APPLE AND GOOGLE TWEET SENTIMENT ANALYSIS

OVERVIEW

***Problem**: To build a model that can rate the sentiment of a Tweet based on its content.*

This project analyzes **tweets** about **Apple** and **Google** products to understand how people feel about them; **positive**, **negative**, or **neutral**. Using machine learning, the system automatically identifies emotions in thousands of tweets. This analysis helps to track public opinion, improve products, and make better decisions based on customer feedback.

BUSINESS UNDERSTANDING

Social media platforms like Twitter provide valuable **insights** into public opinions about brands such as Apple and Google. Analyzing tweets through automated sentiment analysis helps companies understand customer feelings, monitor reputation, and make informed decisions more efficiently than manual review on their products.

DATA UNDERSTANDING

Source: CrowdFlower via data.world

shape: The dataset has 9093 rows and 3 columns.

Target:

is_there_an_emotion_directed_at_a_brand_or_product later renamed sentiment(positive, negative, neutral).

Dependent variables: tweet_text, emotion_in_tweet_is_directed_at

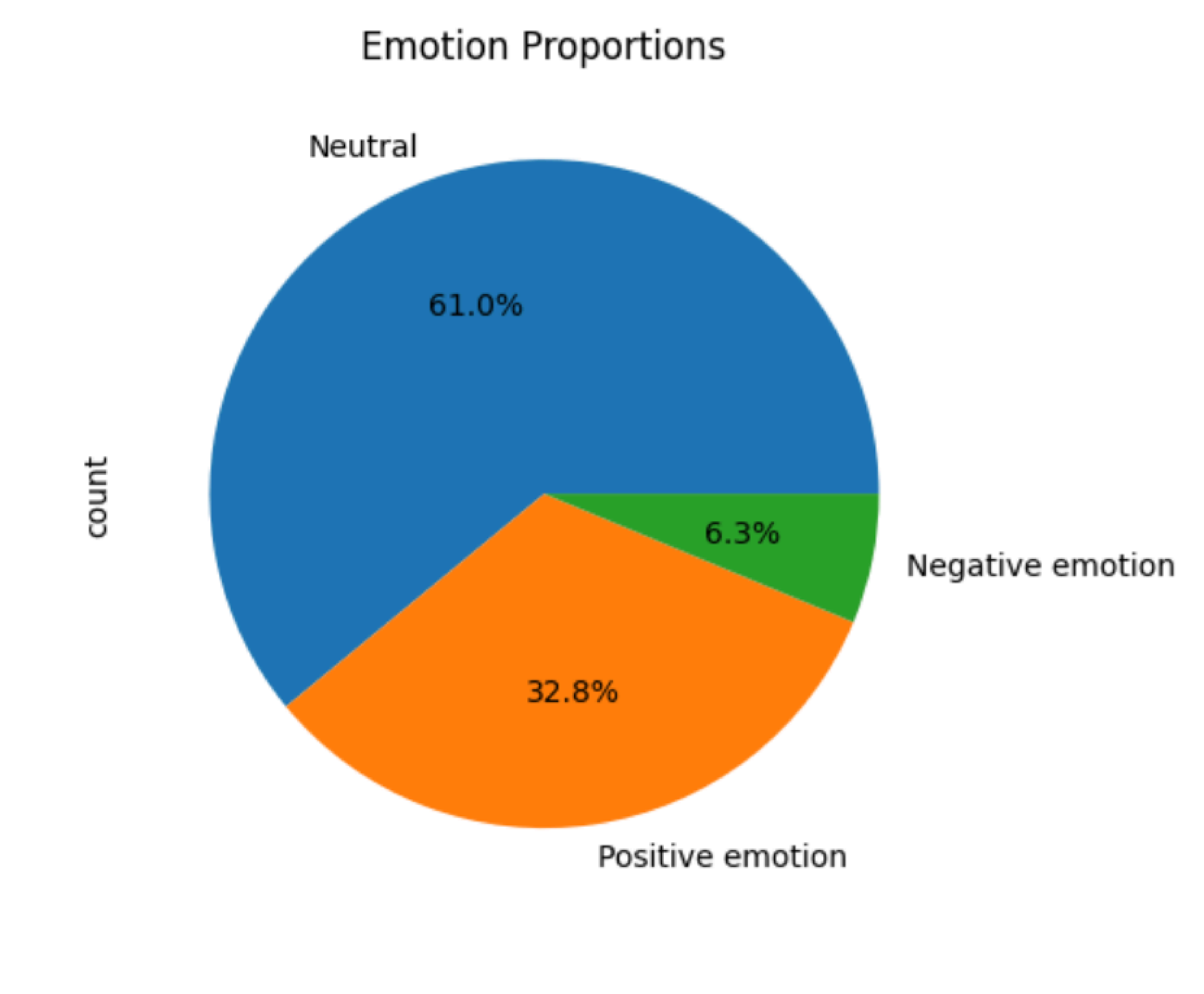
- The dataset contained 22 **duplicates** which were dropped.
- **tweet_text** had one missing value which was dropped.
- emotion_in_tweet_is_directed_at was imputed with **unknown** since we would like to see how the brands affect users.
- Before analysis, the text was cleaned to remove noise such as links, punctuation, and special symbols, leaving only meaningful words.
- After cleaning, the most common meaningful words were brand names such as **Google**, **Apple**, **iPhone**, **Android**, and **iPad**, giving a clear view of which products dominate public discussion.

