TWEET SENTIMENT ANALYSIS

Presented by:
Anthony, Winnie, Jedidah, Navros,Kelvin,Rachel.

PROJECT OVERVIEW

This project leverages Natural Language Processing (NLP) techniques to analyze public sentiment toward Apple and Google products. we aim to develop predictive models capable of accurately classifying sentiment.

The primary objective is to create a proof-of-concept sentiment analysis system that not only predicts sentiment accurately but also provides interpretable insights for decision-making.

STAKEHOLDER AND BUSINESS VALUE

Stakeholders

- Tech Companies: Track customer satisfaction, detect issues, and prioritize product improvements.
- Social Media Analysts & Data Scientists: Monitor trends and generate actionable insights.
- Investors & Strategists: Inform investment decisions and competitive strategies.
- Researchers: Study NLP, social media trends, and consumer behavior.
- Customers: Benefit indirectly through improved products, services, and brand experience.

Business Value

- Identify product satisfaction and potential issues.
- Optimize marketing, engagement, and communication strategies.
- Guide strategic planning and investment decisions.
- Support research in NLP and social media analytics.
- Enhance customer experience through feedback-driven improvements.

ABOUT THE DATASET

- Source: CrowdFlower via data.world (added August 30, 2013 by Kent Cavender-Bares)
- Size: 9,093 labeled tweets spanning multiple brands and products
- Annotation: Contributors manually labeled tweets as positive, negative, or neutral
- Granularity:When sentiment was expressed, the specific brand or product was also tagged

LIMITATIONS OF THE DATASET

- Primarily English tweets, limiting global coverage
- Informal language requires careful preprocessing
- Neutral sentiment dominates → potential class imbalance
- Data collected before 2013 → modern sentiment may differ, but methodology remains relevant

