Agenda

1. Executive Summary
2. Business Problem
3. Dataset Overview
4. Model Creation, Testing , Performance,
5. Business Insights ,recommendations and conclusion
6. Meet the Team

Current Business Problem

1. Thousands of tweets per minute → impossible to read manually
2. No automated way to spot early negative trends
3. No brand-level sentiment comparison

Our Solution

1. Automated text Processing

- Processes tweets automatically
- Real time data extraction

2. Classifies sentiment with strong accuracy

- Classifies sentiment with strong accuracy
- High precision emotional labelling

3. Brand specific Emotional Insight

- Delivers brand-specific emotional insights
- Actionable data driven brand understanding

What we have built

An NLP model that reads a tweet and predicts whether the sentiment is positive, negative, or neutral.

Why Stakeholders Care?

Competitive Benchmarking

Evaluating brand perception against competitors to identify strength and weakness

Early Reaction Detection

Identifying negative feedback promptly to mitigate potential damage

Informed Decision Making

Grading products, PR, and marketing strategies and data driven insights

Scaled Sentiment Analysis

Understanding customer emotions across large audience

Group 4

Dataset overview

Key columns Used

- tweet_text — customer opinions
- emotion_in_tweet_is_directed_at —brands/products mentioned
- sentiment label — positive, negative, or
- neutral

Data Quality Actions

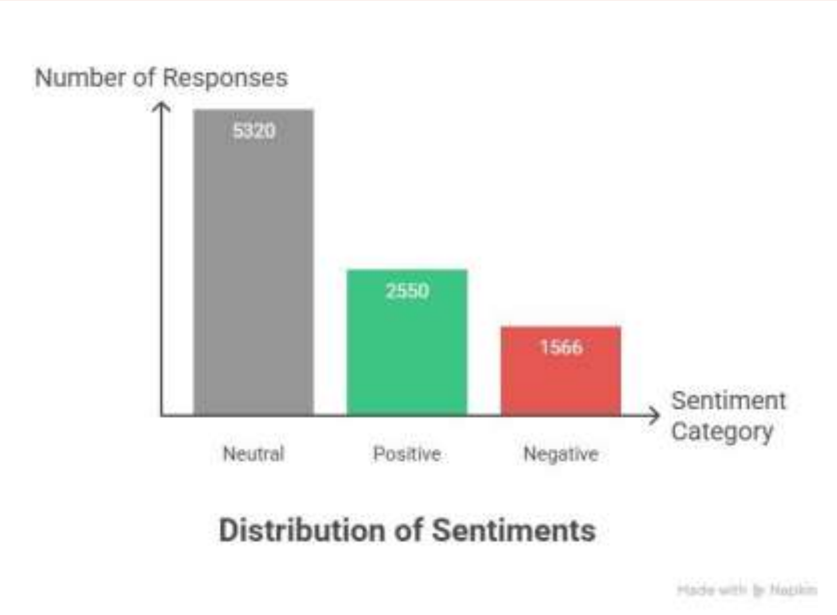
- Removed duplicates
- Fixed encoding errors
- Standardized labels

	tweet_text	emotion_in_tweet_is_directed_at	is_there_an_emotion_directed_at_a_brand_or_product
0	.@wesley83 I have a 3G iPhone. After 3 hrs twe...	iPhone	Negative emotion
1	@jessedee Know about @fludapp ? Awesome iPad/i...	iPad or iPhone App	Positive emotion
2	@swonderlin Can not wait for #iPad 2 also. The...	iPad	Positive emotion
3	@sxsw I hope this year's festival isn't as cra...	iPad or iPhone App	Negative emotion
4	@sxtxstate great stuff on Fri #SXSW: Marissa M...	Google	Positive emotion

Dataset overview

Sentiment Distribution(Before Balancing)

Original Class Distribution



After SMOTE Oversampling, all classes balanced to 5,388 samples each.

