SENTIMENT ANALYSIS ABOUT APPLE AND GOOGLE PRODUCTS
GROUP 7

- MEMBERS:
- 1. TRACY GWEHONA - BUSINESS UNDERSTANDING & DATA UNDERSTANDING
- 2. BOSCO MUKARA - DATA CLEANING & EDA
- 3.EDWIN MAINA - DATA PREPROCESSING & DEPLOYMENT
- 4. STEPHEN MUNGAI - MODELLING
- 5.KELVIN MWANGI - SCRUM MASTER, README.md & PRESENTATION

OVERVIEW

- This project aims to analyze Twitter sentiment about Apple and Google products using Natural Language Processing (NLP). The dataset contains tweets labeled as positive, negative or neutral. By building a sentiment analysis model, we aim to categorize the sentiment of tweets accurately and gain insights into public perception of these tech giants' products.

BUSINESS PROBLEM

- Understanding customer sentiment is critical for businesses to gauge public opinion and improve products or services. For Apple and Google, analyzing Twitter sentiment can provide actionable insights to enhance user satisfaction and market strategies.

STAKEHOLDERS

1. Marketing Teams : Use insights to create targeted campaigns focusing on products with positive sentiment. Address negative feedback to improve brand perception.
2. Product Teams : Identify areas of improvement for specific products (e.g., iPhone or Pixel).
3. Executives : Make data-driven decisions for product launches, pricing strategies, and market positioning.

OBJECTIVES

1. Build a model to classify the sentiment of tweets into positive, negative, or neutral categories.
2. Evaluate model performance using suitable metrics.
3. Provide insights and recommendations based on the analysis results.

DATA UNDERSTANDING AND INSPECTION

- Dataset Overview :

The dataset contains over 9000 tweets related to Google and Apple products. This provides valuable insights into public opinions and perceptions.

- Key Columns:

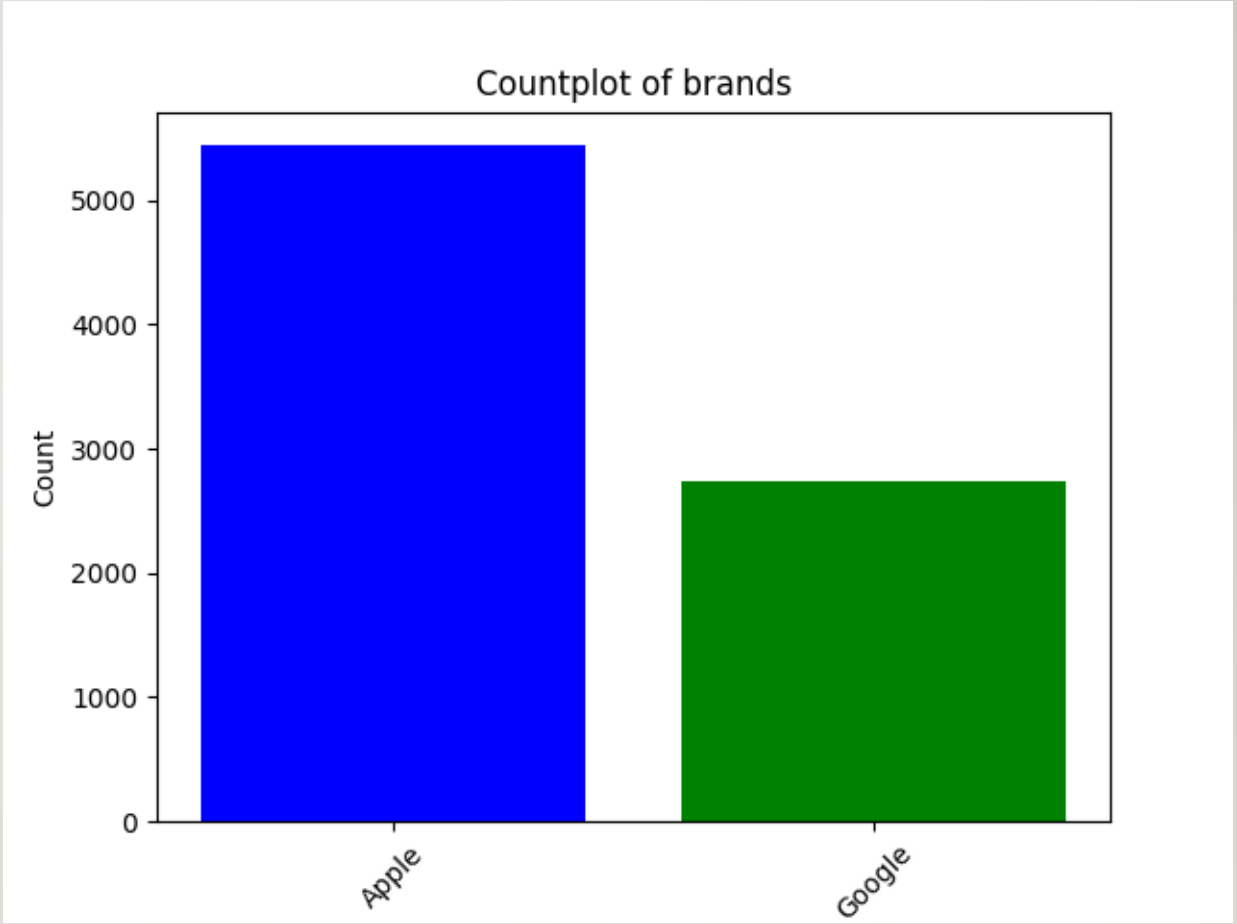
Tweet: The text content of the tweet, capturing user opinions and reviews.

