APPLE TWITTER ANALYSIS DATA REPORT

MORINGA DSF-FT 11- HYBRID

GROUP V PROJECT

PRESENTED BY:

Michael Gatero

Betsy Gitije

Abdirahaman Sheikh

Kelvin Nyawira

Petina Leni

Table of Contents

Apple Twitter Analysis Data Report

1.0 Business Understanding

1.1: Business Context

Apple is one of the most talked-about brands on social media, with millions of people sharing their opinions of its products, services, and company decisions. Understanding public sentiment from these discussions can help the company track brand perception, customer satisfaction, and market trends. Due to this, the project uses Natural Language Processing (NLP) and machine learning to classify Apple-related tweets as positive, negative, or neutral. By identifying the best-performing model, we can help the company and analysts gain valuable insights into public opinion, guiding better decision-making

1.2: Business Problem

Understanding public sentiment toward Apple on Twitter is challenging due to short, informal text and varying contexts. Misclassifying sentiment can lead to inaccurate insights, affecting company's decisions. This project aims to determine the most effective sentiment analysis model by comparing traditional machine learning and deep learning approaches to achieve the highest accuracy.

1.3: Objectives

1.3.1: Main Objective

To develop an accurate sentiment analysis model for Apple-related tweets by comparing traditional machine learning and deep learning approaches.

1.3.2: Specific Objectives

1. To preprocess Apple-related tweets by cleaning, tokenizing, and normalizing text data to ensure high-quality input for analysis.
2. To handle data imbalance and enhance dataset quality using techniques such as SMOTE and other resampling methods to create a well-balanced training set.
3. To develop and compare multiple sentiment classification models, including traditional machine learning such as Logistic Regression, and XGBoost and deep learning approaches such as LSTM and CNN, to identify the most effective model.
4. To evaluate model performance using appropriate metrics such as accuracy ensuring the best-performing model provides reliable sentiment insights.

1.4: Why Machine Learning and Deep Learning

Machine Learning (ML) and Deep Learning (DL) suit sentiment analysis due to their ability to handle large-scale text data, capture language patterns, and generalize across unseen data.

- **ML models** like Logistic Regression and XGBoost are interpretable, efficient, and perform well on structured text features like TF-IDF and word embeddings. They train quickly and serve as strong baselines.
- **DL models** like LSTM and CNN excel in understanding context, capturing sequential dependencies, and leveraging pre-trained knowledge. They enhance sentiment classification by recognizing complex patterns.

Combining both approaches allows for performance, efficiency, and scalability comparisons, ensuring optimal model selection.

1.5: Success Metrics

The model's performance was evaluated using the following key metrics:

1. Accuracy – The percentage of correctly classified sentiments, with a target of above 70%.

2. Overfitting Control – The model was assessed for generalization, ensuring minimal performance gaps between training and test sets.

3. Model Stability – The model's consistency was tested across different subsets of data to confirm its reliability.

Success was defined as achieving these metrics while preventing overfitting and ensuring robust sentiment classification.

1.6: Key Stakeholders

1. **Apple Inc.** – Understands public sentiment to enhance product development, marketing strategies, and customer engagement.
2. **Investors & Market Analysts** – Leverage sentiment insights to predict consumer confidence and potential stock movements.
3. **Marketing & PR Teams** – Optimize branding, crisis management, and targeted advertising based on sentiment trends.
4. **Technology Consumers & Apple Users** – Benefit from improved products, services, and customer support driven by sentiment analysis.
5. **Data Scientists & AI Researchers** – Gain insights into NLP advancements and sentiment analysis techniques for future applications.

2.0: Data Understanding

2.1: Data Source

The data source for this project is accessed in a web page accessible at the URL https://www.kaggle.com/datasets/slythe/apple-twitter-sentiment-crowdflower/data. The Apple Twitter Sentiment (CrowdFlower) dataset comprises tweets referencing Apple Inc., each labeled with sentiment data sourced from data.world. This dataset is valuable for training and evaluating models in sentiment analysis, particularly within the context of social media discussions about Apple products and services.

The dataset consists of **3886 tweets**, each labeled with sentiment and sentiment confidence scores.

Sentiment Distribution

1. **Neutral (3):** 2162 tweets (Largest class)
2. **Negative (1):** 1219 tweets
3. **Positive (5):** 423 tweets
4. **Not Relevant:** 82 tweets
5. **Observation:** The dataset is **imbalanced,** with more neutral and negative tweets.

Sentiment Confidence Scores

1. The scores range from **0.3 to 1.0.**
2. Peaks at **0.7 and 1.0**, indicating varying label reliability.
3. High-confidence **labels** can be prioritized for training to improve model accuracy.

Tweet Length Distribution

1. Most tweets are **between 100 and 140 characters**.
2. A **longer tweet length** trend is observed, likely due to detailed opinions or news articles.

Handling Missing Values

1. sentiment_gold: **Missing in 3783 rows, making it unusable.**
2. _last_judgment_at: **103 missing values, but not critical for modeling.**

Duplicates

There was no duplicate tweets found.

Top Hashtags and Words

Top Hashtags included: #AAPL, #Apple, #trading, #Stocks, #iPhone6.

Top Words: "apple", "aapl", "http", "rt", indicating frequent mentions of Apple products, financial discussions, and retweets.