Data Cleaning and NLP Preprocessing

Text Processing sequence

Steps applied

Lowercasing → Removing URLs, emojis, punctualtons → Tokenization → Stopword removal → Lemmatization → TF-DF Vectorization

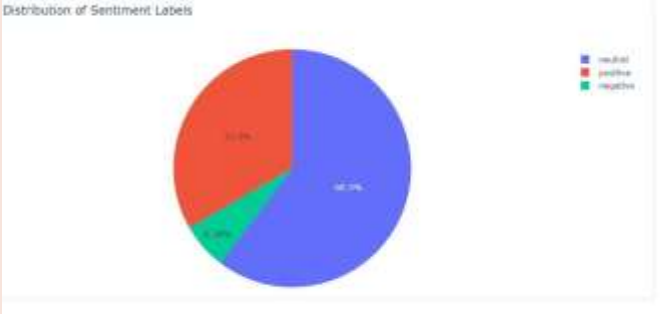
Lowercasing	Removing URLs, emojis, punctualtons	Tokenization	Stopword removal	Lemmatization	TF-DF Vectorization
Remove text to lowercase	Clean text by removing unnecessary elements	Break text into individual words	Eliminate common words like “the” and “and”	Reduce words to their base form	Convert text into numericals vector

Dataset Overview

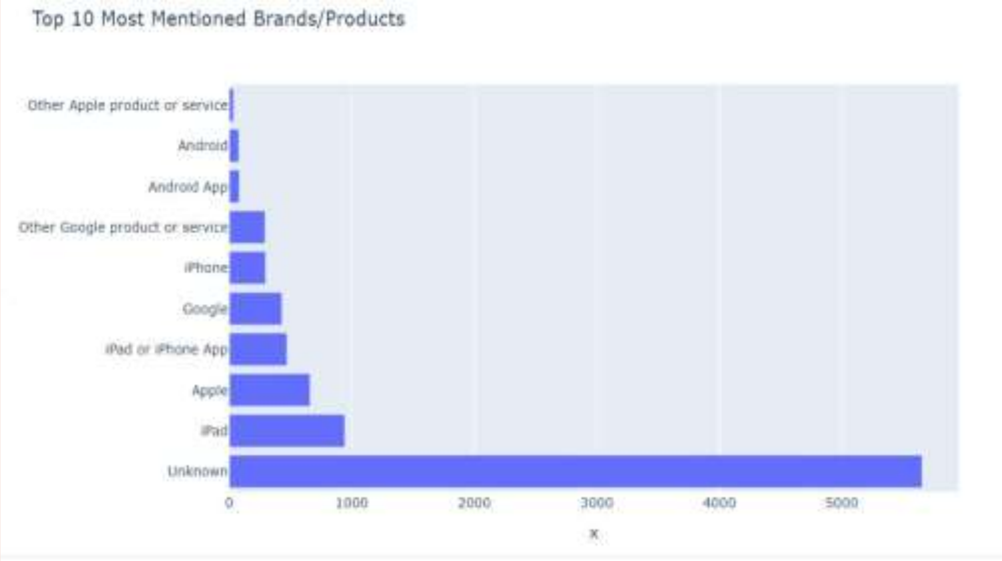
Segment Analysis

Brand mention analysis

01



02

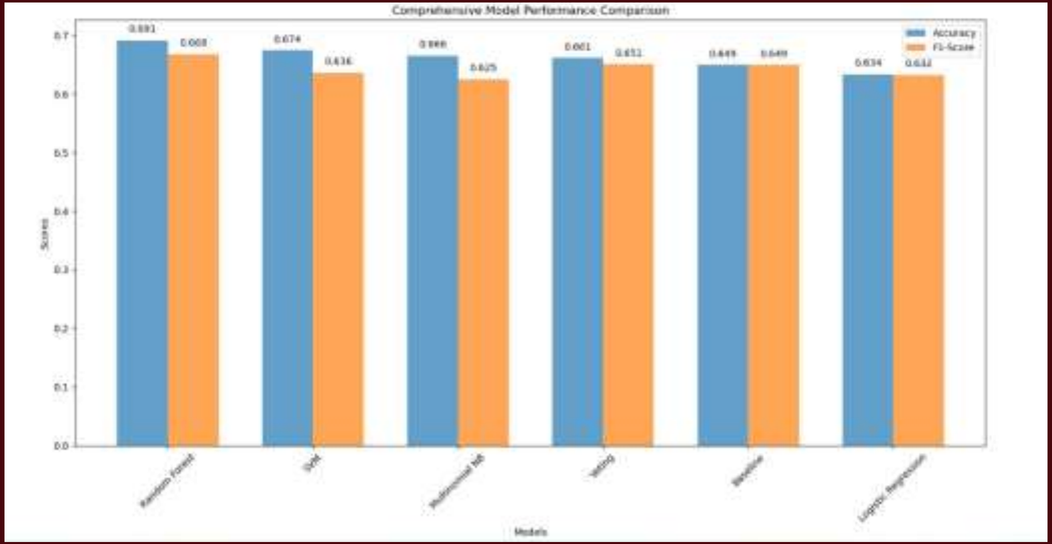


Modeling Approach Models Tested

This is based on the notebook

- Naive Baye
- Logistic Regression Multinomial
- Random Forest
- Voting Classifier (combined models.

