Birdie Inc.

Sentiment Analysis on Tweets about Brands and Products

A new way to engage with customers from across the globe.

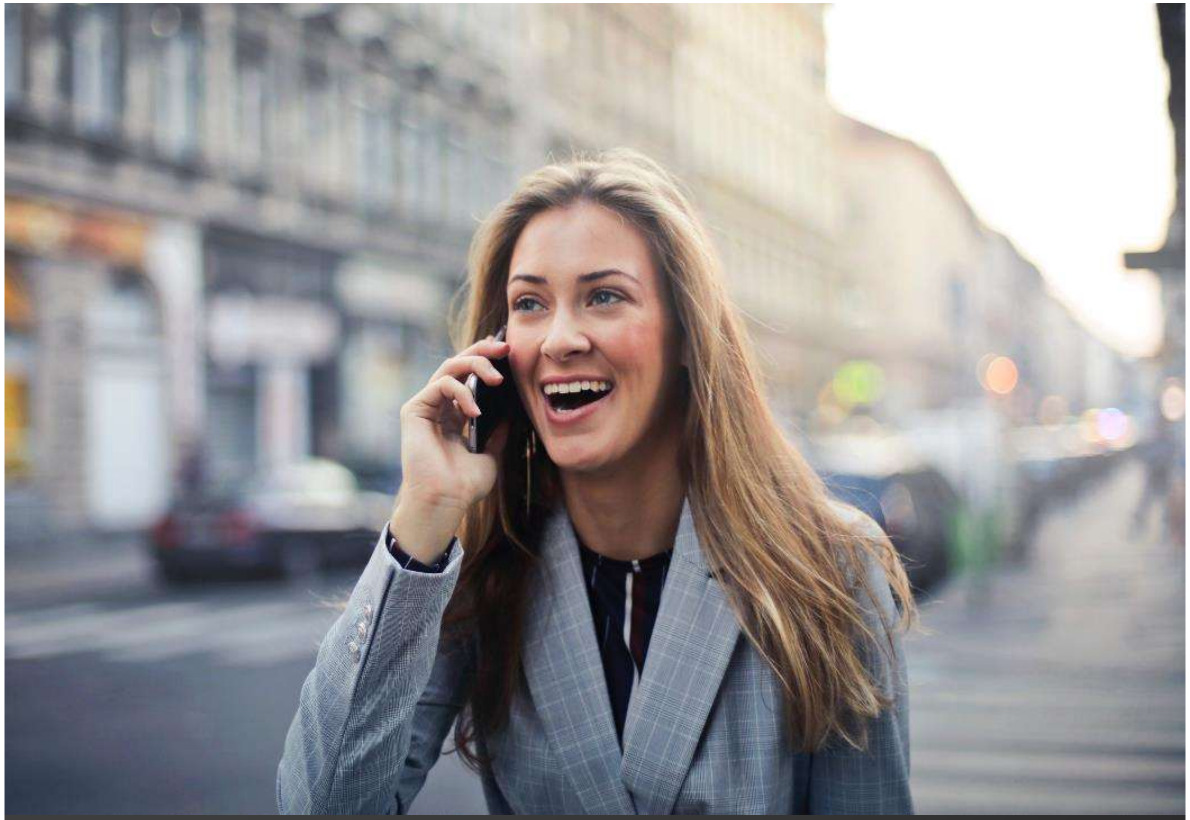
Members:

Ingavi Kilavuka

Calvin Omwega

Alvin Kimathi

Ronny Kabiru



About Us

The Problem

In the age of social media, understanding public sentiment towards brands and products is crucial for businesses

Determine whether the sentiment expressed in each tweet is positive, negative, or neutral.

Identify the specific brand or product that the sentiment is directed towards.

Derive insights and trends regarding public perception of various brands and products.

