Project Title:
Which Headline Works Better? A/B Testing with Real API Data

Name:
Toluwalase Taiwo

Date:
14th June, 2025

1. Introduction

Picture this.
You're scrolling through the news after a long day, and two headlines catch your eye:

"Hope Rises as Community Comes Together After Flood"

vs.

"Flood Leaves Thousands Homeless: Chaos in the Streets"

Be honest. Which one are you more likely to click? And why?

That's the heart of what I set out to explore in this project.

Headlines aren't just short summaries. They're emotional hooks. They influence how we feel, what we read, and what we share. But here's the big question: Which works better, hope or crisis?

To dig into this, I pulled real-time articles from The Guardian's News API and used **TextBlob**, a simple NLP tool, to simulate polarity scores and analyze the sentiment in each headline and summary. Then I grouped them into positive and negative and simulated user engagement by assigning clicks and shares with numpy Ufuncs.

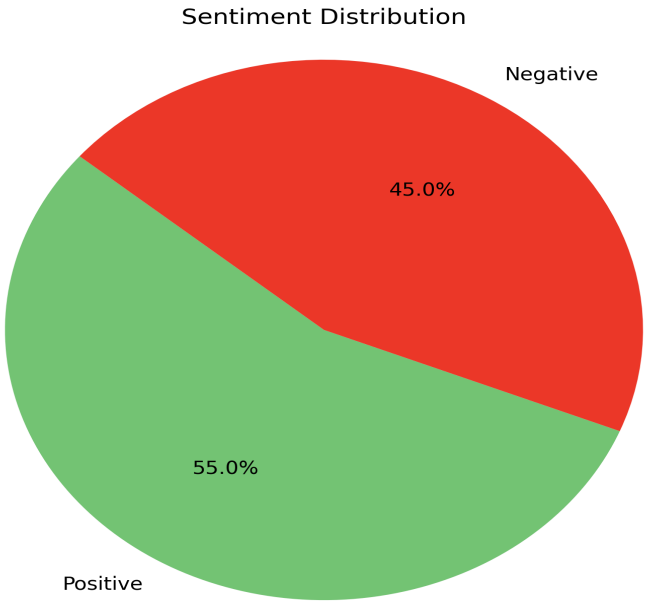
- **Objective**: **To test whether positive or negative sentiments drive more engagement (clicks/shares)**

2. Data Acquisition

- **Source:** The Guardian API
- **Data Fields:** Headline, trailText (summary), published_date, section name(category), full_text(body of the text)
- **Data pulled:** I converted the data pulled into a csv file. Access it here
  Note: Used request to extract and pandas to store/clean

3. A/B Test Simulation

- Grouped articles by Sentiment (Positive or Negative) using TextBlob
- Sentiment calculated based on polarity from headline + summary
  Note: Polarity score in sentiment analysis is a numerical score representing the overall sentiment present in a text. It ranges from -1 to +1. With -1 being very negative and +1 being very positive.