LINK TO THE NOTEBOOK



Training set up

- Train/test split: 70/30
- TF-IDF inputs
- SMOTE balanced training data

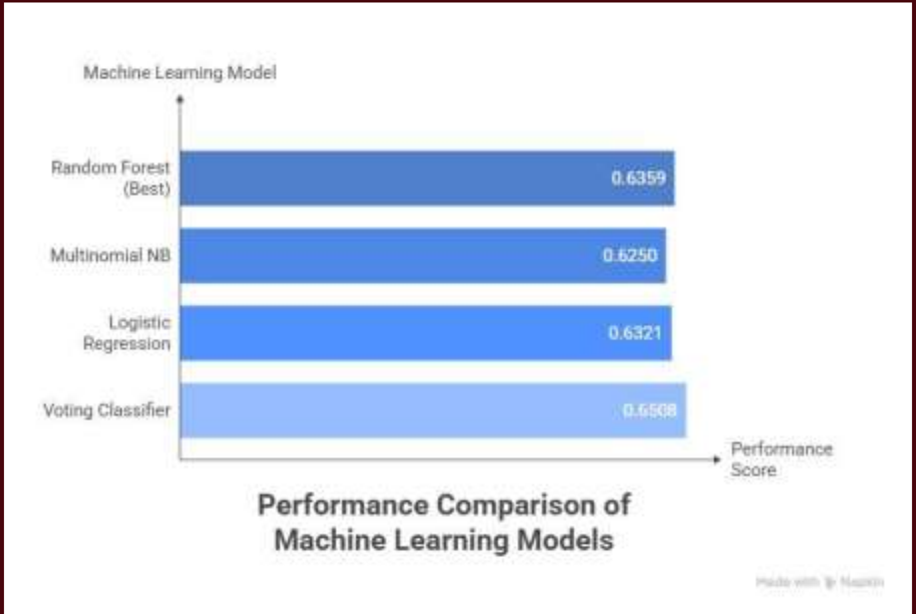
Baseline Model Performance

Metric	Precision	Recall	F1-Score	Support
Negative	0.31	0.34	0.33	114
Neutral	0.73	0.74	0.74	1078
Positive	0.56	0.54	0.55	596

Made with Napkin

Key Model Performance Baseline Model

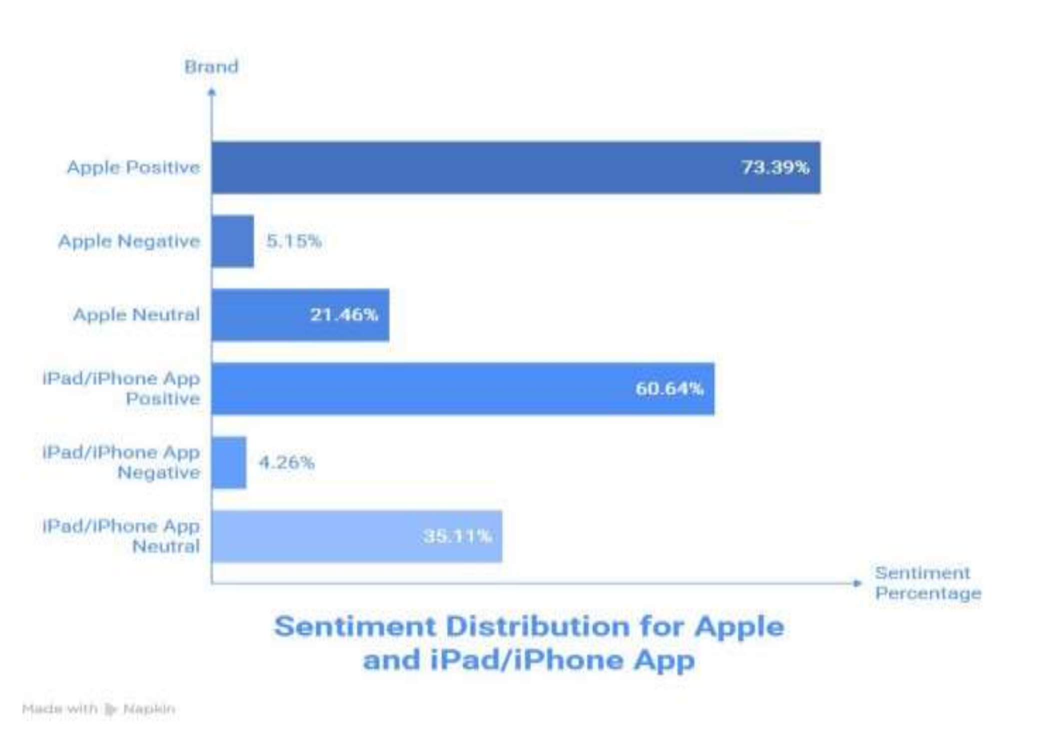
Logistic Regression baseline accuracy: 0.6493



