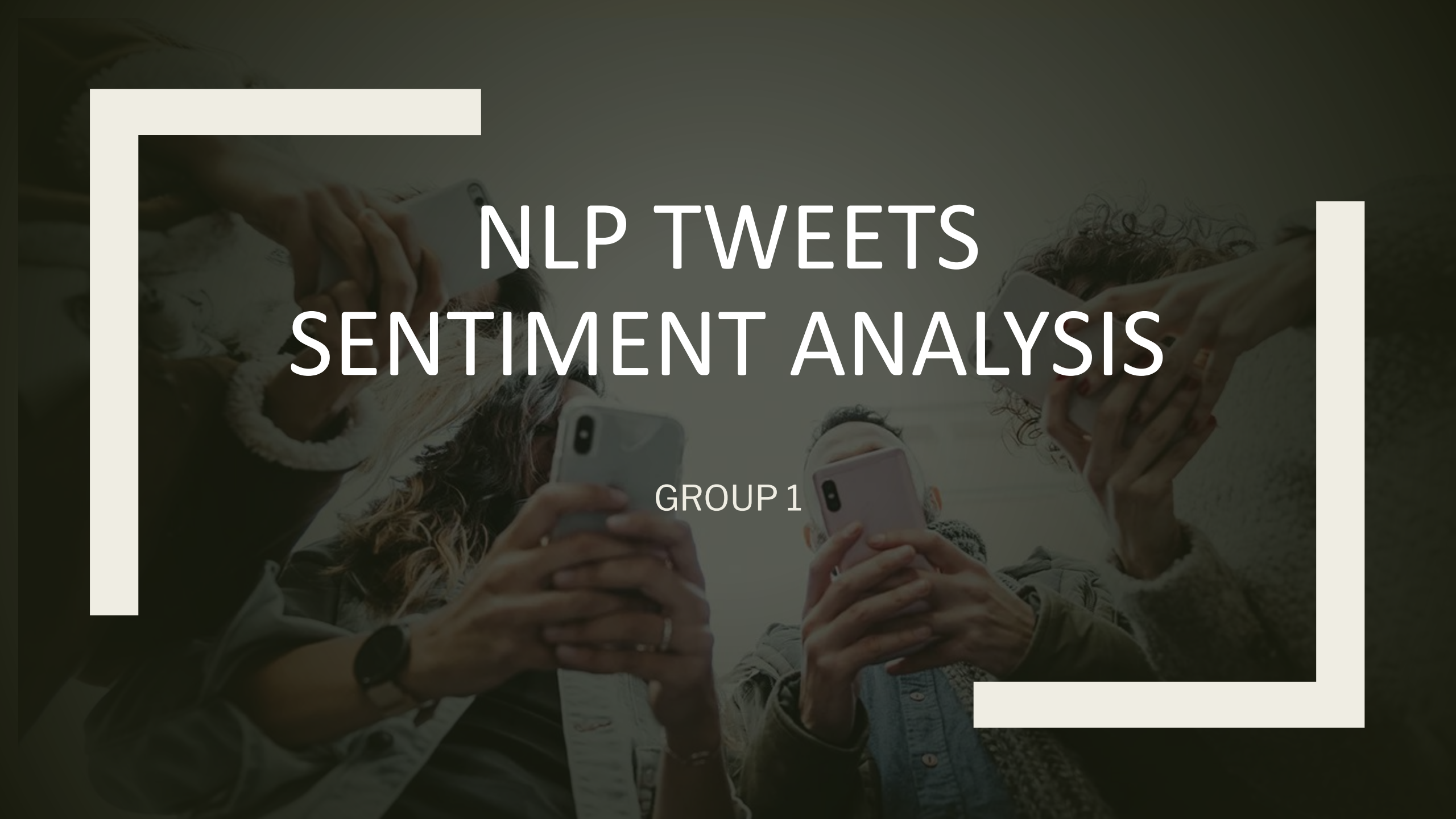


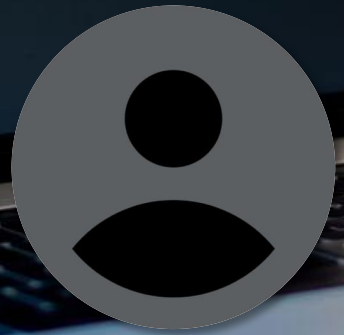
NLP TWEETS SENTIMENT ANALYSIS

GROUP 1

GROUP MEMBERS

GERALD MWANGI
Group Leader

Lynn Komen



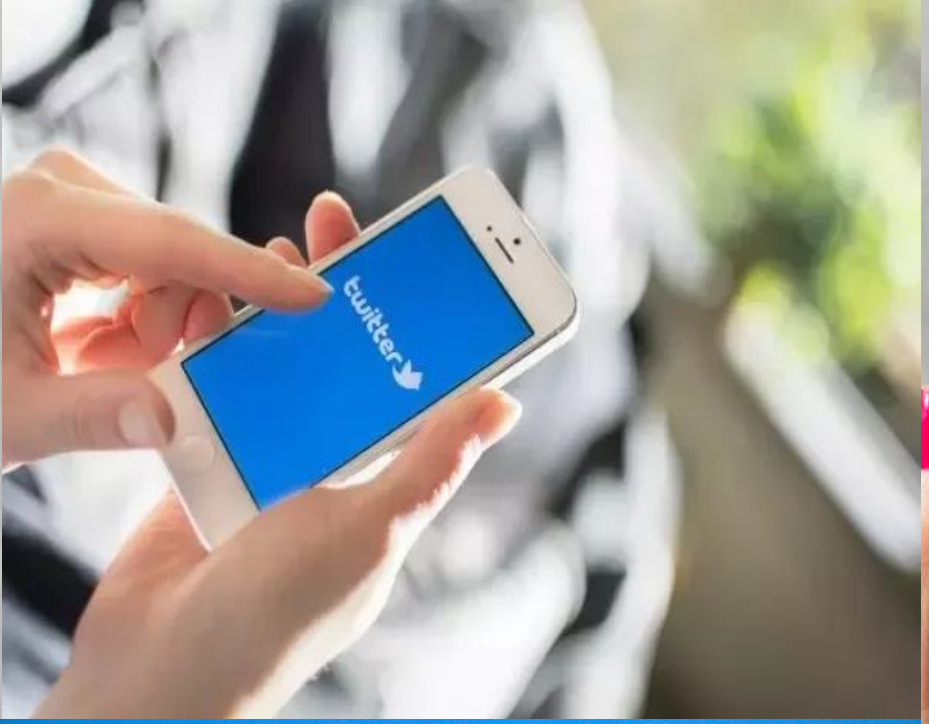
Ian Kiptoo

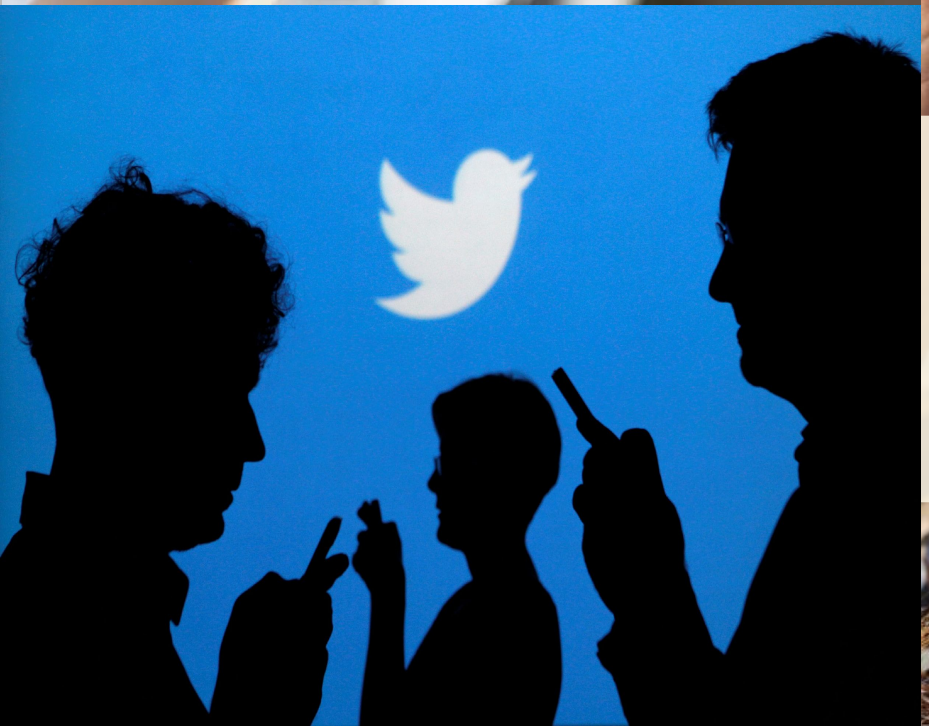
OVERVIEW

In the fast-paced electronics market, Codel Electronics aims to enhance customer satisfaction and brand reputation by understanding consumer sentiments expressed in tweets. By leveraging data analytics and machine learning, Codel Electronics can proactively address negative sentiments and capitalize on positive feedback to improve customer