Research Questions

Objective:
To analyze the sentiment of tweets directed at various brands and products to understand public perception. This will involve determining whether the sentiment expressed in tweets is positive, negative, or neutral and identifying the specific brands and products mentioned.

1. **Sentiment Classification**

   What is the distribution of positive, negative, and neutral sentiments across the dataset?

2. **Brand/Product Association**

   Which brands and products are most frequently mentioned, and what is the associated sentiment for each?

3. **Temporal Trends**

   Are there any notable trends in sentiment over time for specific brands or products?

Data Understanding

Data Type

Columns & Rows

File Type

Microsoft Excel CSV

Data Type

Objects, Integers, Floats(Decimated Numbers)

Columns

3

Rows

8721

Data Preparation & Analysis

Columns:

1. Tweet text
2. Emotion in tweet directed at…
3. Is there an emotion directed at brand/product?

Missing values in Columns (Respectively):

1. 0.01%
2. 63.7%
3. 0%

Null values in Columns: **0**

Dealing with missing values in Columns:
Drop rows with missing values.

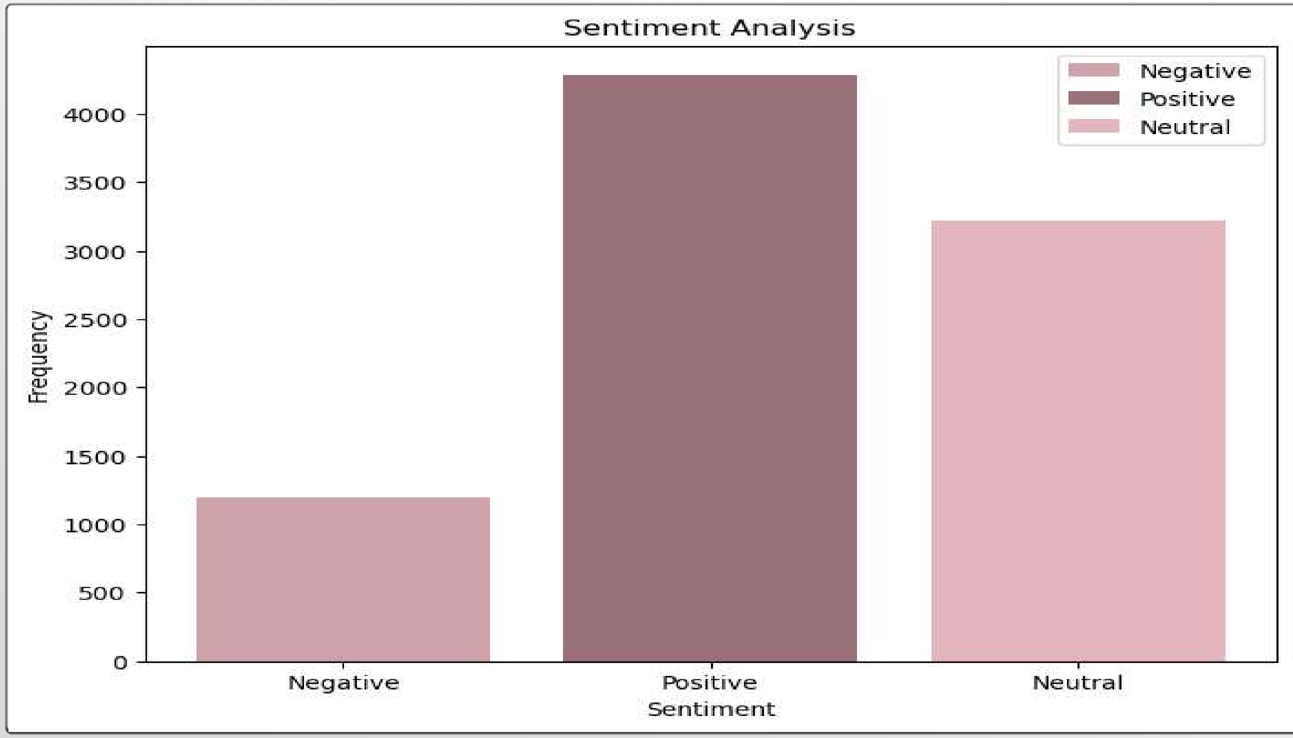
Exploratory Data Analysis: Sentiment Analysis

1. Categorize Sentiments as Positive, Neutral or Negative using TextBlob.
2. Create a frequency table to visualize the findings.