What does Baseline accuracy of 0.6493 mean?

- Models perform consistently around 66% accuracy
- Strong performance on positive & negative tweets
- Neutral tweets are harder due to ambiguous language

Brand-Level Sentiment Insights



Key Takeaways

- **Apple-related products dominated positive sentiment**
- **Foursquare received the highest percentage of negative tweets**
- **Low negative percentages indicate strong brand satisfaction overall**
-

Business Insight

01

Strong Positive Momentum

Marketing teams should amplify user generated excitement and support positive creators

02

Negative Tweets Spikes

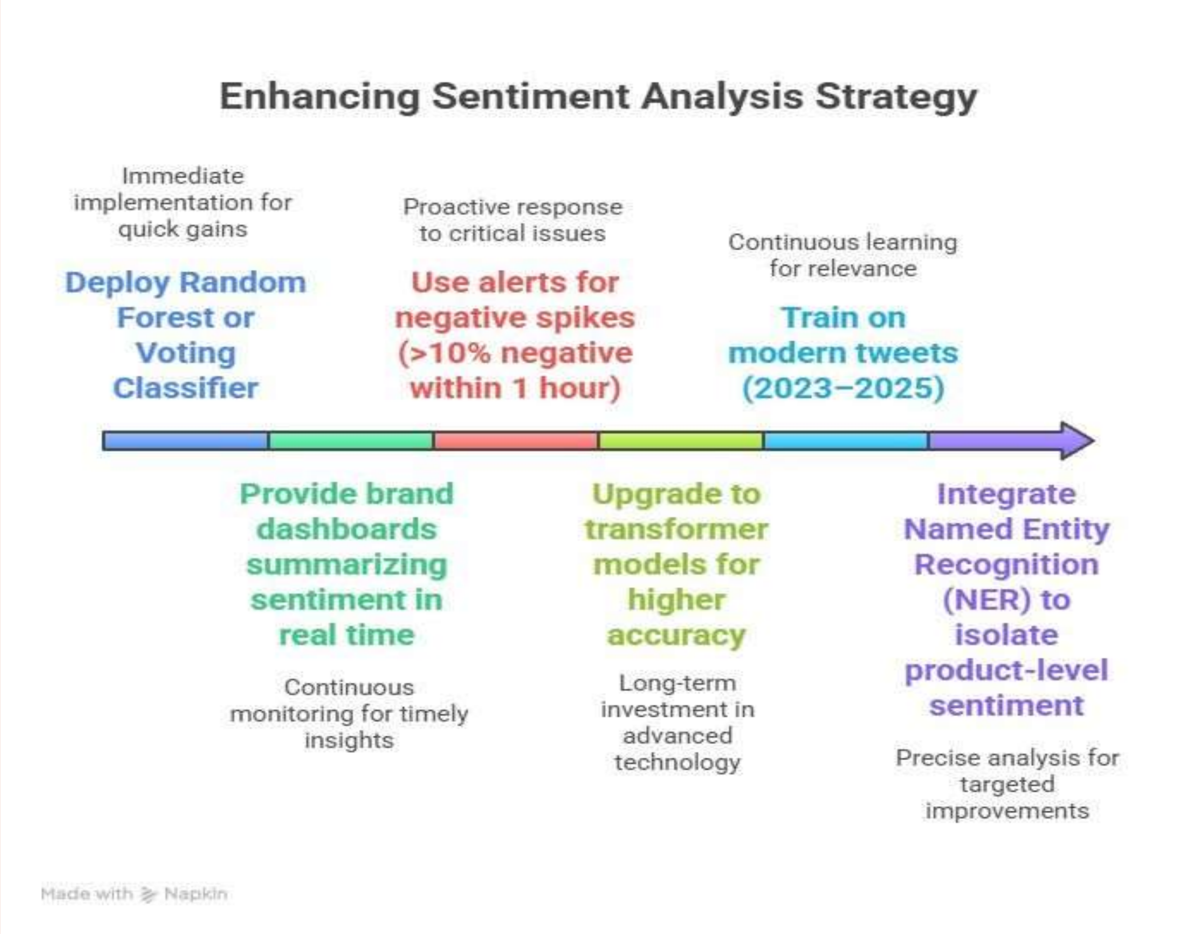
Small negativity matters due to viral amplification and event visibility

