



MORINGA
SCHOOL

Apple_Google sentiment Analysis

DATA SCIENCE: NATURAL
LANGUAGE PROCESSING

Presented by:

- ◆ 1. *Immaculate Kithei Kyalo*
- ◆ 2. *Kemboi Henry*
- ◆ 3. *Wanjiru Damaris*
- ◆ 4. *Muuo Dennis*
- ◆ 5. *Njuguna Jane*



BUSINESS PROBLEM

The business's main issues in sentiment analysis include

1. The tweets are full of slang making machine learning challenging
2. Abbreviations are hard for the stakeholders to interpret thus challenging when making business decisions and insights.
3. Sarcasm makes it challenging to classify and interpret sentiments
4. Spam and irrelevant tweets affect analysis with much noise.

BENEFITS TO STAKEHOLDERS

1. From this project tech companies will generate key insights into public perception of their products
2. The organizational product teams will identify major areas of product improvement
3. Marketers will leverage marketing on the trends to develop targeted marketing campaigns.

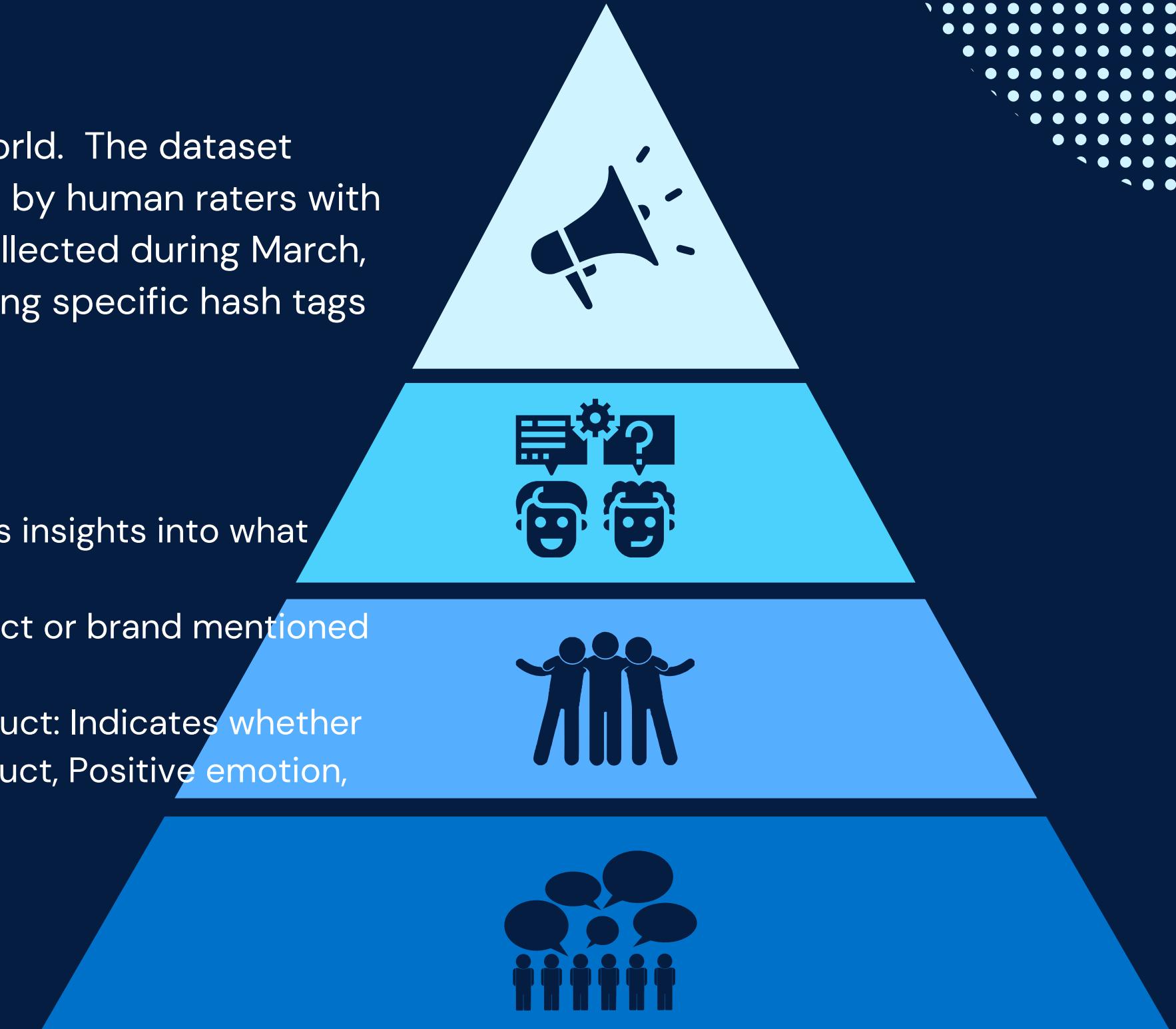


DATA

The research project used data from Crowd Flower via Data.World. The dataset contains 9093 Tweets that have been meticulously annotated by human raters with sentiment labels: positive, negative, or neutral. The data was collected during March, around the time of SXSW event and tweets were filtered by using specific hash tags or keywords.

The dataset has 3 columns:

1. tweet_text: The actual text of the tweet, which provides insights into what users are saying about Apple and Google products.
2. emotion_in_tweet_is_directed_at: The specific product or brand mentioned in the tweet (e.g., iPhone, iPad, Google).
3. is_there_an_emotion_directed_at_a_brand_or_product: Indicates whether the tweet expresses: No emotion toward brand or product, Positive emotion, Negative emotion, I can't tell.



DATA ANALYSIS

UNIVARIATE ANALYSIS

MOST FREQUENT WORDS

The word cloud graph indicates the most frequent words.

In the word cloud graph, the most repeated words include

- SXSW: This word represents the time when the event happens and thus has the potential of being overused
- Google: Most emotions are based between two brands which were Google and iPhone and appear most times
- iPhone: The emotions focused on iPhone too as another frequent brand

