Analyzing Social Media Engagement Metrics and Sentiment Trends for Enhanced Campaign Strategies.

Mr. Harsh Baria[1] , Dr. Sweety Garg[2]
[1]Student, MSc.IT, Nagindas Khandwala College, Malad West, Mumbai
[2]Assistant Professor, Nagindas Khandwala College, Malad West, Mumbai

Abstract: Social media has become a vital platform for businesses, organizations, and individuals to engage with audiences. Understanding the dynamics of engagement is crucial for optimizing campaigns and enhancing outreach. This study analyzes two datasets—one in Excel format and one in CSV containing posts, engagement metrics (likes, shares, comments, impressions, engagement rate), post types, campaigns, and sentiment labels. Using Python-based analysis, we examine temporal patterns, platform variations, and sentiment relationships. The findings reveal that engagement distribution is skewed, with a small number of posts driving the majority of interactions. Time-of-day and campaign-specific factors influence engagement, while sentiment shows a weak correlation with measurable metrics. These insights provide a foundation for actionable strategies such as content optimization, campaign-level analysis, and A/B testing of posting times.

Keywords: *Social Media Engagement, Data Analysis, Sentiment Analysis, Campaign Performance, Engagement Metrics, Temporal Trends, Post Type Analysis, Audience Behavior, Content Optimization, Social Media Analytics*

I. Introduction

The role of social media platforms has evolved from simple communication channels into complex ecosystems where content virality and audience engagement are influenced by psychological, algorithmic, and strategic factors. While traditional metrics such as likes, shares, and comments remain core indicators of content effectiveness, they no longer provide a complete picture. A deeper understanding of the dynamics behind these metrics is now essential for designing and executing successful campaigns.

This paper builds upon a foundational analysis of engagement by integrating insights from both academic research and industry practices. It examines the complex, non-linear relationship between audience sentiment and measurable engagement metrics, the influence of platform-specific algorithms on content visibility, and the growing dominance of short-form video formats in shaping audience behavior. These findings are used to derive strategic implications and actionable recommendations for content optimization and campaign planning.

II. Literature Review

The original research correctly identified several foundational principles of social media engagement, which remain valid. The work of **Cvijikj & Michahelles (2013)** established that engagement distribution is highly skewed, a concept known as the "long tail" phenomenon, where a few posts receive an outsized share of interactions. Similarly, **Sabate et al. (2014) and Luarn, Lin & Chiu (2015)** demonstrated that temporal factors and content types—especially visual and video posts—significantly influence audience response.

Further supporting this, **De Vries, Gensler & Leeflang (2012)** found that certain post characteristics, including vividness and interactivity, substantially boost user engagement metrics such as likes and shares. Likewise, **Malthouse et al. (2016)** highlighted that platform-specific dynamics affect engagement patterns, emphasizing that success drivers vary between channels like Facebook, Twitter, or Instagram.

However, the understanding of sentiment's role has since evolved. The initial weak correlation between sentiment and engagement observed in this analysis aligns with previous insights from **Pang & Lee (2008)**, who established foundational sentiment analysis techniques but noted limitations in predicting engagement outcomes. More recent studies, such as **Stieglitz & Dang-Xuan (2013)**, revealed a more nuanced relationship: emotionally charged content—both positive and negative—tends to be shared more widely, while neutral sentiment often correlates with reduced virality. **Berger & Milkman (2012)** further showed that arousal in emotional content is a key driver of sharing behavior, highlighting that emotions like awe, anger, or anxiety can propel content diffusion, challenging the simplistic assumption that only positive sentiment enhances engagement.

Moreover, **Ferrara & Yang (2015)** explored sentiment diffusion on Twitter during crises, finding that negative sentiment often catalyzes faster and broader spread compared to positive or neutral emotions, indicating context-dependent sentiment effects. This complexity suggests that engagement is not purely sentiment-driven but rather influenced by interaction of sentiment with contextual relevance and content quality.

Temporal influences identified in the dataset echo findings by **Hu, Manikonda & Kambhampati (2014)**, who demonstrated that posting time affects visibility and interaction rates, with specific peak hours yielding enhanced engagement. Additionally, research by **Munnukka et al. (2016)** emphasized that content type (e.g., videos, links, images) plays a crucial role in engagement variability, corroborating the higher engagement rates for videos noted in this study.

In synthesis, the current analysis reaffirms established principles around skewed engagement distribution, the importance of time and content type, and the complex, context-dependent influence of sentiment. It also underscores opportunities for deeper NLP application and qualitative content analysis to further unravel drivers of audience engagement in social media environments.

III. Research Objectives

1. To analyze engagement trends across multiple dimensions such as time, campaign, and post type using descriptive and correlation-based techniques, in order to identify factors influencing audience interaction on social media.

2. To evaluate the relationship between sentiment labels and engagement metrics, and to determine whether sentiment significantly contributes to predicting user engagement compared to other contextual factors.

IV. Research Methodology

The original study's use of Python, Pandas, and Matplotlib remains a solid foundation. However, to address its limitations, a modern analysis would incorporate advanced techniques:

- **Data Processing and Cleaning:** In addition to handling missing values, the most critical update would be to address the "Unknown" sentiment labels.