3.0: Exploratory Data Analysis

3.1: Univariate Analysis

1. Word Cloud

i) Overall Word cloud:

- A mix of positive and negative words related to Apple products, such as batteries, studio, protests, and future. Some dissatisfaction is apparent (e.g., misplaced, anger), but general topics include technology and Apple-related issues.

Overall Word Cloud

ii) Positive Word Cloud

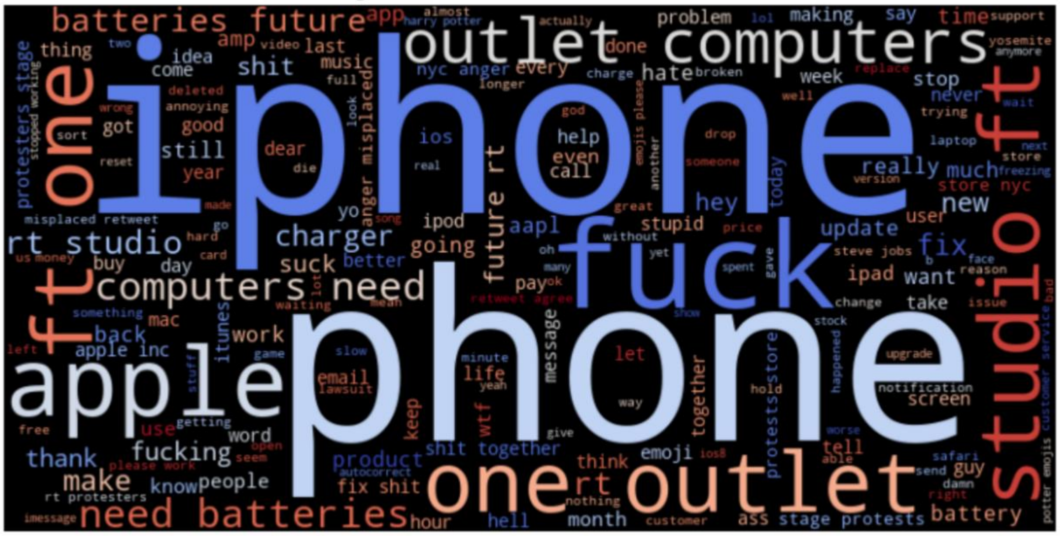
- More positive sentiment with words like thank, new, great, and love. This suggests that many users are expressing appreciation for Apple products or services.

Positive Sentiment Word Cloud

iii) Negative Word Cloud

- More negative sentiment, with words like fuck, suck, and fix. This cloud highlights frustration with Apple, possibly related to product issues or customer service complaints.