Data Analysis

UNIVARIATE ANALYSIS

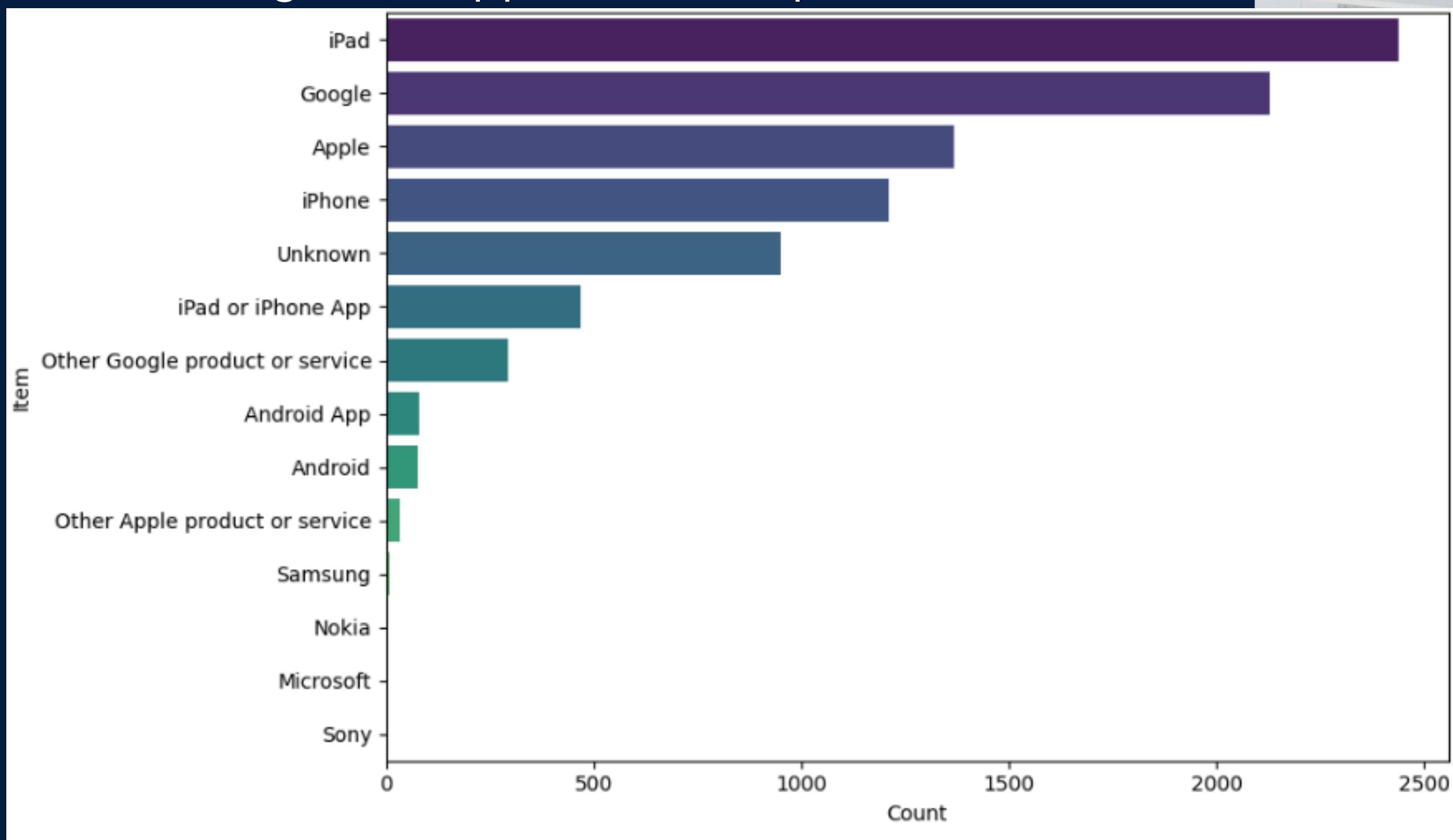
Most Frequent Words



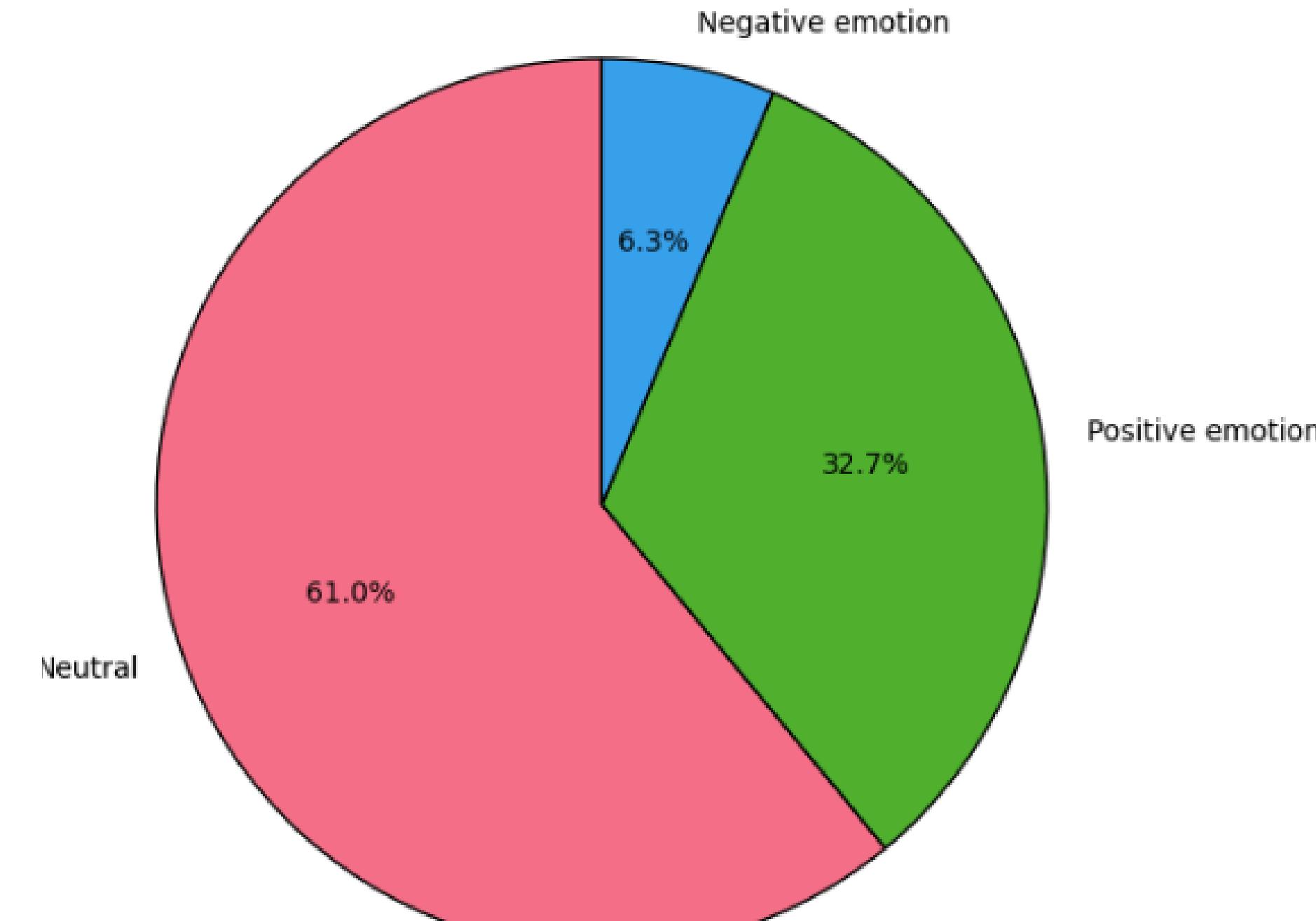
DATA ANALYSIS: MOST OCCURRING ITEMS

1. Which is the most common product?:

iPad, Google and Apple are the top most items



Distribution of Sentiments in Tweets

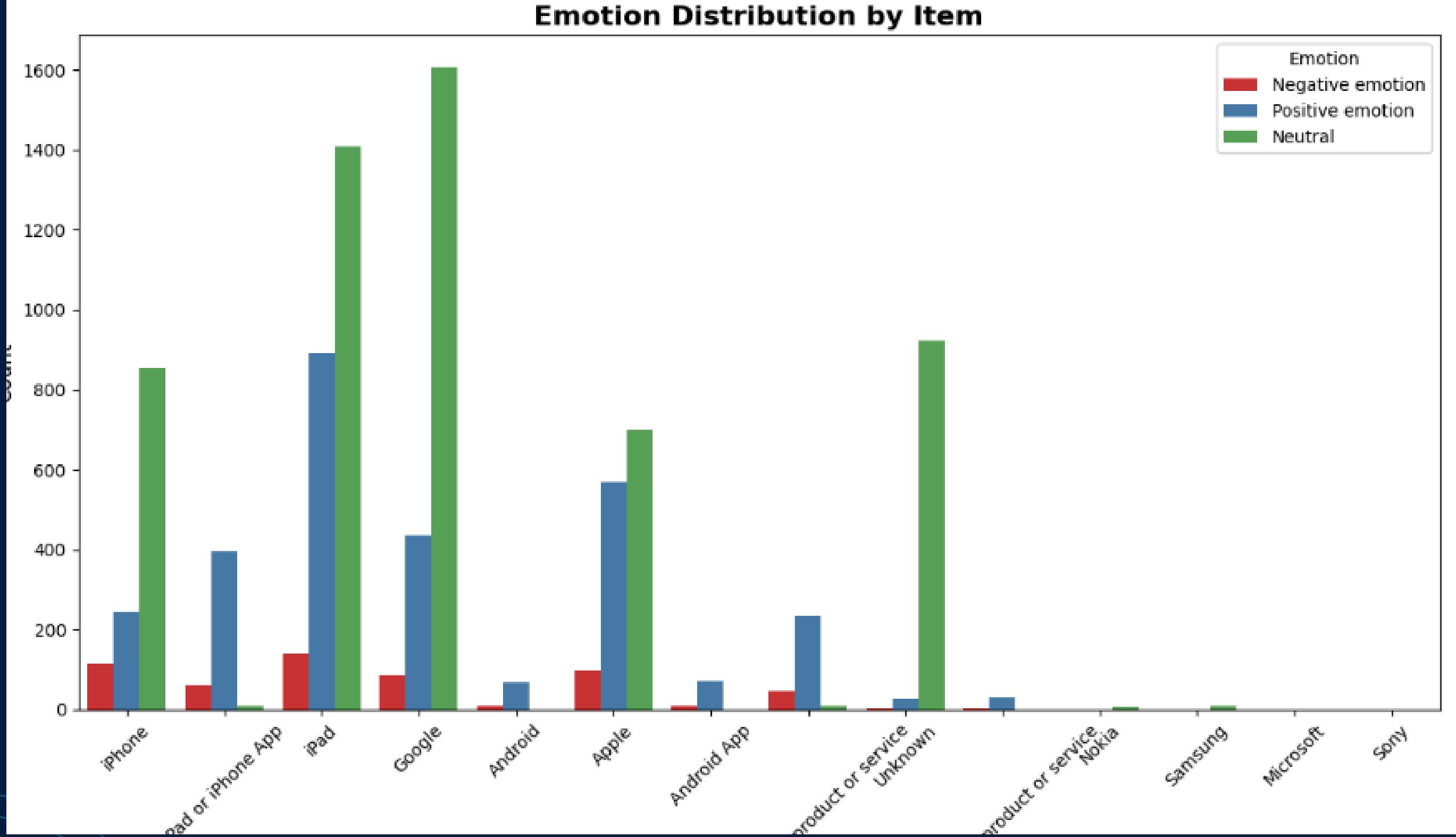


The visualization reveals that the majority of responses exhibit no emotion toward the brand or product(Neutral) followed by a significant proportion of positive reviews, with negative reviews being the least frequent. This suggests overall neutrality with a leaning toward positivity in customer sentiment.

