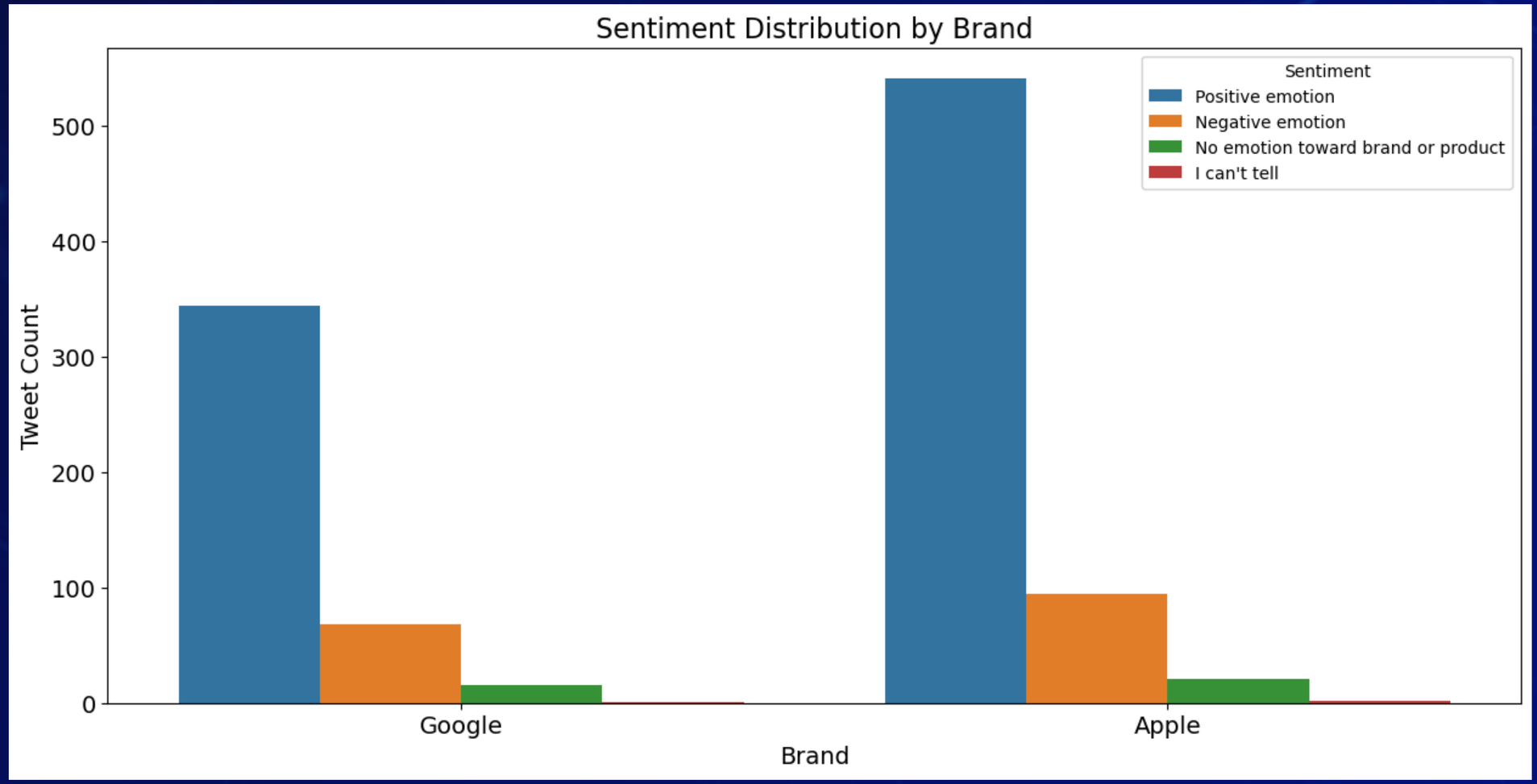
Social media platforms like twitter/X are where people openly share their thoughts, complaints, and praise about products and brands.

For large companies like Apple and Google, understanding this feedback is crucial to enhancing products and maintaining a robust brand image.