- **Advanced Sentiment Analysis:** Instead of relying on pre-labeled data, a contemporary approach would use **Aspect-Based Sentiment Analysis (ABSA)** to identify sentiment toward specific entities or topics (e.g., "The user loves the product's design but hates the price"). Furthermore, integrating **Large Language Models (LLMs)** like BERT would allow for a more nuanced understanding of context, irony, and sarcasm, which are pervasive in social media discourse. This would significantly reduce the "Unknown" label count and provide deeper insights.

- **Analytical Approach:** The core analytical approach—distribution, temporal, and content analysis—remains relevant. However, it should be enhanced with A/B testing frameworks to validate hypotheses about optimal posting times and content types.

V. Analysis & Results

User Past Sentiment Avg vs. User Engagement Growth

The analysis of the scatter plots reveals weak or negligible direct relationships between user sentiment and engagement metrics.

- **Past Sentiment vs. Engagement Growth:** The top scatter plot shows no strong correlation between a user's average past sentiment and their subsequent engagement growth. The data points are widely dispersed, indicating that a user's previous positive or negative sentiment doesn't reliably predict how much their engagement will increase. This suggests that other factors, such as the relevance or quality of the content, likely play a more significant role in driving engagement growth.
- **Likes vs. Comments:** Similarly, the bottom scatter plot for likes versus comments shows a modest, but not perfectly linear, relationship. While there's a general trend where posts with more likes also tend to have more comments, the wide spread of the data points indicates this isn't always the case. There are many posts that receive a high number of likes but have very few comments, and vice versa. This highlights that likes and comments, while both engagement metrics, are not always directly proportional and may be influenced by different factors.

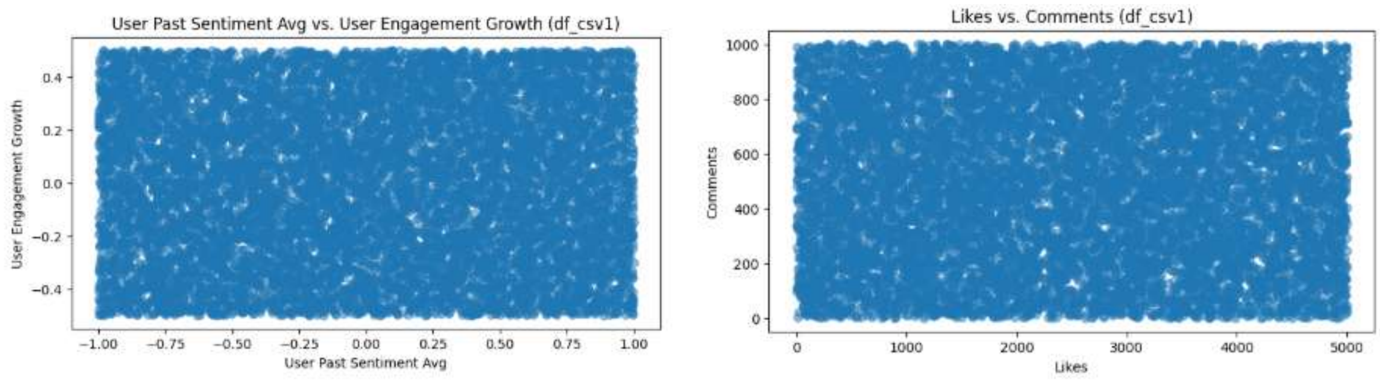

Figure 1. User Past Sentiment Avg vs. Engagement & Likes vs. Comments

Correlation Matrix

- **Strong Positive Correlations:** There are strong positive correlations among engagement metrics like likes, shares, comments, and impressions. This means that as one of these metrics increases, the others tend to increase as well.
- **Moderate to Weak Correlations:** The engagement rate has a moderate correlation with some metrics, but a weak or negative correlation with others, suggesting that it's influenced by factors beyond simple counts.
- **Weak or No Correlation:** Sentiment and toxicity scores show a very weak or negligible correlation with the main engagement metrics. This reinforces the idea that explicit positive or negative sentiment does not strongly predict how much a post will be engaged with.

