SENTIMENT ANALYSIS WITH NLP

Team Members

- Royce Bett
- Catherine Wangui
- Wilson Mutungu
- Paulette Ndeda

BUSINESS OVERVIEW

BUSINESS UNDERSTANDING

Twitter is an invaluable resource for companies for gathering data on public sentiment. Apple and Google are two such companies that stand to benefit from this. They are constantly and consistently designing and rolling out new products and services, and public feedback is crucial to the optimization of products and rollout.

DATA UNDERSTANDING

Contributors evaluated tweets about multiple brands and products. The crowd was asked if the tweet expressed positive, negative, or no emotion towards a brand and/or product. If some emotion was expressed they were also asked to say which brand or product was the target of that emotion.

02 METHODOLOGY

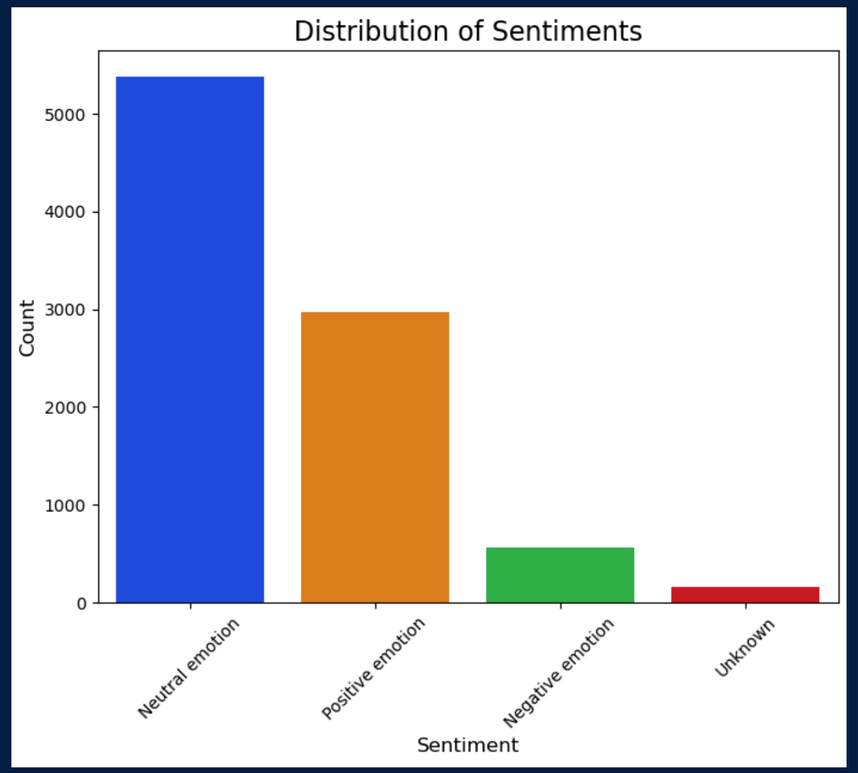
01 Data preparation and cleaning using python and packages such as nltk.

02 Exploratory data analysis to improve comprehension and visualization of the data.

03 Modeling using methods such as Logistical regression, Naive Bayes and Random Forest and evaluating the effiecacy of the models

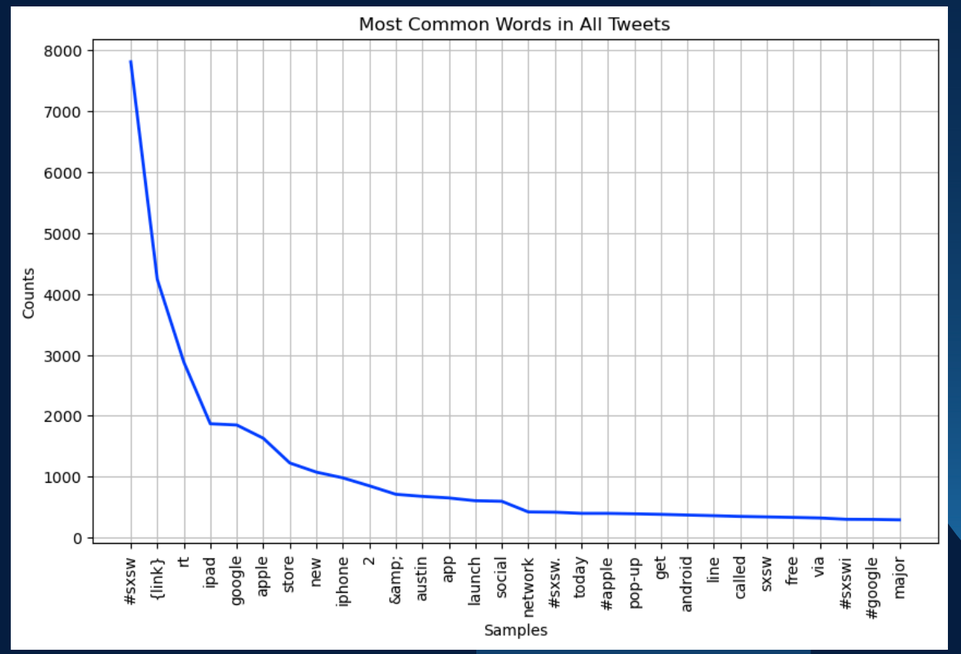
04 Draw conclusions and give recommendations based on the findings.