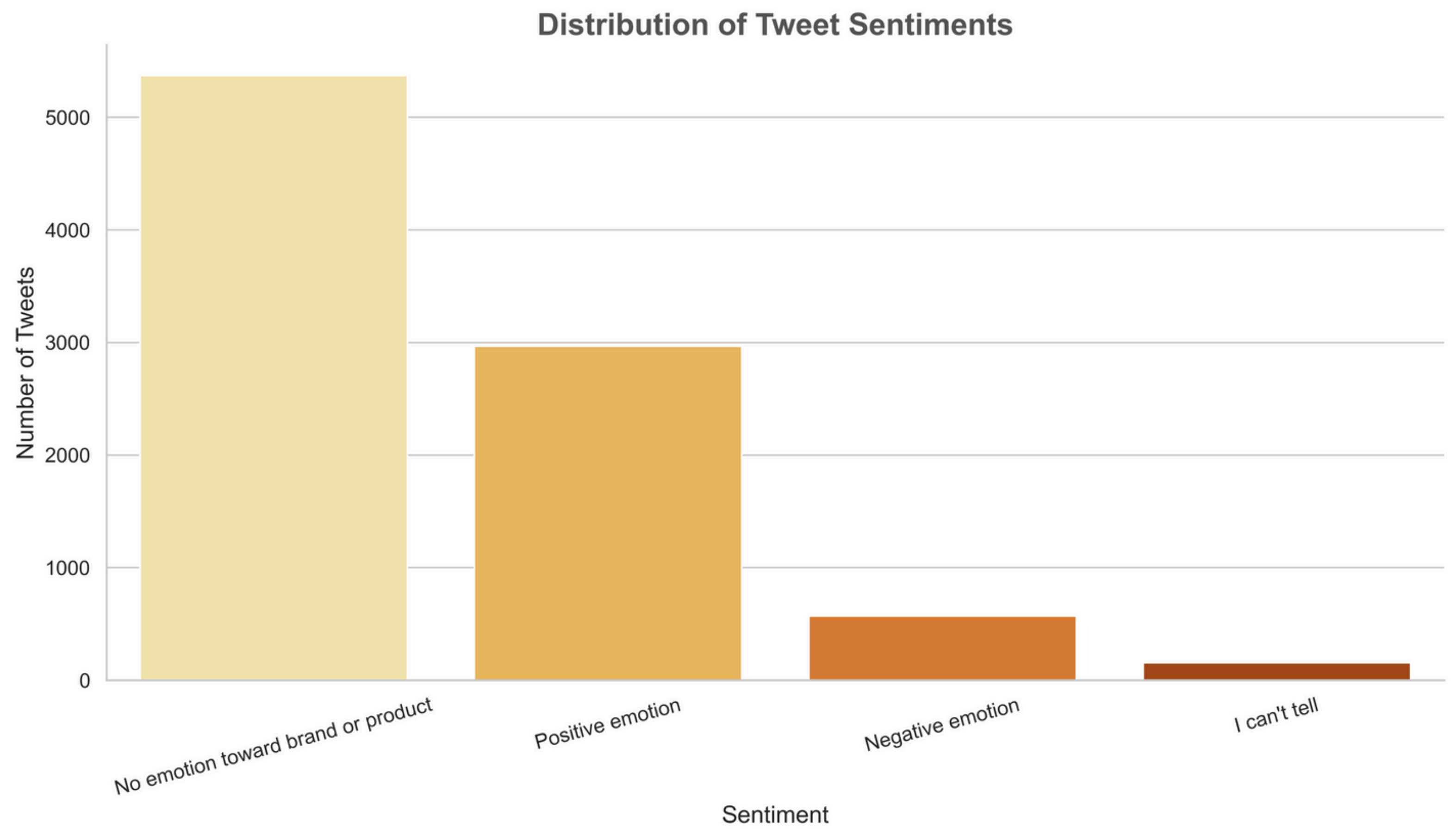
DATA CLEANING AND PRE- PROCESSING

- Removed Missing & Duplicate Data → ensured accuracy and relevance.
- Handled Missing Brand References → assigned “Unknown” to keep all tweets.
- Standardized Text → lowercasing, removing punctuation, URLs, hashtags, mentions, and numbers.
- Prepared for NLP → tokenization, stopword removal, and lemmatization.
- Checked Class Balance → noted imbalance (Negative underrepresented).
- Ensured Data Quality → reliable input for feature engineering & model training.

