



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*



OVERVIEW

- Sentiment analysis is an NLP technique that helps determine the emotional tone of text, categorizing it as positive, negative, or neutral.
- It's commonly used to analyze social media posts and product reviews, providing businesses with insights into customer feelings.
- Twitter is a key platform where people share their opinions on products and services.
- By analyzing tweets, companies like Google and Apple can track public sentiment in real time, using this data to improve their products and boost customer satisfaction .



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.

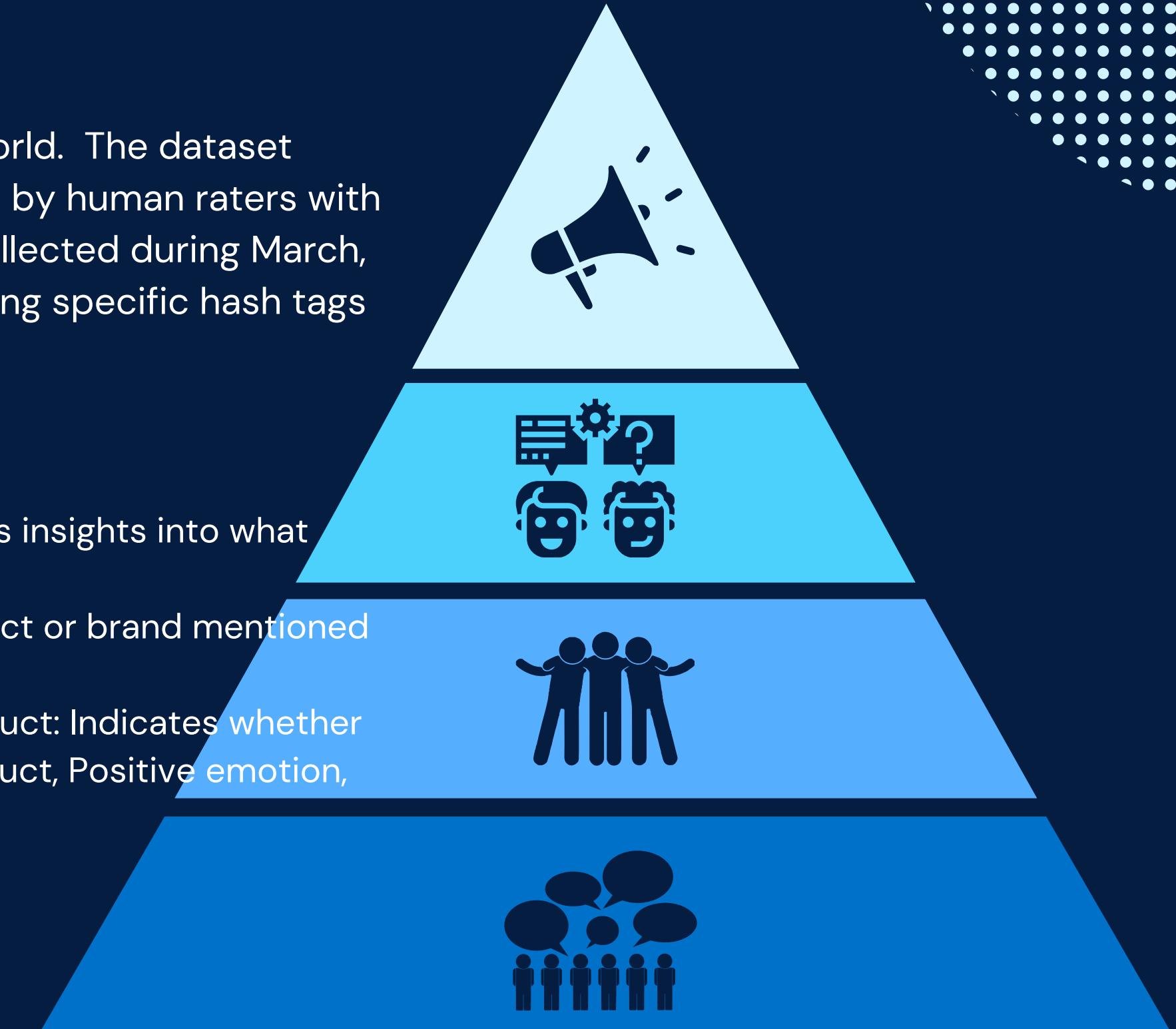


DATA

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

DATA PREPARATION

To prepare a dataset for NLP, efficient cleaning process should be conducted. The data cleaning involved

1. Column Renaming
 2. Hand;ing null or missing values
 - 3.Category Merging
 - 4.Removing duplicate records
 - 5.Text Preprocessing.
-

A photograph showing a man with dark hair and glasses, wearing a dark long-sleeved shirt and jeans, sitting in a white chair and looking at a green tablet. A woman in a white t-shirt and grey pants is seated next to him. They are in a room with grey tiled walls.