Data Understanding

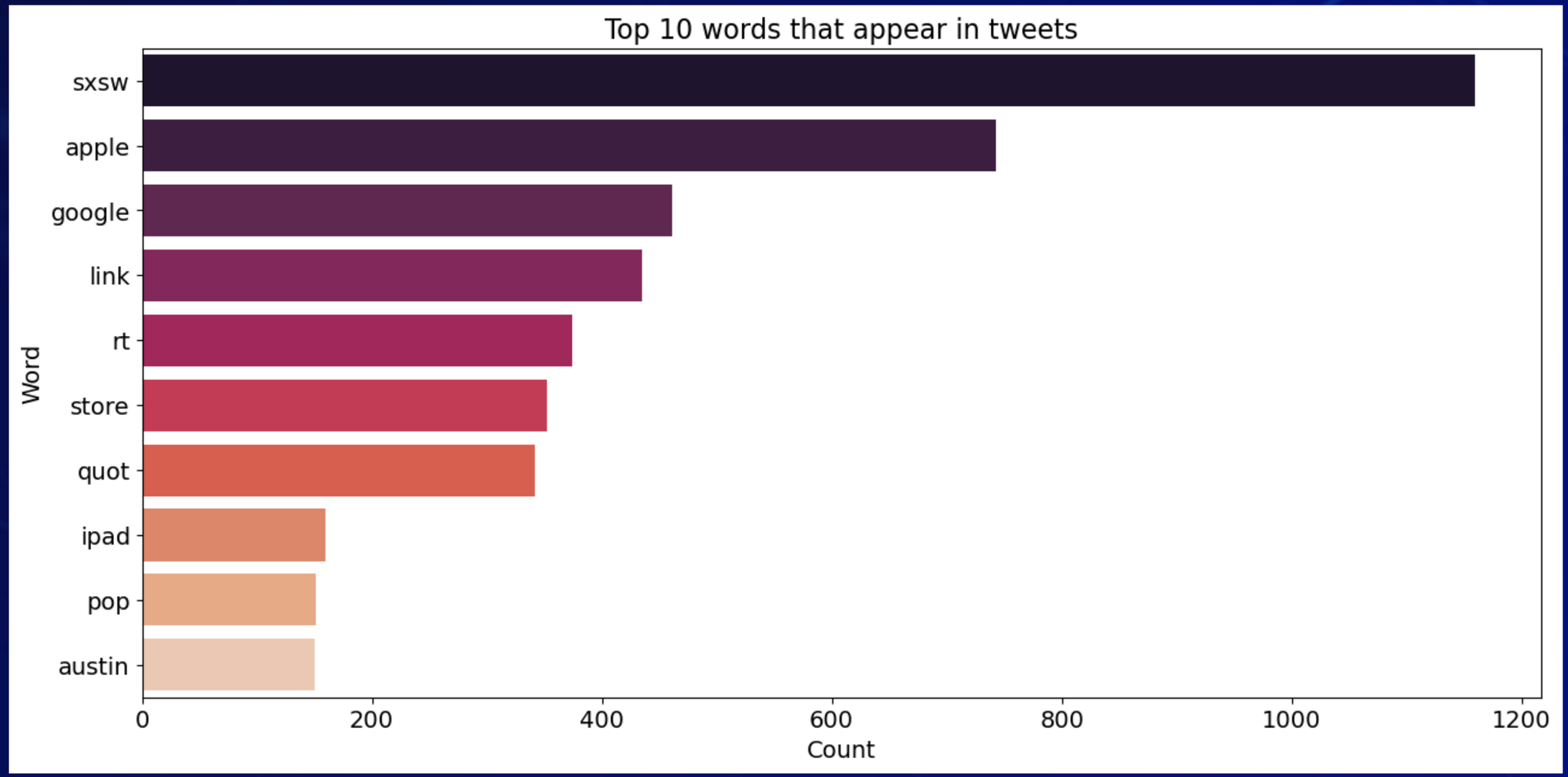
The dataset contains 9,093 tweets collected from crowdFlower, to identify whether the emotion in a tweet is directed at either apple or google products and the sentiment it carries.