KEY INSIGHTS FROM EDA

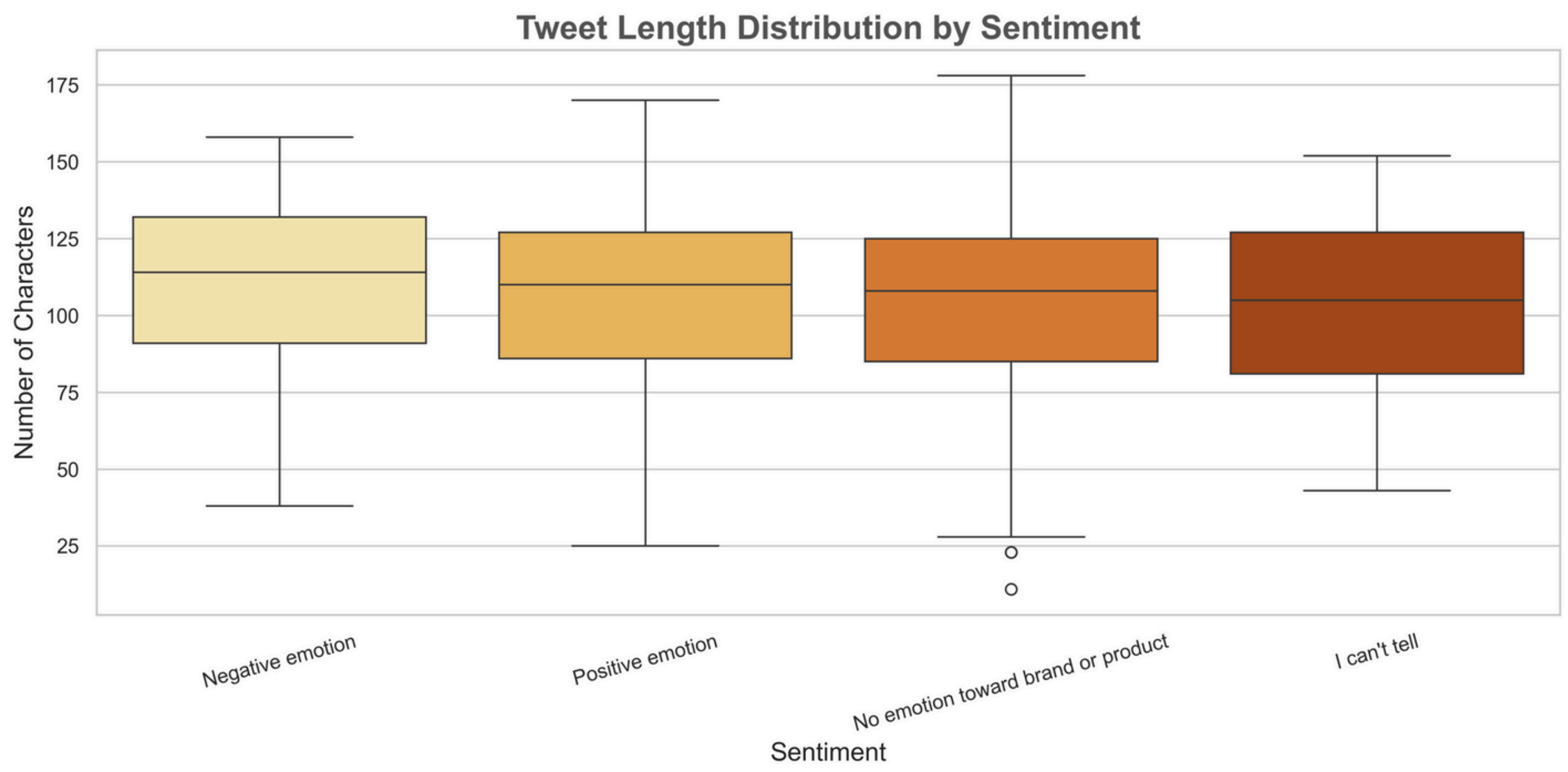
- Most tweets are neutral → over half contain no clear emotion .
- Positive tweets are significant → showing excitement and approval, especially for Apple products.
- Negative tweets are fewer → but still valuable for understanding pain points.
- Very few ambiguous tweets → hard to classify tone.

👉 Takeaway: Most online chatter is neutral, but positive mentions dominate over negatives — a good signal for brand perception



KEY INSIGHTS FROM EDA

- This chart shows the distribution of tweet lengths by sentiment.
- Negative emotion tweets are generally longer, with a higher median length.
- Positive emotion tweets tend to be slightly shorter than negative ones.
- Neutral/no emotion tweets and uncertain sentiment (“I can’t tell”) are the shortest on average.
- 👉 Key takeaway: People expressing strong emotions—especially negative ones—tend to write longer tweets.



