

Twitter Sentiment Analysis



Google and Apple Products

MORINGA SCHOOL PROJECT

1. Elvis Wanjohi
2. Jessica Gichimu
3. Jesse Ngugi
4. Stephen Gachingu
5. Latifa Riziki

Purpose of the Project

Context

- Google's and Apple's reliance on maintaining customer satisfaction
 - Analyze data from data world and classify sentiments

Our goal

- The main objective of this project is to develop a sentiment classification model that analyzes tweets about Apple and Google products and classifies them as positive, negative or neutral.

Business Understanding

Business Overview

The main objective of this project is to develop a sentiment classification model that analyzes tweets about Apple and Google products and classifies them as positive, negative or neutral.

Key questions

1. Which products and services from Apple or Google have the largest negative, positive and neutral feedback?

Data Understanding

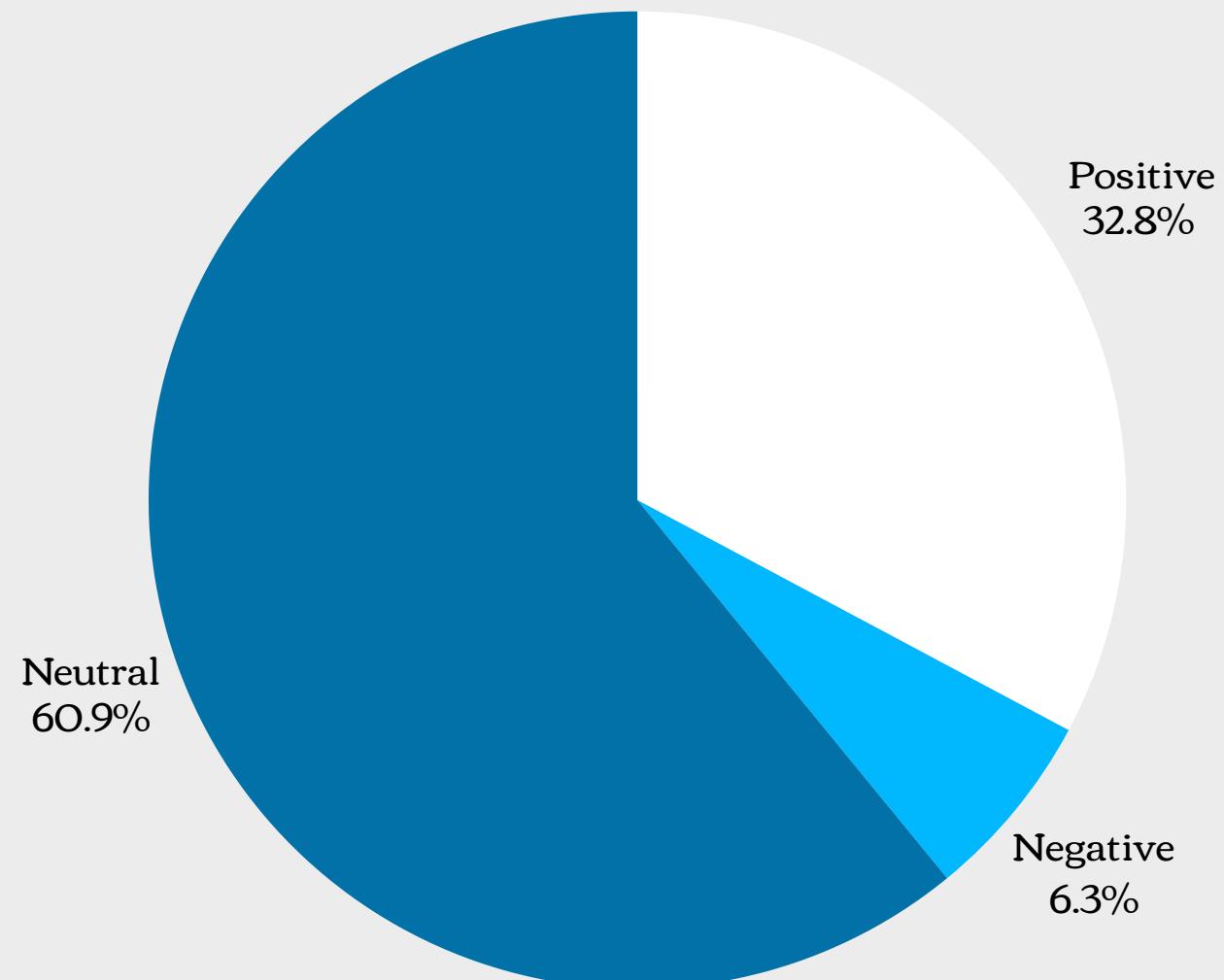


Description of Data:

- Data format: CSV file (judge-1377884607_tweet_product_company.csv).
- Number of records (rows): 9093 rows (varies depending on version of dataset).
- Number of fields (columns): 3 columns (Tweet, tweet_directed_at, sentiment).