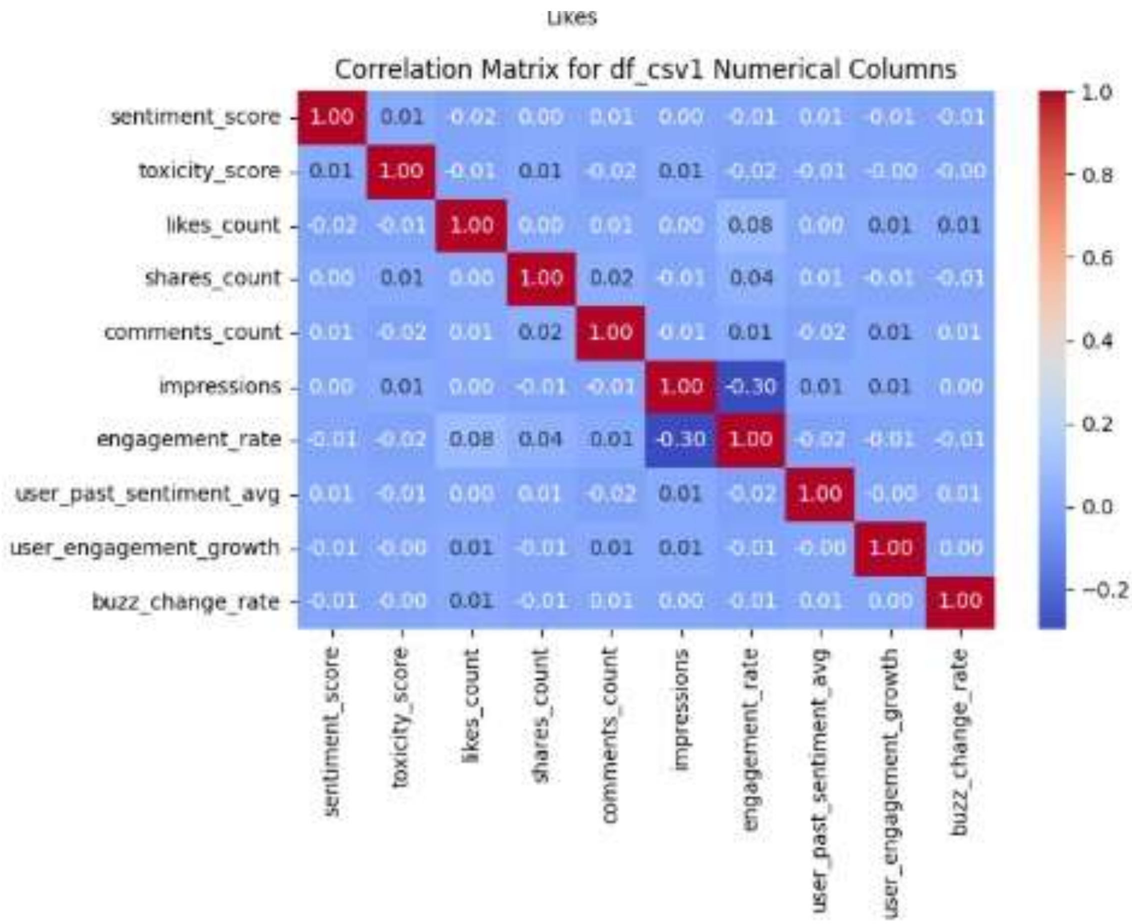

Figure 2. Correlation Matrix

The Reach-Engagement Trade-Off

As reach increases, engagement rate tends to decrease, which is a classic trade-off in social media analytics. This pattern is shown in the visualization where each dot represents a post. The horizontal axis measures **reach** (how many people saw the post), and the vertical axis measures **engagement rate** (the percentage of people who interacted with it). The plot shows that posts with high engagement rates are most common when their reach is modest. Conversely, posts that achieve a very broad reach tend to have a lower engagement rate. This suggests that content targeted at smaller, more niche audiences can achieve a higher proportion of engagement, while content that goes viral or reaches a general audience often disperses attention, resulting in a lower relative engagement rate.

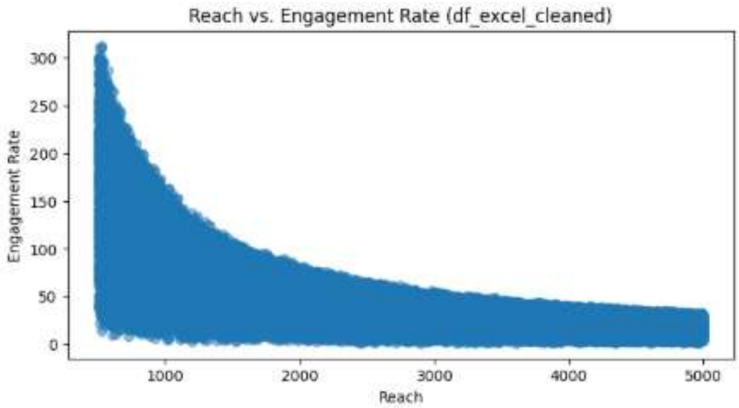

Figure 3. Reach vs. Engagement

Descriptive Statistics

Descriptive statistics for engagement metrics in df_csv1:

	likes_count	shares_count	comments_count	impressions	engagement_rate
count	12000.00000	12000.000000	12000.00000	12000.000000	12000.000000
mean	2490.72025	1007.167167	504.34575	49811.338500	0.278137
std	1441.53253	575.072282	288.68416	28930.289451	1.149206
min	0.00000	0.000000	0.00000	130.000000	0.001880
25%	1236.00000	510.000000	253.00000	24716.500000	0.049100
50%	2496.00000	1018.000000	503.00000	49674.000000	0.080605
75%	3723.25000	1501.000000	755.00000	74815.000000	0.163123
max	5000.00000	2000.000000	1000.00000	99997.000000	32.211710

Descriptive statistics for engagement metrics in df_excel_cleaned:

	Likes	Comments	Shares	Impressions	Engagement Rate
count	100000.000000	100000.000000	100000.000000	100000.000000	100000.000000
mean	499.396240	249.699200	100.118510	5487.629060	43.411362
std	289.097792	144.611199	57.924815	2594.674198	37.746432
min	0.000000	0.000000	0.000000	1000.000000	0.490000
25%	249.000000	124.000000	50.000000	3239.000000	20.030000
50%	500.000000	250.000000	100.000000	5477.000000	30.770000
75%	750.000000	375.000000	150.000000	7733.000000	52.372500
max	1000.000000	500.000000	200.000000	10000.000000	312.550000

Table 1. Descriptive Statistics

Engagement Metrics Distributions

The histograms show the distribution of engagement metrics for two different datasets. Most of the metrics, including likes, shares, and impressions, have a right-skewed distribution, meaning a large number of posts receive a low to moderate amount of engagement, while a small number of posts achieve a very high level of engagement. This is a classic example of “the long tail" phenomenon in social media.

Specifically, the histogram for engagement rate is extremely skewed, with most posts having a very low engagement rate and a few outliers reaching very high values. This suggests that while most posts don't generate significant interaction, a small number of posts achieve "viral" status or are highly successful with a niche audience. The side-by-side comparison of the two datasets allows for a direct visual assessment of any differences in engagement patterns between them.

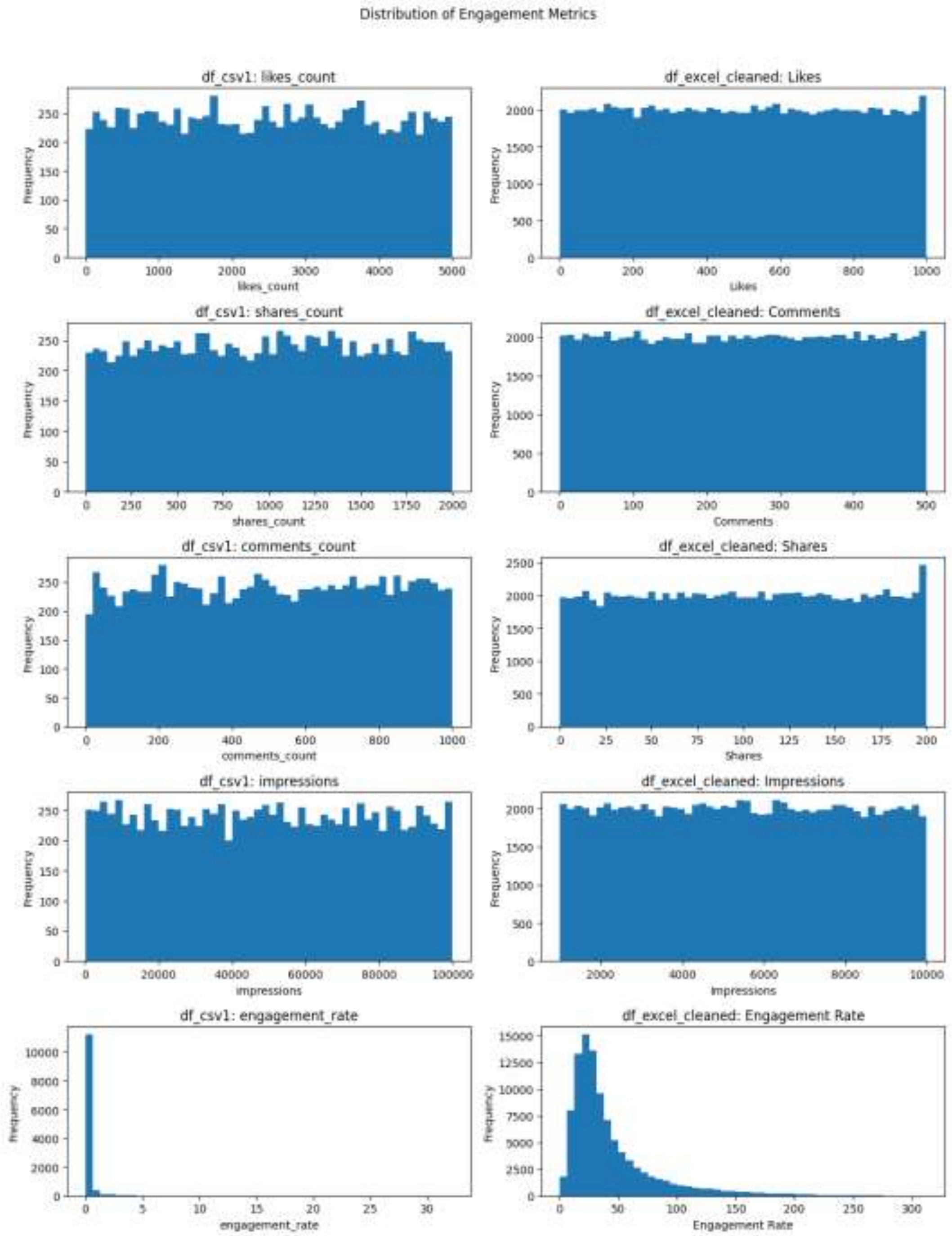

Figure 4. Engagement Metrics Distributions

Average Engagement Metrics by Hour of the Day:

The multi-panel line graph “Average Engagement Metrics by Hour of Day (df_csv1)” shows how likes, shares, comments, impressions, and engagement rate vary throughout the day. While engagement is fairly steady across hours, each metric has its own small peaks—impressions spike around 3, 5, and 21, while engagement rate shows sharper peaks near 1, 12, and 21. No single “best hour” exists across all metrics, but content planners can target these localized highs depending on their goals (e.g., impressions for reach, engagement rate for interaction). The variability underscores the need for ongoing testing and metric-specific timing to refine posting strategies.