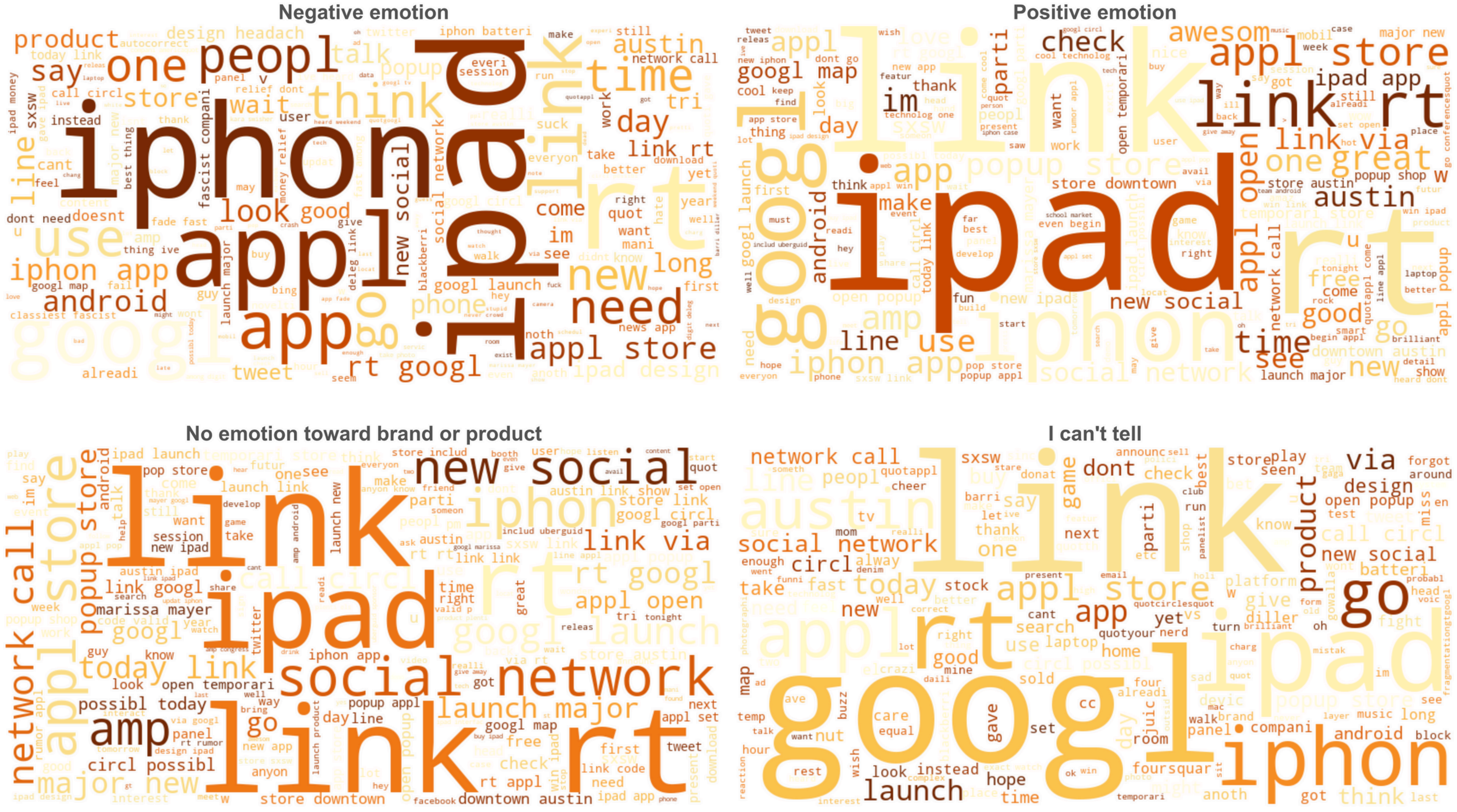
KEY INSIGHTS (WORDCLOUD)

Neutral (No Emotion)
- Keywords: link, iPad, iPhone, app, Google
- Mostly factual mentions (events, launches, links).

Positive Emotion
- Keywords: great, good, free, iPad, iPhone
- Excitement around Apple products (esp. at SXSW).

Negative Emotion
- Keywords: suck, long, Google, iPhone
- Complaints about usability, design, or issues.

Ambiguous
- Keywords: think, new, check, app, Austin
- Mixed tone, hard to classify.



MODELING APPROACH (BINARY CLASSIFICATION)

- We tested several models to classify tweets as Positive or Negative.
- Models included: Logistic Regression, Naïve Bayes, Random Forest, SVM, Extra Trees, AdaBoost, XGBoost, and BERT.
- Tweets were converted into numerical features using TF-IDF (text weighting method)