DATA: TEXT PREPROCESSING

Text Preprocessing involved the following for the success of NLP:

1. Converted all text to lowercase for consistency.
2. Removed special characters, numbers, and non-alphabetic content.
3. Tokenized text into individual words.
4. Removed stopwords using the NLTK library.
5. Applied lemmatization to reduce words to their root forms.
6. Joined the cleaned tokens back into cohesive strings.

DATA ANALYSIS

UNIVARIATE ANALYSIS

MOST FREQUENT WORDS

The word cloud graph indicates the most frequent words.

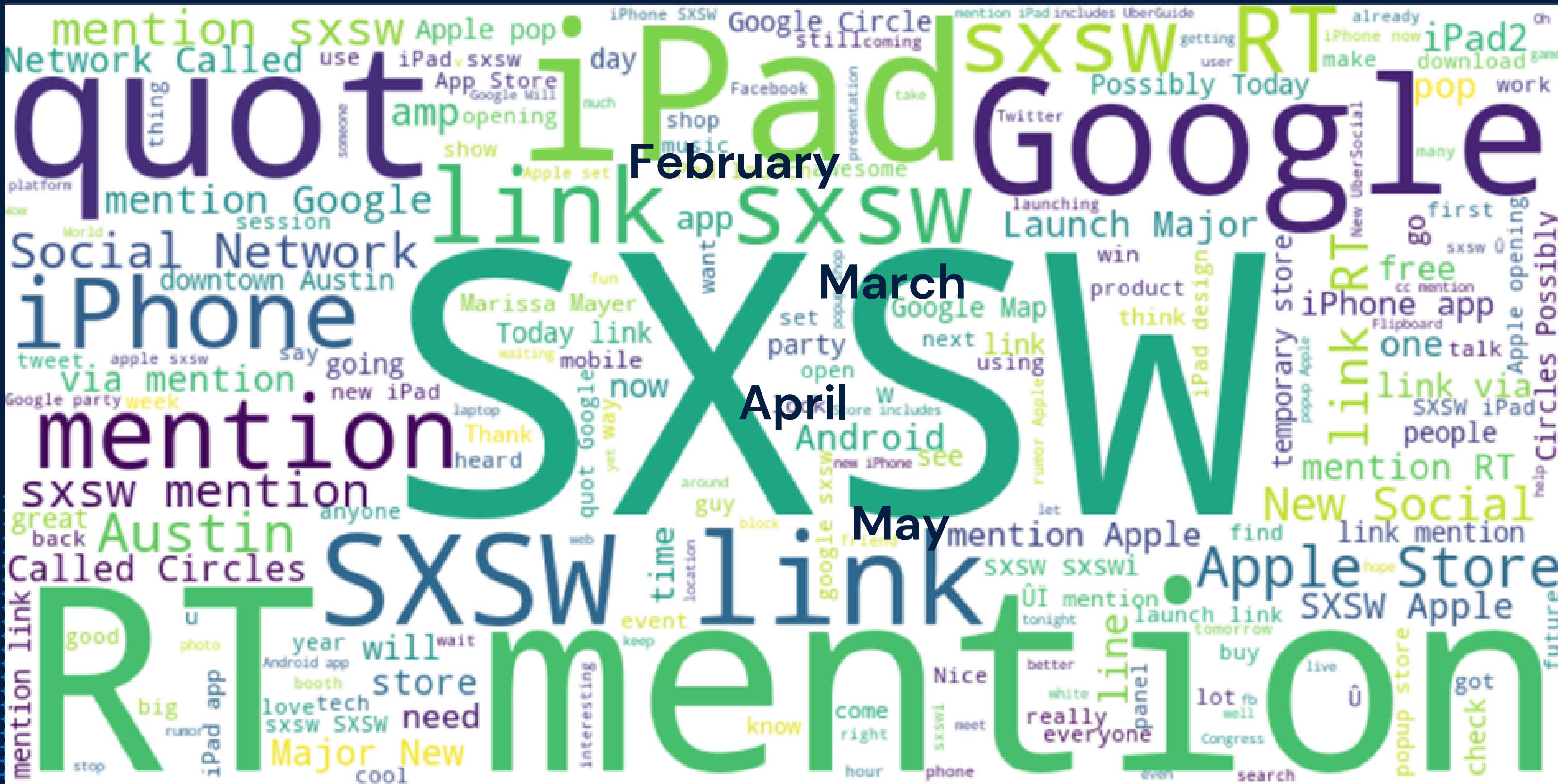
In the word cloud graph, the most repeated words include