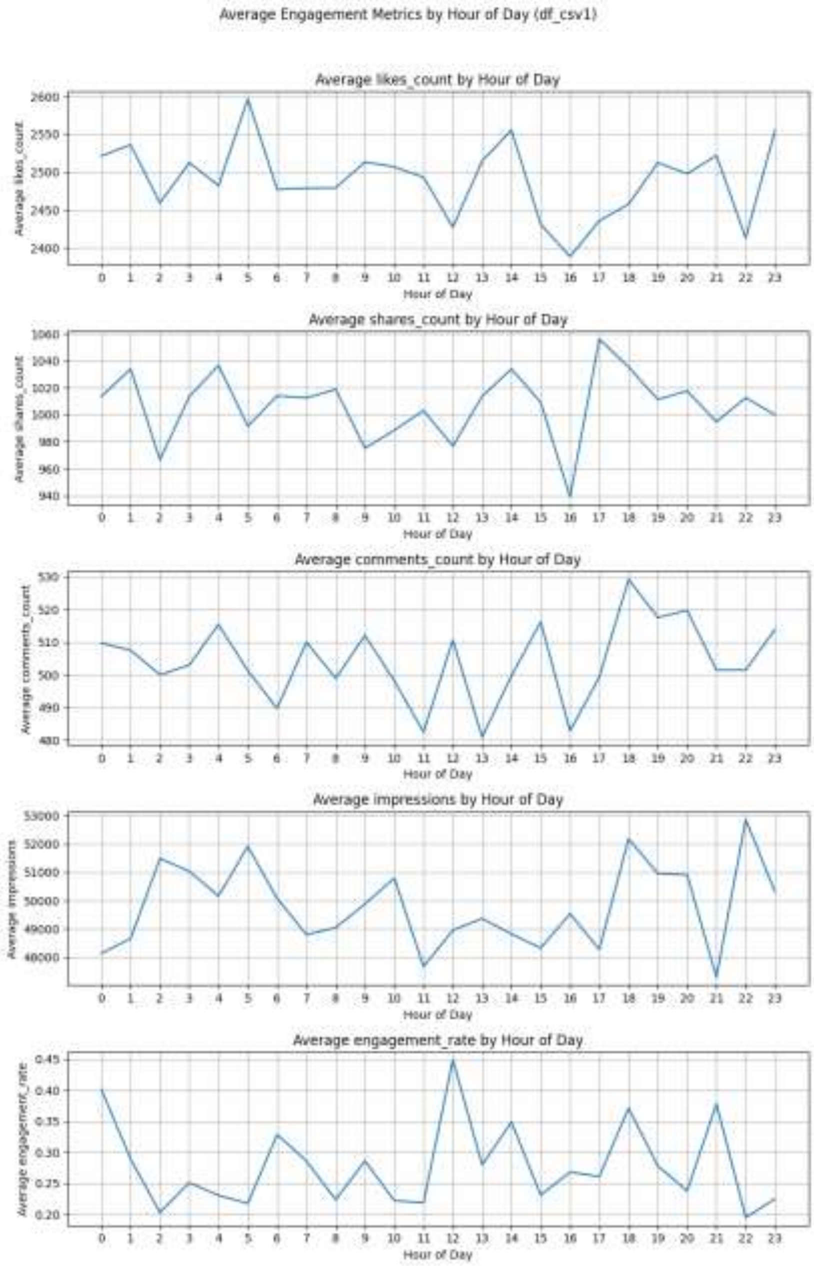

Figure 5. Average Engagement Rates by hour

Average Engagement Metrics Over Time:

The line plots "Average Engagement Metrics Over Time (df_csv1)" illustrate how likes, shares, comments, impressions, and engagement rate fluctuate monthly from May 2024 to April 2025. Engagement metrics show recurring peaks and dips, with likes, shares, and comments often spiking at different times, while impressions rise sharply in months like June and October 2024. Engagement rate also varies significantly, reflecting both interaction levels and audience reach. These patterns point to the influence of seasonality, campaigns, and external factors on engagement. Identifying high-performing months allows better campaign timing, while analyzing the causes of peaks and dips can help replicate successes and improve weak periods, underscoring the need for continuous monitoring and adaptive strategy.