Sentiment distribution by brand



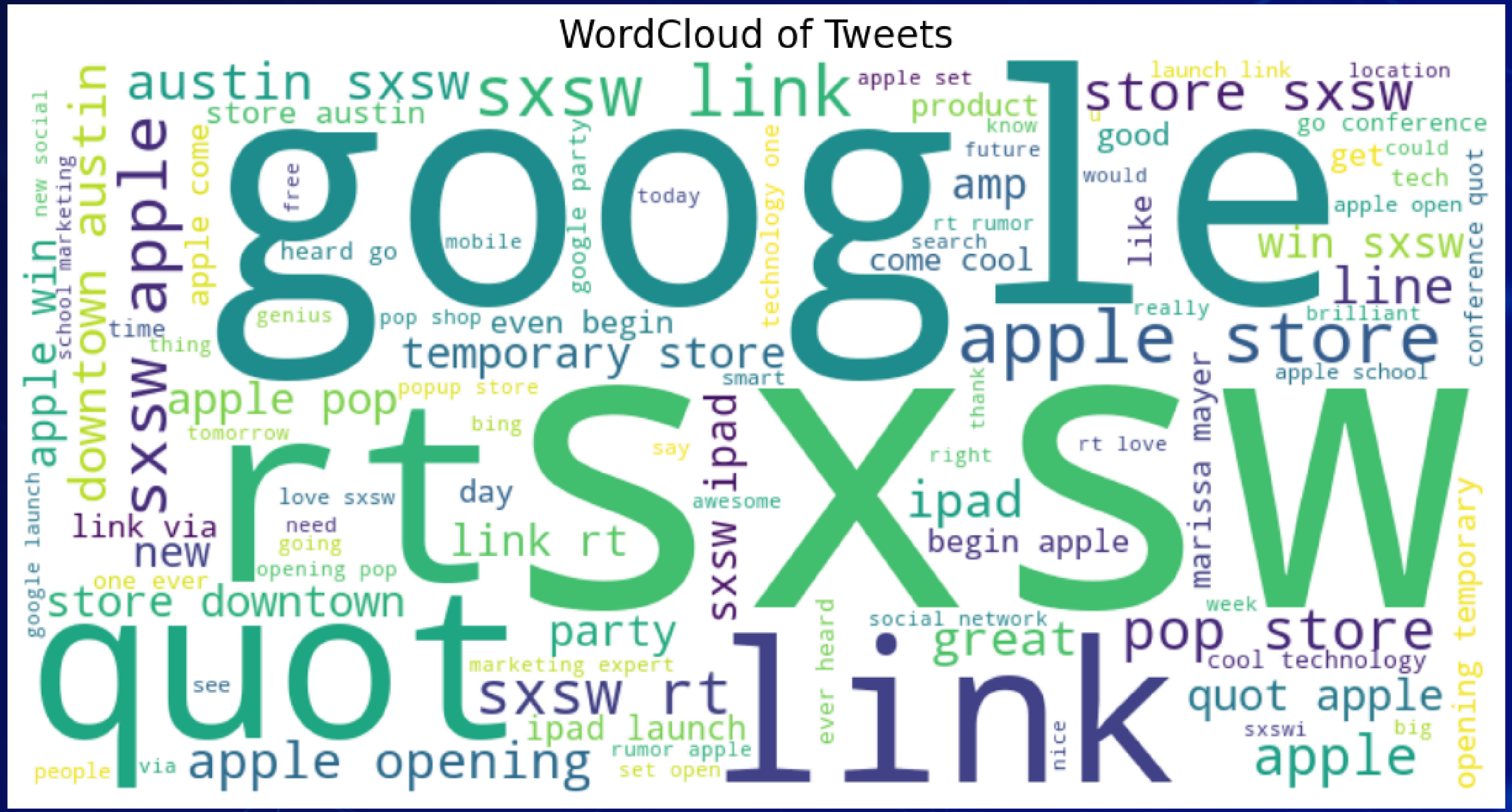
The positive sentiments dominates in both brands but the Apple brand recieves more positive sentiments compared to the Google brand.