loyalty and attract new clients. Without these insights, Codel risks falling behind competitors who use sentiment analysis to drive better customer experiences and brand loyalty.

PROBLEM STATEMENT

Codel Electronics Company lacks comprehensive insights into consumer emotions and public opinion towards its products and brands. This hinders their ability to identify areas for improvement, enhance customer satisfaction, and develop effective market strategies. A systematic approach to analyzing Twitter sentiment is needed to gain actionable insights and stay competitive.





OBJECTIVES

1. To develop predictive models to classify the sentiment (positive, negative, or neutral) expressed in tweets about brands and products.
2. To identify the best-performing product, in terms of positive and negative.
3. To understand sentiment (emotion) distribution regarding Apple and Google brand.
4. To provide actionable insights from analyzing Twitter sentiments.

DATA UNDERSTANDING

Source

The dataset is sourced from crowdflower via data.world.com.

Columns

1. Tweets: Tweet content (predictor variable).

2. Product: Target brand/product.

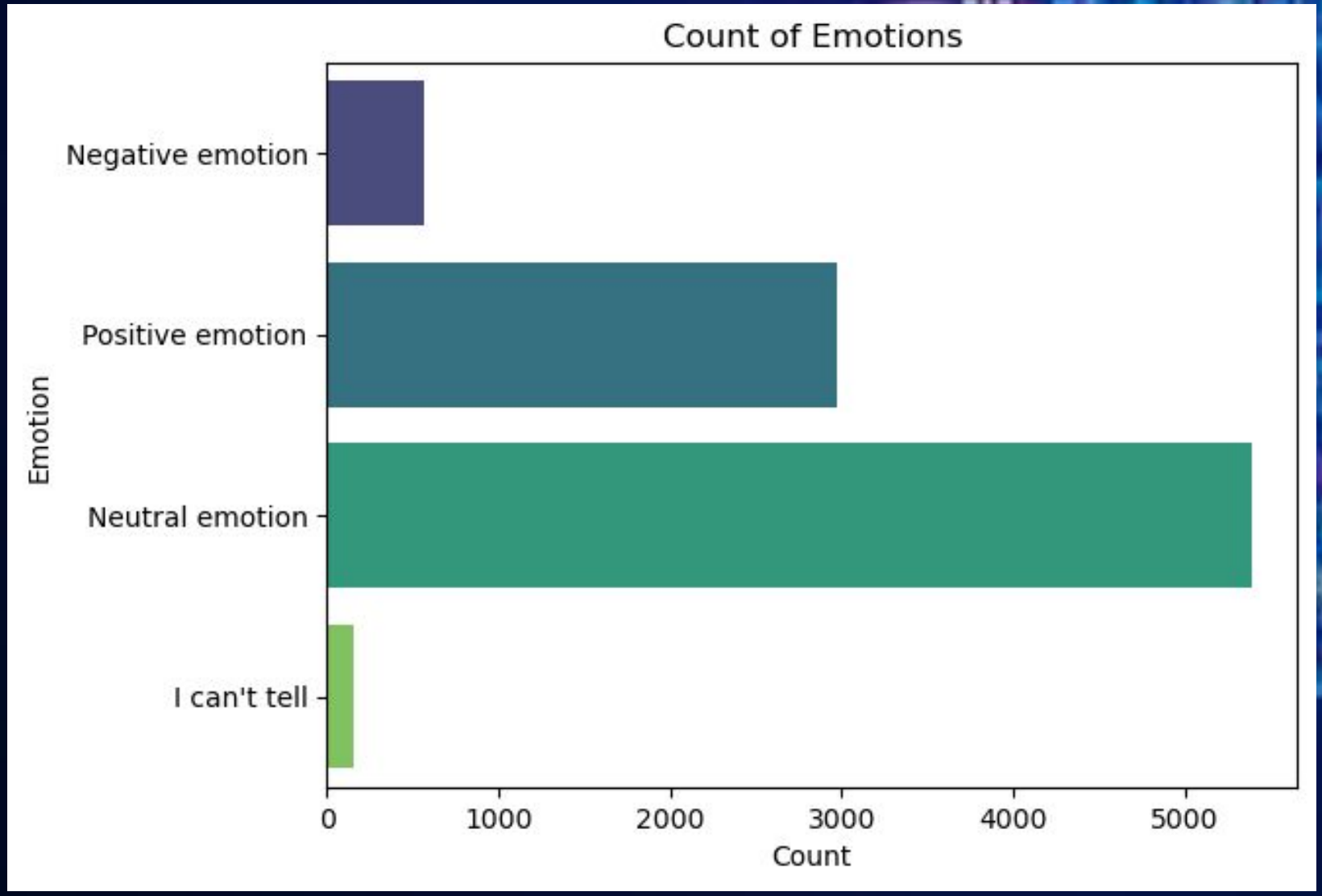
3. Emotion - Emotion type (target): Positive, Negative, Neutral.