- 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

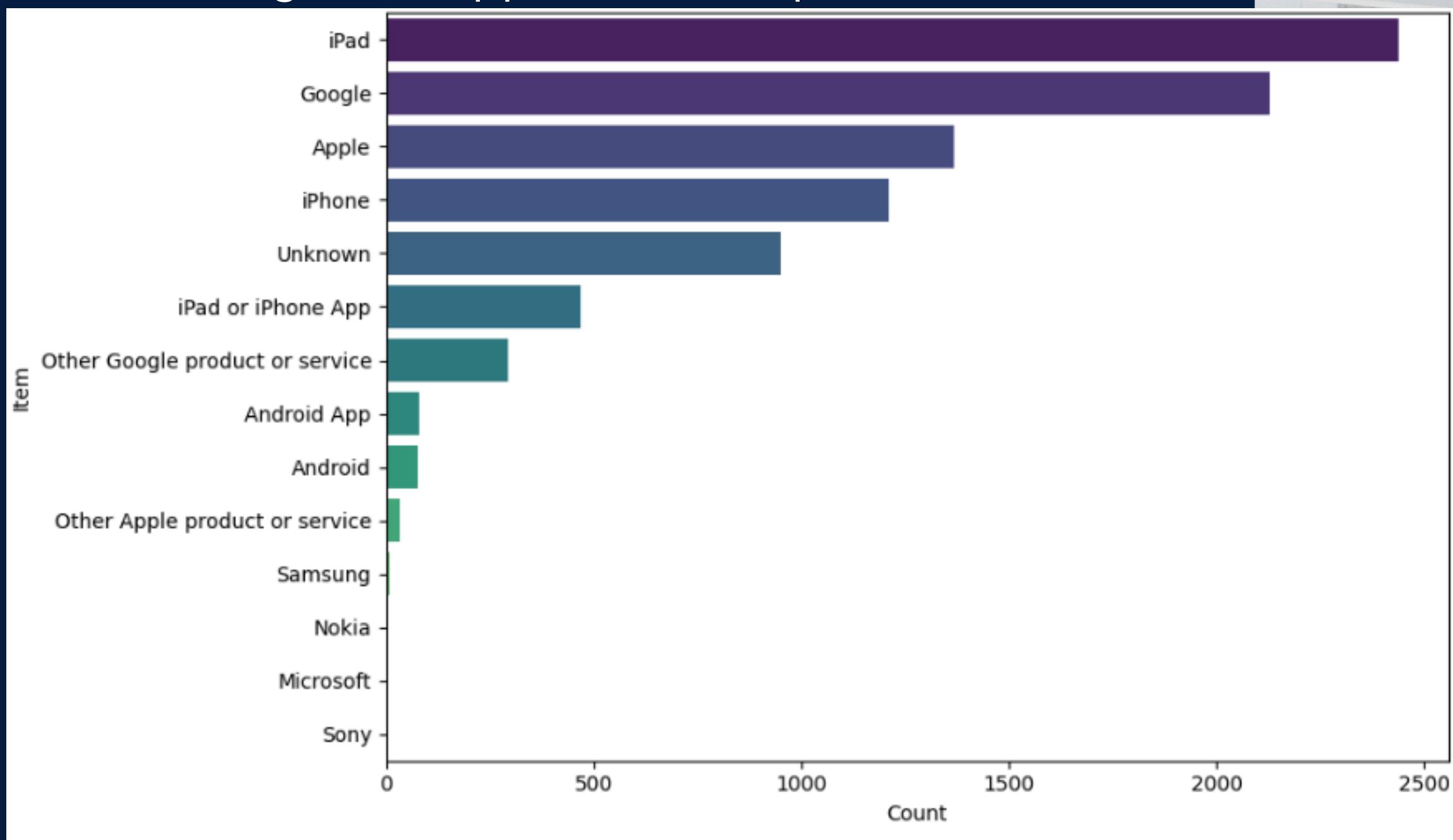
Most Frequent Words



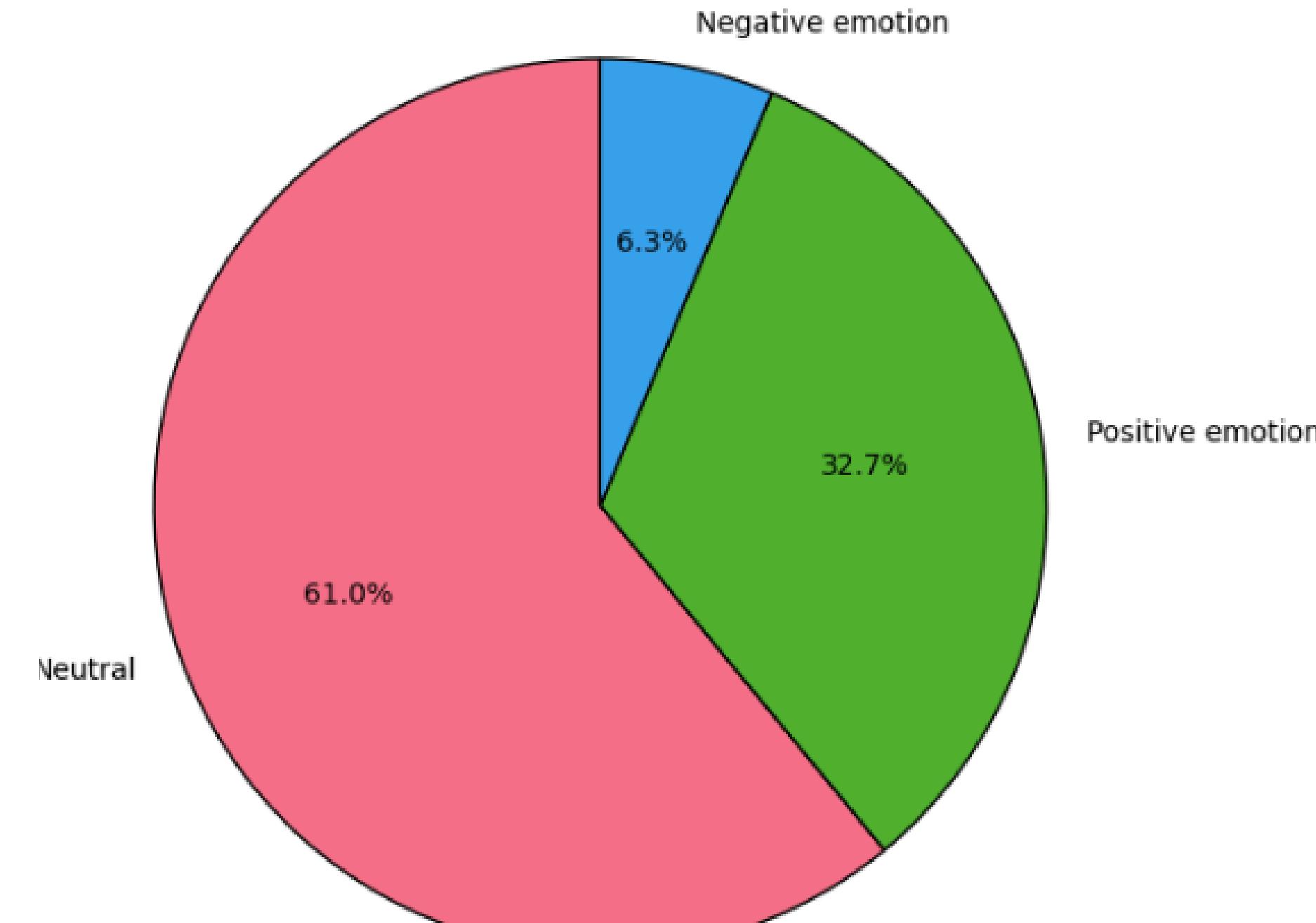
DATA ANALYSIS: MOST OCCURRING ITEMS

1. Which is the most common product?:

iPad, Google and Apple are the top most items



Distribution of Sentiments in Tweets

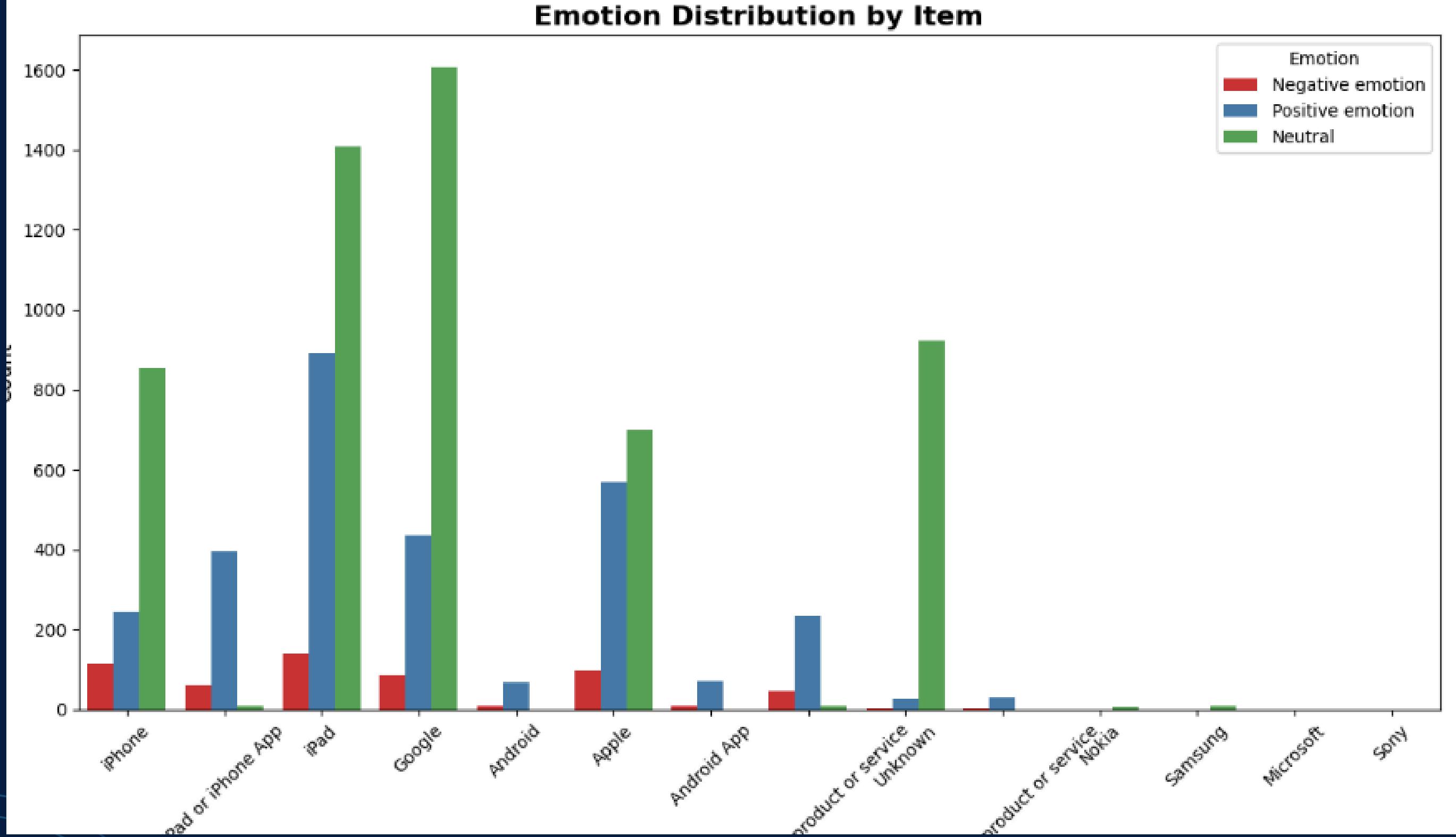


The visualization reveals that the majority of responses exhibit no emotion toward the brand or product(Neutral) followed by a significant proportion of positive reviews, with negative reviews being the least frequent. This suggests overall neutrality with a leaning toward positivity in customer sentiment.