Top 10 Words in tweets

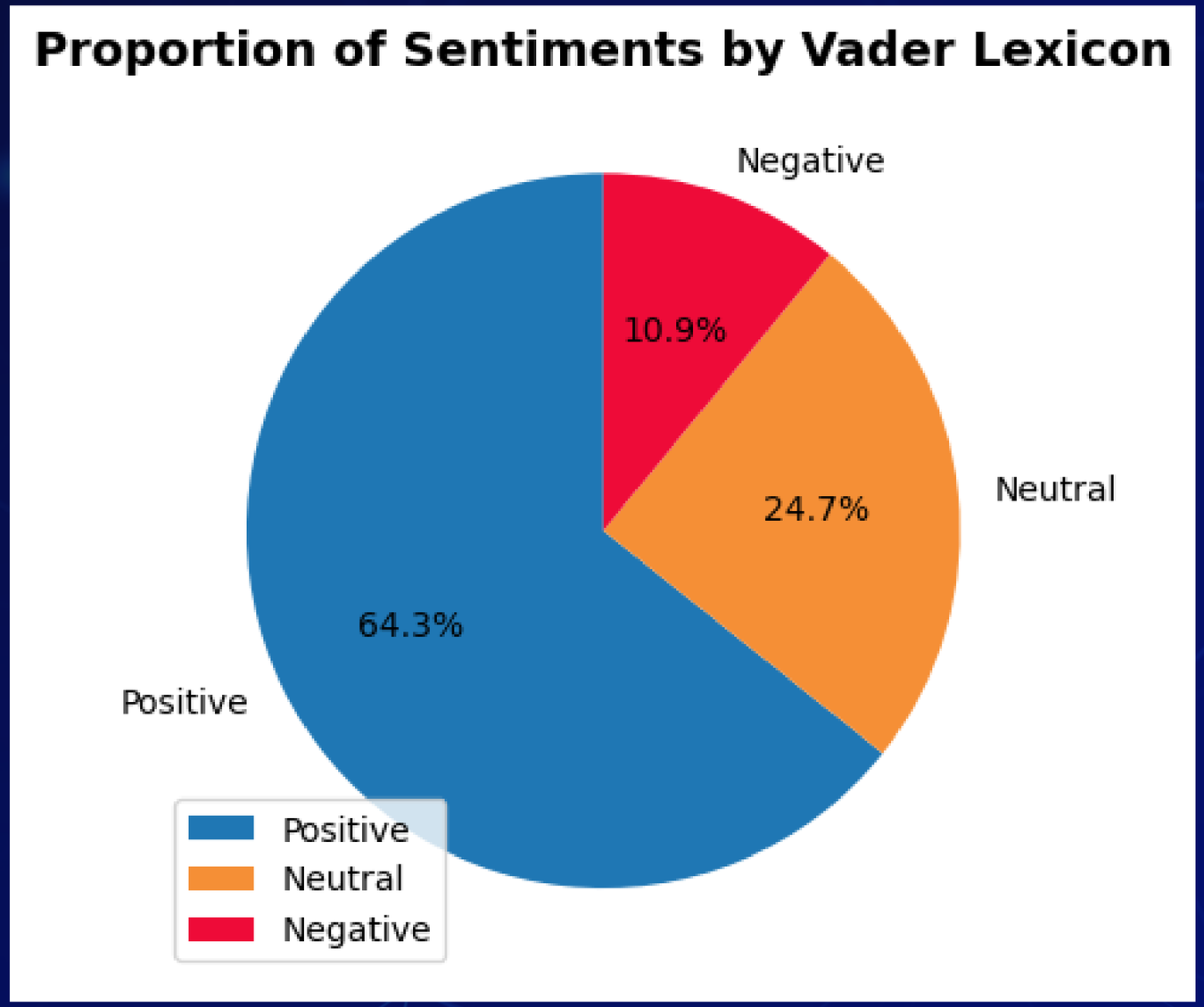


sxsw appears the most. After apple and google, link and retweet are the most common words in tweets that relate to the two brands.