Summary

1. The datasets contain more than: 9000 records t.

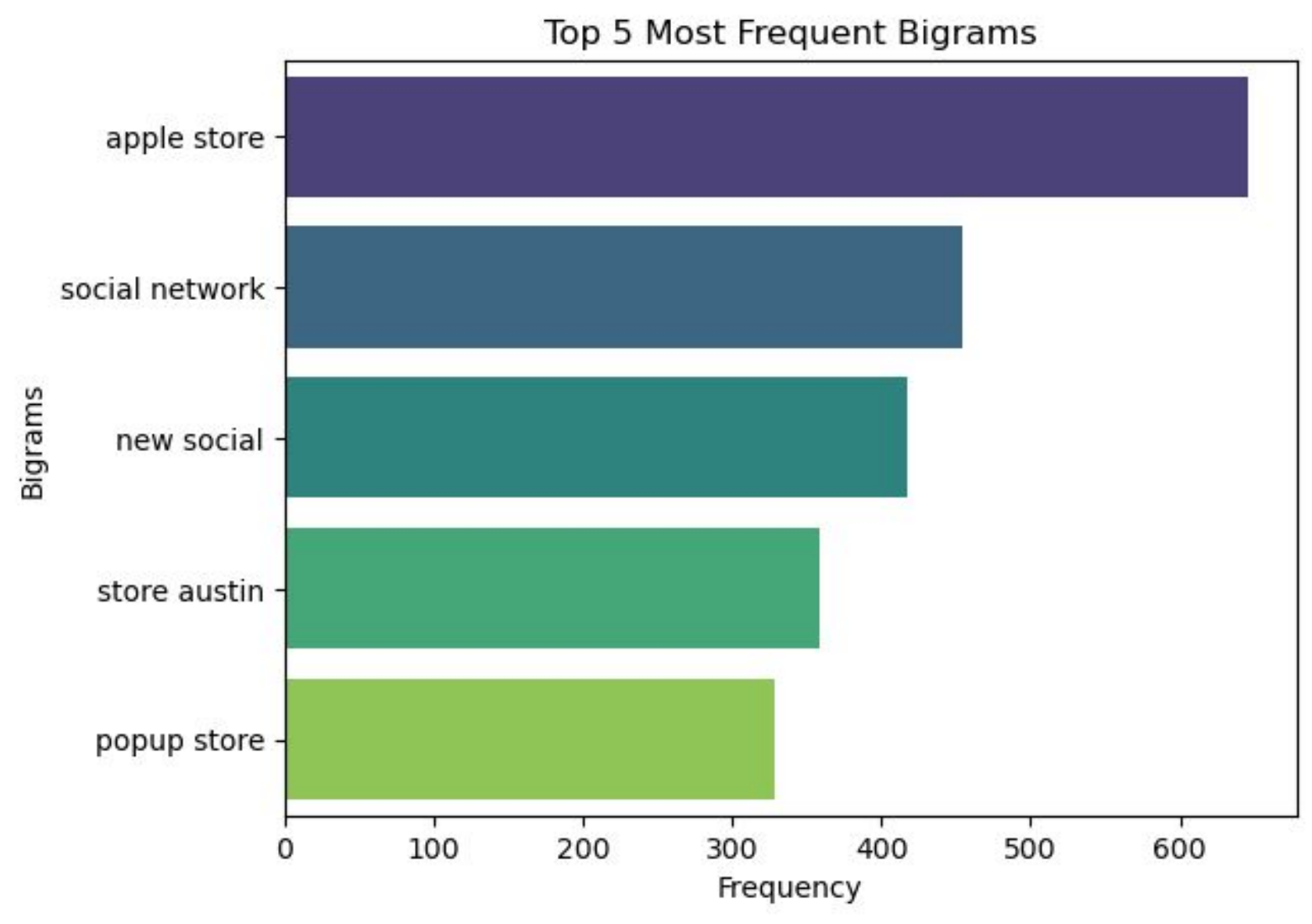
2. It has 3 Columns: All the data type of the dataset is categorical.