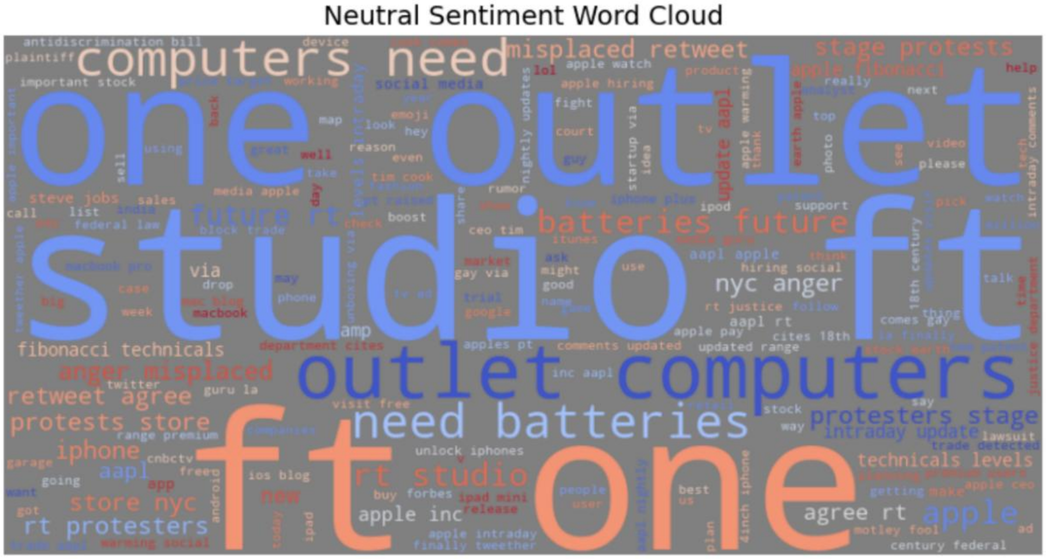
Negative Sentiment Word Cloud



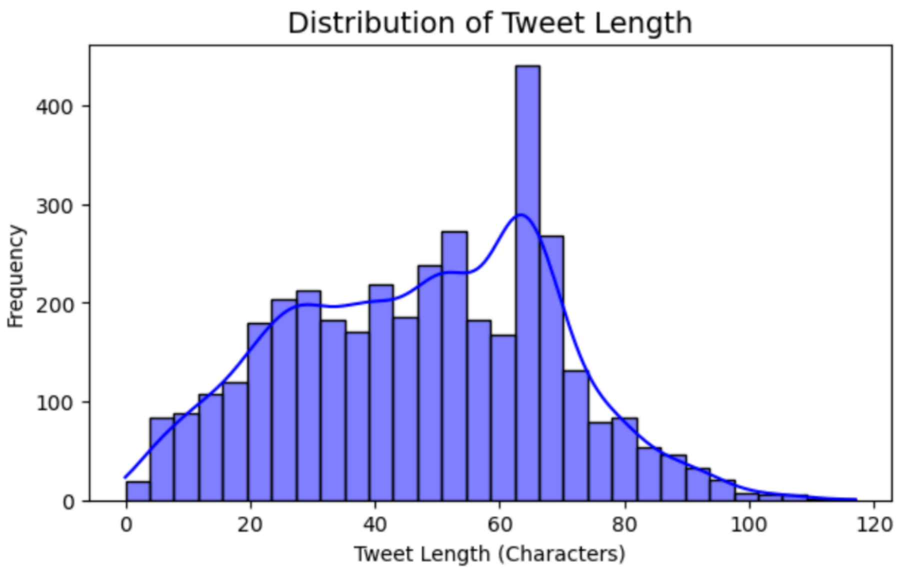
iv) Neutral Word Cloud

- A more neutral cloud focusing on keywords like studio, outlet, computers, and batteries. This indicates general discussions about Apple products without a strong emotional tone.

Neutral Sentiment Word Cloud

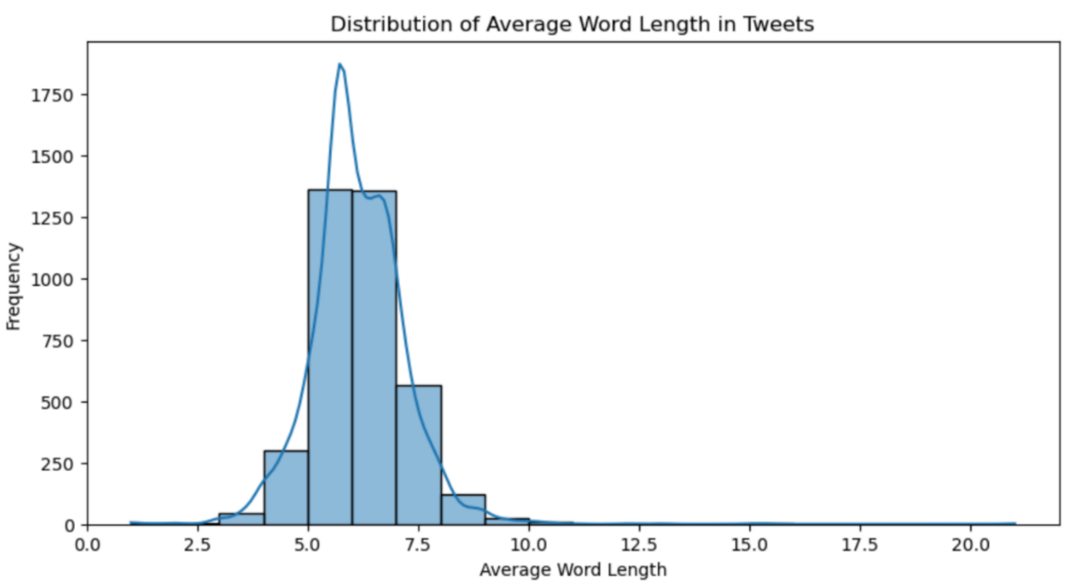


2. Tweet Length Distribution



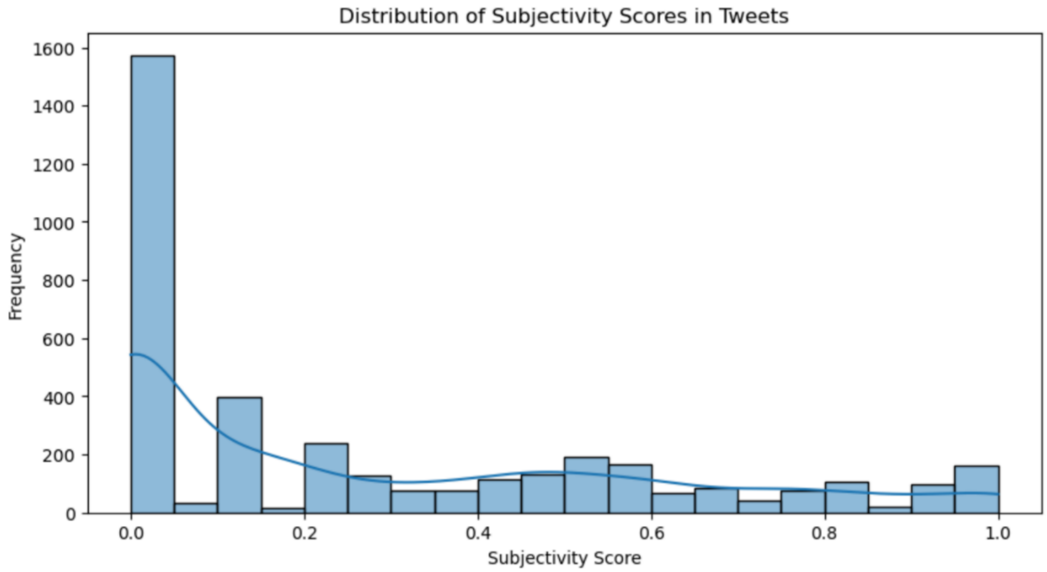
- It ranges between 60- 70 characters

3. Average word Length



- The average word length is mostly around 5-7 characters, indicating that most words in the dataset are relatively short.