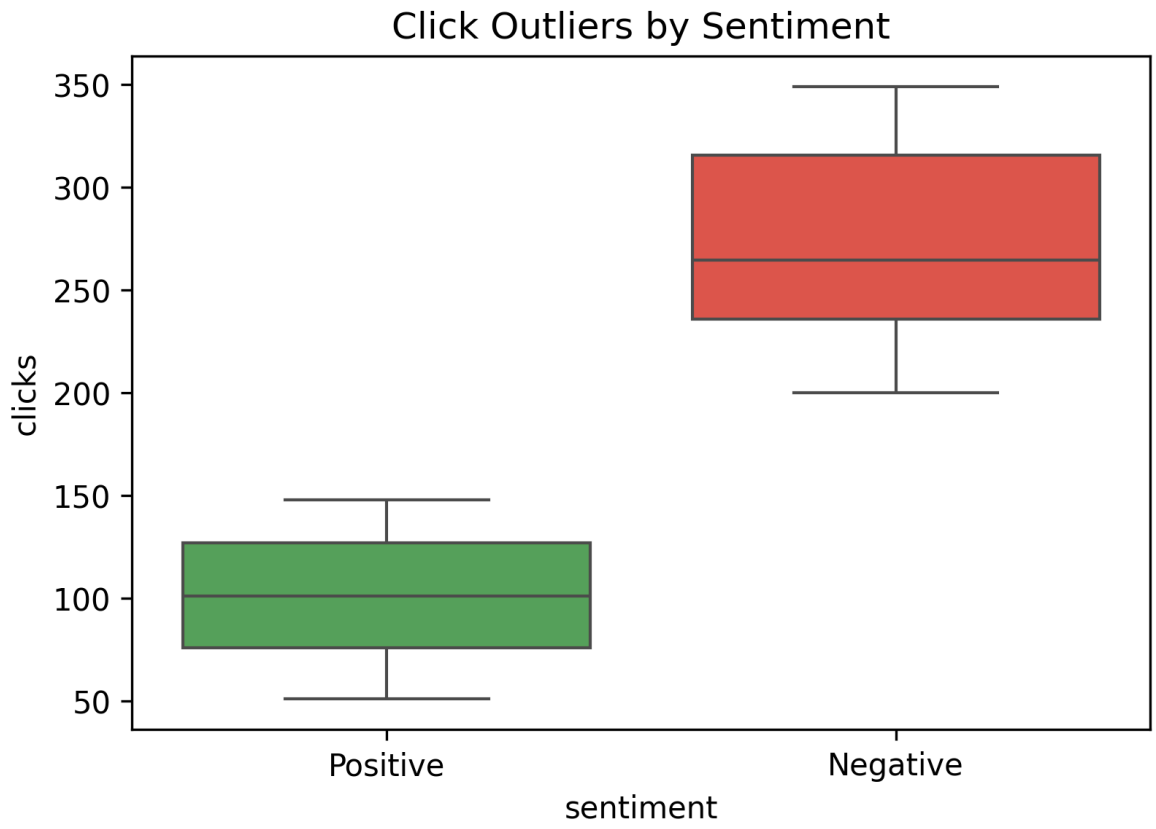


- Removed neutral articles to simplify the A/B test.

- **Why this method:** I chose sentiment-based A/B testing because I wanted to explore the clear emotional distinction between articles. As someone who understands how readers connect with words, it felt more meaningful to group headlines by how they make people feel. This approach provided a more accurate and human-centered analysis of what truly drives clicks and shares.

4. Data Cleaning & Exploration

- Checked for missing values and duplicates and outliers— none found in key fields.
- Explored data types and changed columns like published_date to datetime.
- Explored text length:
    - Headline length
    - Summary length



5. Data Analysis and Visualization

Simulated Engagement Metrics
Simulated Clicks and Shares based on Sentiment:

- **Negative articles → higher values**

Clicks: Random values between 200 and 350
Shares: Random values between 150 and 300

- **Positive articles → lower value**

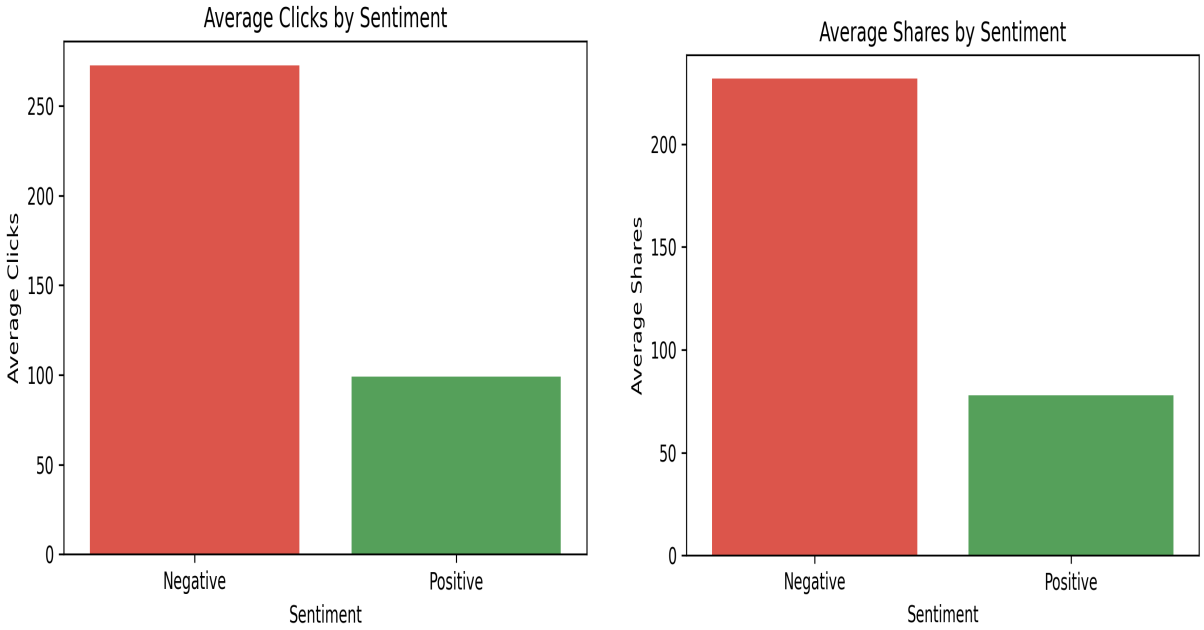
Clicks: Random values between 50 and 150
Shares: Random values between 40 and 120

