SENTIMENT ANALYSIS FOR WALMART REVIEWS

- Team Members:Gideon Ochieng,Ann Mwangi,Victor Masinde,Lorna Gatimu,Charles Odhiambo
- Technical Mentor: Maryann Mwikali

Introduction

BUSINESS UNDERSTANDING

Problem Statement:

- Manual extraction of insights from thousands of reviews is challenging.
- Traditional rating systems may not reflect true sentiment.

Solution:

- Sentiment analysis provides a more accurate understanding of customer feedback.

Stakeholders:

- E-commerce Businesses & Product Sellers
- Consumers & Online Shoppers
- Marketing & Customer Support Teams

PROJECT OVERVIEW

- Online platforms like Walmart generate millions of customer reviews daily.
- Reviews influence purchasing decisions and shape brand reputations.

Goal:

- Use sentiment analysis to interpret customer emotions from text data.
- Automate the process to enhance customer experience, improve products, and drive sales.

PROJECT OBJECTIVES

1. **Develop Sentiment Analysis Model:** Automated system to classify Walmart reviews (positive, negative, neutral).
2. **Enhance Customer Insights:** Extract key trends to understand consumer preferences and satisfaction.
3. **Improve Product & Service Quality:** Identify patterns in negative reviews for improvement.
4. **Automate Review Analysis:** Efficiently process thousands of reviews to reduce manual effort.
5. **Refine Strategies:** Track sentiment trends to improve marketing and customer support.

DATA UNDERSTANDING

- **Data Collection:** Merged customer reviews from multiple products into a unified dataset.
- **Dataset Overview:** Included columns like product names, review texts, ratings, dates, and average ratings.
- **Initial Insights:** Ratings ranged from 1 to 5 stars, with an average rating of 4.24, indicating overall customer satisfaction.
- **Correlation Analysis:** Positive correlations between product ratings and average ratings suggested trends in customer sentiment.
- **Dataset Dimensions:** The dataset consisted of 7095 rows and 6 columns, providing a solid foundation for comprehensive analysis.

DATA CLEANING

- Prepare the dataset by addressing missing values, duplicates, and inconsistencies.
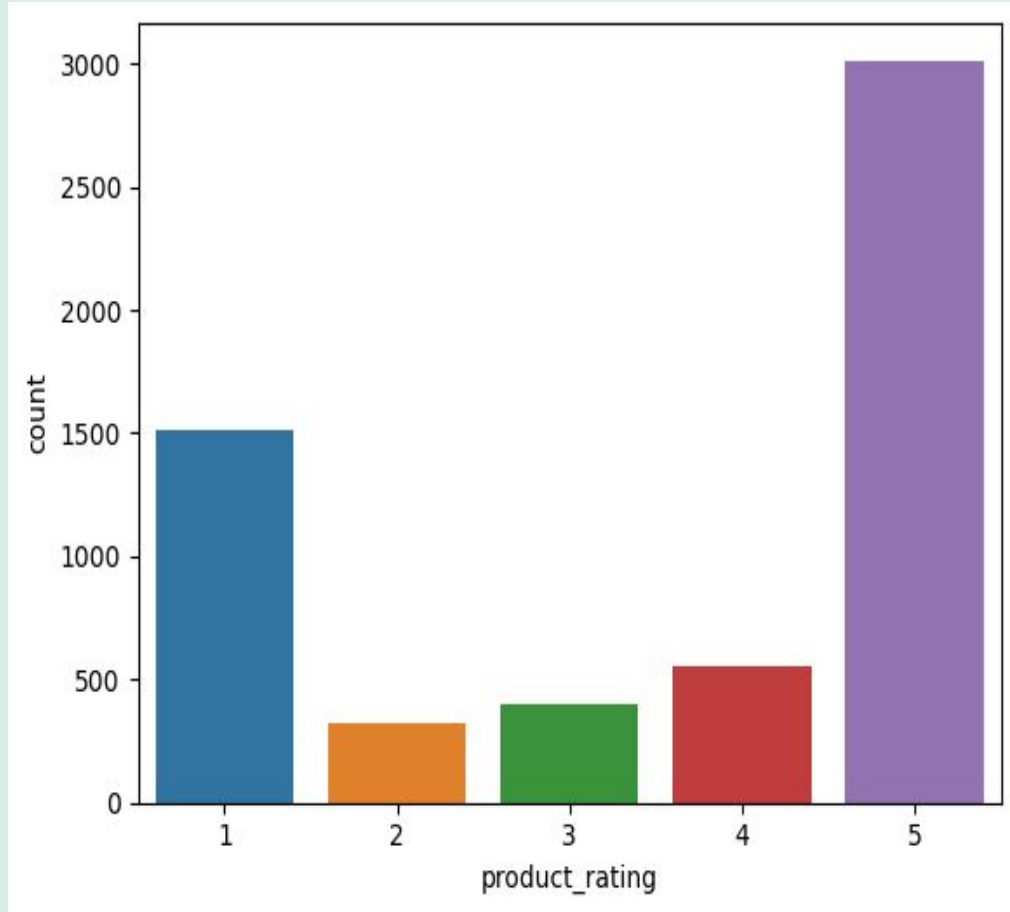
Key Steps:

- **Handling Missing Values:** Addressed missing values in critical columns like `review_text`.
- **Removing Duplicates:** Eliminated duplicate entries to ensure data quality.
- **Ensuring Consistency:** Standardized formats and cleaned inconsistencies in columns like `review_date`.
- **Text Preprocessing:** Applied techniques such as removing stopwords, lemmatization, and cleaning special characters in `review_text`.

EXPLORATORY DATA ANALYSIS

DISTRIBUTION OF PRODUCT RATINGS

● The graph highlights a polarized customer satisfaction, with the majority of ratings being either very high (5) or very low (1)

FREQUENCY OF WORDS

- This word cloud shows which words appear most often in a set of text. Larger words were mentioned more frequently. For example, "walmart," "love," "shoe," and "bottle" are some of the biggest words, indicating they are very common in the text data.

TRENDS OF PRODUCT RATINGS

- The number of reviews for all ratings is low and stable from 2020 to 2022.
- In 2023, there's an increase in reviews for all ratings, especially for Rating 5.
- 2024 sees a peak in reviews, with Rating 5 reaching the highest point at around 1750 reviews.
- By 2025, the number of reviews decreases but remains higher than the initial years.

Trends of Each Rating Over the Years
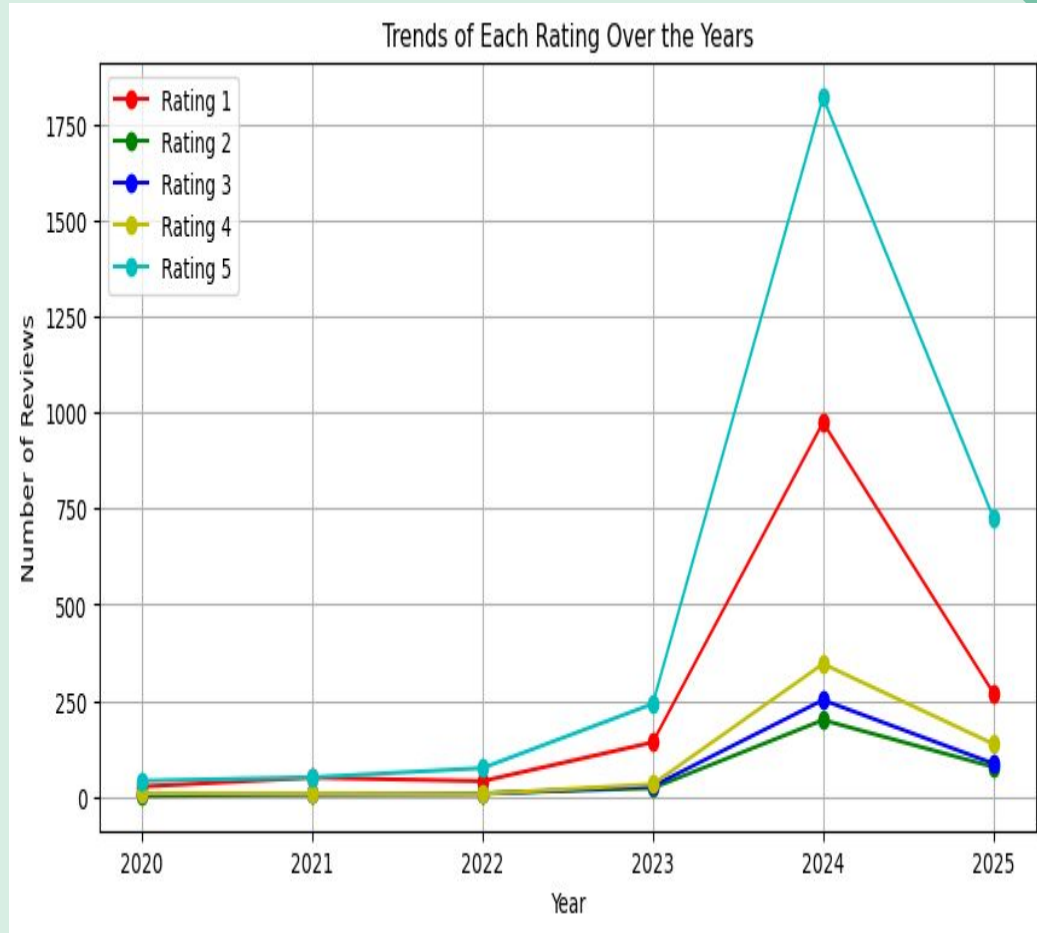
# MODELLING