Positive: 4283

Neutral: 3219

Negative: 1191



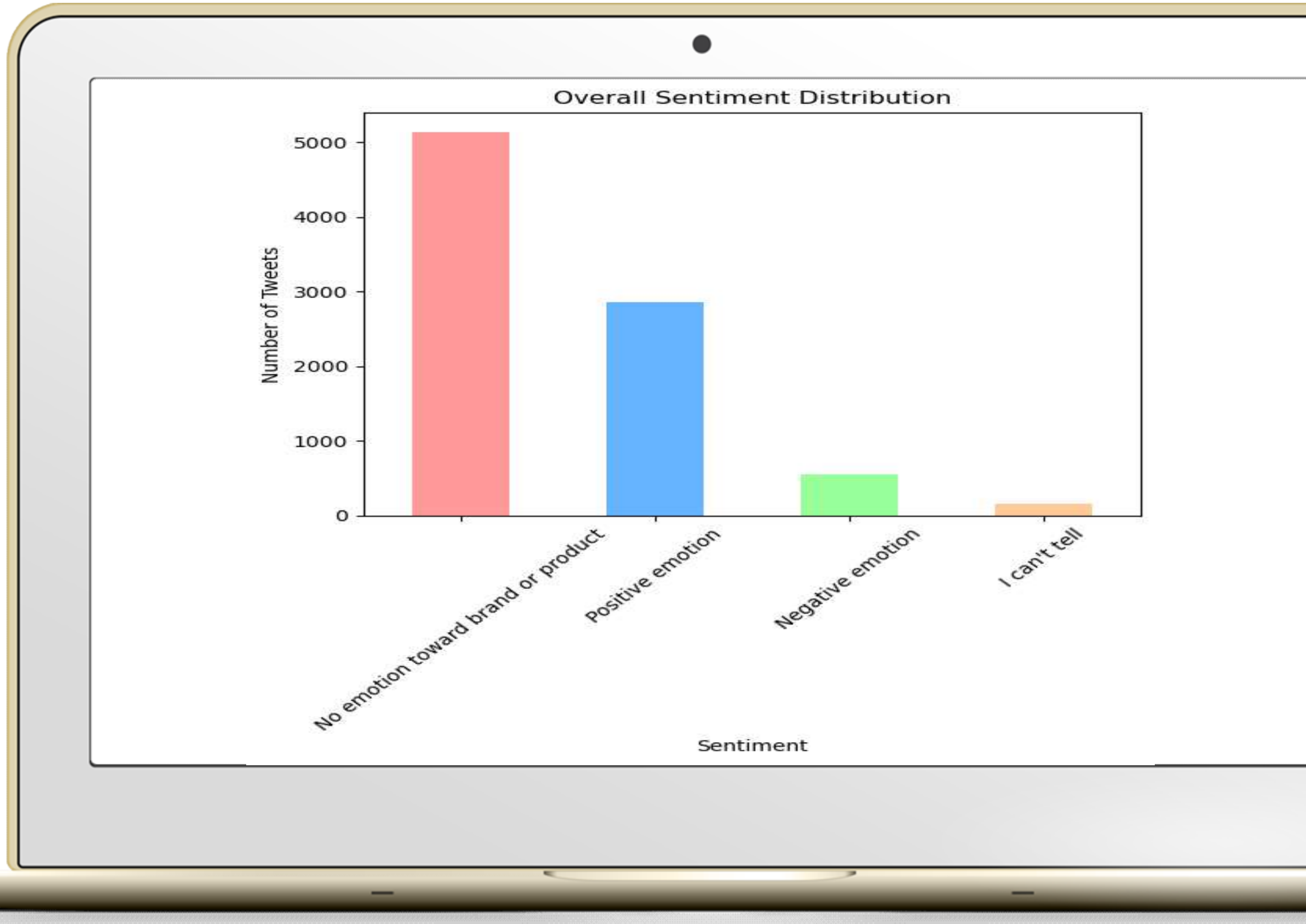
Exploratory Data Analysis: Sentiment Distribution