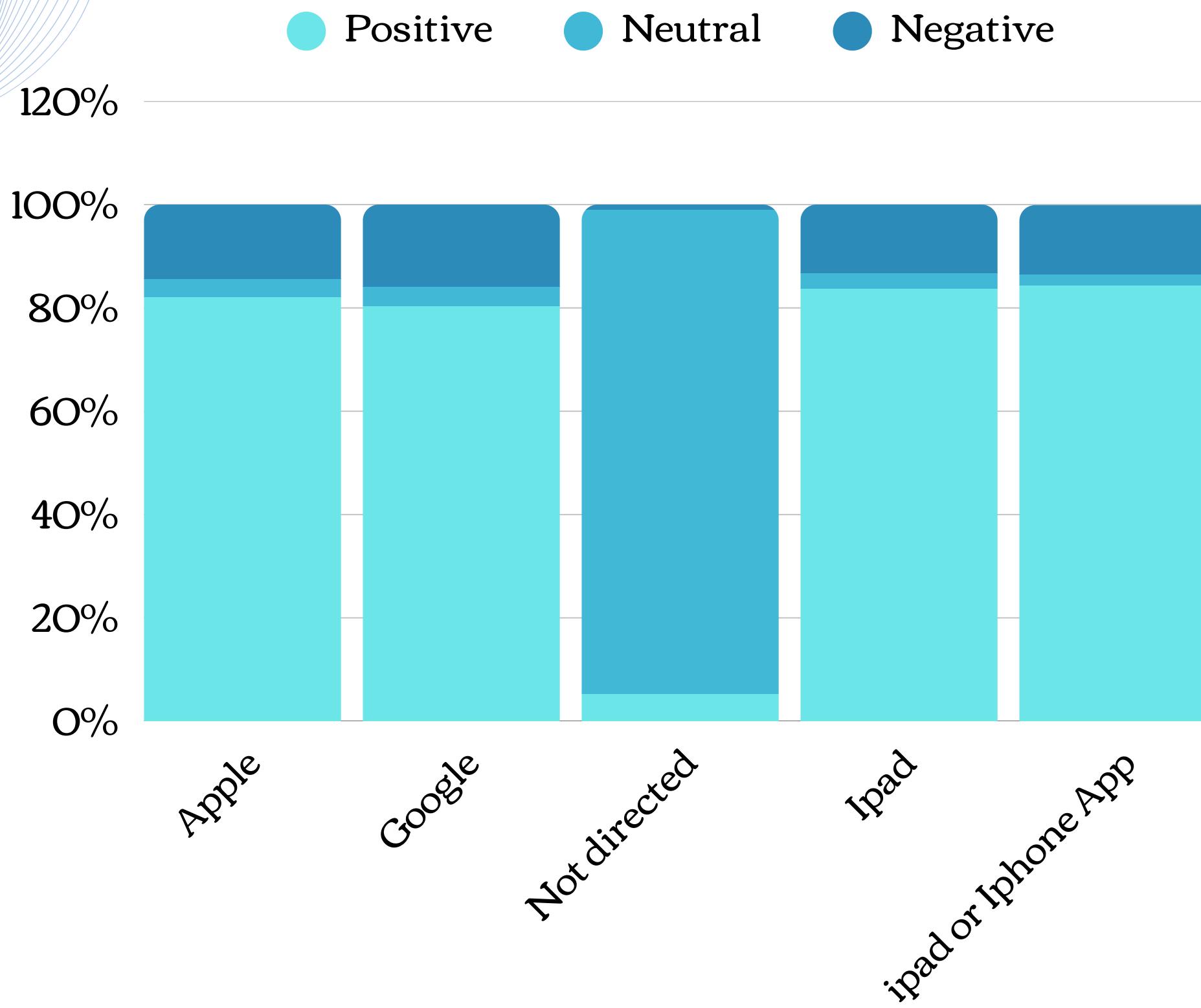
DATA ANALYSIS

SENTIMENT DISTRIBUTION



- Out of the 9070 tweets in the dataset, 5531 tweets express a neutral emotion. This is about 60.98% of the total tweets, 2970 tweets express a positive emotion, which is about 32.75% of the total tweets and 569 tweets express a negative emotion, which is about 6.27% of the total tweets.