SENTIMENT DISTRIBUTIONS

Sentiment Analysis Results for Each Product:

- We conducted sentiment analysis on reviews for 9 different products, classifying them as Positive, Neutral, or Negative.
- For each product, we counted the number of reviews that fell into each sentiment category.
- Example: For "sentiment_computer_reviews.csv," we found 215 Positive, 45 Neutral, and 106 Negative reviews.

**Overall Sentiment Distribution:**

- We summed up the sentiment counts across all products to understand the overall customer sentiment.
- The majority of the reviews were Positive, followed by Negative and Neutral.
- This gives us a holistic view of customer sentiment across all products.

Overall Sentiment Distribution of Walmart Reviews

Negative 20.6%

Neutral 8.1%
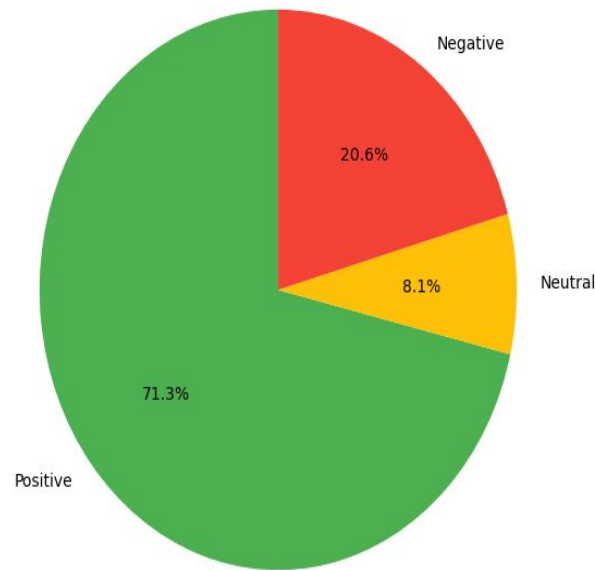
Positive 71.3%

Sentiment Distribution Comparison by Product:

- We compared the sentiment distribution for each product side-by-side.
- This comparison helps identify which products have higher proportions of Positive or Negative reviews.
- It highlights products that may need attention due to a higher number of Negative reviews.