1. This bar chart shows the number of tweets expressing each type of sentiment (Positive, Negative, No Emotion, I Can't Tell).

2. The **most frequent** is "No emotion toward brand or product," indicating that many tweets do not express a clear sentiment about a brand or product.

3. "Positive emotion" is the **second** closely, **followed by** "Negative emotion." This suggests that, among tweets with a clear sentiment, positive emotions are more prevalent than negative ones.

4. * The "I can't tell" category represents tweets where the sentiment could not be determined.



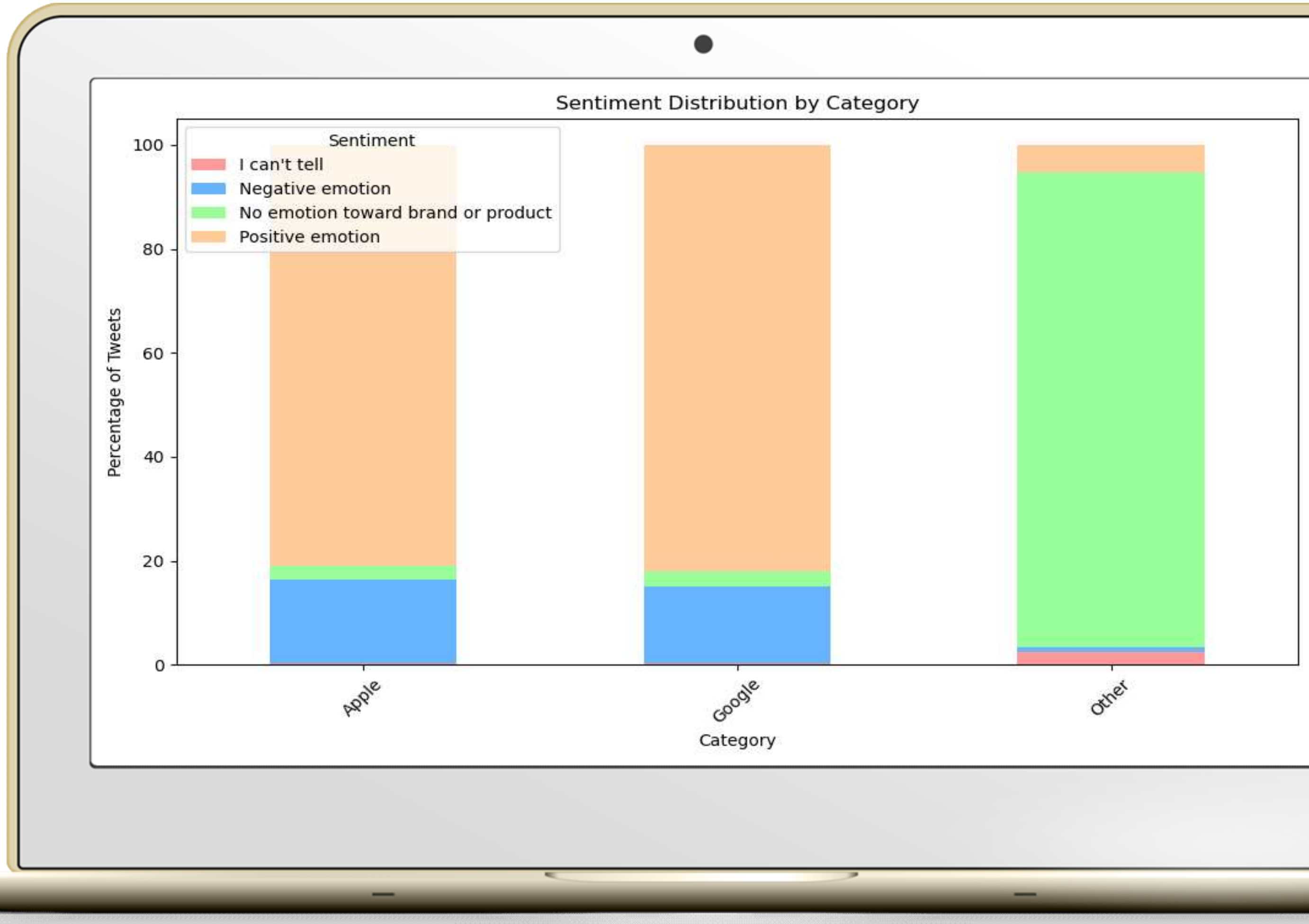