- Used numpy.random.randint() for variability

Why this method:
Real click and share data weren't available through the API, so I simulated these metrics to explore possible engagement patterns. By assigning higher values to negative news, I mirrored real-world trends showing that dramatic or alarming content often gets more attention. This allowed for a realistic A/B test despite data limitations.

Grouping:

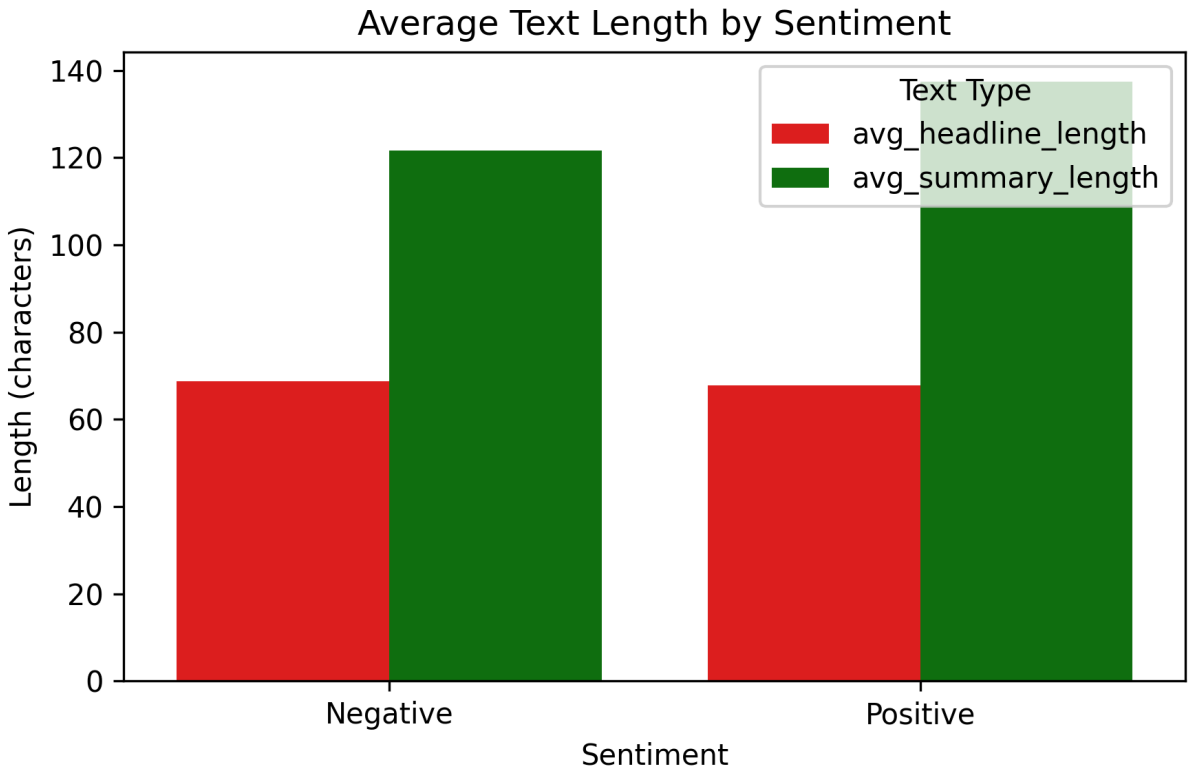
- I grouped the articles by sentiment (Positive/Negative) using `groupby()` and `agg()`.This helped me analyze engagement metrics such as average clicks and shares based on sentiment.