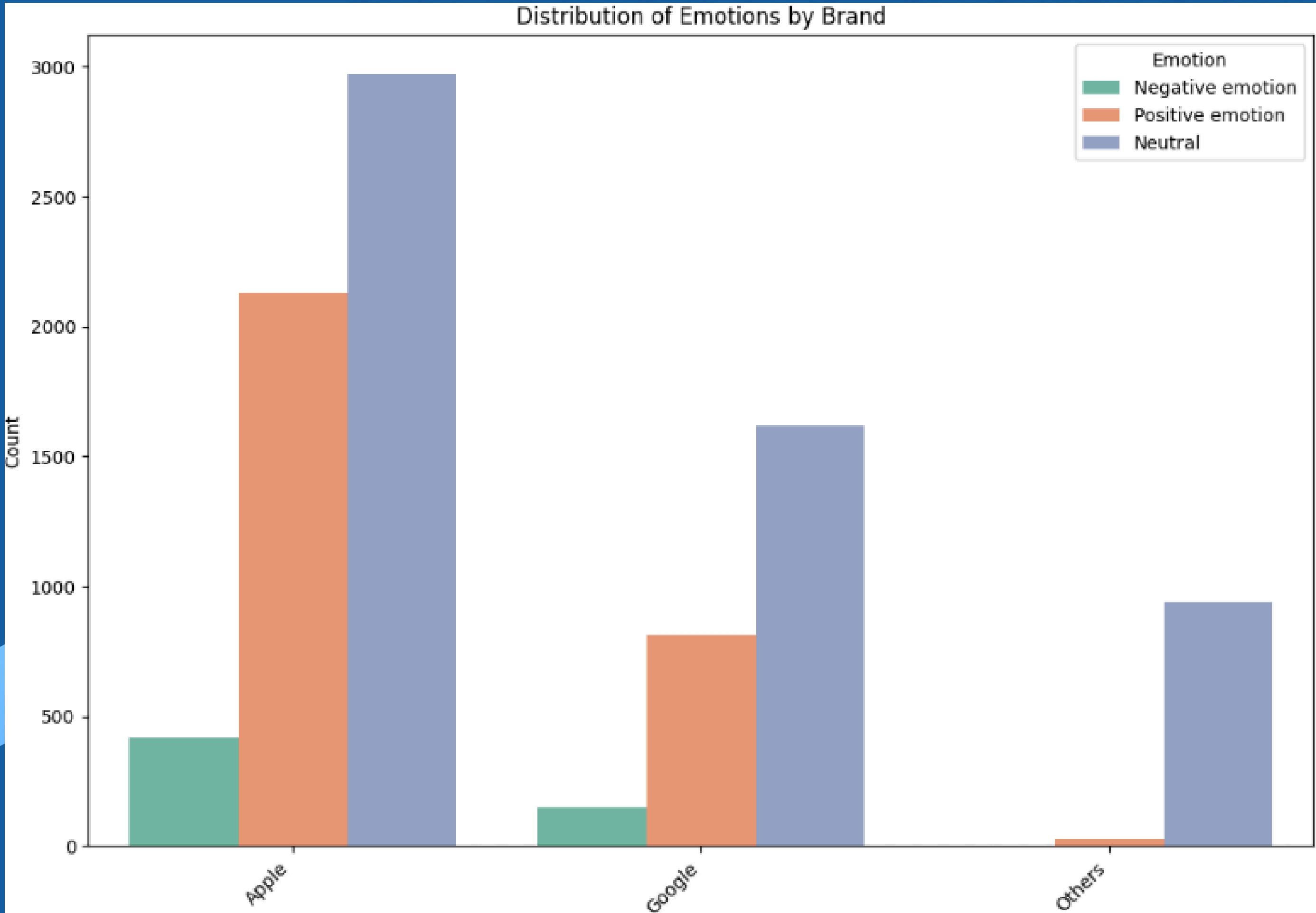
BIVARIATE ANALYSIS

Exploring and analyzing the relationship between two variables.