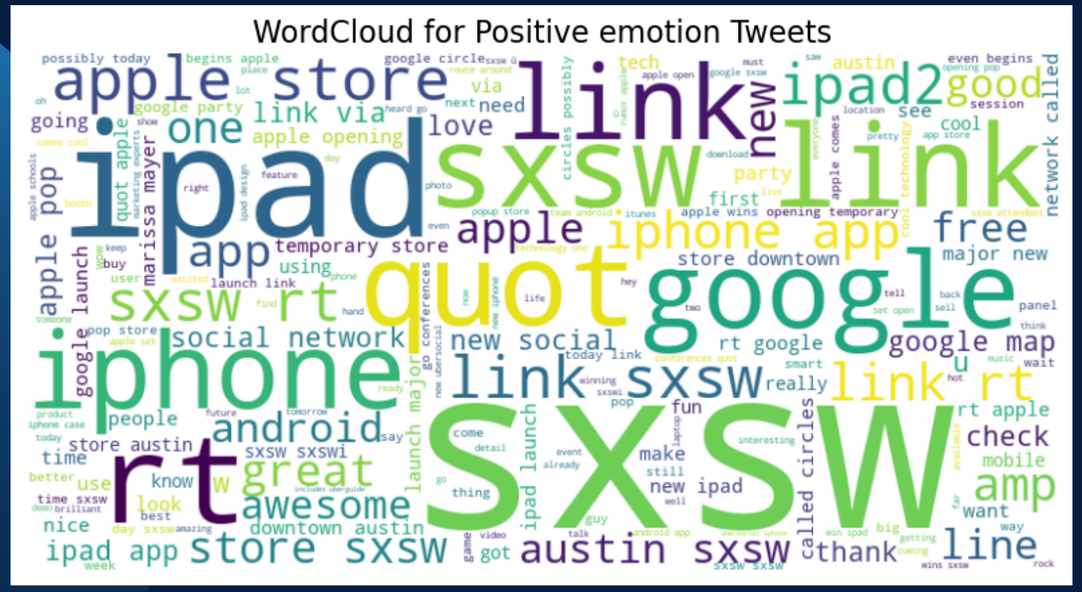
KEY FINDINGS

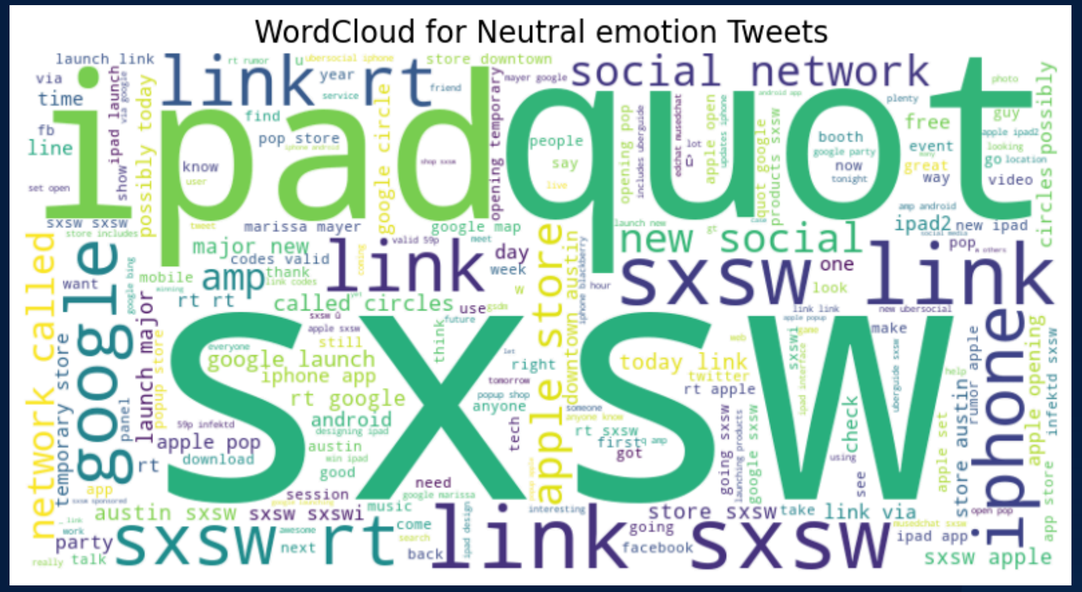


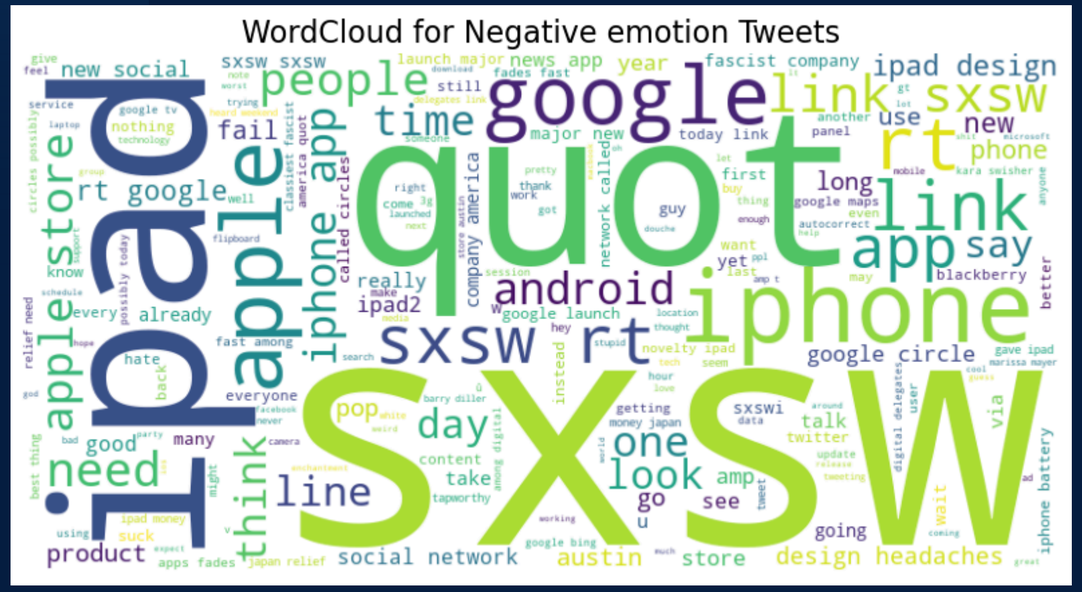
There was a larger proportion of neutral reviews, followed by positive reviews, and finally the negative.

There was also an interest in the distribution of keywords in tweets that could help optimize the model's performance.