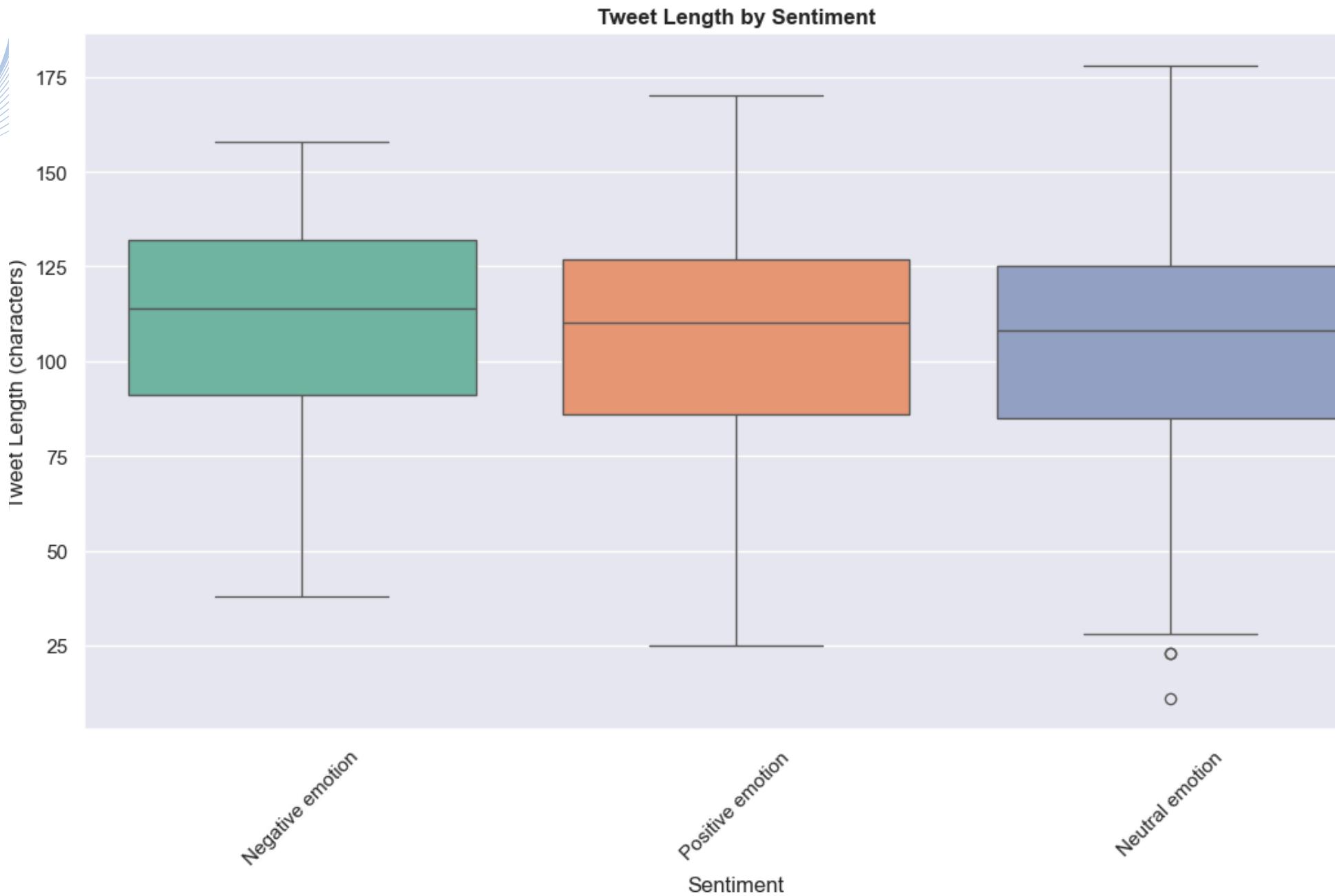
Sentiment proportions by top 5 destinations



This plot shows the percentage distribution of positive, neutral and negative sentiments across the top five tweet destinations:

- Apple, iPad, and iPad or iPhone App show strong positive sentiment, exceeding 80%.
- Google also records high positivity at about 80% but slightly higher negative and neutral shares than Apple destinations.
- Not directed tweets are mostly neutral at 93.8%, showing minimal sentiment variation.
- This shows that tweets targeting Apple and Google products highlight favourable public perception, while non-targeted tweets remain mostly neutral.

Tweet Length Distribution by Sentiment



- Neutral tweets has a wider range of lengths. The median length remains consistent across all sentiment groups.
- Outliers with longer tweet lengths are present within each sentiment category.
- The tweet length shows little variation across sentiments. This shows that sentiment type is not strongly associated with tweet length in this dataset.

Model Results

Best Model

For Binary Classification, LinerSVC had the best accuracy score of 86.44% compared to Logistic Regression with 0.84 and Naives Bayes 0.80

Confusion Matrix

Observation for LinearSVC

- The model has a high True positive rate(818) indicating it correctly identified positive emotions.
- The model has a low False positive rate(81) and low False negative rate(61) indicating the model was able to make correct predictions on the Negative and positive classes since the False negative rate and False positive rate are low



MULTI-CLASS CLASSIFICATION

Accuracy

Linear SVC has the best Validation accuracy with an accuracy of 68%

Classification Report

- The True positive total is 909 these are the values the model correctly predicted for each and every class
- The False positive and False negative for the negative class is (15) and (74) respectively.