Interpretation of the visual:
*Negative sentiment articles had over **250 average clicks** and over **130 average shares**, while positive sentiment articles recorded about **100 average clicks** and around **75 average shares**.*

*This **suggests** that **negative articles received significantly more engagement**, indicating that readers may be more drawn to **urgent, dramatic, or emotionally intense headlines** than to uplifting ones.*

- I compared average text length (headline_length and summary_length across sentiment groups.



Interpretation of the visual:

Headline Length: *Both sentiment groups had similar average headline lengths (~70 characters).*

Summary Length: *Positive articles had longer summaries (~140 characters) than negative ones (~120 characters).*

Insight: *Positive news may need more context to convey meaning, while negative news delivers impact more concisely.*

6. A/B Testing Results

To compare how Positive and Negative articles performed in terms of *clicks* and *shares*, I used a **T-test** — a common statistical method to compare the average difference between two groups.

What's a T-test? It checks whether the difference between two group averages (e.g., average clicks on positive vs. negative headlines) is real and statistically significant — or just happened by chance.

What's a p-value? The p-value tells us how likely it is that the difference is just random noise.

- If the p-value is very small (like less than 0.05), the result is considered statistically significant — meaning the difference is likely *real*.

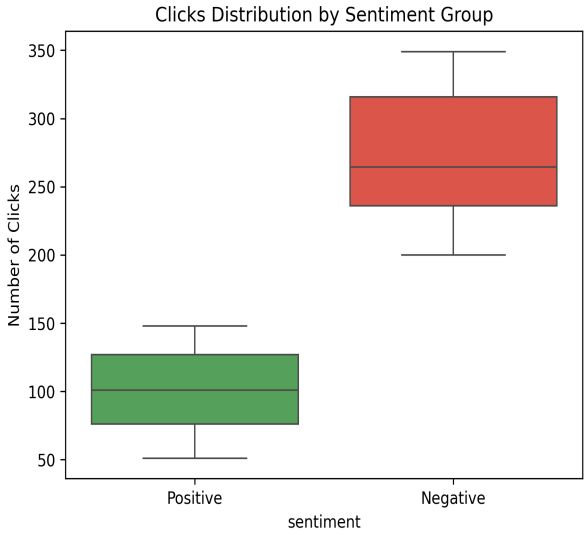
Results

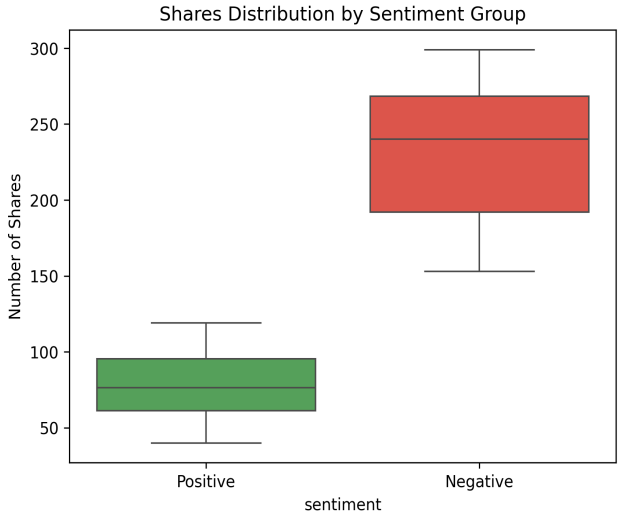
- **Clicks:**
    - T-statistic = -31.74, p-value ≈ 0.000... (e-69)
- **Shares:**
    - T-statistic = -30.05, p-value ≈ 0.000... (e-57)

These values show a very strong difference between how *positive* and *negative* articles perform **— the low p-values mean this is not random**.