UNDERSTANDING THE DISTRIBUTION OF EMOTIONS



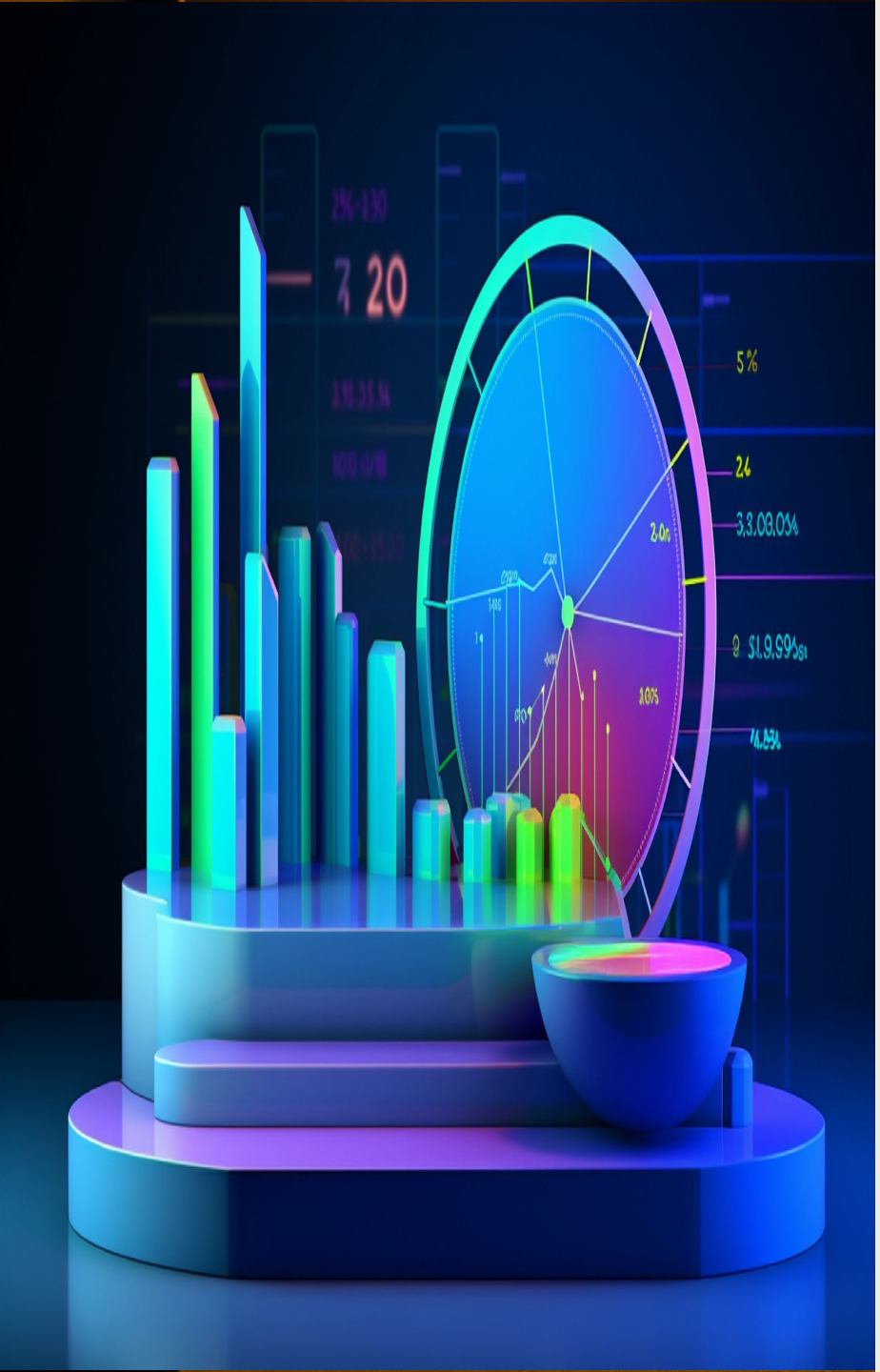
The most frequently occurring emotion is Neutral emotion, with a count of more than 5000, followed by positive emotion with more than 2500 counts. Negative emotions with more than 500 counts and finally I can't tell with below 200 counts.