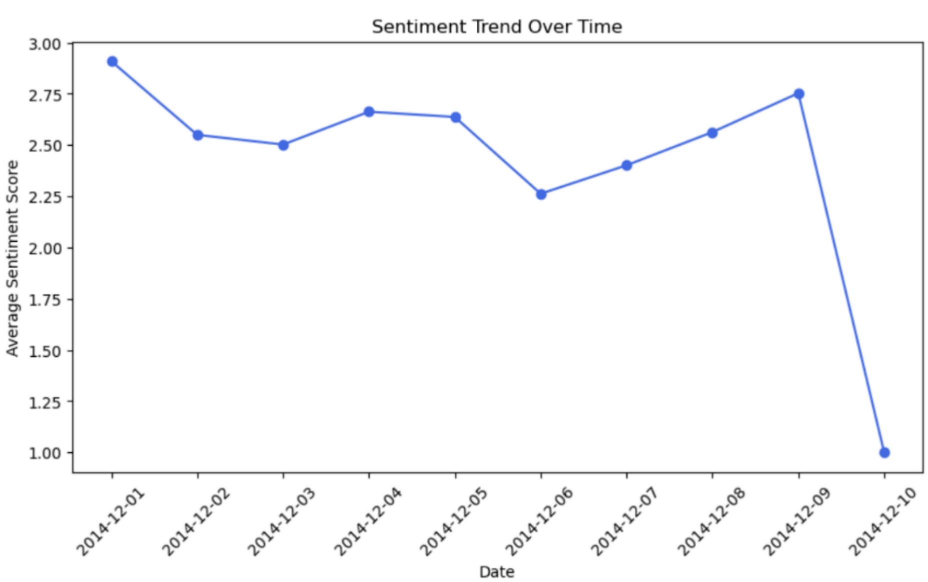
4. Subjectivity Scores



- Most tweets are objective → Subjectivity scores close to 0
- Only a smaller portion are strongly opinionated → Scores near 1
- That suggests many tweets are news, updates, or factual statements rather than personal opinions—useful insight for understanding tone on social media

3.2: Bivariate Analysis

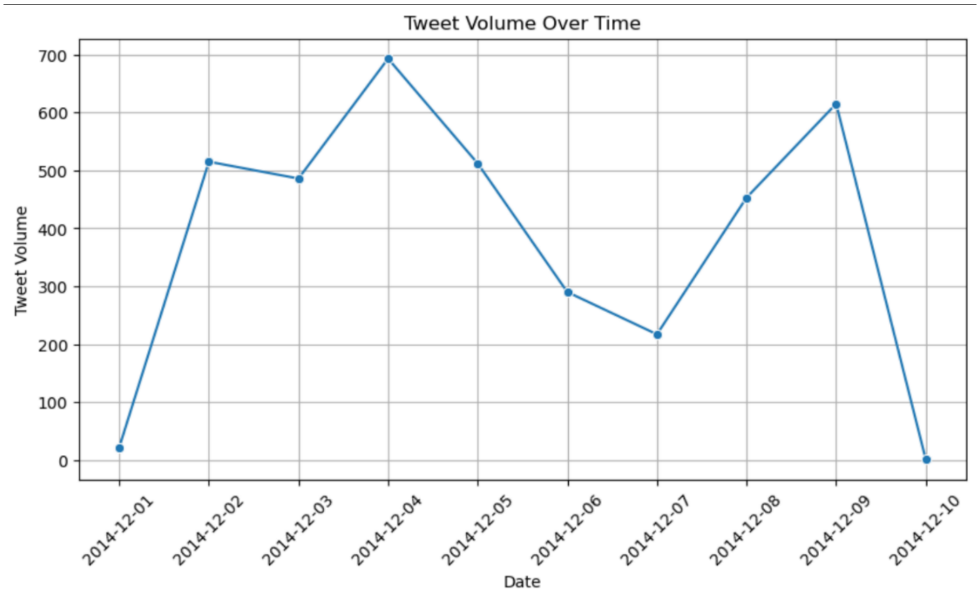
1. Sentiment Scores



Observations:

- The average sentiment score fluctuates over time, indicating variation in user sentiment.
- The sentiment starts high (~3.0) on December 1, 2014, then slightly declines but remains around 2.5 - 2.7 until December 8.
- A sharp drop in sentiment occurs on December 10, 2014, reaching 1.0. This could be due to a significant event or a higher volume of negative tweets on that day.
- The peak on December 8 suggests a temporary increase in positive sentiment before the decline.

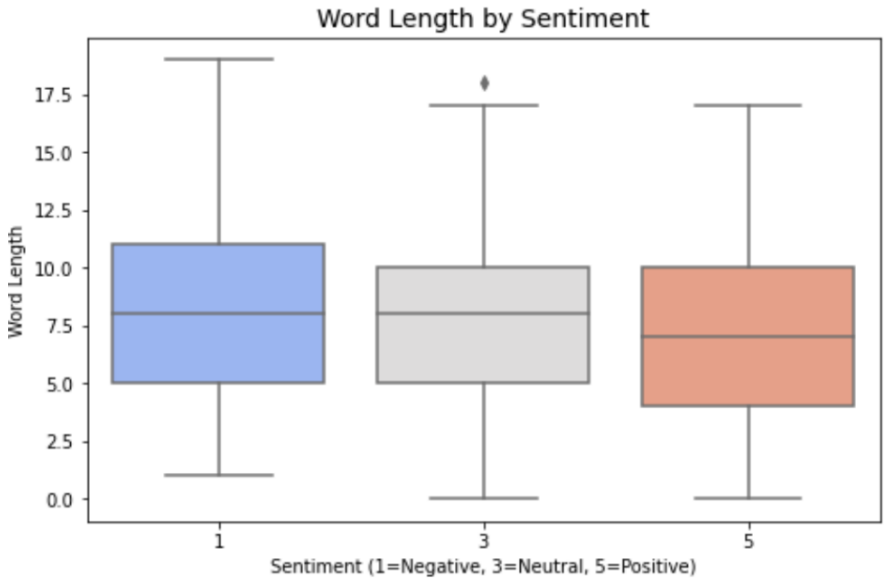
2. Tweet Volume



Observations:

- This confirms that the drastic drop in sentiment on December 10 is likely due to a sharp decrease in tweet volume rather than a genuine sentiment shift.
- This could indicate missing data or a lack of engagement rather than a sentiment anomaly.

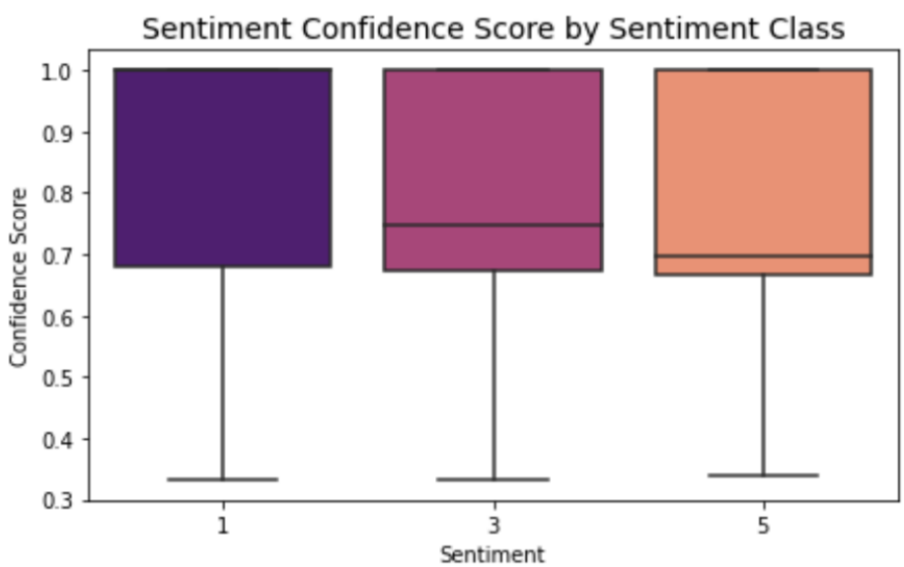
3. Sentiment Vs Word Count



Observations:

- Similar median values across all sentiments, meaning tweet length doesn't vary drastically by sentiment.
- Some outliers, but no extreme differences in distribution.
- Interquartile ranges (IQRs) are quite similar, suggesting tweets in all sentiment categories tend to have comparable word counts.

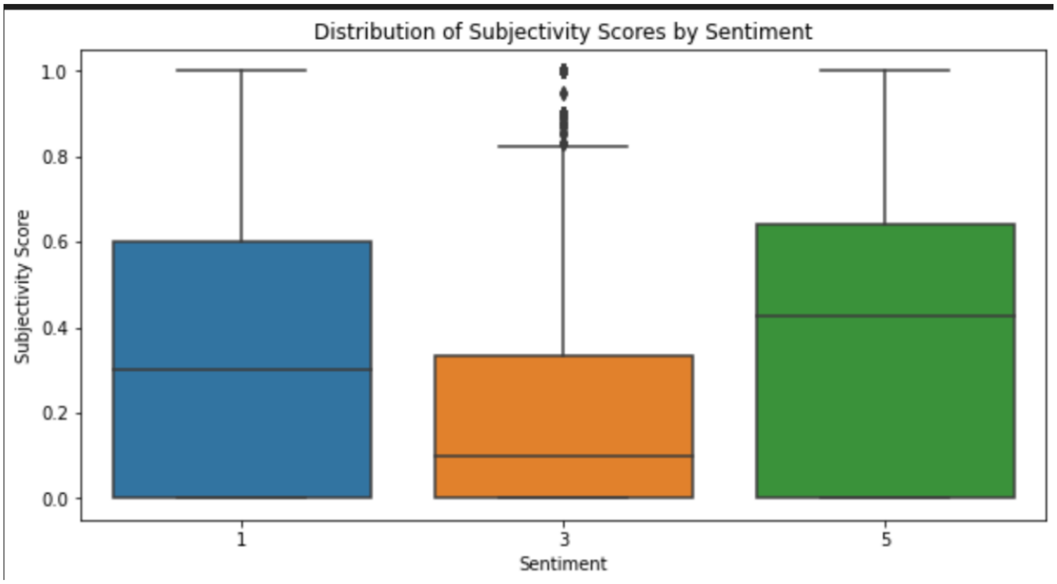
4. Sentiment Vs Sentiment Confidence Score



Observations:

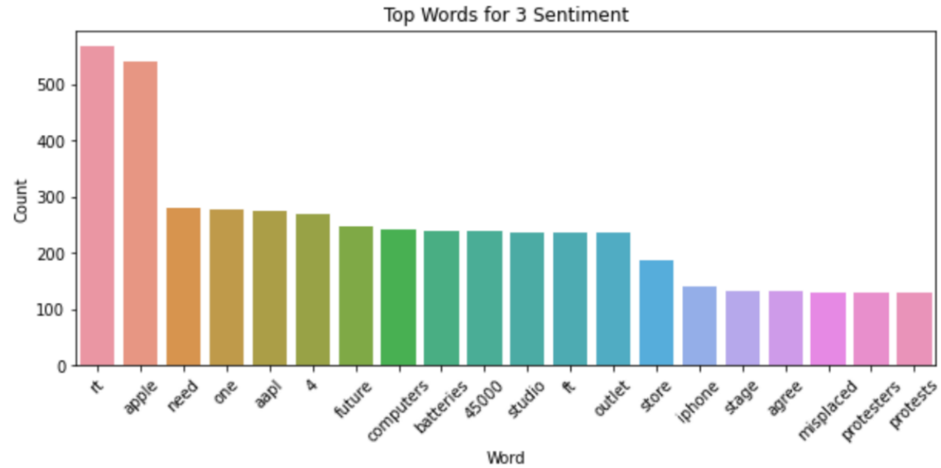
- Confidence is relatively high across all sentiment categories.
- Wide spread in confidence scores.
- No significant differences between sentiment categories.

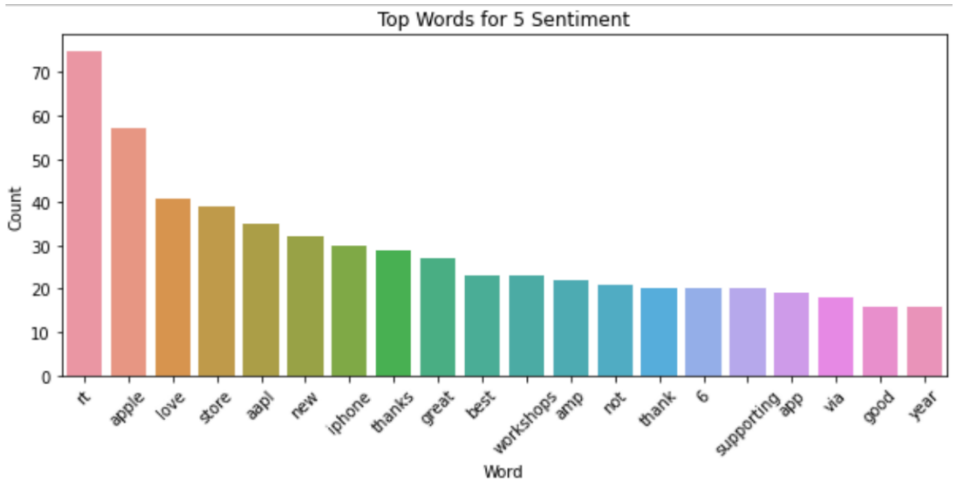
5. Sentiment Vs Subjectivity Scores

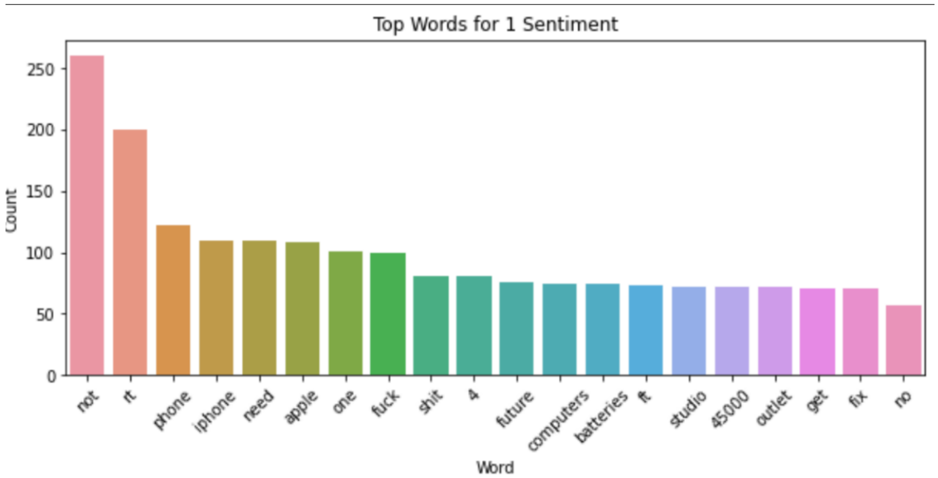


- Negative and positive tweets are often more opinion-based, while neutral tweets are more fact-based.
- This aligns with expectations — neutral tweets tend to state facts, whereas opinions (positive or negative) include emotional language.