DATA ANALYSIS

Target Analysis

The analysis revealed that positive emotions dominate across all brands, especially for Android products, showing high user satisfaction. iPhone received the most negative sentiment (34%), indicating some user dissatisfaction, while neutral tweets mostly conveyed factual information.

Sentiments towards apple and Google are distributed as shown in the pie chart.



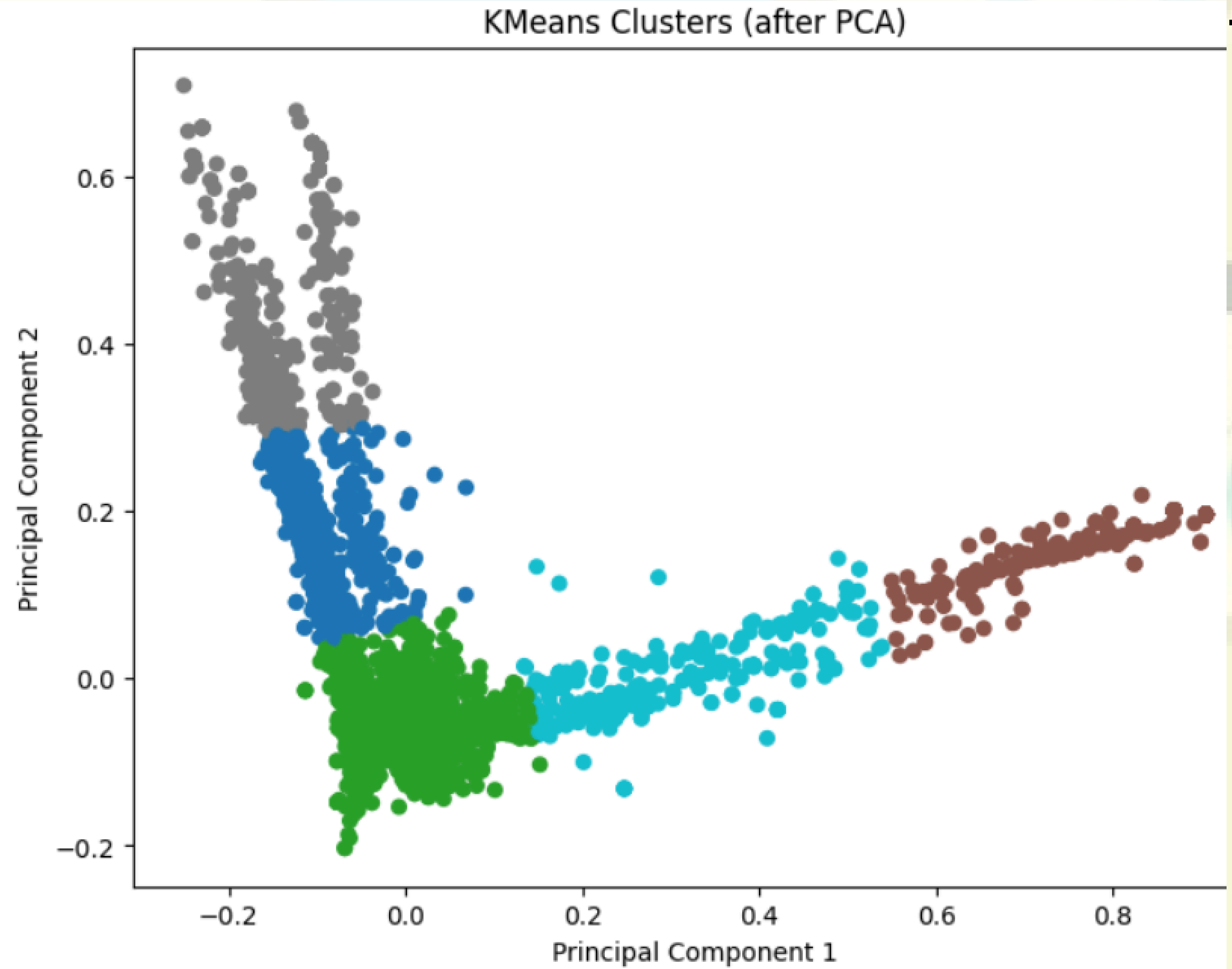
This visualization illustrates the emotional landscape, highlighting the dominance of neutral sentiment while clearly distinguishing the smaller proportions of positive and negative emotions