Emotion Distribution by Item

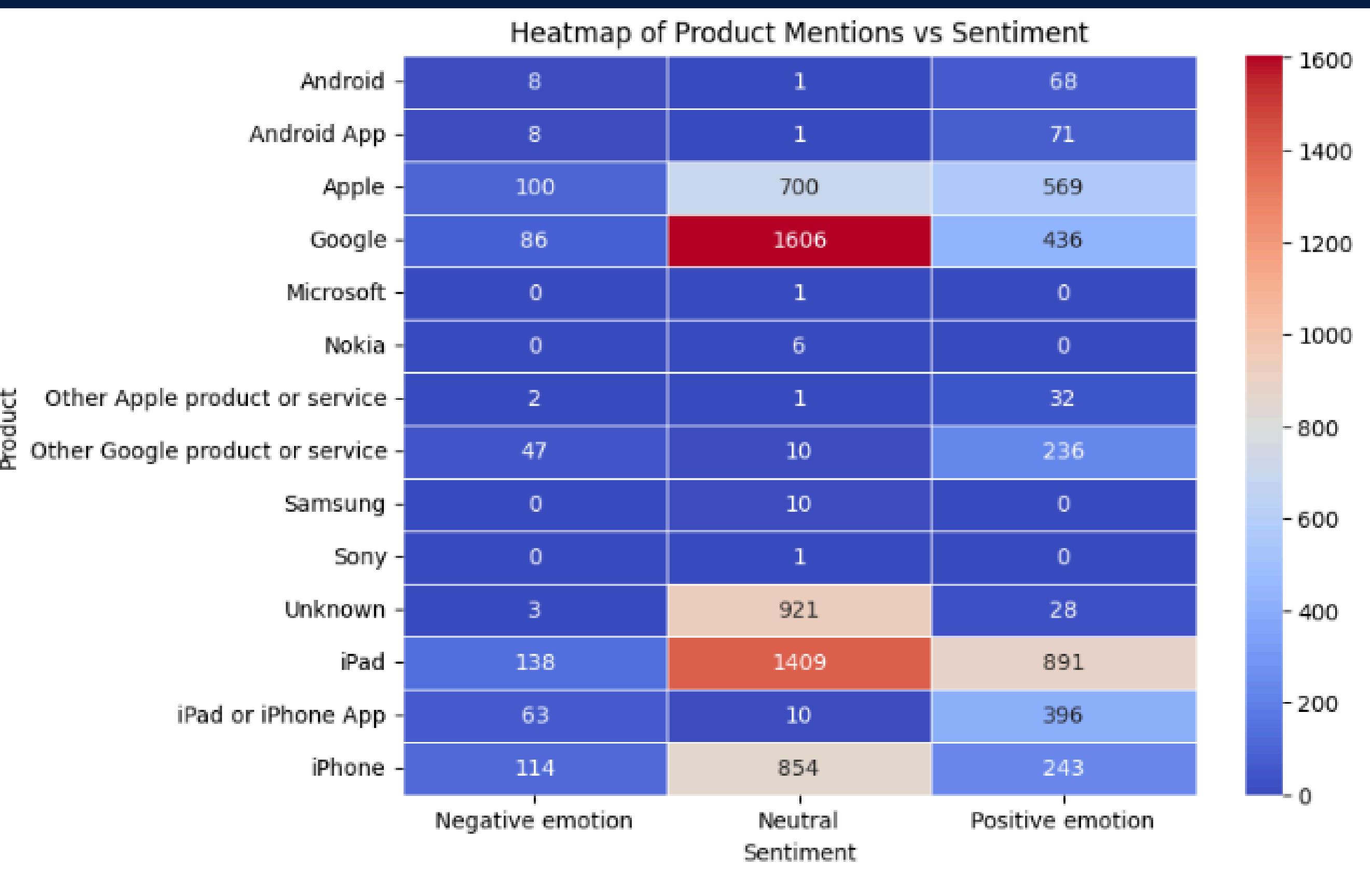


DISTRIBUTION OF EMOTION BY BRAND

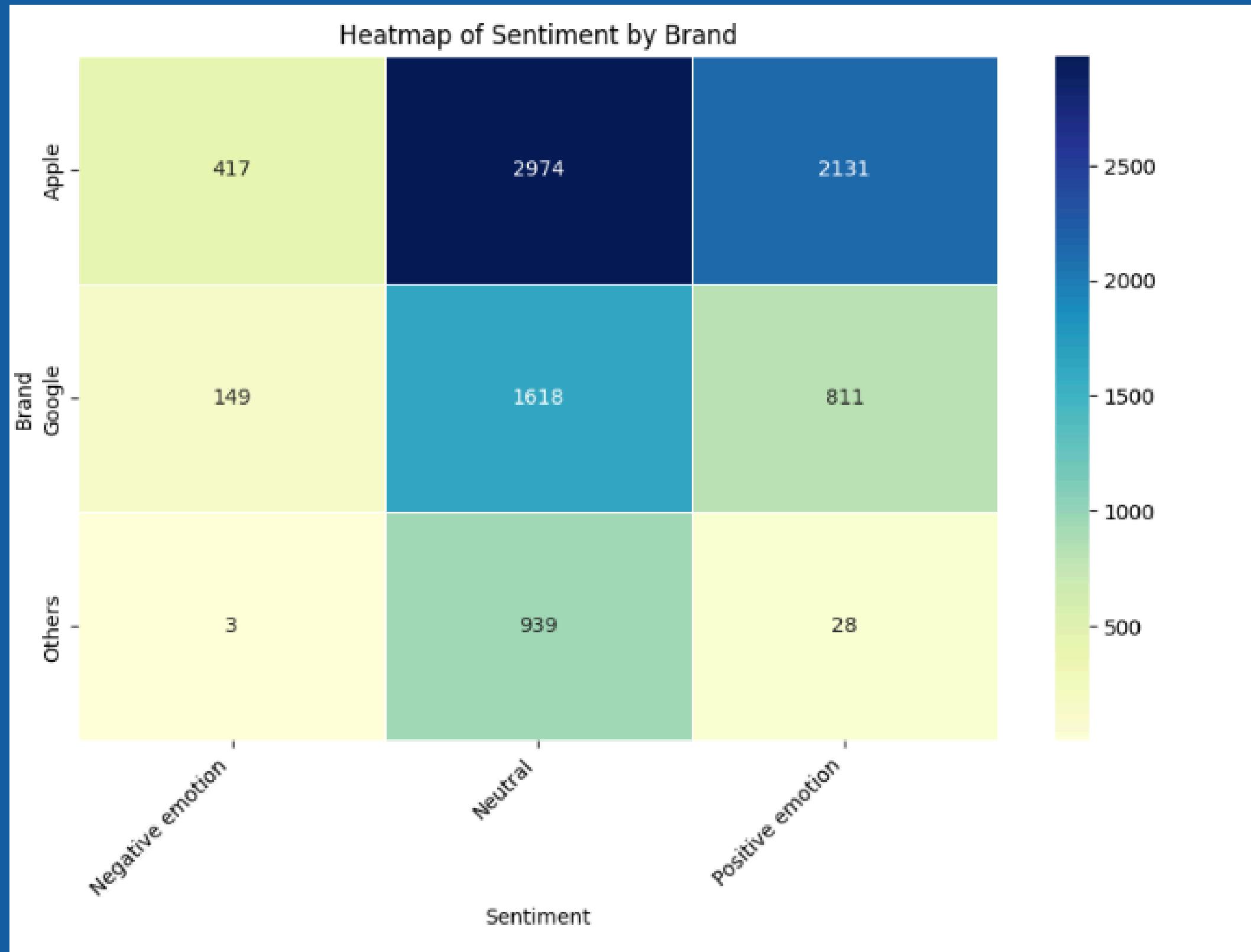


Apple products seem to evoke more neutral or positive feelings, while Google products generate a wider mix of emotions. The rest of the brands, however, don't seem to stir as much emotion overall.

CORRELATION OF PRODUCT MENTIONS AND SENTIMENTS



CORRELATION OF SENTIMENT AND BRAND



- **Positive Sentiment:** Apple has the highest number of positive mentions, reflecting strong customer satisfaction and loyalty.
- **Negative Sentiment:** While Apple also has some negative mentions, they are proportionally lower than its positive sentiment.
- **Neutral Sentiment:** A significant number of neutral mentions indicate widespread discussions about Apple products, not always tied to opinions.
Google:
 - **Positive Sentiment:** Google shows a balanced distribution of positive sentiment, indicating a generally favorable perception.