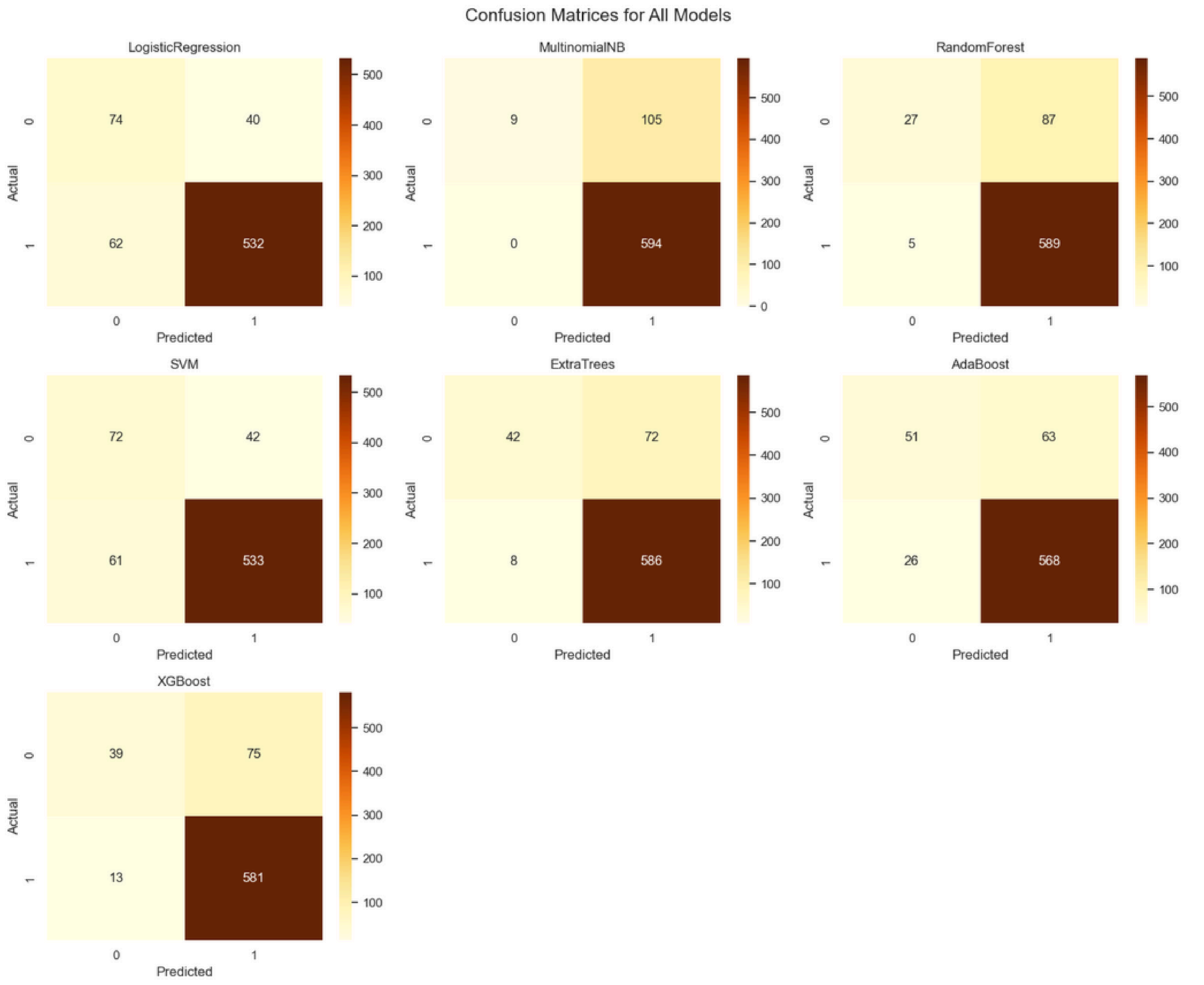
MODEL PERFORMANCE

Model Performance Metrics & Confusion Components

	Accuracy	Precision_Pos	Recall_Pos	F1_Pos	F1_Macro	ROC_AUC	TP	FP	TN	FN
gisticRegression	0.86	0.93	0.90	0.91	0.75	0.88	532.00	40.00	74.00	62.00
MultinomialNB	0.85	0.85	1.00	0.92	0.53	0.81	594.00	105.00	9.00	0.00
RandomForest	0.87	0.87	0.99	0.93	0.65	0.89	589.00	87.00	27.00	5.00
SVM	0.85	0.93	0.90	0.91	0.75	0.87	533.00	42.00	72.00	61.00
ExtraTrees	0.89	0.89	0.99	0.94	0.72	0.90	586.00	72.00	42.00	8.00
AdaBoost	0.87	0.90	0.96	0.93	0.73	0.82	568.00	63.00	51.00	26.00
XGBoost	0.88	0.89	0.98	0.93	0.70	0.86	581.00	75.00	39.00	13.00

CONFUSION MATRICES

- Logistic Regression → Safest, very accurate for positive tweets, but misses some negatives.
- MultinomialNB → Catches all positives, but too many false alarms (labels many negatives as positive).
- Random Forest & SVM → Decent, but still miss some tweets.
- Extra Trees & XGBoost → Best overall balance — accurate and consistent across sentiments.
- AdaBoost → Fair, but not as strong as Extra Trees or XGBoost.