Sentiment Distribution by Product (Grouped Bar Chart)
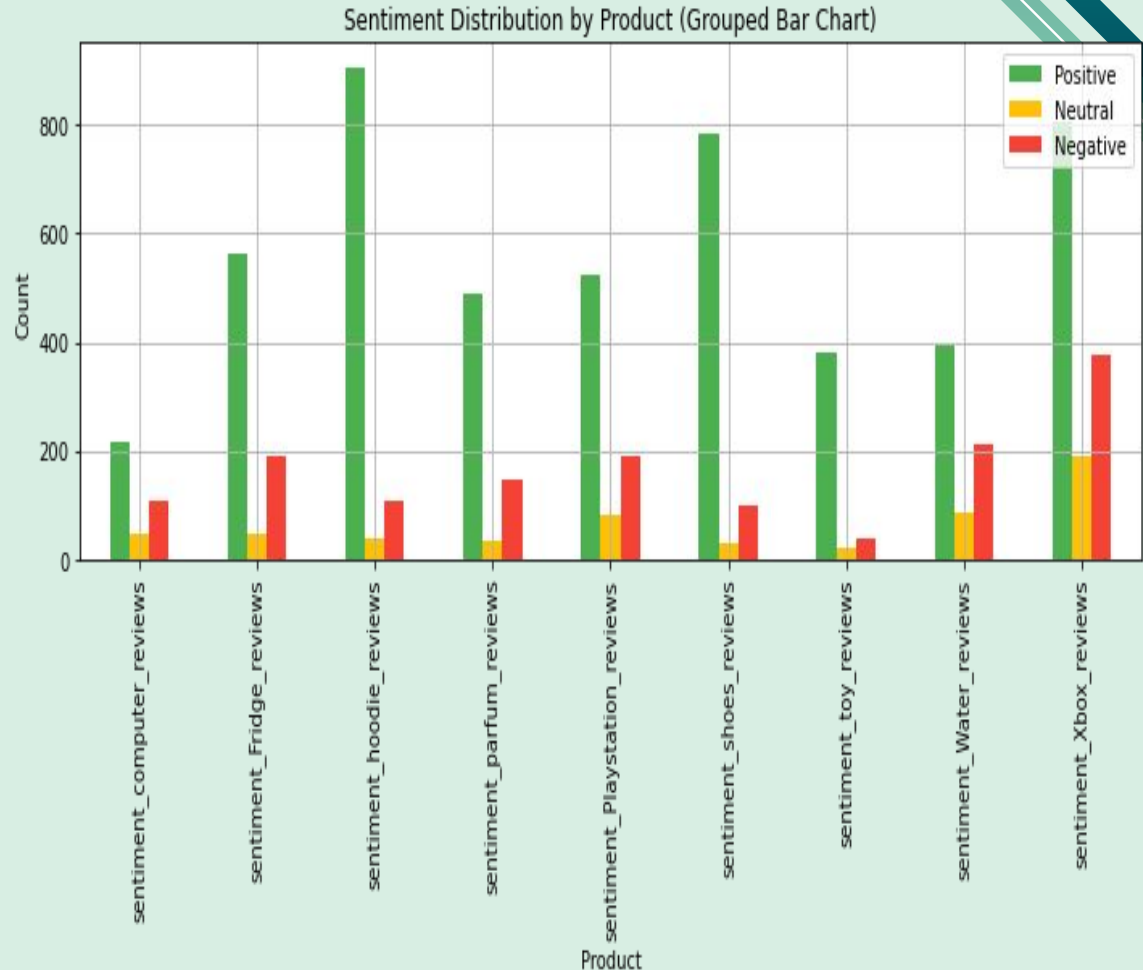
RECOMMENDATIONS

- Implement an automated review monitoring system to flag negative reviews for prompt intervention.
- Provide personalized responses to dissatisfied customers to rebuild trust and enhance brand reputation.
- Continuously update the sentiment lexicon to capture evolving customer language trends.
- Address common negative feedback themes to improve product quality and service delivery.
- Leverage positive sentiment insights to reinforce well-received features and maintain customer satisfaction.

CONCLUSION

- Sentiment analysis reveals valuable insights from Walmart reviews. Using the VADER model, we classified sentiments to help businesses improve satisfaction and strategies.

Future steps:

- Integrate advanced models (e.g., BERT) for deeper analysis.
- Expand datasets and refine techniques for better accuracy.
- Automated sentiment analysis empowers data-driven decisions, enhancing customer engagement and growth.

THANK YOU!!

Name	Github
Victor Masinde	@Masinde10
Gideon Ochieng	@OchiengGideon
Ann Mwangi	@ann-mw
Lorna Wangui	@lorna-creator
Charles Odhiambo	@T-hoveen