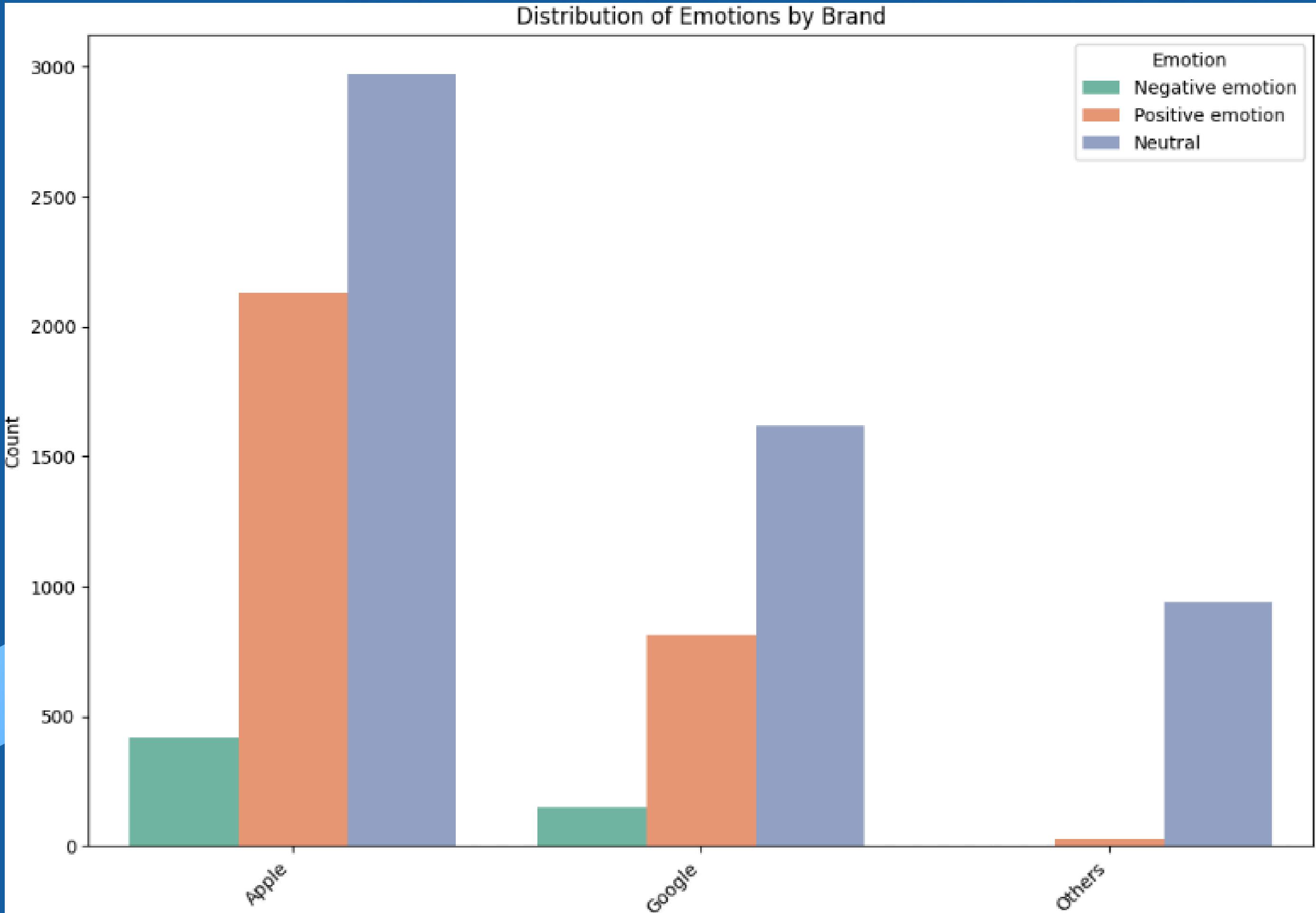
BIVARIATE ANALYSIS

Exploring and analyzing the relationship between two variables.