6. Most Common Words By Sentiment





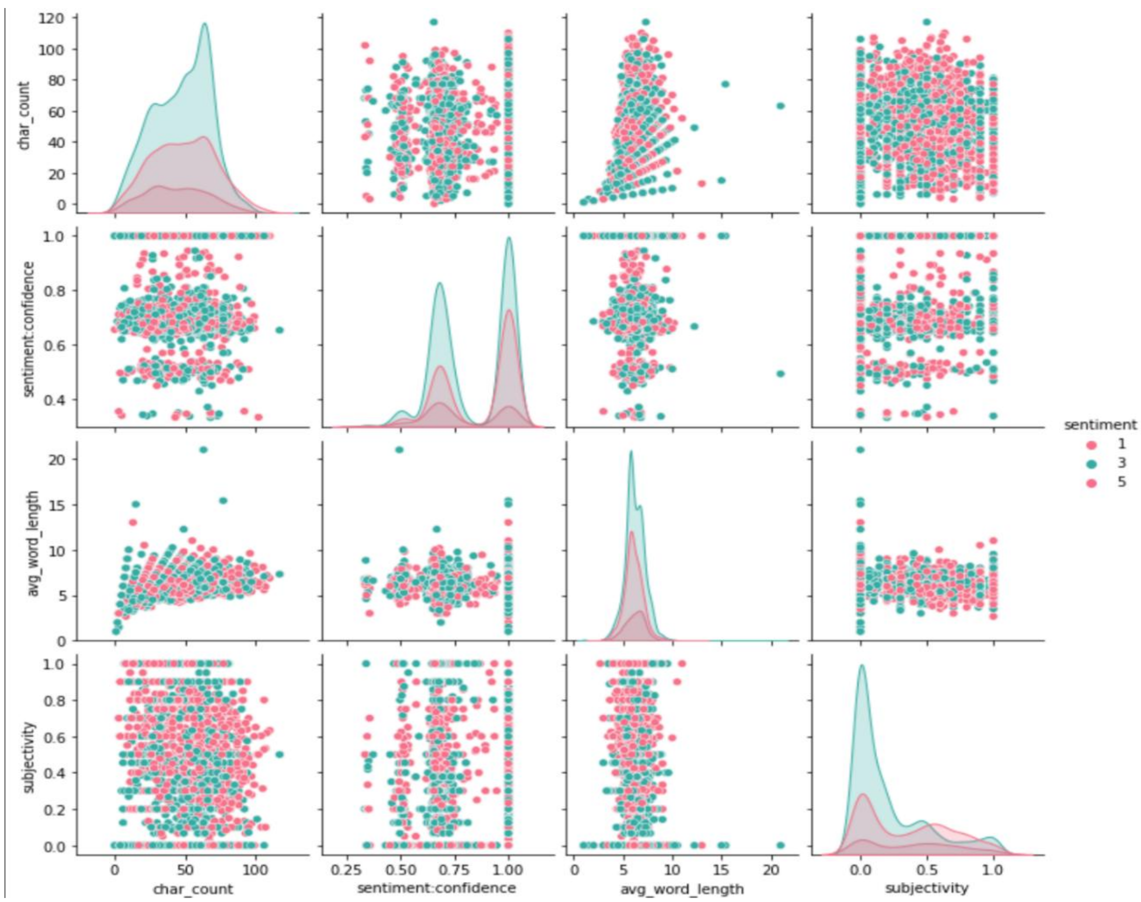


Observation:

- Negative (Score 1): Strong dissatisfaction, often about Apple products. Complaints include technical issues and unmet expectations. Filtering explicit words may help in sentiment analysis.
- Neutral (Score 3): Focused on Apple stock and company updates, mainly from investors or analysts. Less emotional content.
- Positive (Score 5): Praise for Apple products and service. Driven by satisfaction, gratitude, and excitement over new releases

3.3: Multivariate Analysis

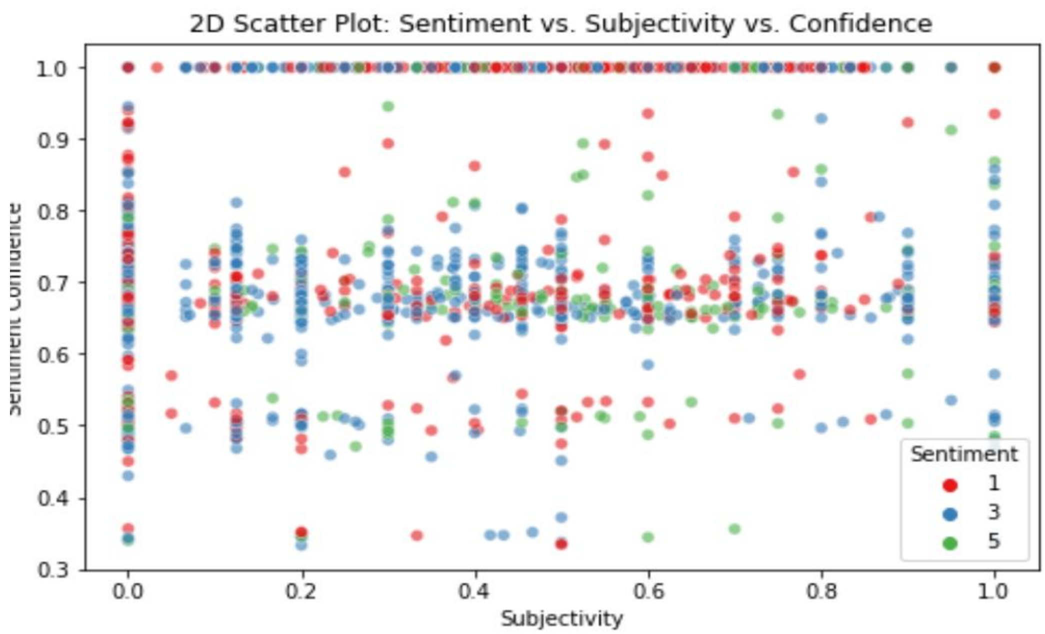
1. Pairplot of Numerical Features



Observations:

- Feature Distributions: Some features (e.g., sentiment confidence, subjectivity) show distinct patterns, but others (e.g., char count, avg word length) have overlapping distributions.
- Feature Relationships: Certain features may help distinguish sentiment classes, but heavy overlap suggests some features may not be strong predictors.
- Class Separation: If sentiment classes form clear clusters, the features are effective. Otherwise, more feature engineering may be needed.