MODEL HIGHLIGHTS

- Logistic Regression → Very good at being correct when it says “negative”, but it missed many actual negative tweets.
- Naïve Bayes → Caught every negative tweet, but also wrongly flagged a lot of positives as negative.
- Random Forest → Good at finding negatives, but often overreacted and flagged too many tweets as negative.
- SVM → Gave a nice balance overall, but still missed some negatives.
- Extra Trees → Showed the best balance between accuracy and reliability.
- AdaBoost → Decent and consistent, but not as strong as the top models.
- XGBoost → Excellent mix of accuracy and reliability, with only a few mistakes.
- Best overall balance came from Extra Trees and XGBoost.

BERT MODEL EVALUATION

- Overall Accuracy: 89.6% (very strong)
- Majority class (positive tweets): Precision 91.5%, Recall 96.5%
- Minority class (negative tweets): Precision 74.4%, Recall 53.5%
- Macro F1: 78.1% (treats classes equally)
- Weighted F1: 88.8% (accounts for imbalance)
- ROC-AUC: 0.898 (excellent separability)
- Most powerful model with Very strong results but struggled a bit with minority negative tweets.

MULTI-CLASS CLASSIFICATION

- In multiclass sentiment classification, the task is to distinguish between Negative, Neutral, and Positive sentiments. Compared to binary classification, this adds complexity as models often confuse similar classes (especially Positive and Neutral). The following slides show how different models perform on this task

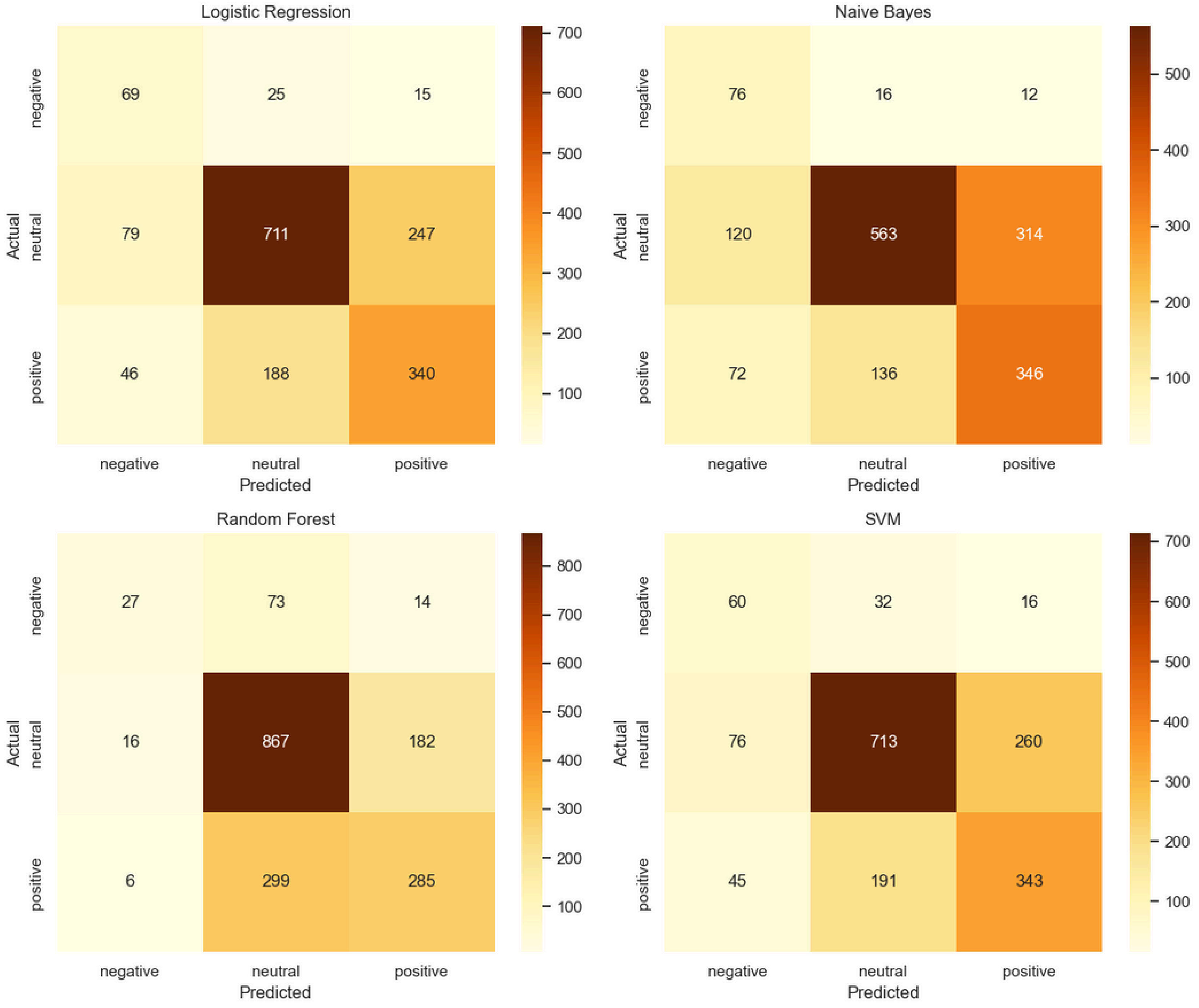
.