Brand: Indicates the product's associated brand (Google or Apple).Sentiment: Labels the tweet as Positive, Negative, or Neutral, reflecting the user's sentiment.

VISUALIZATIONS (COUNT PLOT)

Apple has the most sentiments having more than 5000 counts.

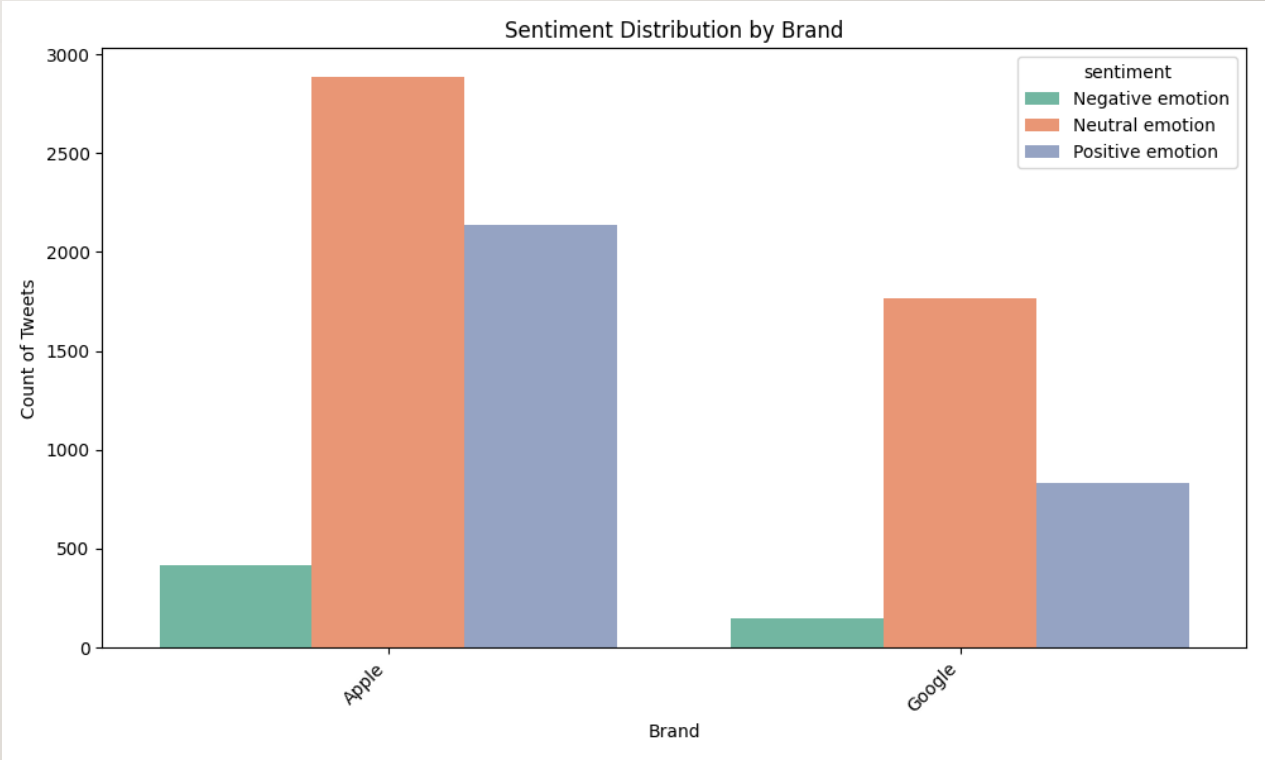
Google follows with slightly short of 2000



SENTIMENT DISTRIBUTION BY BRAND

Here we have the distribution of sentiments by brand.

Apple has the most sentiments in all sections.



MODELS UTILIZED

- Logistic Regression
- Random Forest
- Multinomial Naive Bayes
- Support Vector Machine
- The **Voting Classifier combining Logistic, Random forest and SVM models** is recommended as the best model for sentiment analysis on this dataset as it provides the best of all models

RECOMMENDATIONS

1. Adopt the Best Performing Model The Voting Classifier is the most suitable for sentiment analysis due to its strong performance metrics:

- **F1 Score:** 0.6452 (indicating good balance between precision and recall).
- **ROC AUC:** 0.7483 (showing strong ability to distinguish between sentiments).
- **Consistency:** Low standard deviation of 0.0043, ensuring reliable predictions across data samples.
- **Accuracy: 0.6623**, the highest among all models

RECOMMENDATIONS

2. **Address Challenges with Positive and Neutral Sentiments** The analysis revealed that all models struggled to differentiate between positive and neutral tweets. This is due to the overlap in vocabulary used in these sentiments. To address this:

- **Actionable Insights:** Review common keywords in tweets classified as positive or neutral and refine them. For example, terms like "good" and "okay" may need additional context to ensure correct classification.
- **Enhanced Data Collection:** Better classification on sentiment labels can be achieved by collecting more diverse and detailed data. This can help in distinguishing subtle differences in sentiment expressions. For example, collecting tweets with stronger emotional language can improve model performance.
- **Different Data Source:** Consider using additional data sources beyond Twitter to capture a wider range of sentiments and expressions. This can provide a more comprehensive view of customer feedback and improve sentiment analysis accuracy.
- **Deep learning approaches** use advanced deep learning methods like (BERT, transformers) to improve performance or advanced vectorizers like word2vec to improve pattern recognition in the dataset.

CONCLUSIONS

- The **Voting Classifier combining Logistic, Random forest and SVM models** is recommended as the best model for sentiment analysis on this dataset as it provides the best of all models. While it achieved the highest F1 Score and ROC AUC, the presence of class imbalance and the similarity of words in positive and neutral tweets remained challenges for all models.

THANK YOU

QUESTIONS