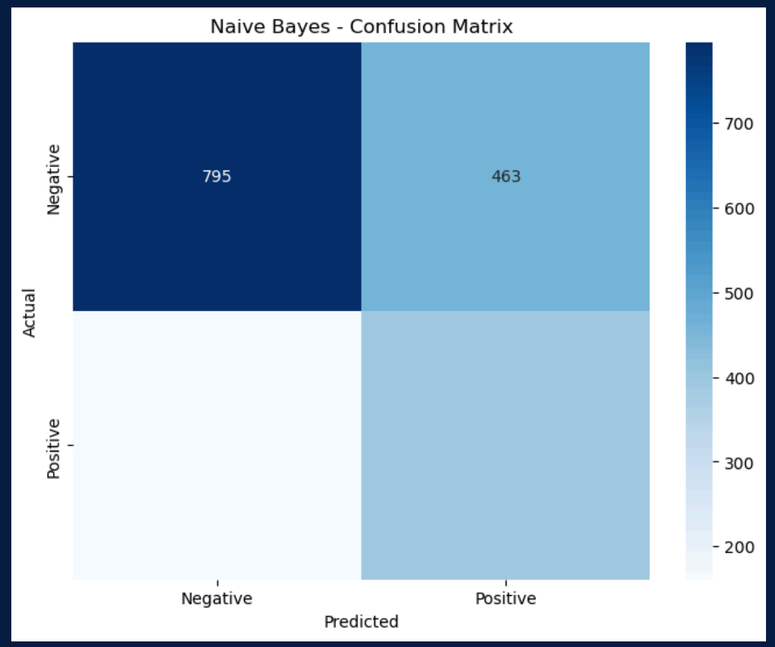


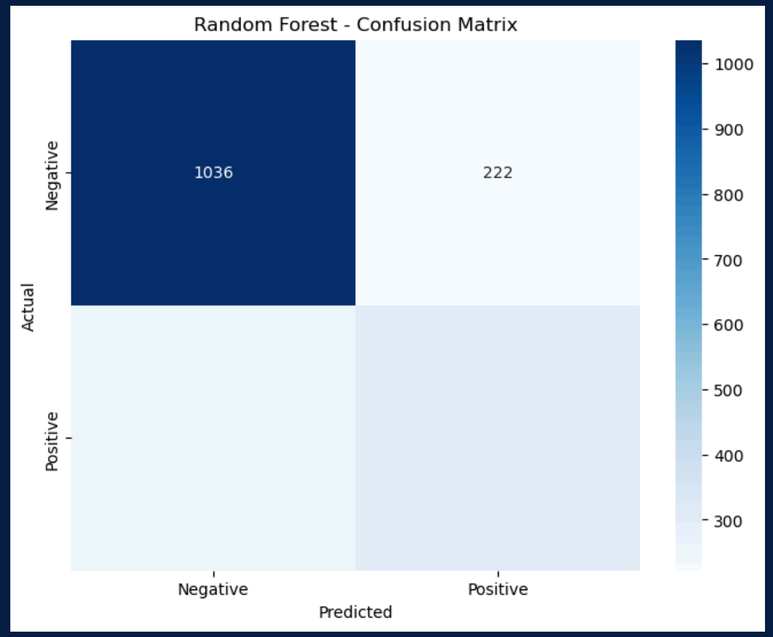




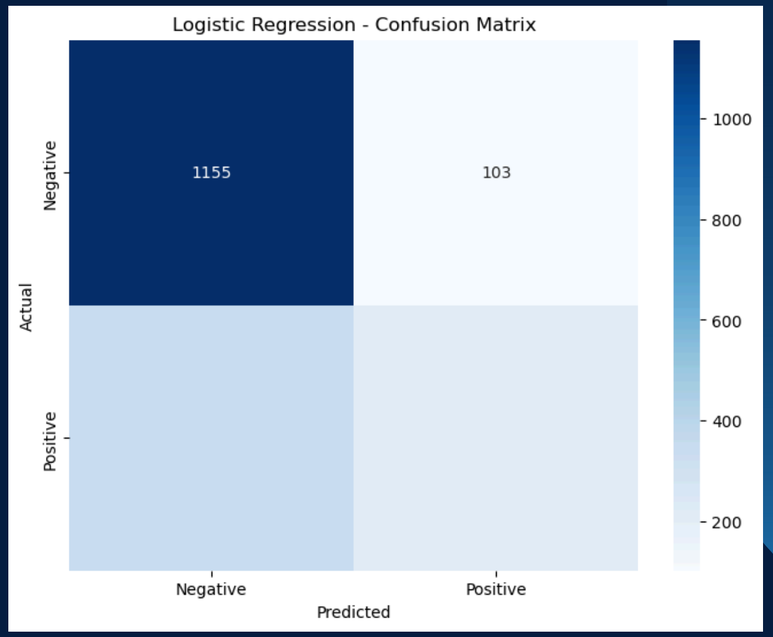
These wordclouds visualize the more prevalent words as per the sentiment