 - **Negative Sentiment:** Google has notable negative mentions, suggesting areas for product or service improvement.
 - **Neutral Sentiment:** The high neutral sentiment count implies many users discuss Google factually, possibly regarding features or updates.

MODELLING

Exploring multiple machine learning models for text classification, including:

- Naive Bayes – A strong baseline model for text classification.
- Logistic Regression – A popular linear model for binary and multi-class classification.
- Support Vector Machine (SVM) – Effective for high-dimensional data.
- Random Forest – An ensemble learning method that enhances accuracy.

BEST MODEL RESULTS: SUPPORT VECTOR MACHINE MODEL

SVM Accuracy: 0.93

Classification Report for SVM:

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

Negative	0.90	0.61	0.72	170
Neutral	0.89	0.98	0.94	770
Positive	0.96	0.94	0.95	874

accuracy		0.93	1814	
macro avg	0.92	0.84	0.87	1814
weighted avg	0.93	0.93	0.92	1814

NEXT STEPS

- Continuously track key metrics (e.g., accuracy, recall for Negative class) post-deployment to ensure consistent performance. Set up alerts to detect significant drifts or anomalies in predictions.

Recommendations

- Google's Neutral Sentiment: A significant proportion of neutral sentiment toward Google products suggests room to build excitement through more engaging marketing strategies.
- Recurring Issues for Google: Neutral or negative sentiment tweets for Google often mention usability issues. Develop targeted support initiatives or FAQs addressing these concerns to preemptively mitigate dissatisfaction.
- Apple's Opportunities: While Apple has strong positive sentiment, monitor any emerging concerns to maintain its leading market position.

Thank You Note



Data Scientist

Moringa School

Connect with us.