CONFUSION MATRICES- MULTICLASS

- Neutral class is easiest → all models perform strongest here.
- Positive vs Neutral confusion → main weakness across all models.
- Negative class struggles → likely due to class imbalance (few samples).

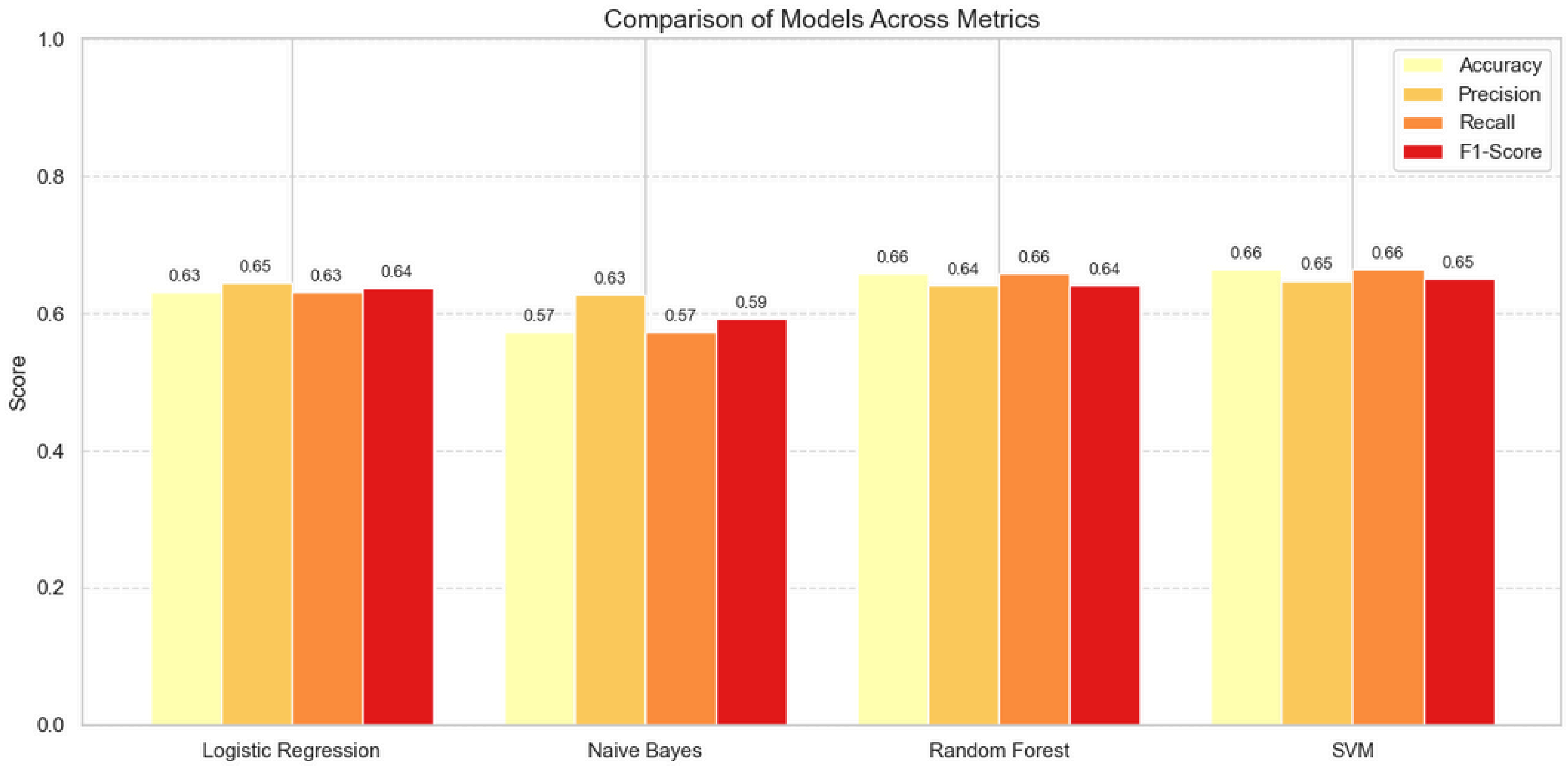
Model-specific:

- Logistic Regression: Good on neutral, confuses positive ↔ neutral.
- Naive Bayes: More errors, many neutral misclassified as positive.
- Random Forest: Best on neutral, weak on negative, biased toward neutral.
- SVM: Similar to Logistic Regression, balanced but still confuses positive ↔ neutral.



MODEL METRICS - MULTICLASS

- Model Comparison (Accuracy, Precision, Recall, F1)
- Random Forest & SVM achieve highest scores (~0.66 across metrics).
- Logistic Regression: Stable mid-range performer (~0.63–0.65).
- Naive Bayes: Lowest performance (F1 ≈ 0.59).
- Best balance: Logistic Regression & SVM.



BERT MODEL PERFORMANCE - MULTICLASS

- Overall Accuracy: ~70% (weighted F1 ≈ 0.70).
- Strength: Neutral sentiment detection (F1 = 0.77).
- Weakness: Negative detection (F1 = 0.44).
- Positive class: Moderate (F1 = 0.61).
- Training/Validation: Stable → no overfitting.

.

FINDINGS

Best Traditional Model: Random Forest

- Fast predictions, interpretable outputs
- Quick deployment
- Feature importance for interpretability
- Low compute cost

Best Deep Learning Model: BERT

- Higher accuracy, nuanced predictions
- Contextual understanding in production
- Tolerates false positives in neutral class
- Can still be improved

KEY TAKEAWAYS

- Neutral sentiment is well-detected by all models.
- Positive ↔ Neutral misclassification is the main limitation.
- Negative class under-represented → hurts recall.

- **Recommended next steps:**
- Address class imbalance (data augmentation, reweighting).
- Explore advanced models beyond baseline BERT.
- Deploy BERT as a strong starting point, refine iteratively.

CONCLUSION

- Multiclass sentiment classification (Negative, Neutral, Positive) is more challenging than binary due to class overlap.
- Neutral sentiment is detected most reliably across models.
- Positive vs. Neutral confusion remains the main weakness.
- Negative sentiment is hardest to capture → likely due to fewer training examples.
- BERT baseline model (~70% accuracy) provides a strong foundation for deployment and future improvements.

THIS PROJECT LAYS A SOLID FOUNDATION FOR UNDERSTANDING CUSTOMER SENTIMENT AND GUIDING SMARTER BUSINESS DECISIONS.

QNA

THANK YOU