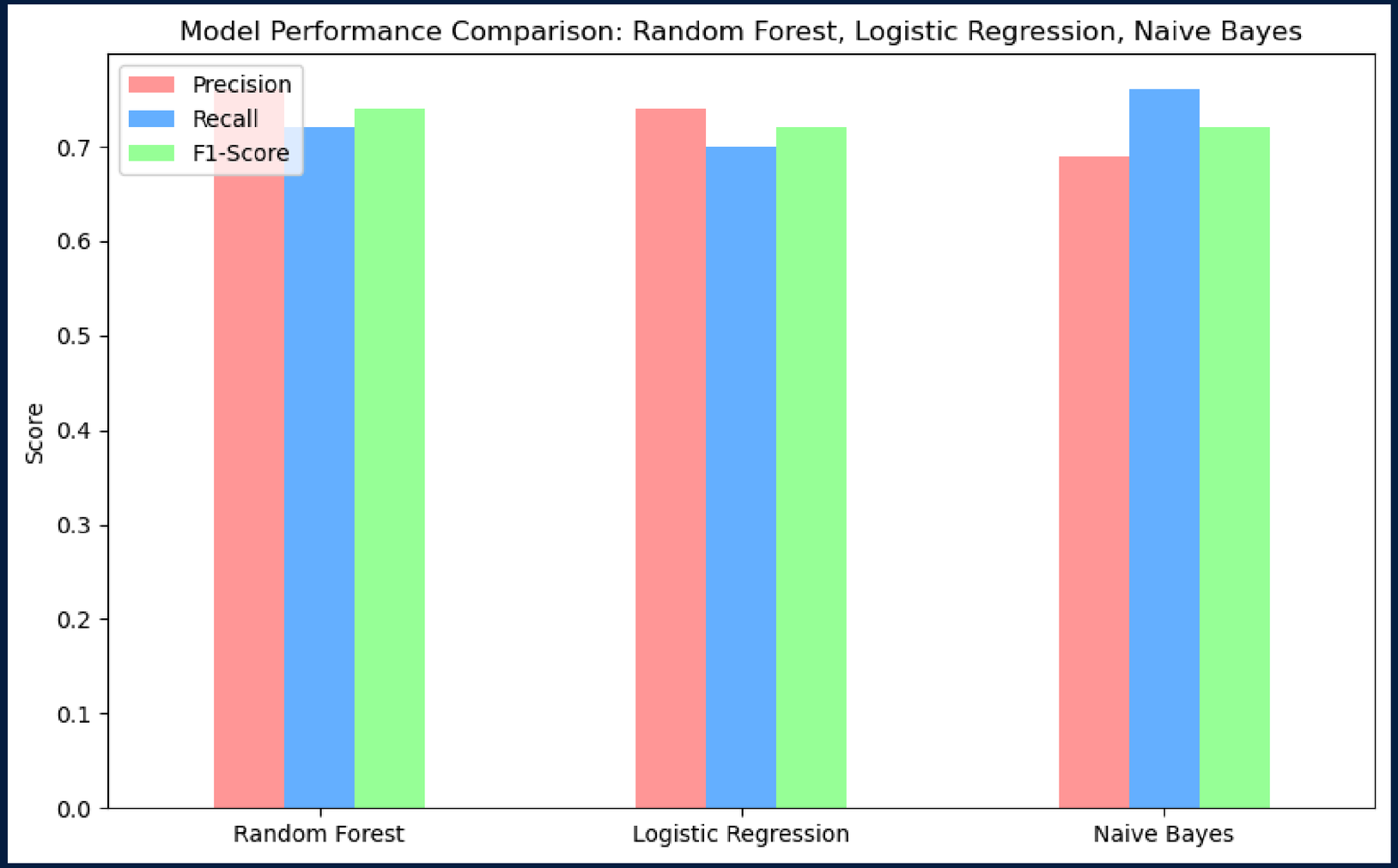
MODELLING





As stated earlier, three models were used,
yielding the results in these confusion matrices





This then depicts the performances of the different models against each other in the different metrics(precision, recall & F1 score)

CONCLUSIONS

1. The majority of tweets related to tech products exhibited positive sentiments, reflecting strong brand perceptions for the products analyzed.
2. The Random Forest classifier emerged as the best-performing model with balanced precision and recall, making it well-suited for real-world application in sentiment classification.
3. Although the model performed well, the class imbalance (more neutral and positive tweets) suggests the need for careful data handling to maintain performance across all classes, particularly for the minority negative class.

RECOMMENDATIONS

1. Given that positive sentiment dominates the data, brands can leverage this in marketing campaigns by highlighting the most favorable topics (e.g., product innovation, customer satisfaction).
2. Negative sentiment still exists and revolves around specific issues. Addressing common negative themes like battery performance for Apple or map service issues for Google can help improve overall perception.
3. Expand the analysis to other product categories or competitor products to gain broader insights into customer sentiment across the tech industry.

THANK YOU