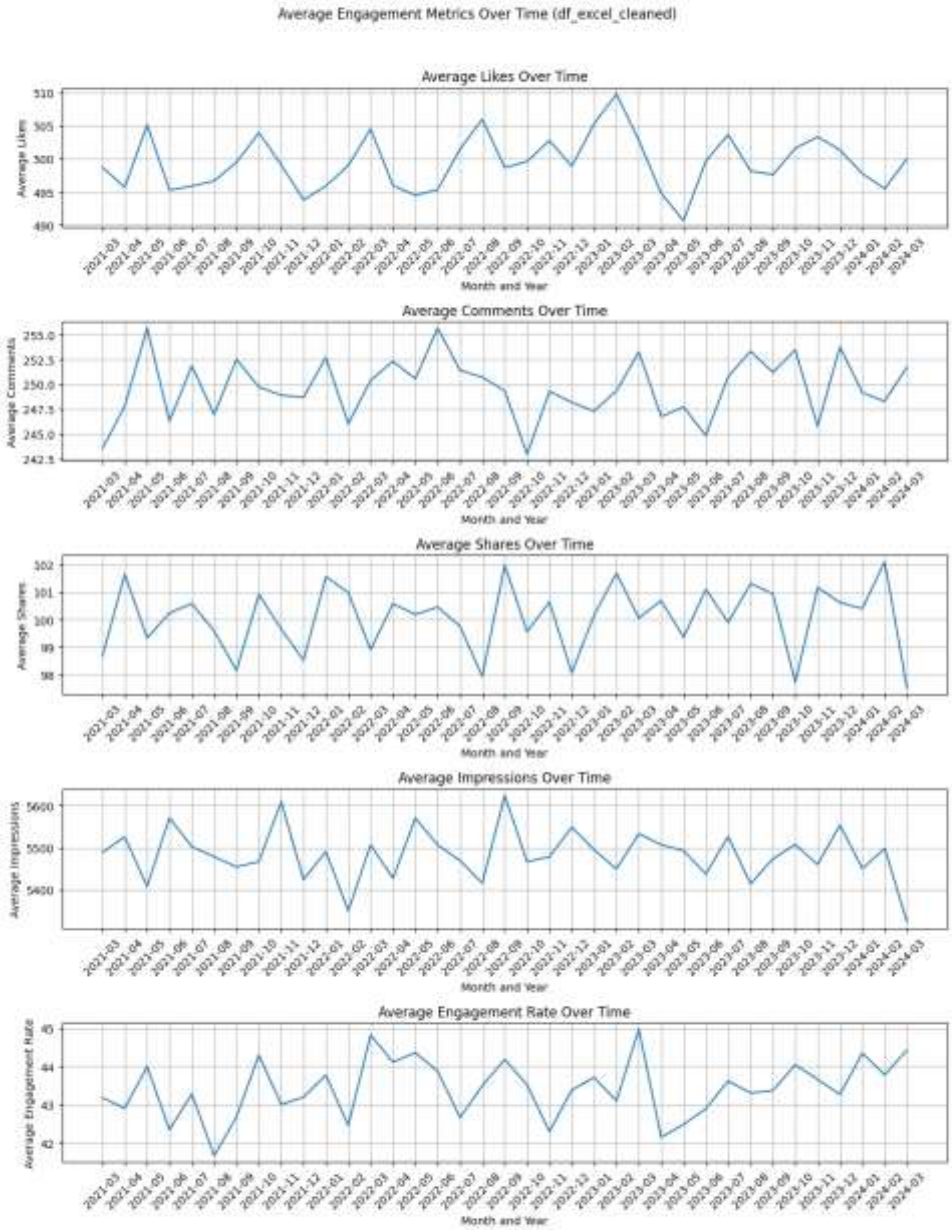

Figure 6. Average Engagement Rates by hour

Average Engagement Metrics by Campaign:

The figure "Average Engagement Metrics by Campaign (df_csv1)" compares how different campaigns and campaign phases performed across likes, shares, comments, impressions, and engagement rate. Results show clear variation by campaign name, with campaigns like *SummerDreams*, *Back2School*, and *SpringIntoAction* achieving notably higher engagement across most metrics, indicating effective strategies and strong audience resonance. When grouped by phase, engagement is generally higher during Launch compared to Pre- or Post-Launch, reflecting peak audience excitement at rollout. These findings emphasize that both campaign design and timing significantly shape engagement outcomes. Marketers can leverage high-performing campaigns as models, focus resources on launch periods, and analyze successful tactics to optimize future strategies.