UNDERSTANDING THE MOST USED WORDS IN THE TWEETS



We can observe that the bigrams such as Apple store are mostly related to dominating the Google and Apple products tweets.

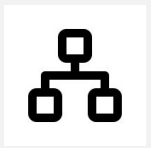
Bigrams such as "rt mention" dominating the dataset suggest that many tweets are retweets or mentions, typically used to highlight trending topics, news, and user opinions.



MODELLING

Classification Tasks

In this project, we analyzed tweet sentiments using two main approaches.



Binary Classification: We aimed to identify tweets as either 'Positive' or 'Negative'.

Best Model: Random Forest Classifier (Accuracy: 86%).

Multiclass Classification: We included 'Neutral' along with 'Positive' and 'Negative' emotions.

Best Models: Voting Classifier and Stacking Classifier (Accuracy: 68%).

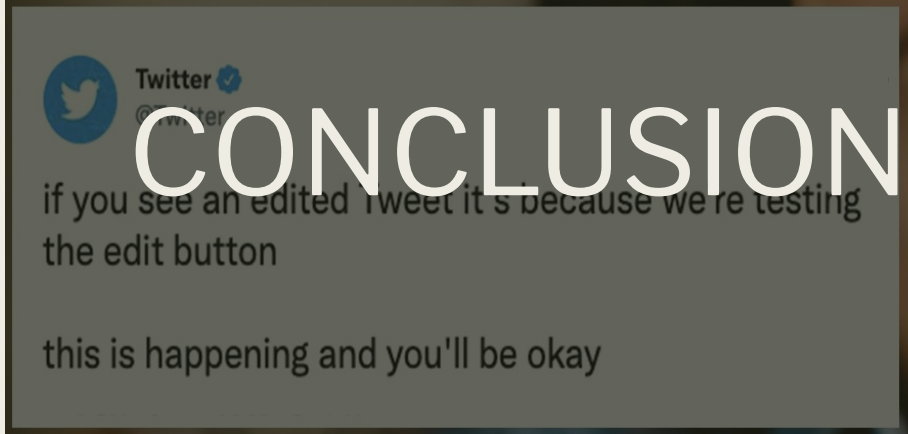
MODELLING BINARY CLASSIFICATION (POSTIVE AND NEGATIVE)

Model	Accuracy	Precision	Recall	F1score	Roc-Auc
logistic	0.85	0.93	0.89	0.91	0.87
Naïve baye's	0.85	0.91	0.91	0.91	0.83
Random forest	0.86	0.89	0.95	0.92	0.82

MODELLING MULTICLASSIFICATION(POSTIVE ,NEUTRAL AND NEGATIVE)

Model	Accuracy	Precision	Recall	F1 score	Roc-Auc
Naïve bayes	0.64	0.52	0.50	0.51	0.62
Gradient Boosting	0.63	0.57	0.55	0.56	0.68
Voting Classifier	0.68	0.60	0.61	0.60	0.70
Stacking Classifier	0.68	0.60	0.61	0.60	0.70

CONCLUSION



1. Most of the sentiments show no emotion towards the brand.
2. The most mentioned product brand was Apple products.
3. The best brand in terms of positive emotion is Google.
4. The worst brand in terms of negative emotion is Apple.
5. Best Binary classification model is Random forest classifier and in the multiclass is Stacking classifier.

RECOMMENDATIONS

1. **Develop strategies to convert neutral sentiments into positive ones.**
2. **Create targeted campaigns to address specific issues in the negative sentiments.**
3. **Analyze why Apple products dominate mentions and benchmark against Apple.**
4. **Stock google products.**

THANK YOU

GROUP 1