03

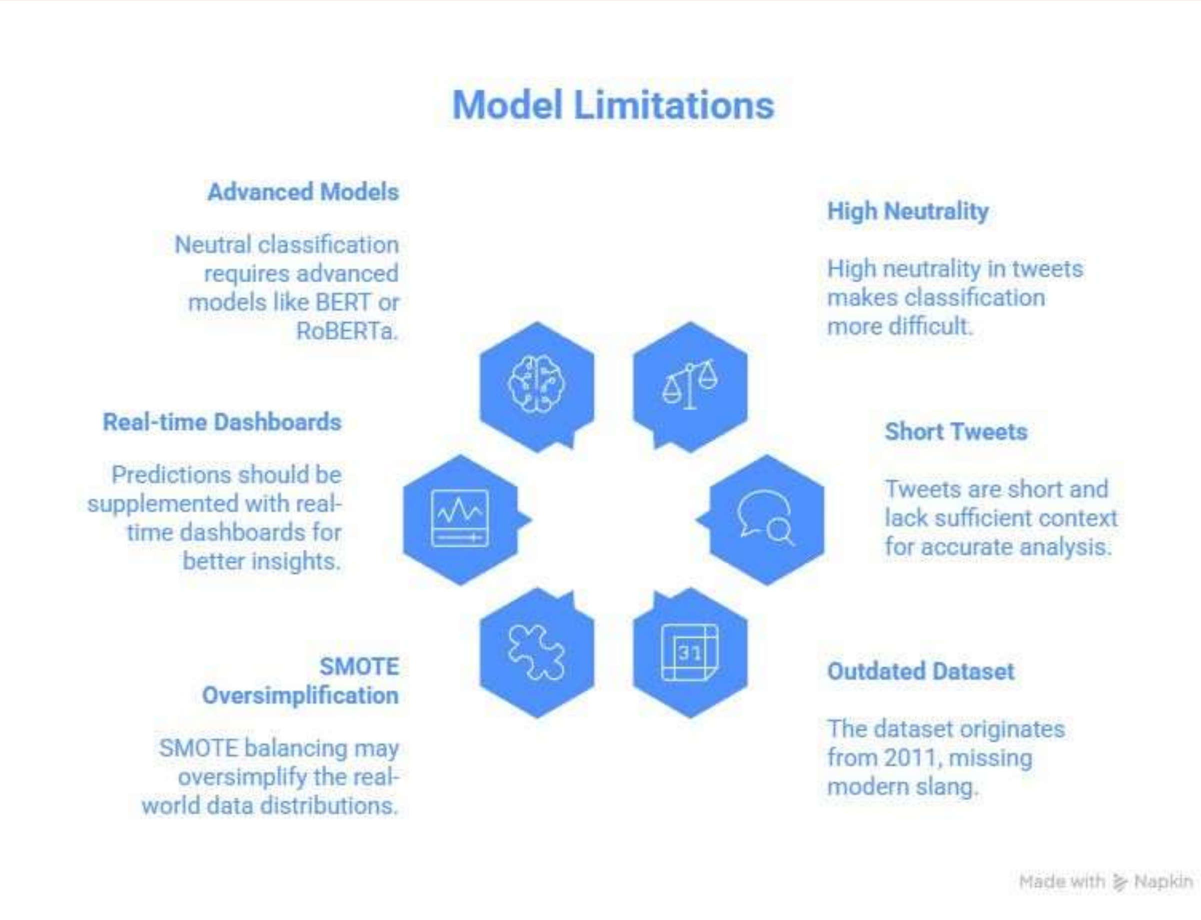
Neutral Sentiment Opportunity

High neutral rates indicate unclear messaging or features not resonating

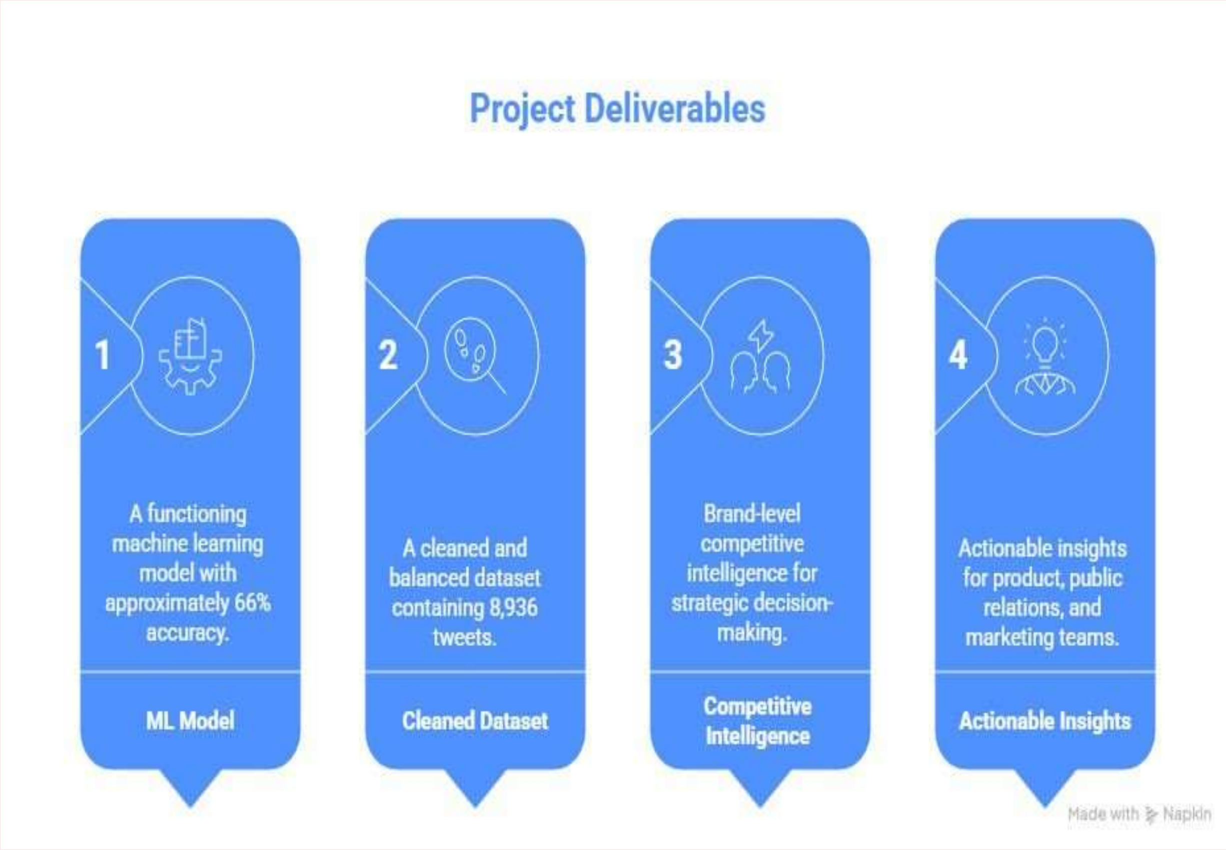
Recommendations for Product Managers



Model Limitations And its Impact on Stakeholders



Conclusion


Project Deliverables
1
A functioning machine learning model with approximately 66% accuracy.
ML Model
2
A cleaned and balanced dataset containing 8,936 tweets.
Cleaned Dataset
3
Brand-level competitive intelligence for strategic decision-making.
Competitive Intelligence
4
Actionable insights for product, public relations, and marketing teams.
Actionable Insights
Made with Napkin

Meet the Group Members

Cindy Akinyi
GitHub Name:https://github.com/cindyakinyi

Abigael Musyoka
Github Name: https://github.com/Opulent-Abby

Kelvin Sesery
Github Name:https://github.com/Sesery

Joan Omanyo
Github Name: ttps://github.com/Joan-omanyoh

Any questions? Ask away!

Thank you!