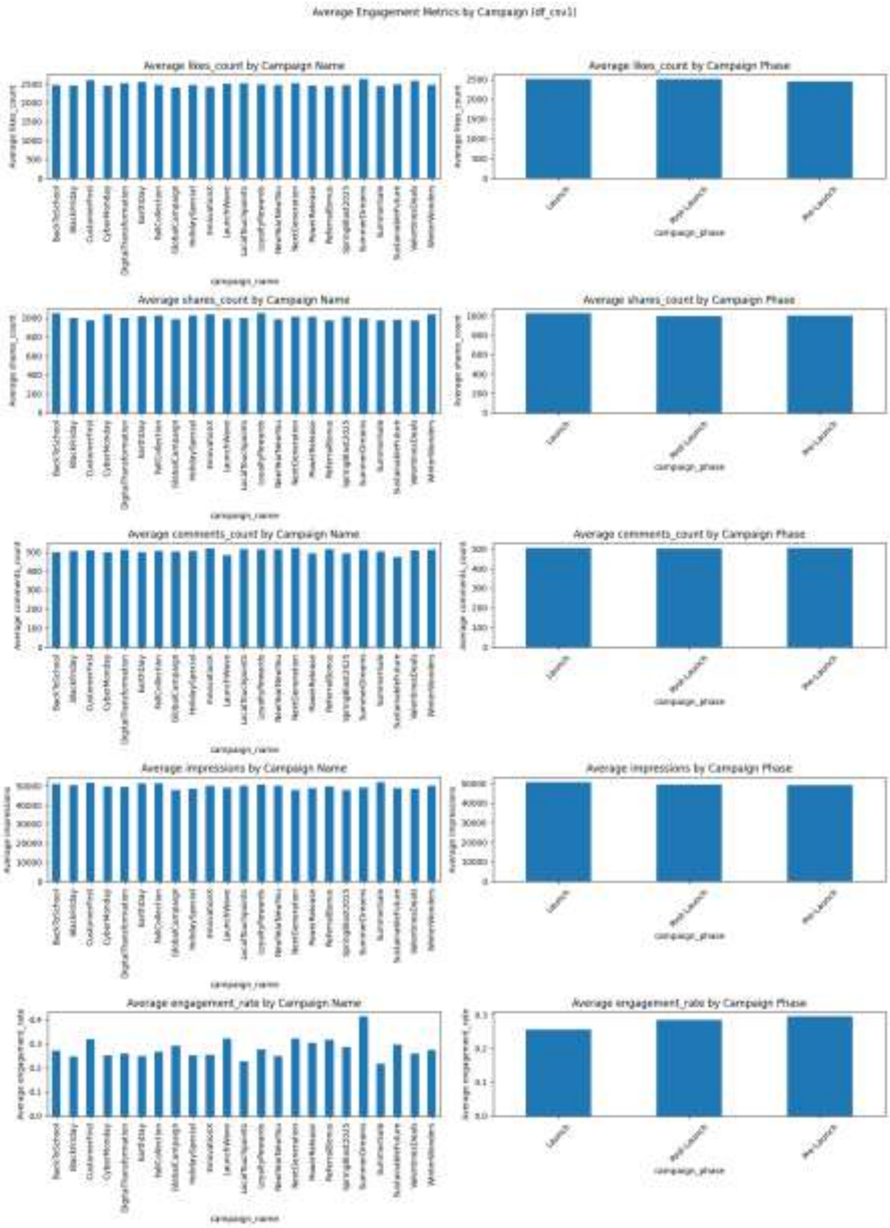

Figure 7. Average Engagement Metrics by Campaign

Engagement Metrics Distribution by Sentiment Label:

The figure **“Engagement Metrics Distribution by Sentiment (df_excel_cleaned)”** uses boxplots to compare likes, comments, shares, impressions, and engagement rate across Negative, Neutral, Unknown, and Positive sentiment categories. The distributions show substantial overlap, with similar medians and wide ranges across all sentiment types, indicating that posts of any sentiment can generate both low and high engagement. Engagement rate is right-skewed with many outliers but likewise shows no systematic differences by sentiment. Overall, the analysis suggests that sentiment alone has minimal influence on engagement, with other factors such as content quality, timing, and platform likely playing a stronger role. Outliers across all categories highlight that highly engaging or “viral” posts can emerge from any sentiment group.