Sentiment Distribution by Category

Apple Products: A significant majority of the tweets express positive sentiments (81.02%), with a smaller proportion being negative (16.01%). This indicates a generally favorable public perception.

Google Products: Similar to Apple, the majority of tweets are positive (82.15%), with a smaller percentage being negative (14.66%). This suggests that Google products are also well-regarded.

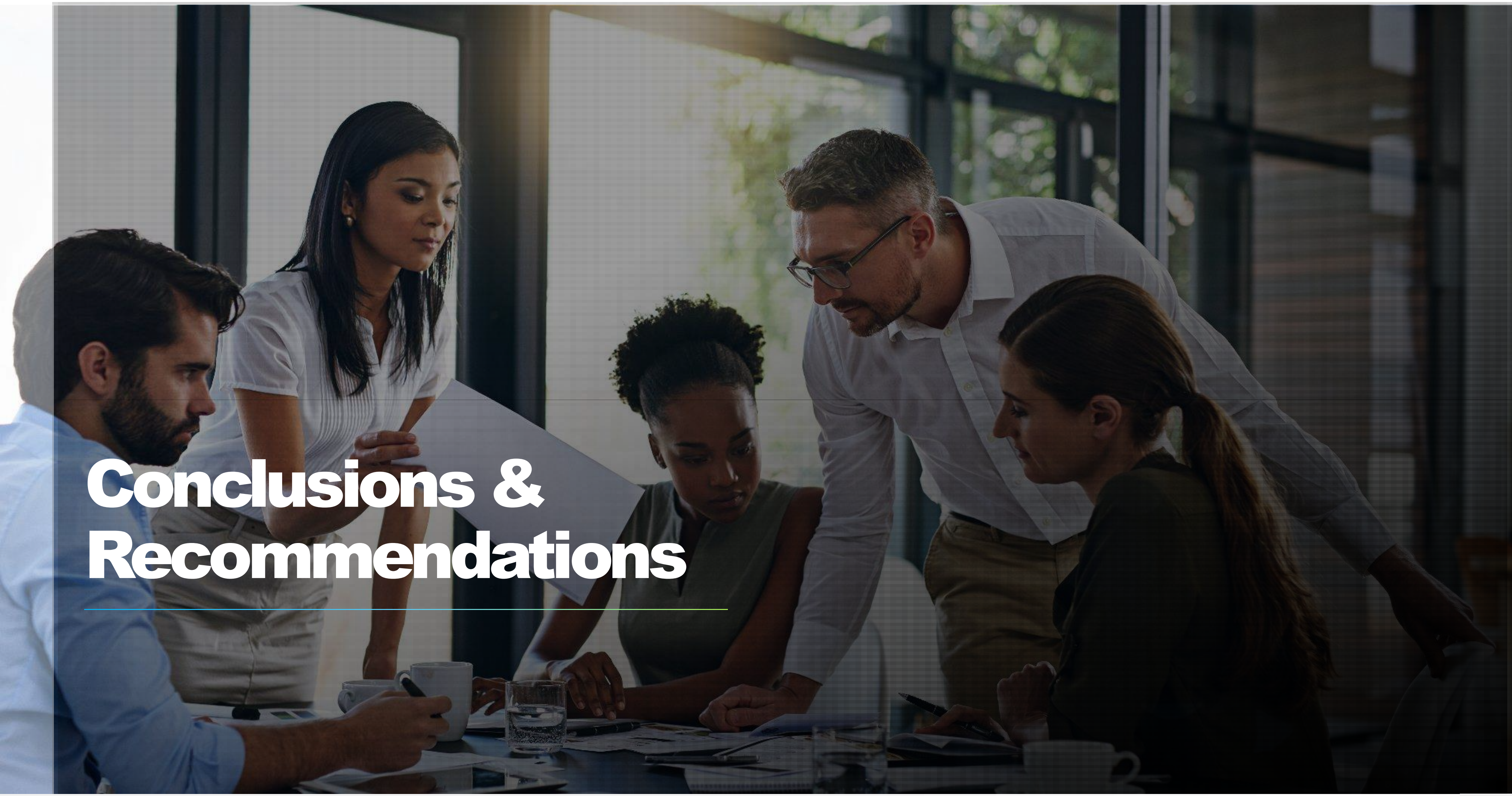
Unknown Category: Most tweets in this category have no specific emotion directed at a brand/product, which may include generic comments or non-specific discussions.