Emotion Distribution by Item

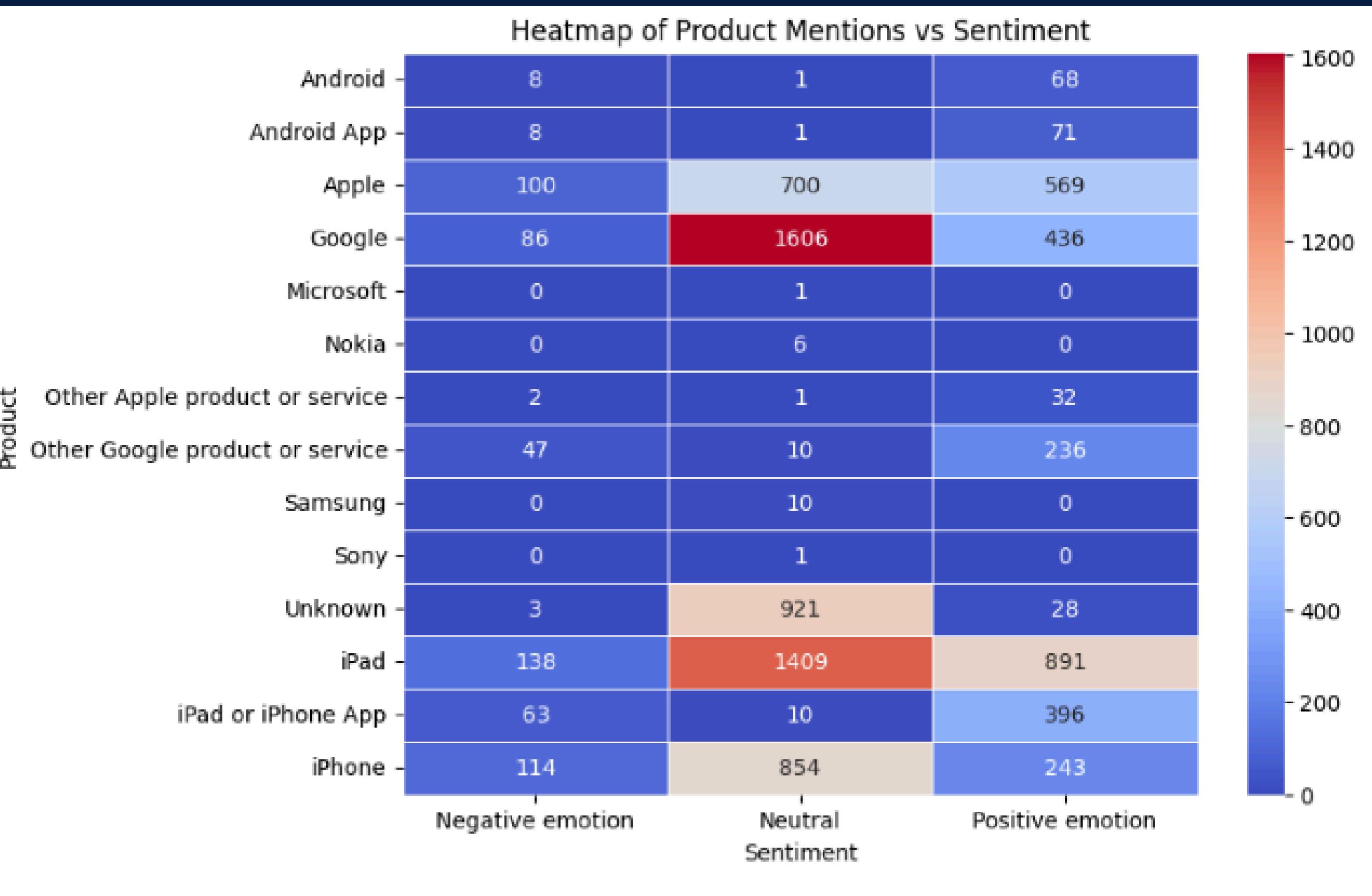


DISTRIBUTION OF EMOTION BY BRAND



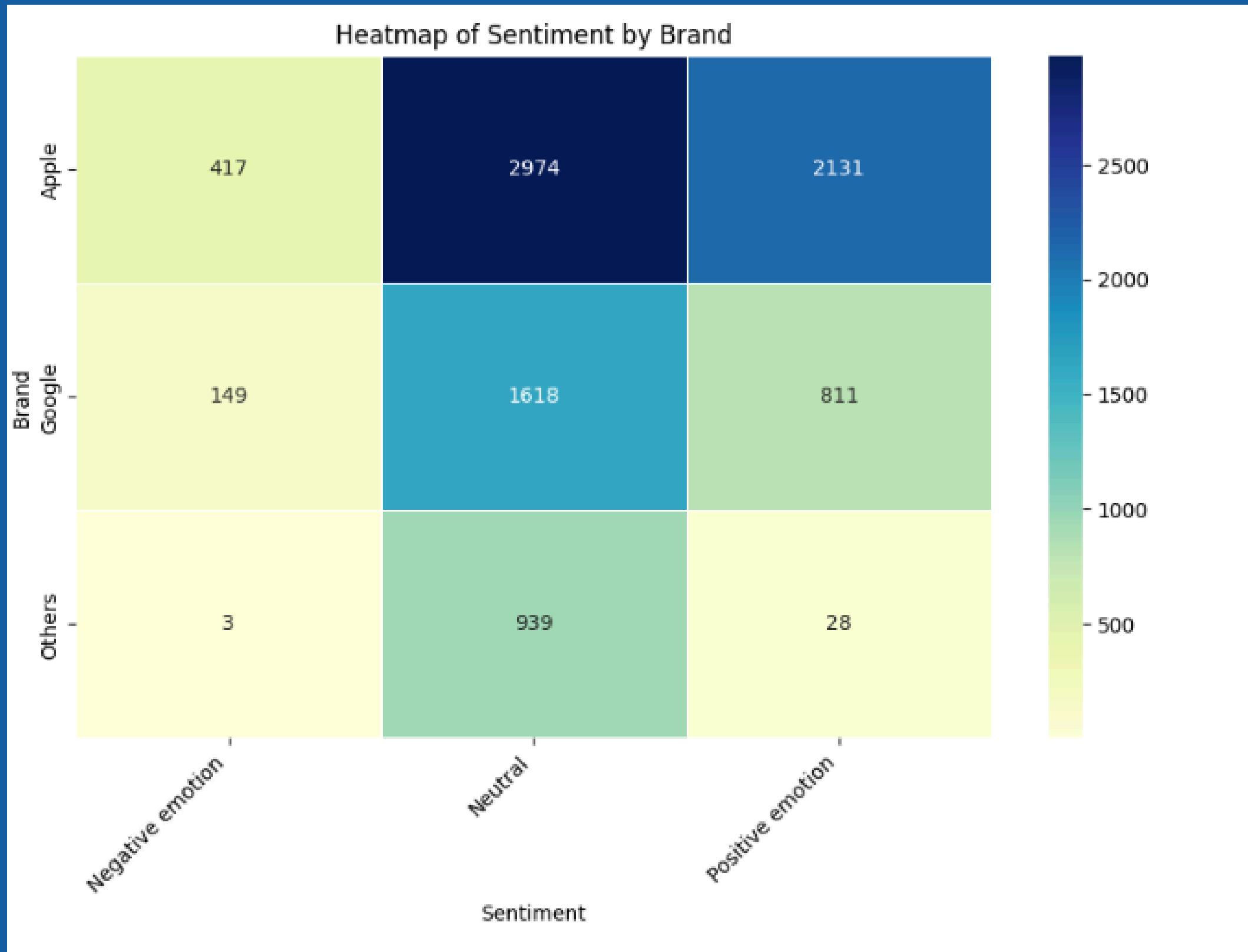
Apple products seem to evoke more neutral or positive feelings, while Google products generate a wider mix of emotions. The rest of the brands, however, don't seem to stir as much emotion overall.

CORRELATION OF PRODUCT MENTIONS AND SENTIMENTS



- **iPhone:** Strong customer satisfaction with the highest positive mentions, while negative feedback is minimal in comparison.
- **Google:** High neutral mentions suggest factual discussions, but significant negative feedback highlights areas for improvement.
- **Other Products (iPad, Chromebook):** Mixed sentiments present opportunities to address negative feedback and improve perception.

CORRELATION OF SENTIMENT AND BRAND



NEXT STEPS

Thank You Note



Data Scientist

Moringa School

Connect with us.