Word cloud on the common words



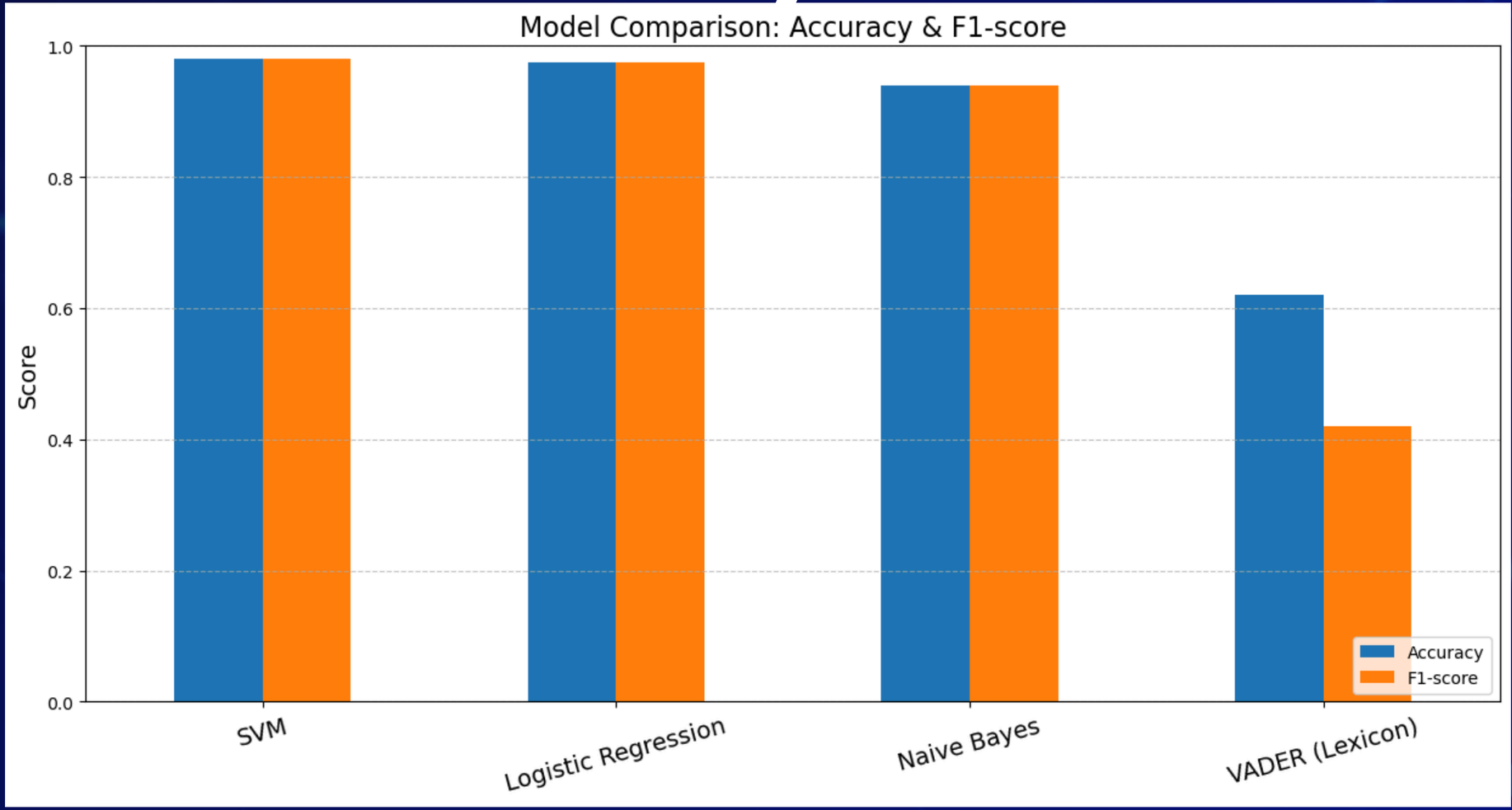
Sentiment Proportion



The positive class is dominating with 64.3%

The Neutral has 24.7% while the Negative is the least with 10.9%.

Models Accuracy



SVM achieved the highest accuracy of 98.12%, slightly outperforming Logistic Regression of 97.55% and Naive Bayes of 93.97%.

Sentiment Analysis App

A streamlit app was generated to predict the sentiment based on the tweet.

Conclusion

This project explored different models for sentiment analysis, including VADER, Naive Bayes, Logistic Regression, and SVM.

The SVM model performed best, achieving an accuracy of 0.9812, showing a strong ability to understand patterns in the text.

Overall, machine learning models, especially those using TF-IDF features, performed much better than rule-based approaches like VADER.

Recommendation

- SVM to be used as the main model for sentiment classification as it performes better than other models with a 0.98 accuracy.
- Logistic Regression performed strongly with an accuracy of 0.98 and can be used alongside SVM when transparency and stakeholder understanding are priorities
- Naive Bayes achieved an accuracy of 0.94, making it a suitable lightweight analysis tool, but not for sentiment monitoring

Thank You