Positive

Negative

Word clouds showed that positive tweets used words like beautifully and simple, negative ones included crashing and dead, and neutral tweets mentioned general terms like app and Android.

Grouping Similar Tweets Based on Emotion



This chart groups tweets with similar language or tone into five color-coded clusters. Each color represents a different theme or emotion. For example, positive, negative, or neutral discussions, showing how people express different opinions about brands.

Model Performance

- The Logistic Regression model achieved an accuracy of 64% and a weighted F1 score of 0.65, showing moderate but useful predictive power for a 3-class sentiment problem.
- The model performed best on Neutral tweets but was less confident distinguishing between Positive and Negative.
- A train-test accuracy gap of ~18% indicated mild overfitting, suggesting the model could still improve with better preprocessing and balance handling.

RECOMMENDATIONS

1. Enhance Data Balance and Contextual Understanding
   - Collect more labeled data for Positive and Negative tweets to help the model learn minority sentiment patterns better.
   - Incorporate techniques like lemmatization, bigrams, and deep contextual embeddings to improve understanding of tone and nuance.
2. Use Sentiment Insights for Strategic Decisions
   - Map positive and negative sentiment trends to specific features, services, or campaigns.
   - Identify what drives loyalty (positive clusters) and where dissatisfaction arises (negative clusters).
3. Continuously Monitor and Update the Model
   - Re-train periodically to adapt to evolving language (slang, emojis, abbreviations).
   - Maintain model performance to ensure reliable, real-time sentiment tracking.

NEXT STEPS

1. Collect more balanced data across sentiment categories for better model generalization.
2. Develop a Tableau or Power BI dashboard to visualize sentiment by time, brand, and region.
3. Experiment with advanced models such as CNN-LSTM or transformer-based embeddings (e.g., BERT) to capture deeper text meaning.
4. Integrate model outputs into business workflows, allowing product or marketing teams to respond quickly to customer sentiment changes.

THANK YOU

This project demonstrates how analyzing tweets can transform scattered online opinions into valuable, actionable insights — empowering businesses to understand customer emotions, enhance their products, and strengthen their brand reputation. Thank you.