Interpretation

- *A very small p-value (like e-69 or e-57) means the result is statistically significant. There is a real difference in engagement between positive and negative articles.*
- *Negative articles received much more engagement — higher average clicks and shares.*
- *This supports the idea that emotionally intense or urgent headlines attract more attention than uplifting ones.*





Boxplot Interpretation

The boxplots visually compare engagement (clicks/shares) across sentiment groups:

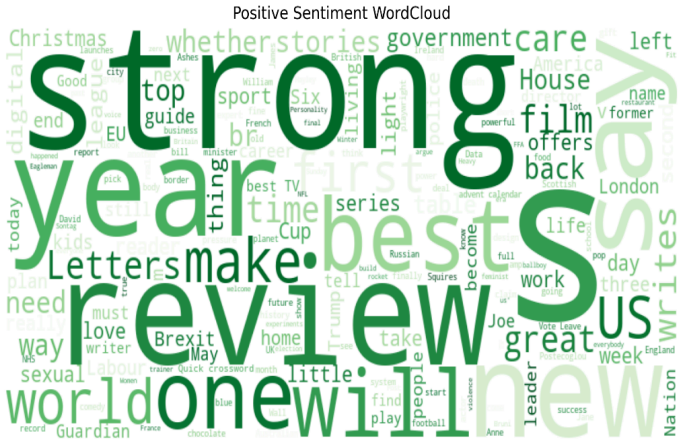
- *The **wider spread** and **higher median** in the Negative group show greater and more consistent engagement.*

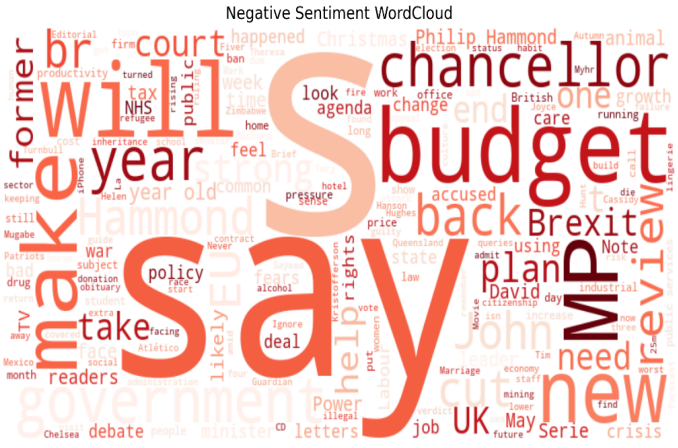
- *You might also notice that there are no outliers, that is, individual articles that performed unusually well or poorly. These would have been shown as **points outside the box or whiskers as I prefer to call them.***
  *In statistical terms, **outliers appear when some values are much higher or lower than the rest**, usually beyond 1.5 times the interquartile range (IQR).*
- *Since the engagement values were **simulated within defined ranges** (e.g., 50–150 for positive clicks), the data lacks extreme variations — **so no outliers are detected.***

7. WordCloud Analysis

A **word cloud** is a visual representation of text data where the size of each word reflects how frequently it appears. The more often a word shows up in the dataset, the bigger and bolder it appears in the cloud.

- Generated separate word clouds for:
    - Positive sentiment articles
    - Negative sentiment articles





In this project, I used word clouds to:

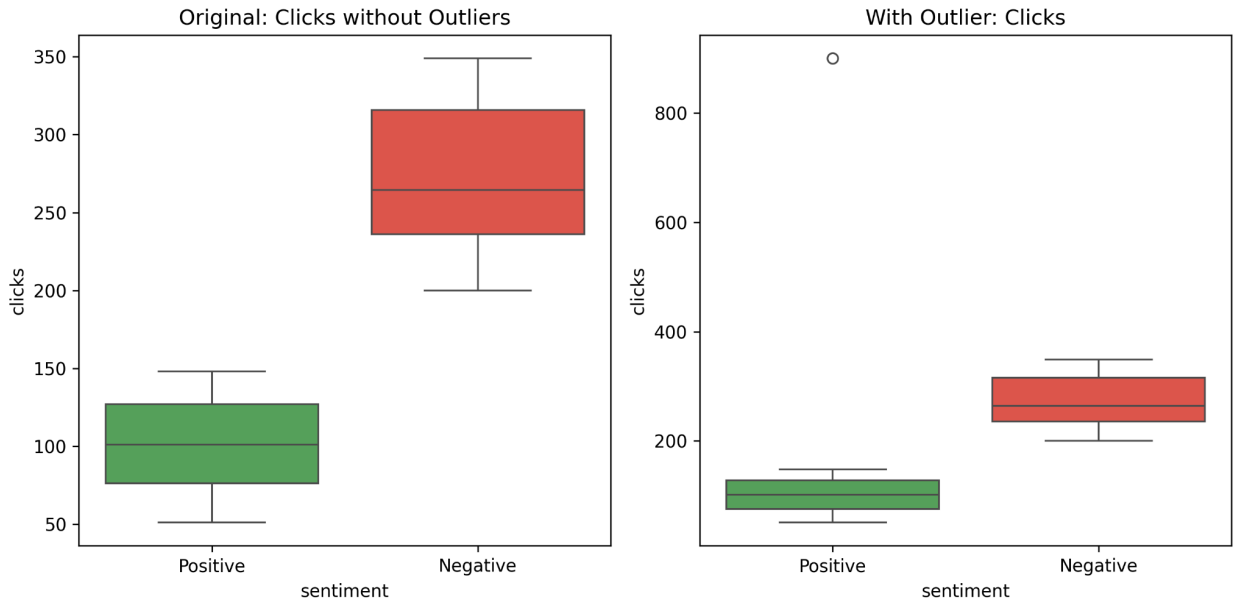
- **Quickly visualize the common language** used in positive vs. negative articles.

- **Highlight emotional tone**: positive articles leaned toward hopeful or uplifting words, while negative ones showed terms related to conflict or urgency.

- Make the sentiment groups **more relatable** by showing the **vibe** of each group at a glance.

8. Outliers & Observations

To illustrate the impact of outliers, I injected extreme value (900) into a copy of the guardian news dataset. This helped demonstrate how even a few unusually high-performing articles could distort averages, emphasizing the value of using median or visual tools like boxplots for clearer insights.



Interpretation

After injecting outliers into the **Positive sentiment group only**, here's what happened:

- ***Clicks (Positive)** average increased from **99.15** to **106.41***

- ***Shares (Positive) average** increased from **77.95** to **84.57***

- ***Negative group** stayed the same because no outliers were added there.*

After introducing outliers, the average clicks and shares for Positive articles increased significantly, while Negative sentiment remained largely unchanged. This highlights how outliers can skew engagement metrics, making visual tools and median values more reliable in some analyses. As we see that the median remained unchanged in the box plot.

9. Key Insights & Recommendations

Which Sentiment Performed Better?

- Articles with **negative sentiment** had higher average clicks and shares than positive ones.
- Negative: around **272 clicks** and **231 shares**
- Positive: around **99 clicks** and **78 shares**

What This Means for the Guardian News Platform

- Negative headlines may get more attention.
- But it's important to balance — using only a negative tone might affect how people feel about the platform.

  If the goal is reach, headlines with negative tone might attract people

  If the goal is brand impact or perception, it's best there should be a balance with positive headlines.
- Testing sentiment can help the team make better decisions when publishing stories.

Takeaway for Everyone

- This project shows that **what we feel when we read a headline** can predict how likely we are to click or share it.
- People seem to react more to headlines that feel urgent or emotional.

10. Appendix

- Link to Jupyter Notebook: Click Here

- Personal Reflection:

Working on this project helped me understand how to apply sentiment analysis to real-time data using Python. I also learned how to simulate and test user engagement using basic statistics like t-tests, and how to communicate findings clearly using visualizations. It was a practical and insightful experience that made data feel more human.

Summary of Tools Used

- **Python**: Main programming language for data extraction, analysis, and simulation.
- **Pandas**: Used for data cleaning, manipulation, and grouping.
- **NumPy**: Helped with simulating numeric data like clicks and shares.
- **TextBlob:** Used for sentiment analysis to classify articles as positive or negative.
- **Matplotlib & Seaborn**: For creating clear and engaging visualizations (bar plots, box plots, etc.).
- **WordCloud:** To visualize commonly used words in positive vs. negative headlines.
- **Scipy (ttest_ind):** Used to compare group performance and test for statistical significance.