2. Sentiment Vs Subjectivity Vs Condidence



Observations:

- The points appear scattered across the graph, indicating sentiment values are spread across different input features.
- The high density of blue and red points suggests that neutral and negative sentiments are more frequent in certain regions.
- Some sentiment clusters appear along the top and bottom, which might indicate edge cases or outliers.

4.0: Text Preprocessing

4.1: Tokenization

- Breaking text into individual words (tokens).This is the first step in text preprocessing, where we split a sentence or paragraph into smaller units (words or subwords).

4.2: Lemmatization

- Reducing words to their base (dictionary) form.Lemmatization ensures words like *running*, *ran*, and *runs* are all converted to *run*, making text easier to analyze.

4.3: TF-IDF Vectorization

- TF-IDF (Term Frequency-Inverse Document Frequency) was used in the initial modeling phase to convert text data into numerical representations. This technique assigns higher importance to words that appear frequently in a document but are less common across the entire dataset, helping distinguish key terms.
- TF-IDF was effective for traditional machine learning models but lacked the ability to capture semantic relationships between words.

4.4: Word Embeddings (Word2Vec)

- For deep learning models, Word2Vec embeddings were utilized to represent words in a continuous vector space, capturing semantic relationships between similar words. Unlike TF-IDF, Word2Vec preserves contextual meaning, making it well-suited for deep learning approaches like LSTMs and CNNs.
- This method improved sentiment classification by allowing models to understand word associations and contextual nuances within Apple-related tweets.

5.0: Modelling

To prepare our data for modelling we will handled:

- Scaled numeric columns TF-IDF to convert texts into numerical data , and used SMOTE to address class imbalance in the target variable.

5.1: Best performing model

Stacked Models with class weights

A stacked model (stacking ensemble) is a technique where multiple models work together to improve accuracy. Instead of using just one classifier, we combine different models and use a meta-model to make the final prediction.

How Does Stacking Work?

1. **Base Models (Level 1 Models)**
   - Train multiple models (**Logistic Regression** and **Random Forest**) on the same dataset.
   - Use class weights to handle class imbalance.
2. **Meta-Model (Level 2 Model)**
   - The predictions from the base models are used as **input features** for a final model.
   - A simple model (often **Logistic Regression**) learns to combine their outputs optimally.

- **Model Performance**
  Accuracy:75%
  Precision:74%
  Recall:89%
  F1-Score: 81%

5.2: Other Models:

1. **XG Boost**

XGBoost, an efficient and scalable gradient boosting framework, was employed to predict match outcomes. This model, renowned for its speed and performance, iteratively builds multiple decision trees, allowing it to capture complex relationships in the data.

- **Model Performance**
  Accuracy:72%
  Precision:75%
  Recall:87%
  F1-Score: 80%

2. **LSTM**

LSTM (Long Short-Term Memory) is a type of Recurrent Neural Network (RNN) designed to handle sequential data (like text, time series, and speech). It was introduced to solve the vanishing gradient problem that traditional RNNs face when learning long-term dependencies.

- **Model Performance**
  Training Accuracy: 61%
  Validation Accuracy: 63%
  Training Loss: 0.86
  Validation Loss: 0.86

2.0: Bidirectional LSTM

A Bidirectional LSTM (BiLSTM) is an extension of the standard LSTM (Long Short-Term Memory) network that processes data in both forward and backward directions.

- **Model Performance**
  Training Accuracy: 67%
  Validation Accuracy: 64%
  Training Loss: 0.76
  Validation Loss: 0.86

3. **CNN**

A Convolutional Neural Network (CNN) is a deep learning model originally designed for image processing, but it can also be applied to text data (NLP).

- **Model Performance**
  Training Accuracy: 63%
  Validation Accuracy: 62%
  Training Loss: 0.88
  Validation Loss: 0.92

Recommendations

1. **Address negative sentiment themes** by analyzing concerns and making targeted improvements to enhance brand perception.
2. **Leverage positive sentiment in marketing** by engaging satisfied customers and amplifying feedback to boost loyalty.
3. **Engage with neutral sentiment tweets** to convert passive opinions into positive experiences through personalized support.
4. **Optimize marketing by tracking peak discussion times** and aligning promotions with high-engagement periods.
5. **Monitor sentiment trends at a product level** to quickly identify and resolve issues, improving satisfaction.
6. **Use the Stacked Model with Class Weights** for real-time sentiment classification and accurate tracking.
7. **Improve sentiment detection accuracy** by integrating metadata, refining preprocessing, and fine-tuning models.
8. **Benchmark sentiment against competitors** to compare Apple's brand perception and find differentiation areas.
9. **Strengthen brand advocacy via influencers and communities** by fostering discussions and partnerships.

Conclusion

Sentiment analysis of Apple-related tweets provides real-time insights to enhance decision-making. The Stacked Model with Class Weights offers a reliable and scalable solution for sentiment classification, helping Apple track customer sentiment effectively. By leveraging these insights, Apple can improve brand perception, refine marketing strategies, and enhance customer experience. Addressing negative sentiment, amplifying positive engagement, and optimizing responses to neutral sentiment will strengthen customer loyalty. Further improvements, such as expanding the dataset and integrating external sentiment trends, can enhance accuracy and business impact.