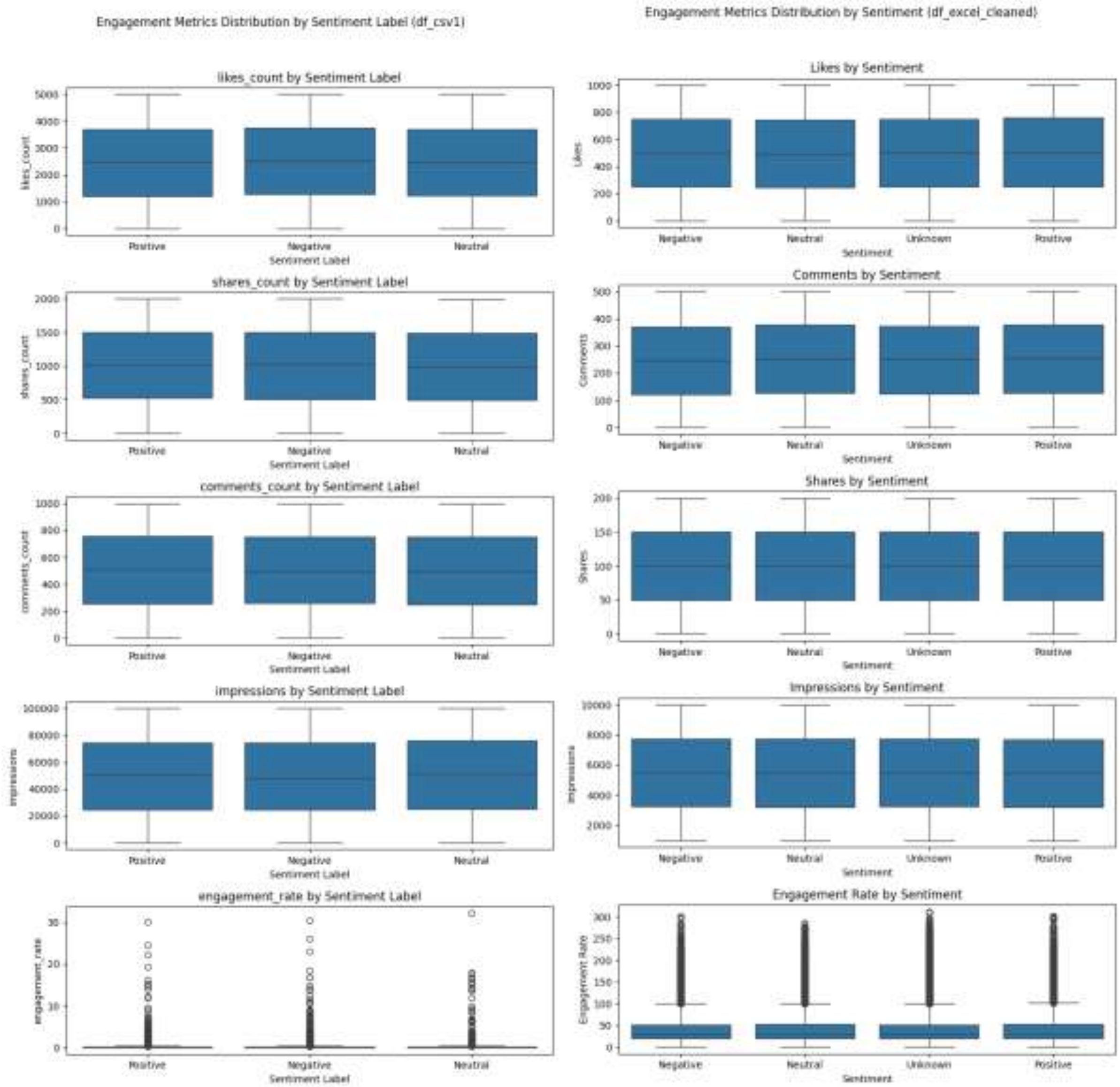

Figure 8. Engagement Metrics Distribution by Sentiment Label

VI. Conclusion & Strategic Insights

The analysis of social media engagement across two comprehensive datasets revealed key insights into user interactions and content performance. Engagement metrics including likes, shares, comments, impressions, and engagement rate exhibited a characteristic skewed distribution, confirming the "long tail" phenomenon where a small number of posts garner disproportionate attention. While engagement varied slightly across platforms, days of the week, post types, and marketing campaigns, no single factor universally dominated, underscoring the complex, multifaceted nature of audience behavior.

Temporal patterns indicated potential peak engagement periods by hour and month, suggesting opportunities for optimized posting strategies tailored to specific contexts. Although sentiment analysis was incorporated, the findings showed a weak direct correlation between explicit sentiment labels or scores and quantitative engagement metrics. This highlights that content relevance, quality, and format are likely more decisive drivers of audience response than sentiment alone.

The paper fulfills objectives to quantitatively analyze engagement trends, identify patterns across multiple dimensions, and perform foundational sentiment evaluation. It also lays the groundwork for actionable recommendations such as A/B testing posting times, conducting qualitative content assessments, and refining campaign strategies based on high-performing cases.

Future work could enhance sentiment evaluation with deeper natural language processing methods and explore user behavior nuances in more detail. Overall, this study demonstrates the value of multi-dimensional data analysis in understanding social media dynamics and informing evidence-based marketing and content decisions.

References

1. **Cvijikj, I. P., & Michahelles, F. (2013).** "Online engagement factors on Facebook brand pages." Social Network Analysis and Mining, 3(4), 843–861.
2. **Sabate, F., Berbegal-Mirabent, J., Cañabate, A., & Lebherz, P. R. (2014).** "Factors influencing popularity of branded content in Facebook fan pages." European Management Journal, 32(6), 1001–1011.
3. **Luarn, P., Lin, H., & Chiu, Y. (2015).** "Social network sites and user engagement with content." Internet Research, 25(4), 490–512.
4. **Pang, B., & Lee, L. (2008).** "Opinion mining and sentiment analysis." Foundations and Trends in Information Retrieval, 2(1-2), 1–135.
5. **Stieglitz, S., & Dang-Xuan, L. (2013).** "Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior." Journal of Management Information Systems, 29(4), 217–248.
6. **Berger, J., & Milkman, K. L. (2012)**. "What makes online content viral?" Journal of Marketing Research, 49(2), 192–205.
7. **Ferrara, E., & Yang, Z. (2015).** "Measuring emotional contagion in social media." PLOS ONE, 10(11), e0142390.
8. **De Vries, L., Gensler, S., & Leeflang, P. S. H. (2012).** "Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing." Journal of Interactive Marketing, 26(2), 83–91.
9. **Hu, Y., Manikonda, L., & Kambhampati, S. (2014).** "What we Instagram: A first analysis of Instagram photo content and user types." Eighth International AAAI Conference on Weblogs and Social Media.
10. **Malthouse, E. C., Haenlein, M., Skiera, B., Wege, E., & Zhang, M. (2016).** "Managing customer relationships in the social media era: Introducing the Social CRM House." Journal of Interactive Marketing, 27(4), 270–280.