These findings show a generally positive sentiment towards both Apple and Google products, with a notable proportion of neutral or non-specific tweets.



Modeling

Model Used					
	SVM with GridSearchCV	Random Forest	Logisitc Regression (Baseline)	Gradient Boosting Classifier	XGBoost
Why?	-Effective in High-Dimensional Spaces -Robust to Overfitting & Outliers -Margin Maximization -Memory Efficiency	-High Accuracy -Feature Importance -Handles High Dimensionality	-Baseline Model -Multicollinearity Handling -Simplicity and Interpretability	-High Predictive Accuracy -Regularization -Handles Complex Data	-High Performance -Regularization -Cross-Validation
Accuracy	75.1%	74.4%	73.5%	72.3%	71.4%
Metrics of Success	Accuracy 70%+ F1 score 80%+ Recall 85%+	Accuracy 70%+ F1 score 80%+ Recall 85%+	Accuracy 70%+ F1 score 80%+ Recall 85%+	Accuracy 70%+ F1 score 80%+ Recall 85%+	Accuracy 70%+ F1 score 80%+ Recall 85%+



Conclusions & Recommendations

Conclusions and Recommendations

Conclusions	Recommendations
The project has addressed the problem statements to a certain extent: 1. **Sentiment Identification**: Achieved with varying degrees of accuracy. Improvements are needed, particularly in handling class imbalance and improving positive sentiment detection. 2. **Target Identification**: Partially addressed through brand/product association with sentiment. More detailed analysis could provide deeper insights. 3. **Trends and Insights**: Initial insights were provided, highlighting the need for further analysis to uncover detailed trends and public perceptions. Overall, while the project successfully built and evaluated models for sentiment analysis, further work is needed to refine the models, address identified issues, and explore additional dimensions of the data to fully meet the project objectives and problem statements.	1. **For Brands**: Use the insights to address negative sentiments and capitalize on positive ones. Engage with customers on social media to improve brand perception. 2. **For Further Analysis**: Extend the analysis to include additional data sources, such as other social media platforms, to get a broader view of public sentiment. 3. **Model Improvement**: Experiment with different model architectures or additional features (e.g., tweet metadata) to improve classification accuracy.

Thank You!

GitHub Repository