Data-Driven Influencer Selection for Marketing Campaigns on YouTube

Cui Zhiming Li Huiyun Li Xinying Ju Yuting Zhang Xiaoyu

BT5153 Applied Machine Learning for Business Analytics - Group 11
National University of Singapore
Github Link: https://github.com/AmoreLi25/BT5153-Group-11-Project

Abstract

Digital marketing on platforms like YouTube often relies on superficial metrics such as follower counts or view numbers to guide influencer advertising placement, overlooking deeper indicators of audience quality and sentiment. This study introduces a data-driven framework for evaluating YouTube influencer effectiveness through a self-defined influencer performance source (IPS) which combining traditional quantitative metrics (e.g., views, likes, comments) with qualitative user comments sentiment analysis of audience interactions. To further enrich our analysis from the large language model side, we leverage GPT-4 for comment interpretation and did a case study where we combined ChatGPT with IPS to help cosmetic industry stakeholders determine target influencers. Our findings offer actionable insights for campaign designers and brand managers, supporting the identification of authentic, high-impact influencers and enabling more targeted and effective advertising strategies.

1. Problem Description

In the rapidly evolving landscape of digital marketing, brands face increasing challenges in identifying the suitable influencers to drive meaningful business improvement. Traditional selection methods that rely heavily on subscriber counts and video views often ignore critical factors such as engagement quality and audience sentiment. This can lead to biased influencer choices and suboptimal campaign outcomes. To address this gap, we propose a data-driven, sentiment-aware evaluation framework for YouTube influencers. We propose the Influencer Performance Score (IPS), a metric designed to evaluate influencer effectiveness by combining both quantitative and qualitative signals. It integrates traditional engagement statistics, such as total views, likes, and comment counts, with sentiment polarity scores extracted from user-generated comments using tools like VADER (a rule-based sentiment analysis tool suited for social media text) and BERT (a pre-trained deep learning model capable of capturing more complex language nuances). IPS gives a complete picture of an influencer's real impact and connection with followers.

To further enhance the relevance of influencer recommendations, we leverage large language models (LLMs) such as ChatGPT to identify industry alignment and extract nuanced audience insights from comment data.

By integrating engagement data with sentiment and large language model understanding, this framework enables brands to select influencers who not only reach large audiences but also foster strong, positive connections, and ultimately improve campaign targeting and return on investment.

2. Dataset Preparation

2.1. Data Collection

We work with two Kaggle YouTube datasets from Great Britain (GB): GBvideos and GBcomments. The GBcomments dataset contains information about user comments on videos, while the GBvideos dataset provides video statistics such as likes, views, and tags. To enable interactive and comprehensive analysis, we merged these datasets into a unified dataset, GB Dataset. This dataset includes a rich collection of 4,208,923 user comments for 1,692 unique videos, with the time period from January 1, 2018 to September 18, 2018.

2.2. Explanatory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a foundational step in data science that allows analysts to understand the patterns, relationships, and anomalies within their data. For our YouTube video data set, we performed an exploratory analysis in various dimensions:

2.2.1. Like vs View Relationship

From the scatter plot, we could see a positive correlation between views and likes, indicating that engagement grows as visibility increases. To maximize marketing impact, brands should work with influencers whose videos not only get many views but also a high number of likes. In this way, they reach an active, engaged audience instead of just passive viewers.

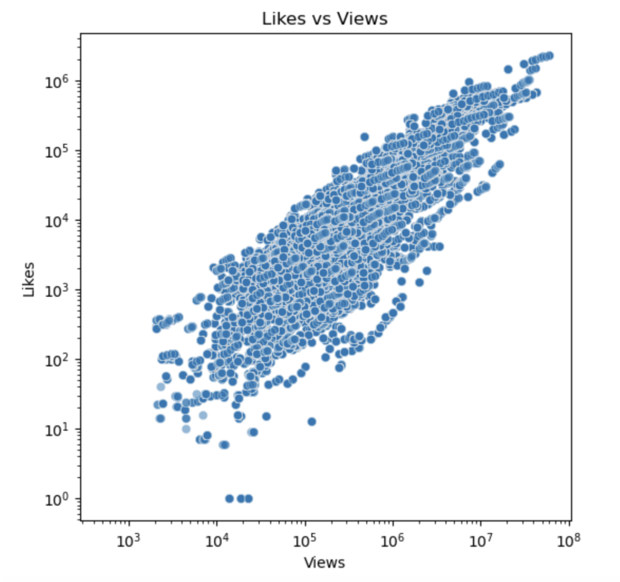

Figure 1. Like vs Views

2.2.2. Tag Distribution

By studying the common video tags, we found that entertaining content, such as comedy, vlogs, and funny, gets the most attention. To reach the widest audience, brands should partner with influencers who specialize in these popular categories.

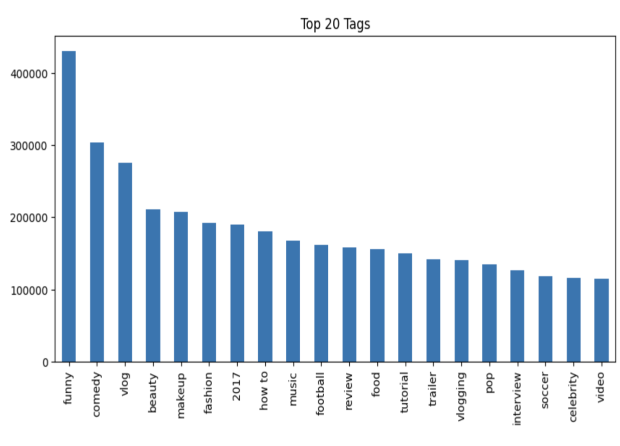

Figure 2. Top 20 Tags

2.2.3. Tag Word Cloud

The word cloud visualization highlights the diversity of popular tags, with strong emphasis on lifestyle, beauty, fashion, and education-related keywords. Brands can choose influencers from these niches based on their campaign goals.

Figure 3. Tags Word Cloud

2.2.4. Top Viewed Videos

By examining the most viewed videos, we observed common traits among highly successful content, such as relatability, humor, and topical relevance. By partnering with influencers who produce similar high-performing content, brands can increases the likelihood of reaching broader audiences and achieving better campaigns results.

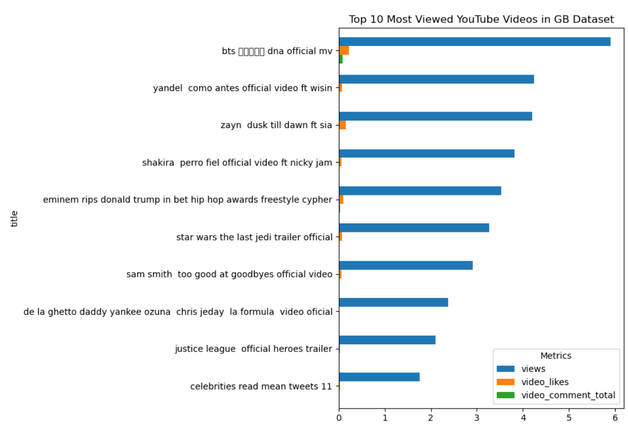

Figure 4. Top 10 Most Viewed Video

2.3. Feature Engineering

To quantify user engagement beyond raw counts, we analyzed deeper engagement metrics, and created new features that better capture viewer interaction and emotional response.

Comment Length: Measures the number of characters in each user comment, reflecting the richness of viewer feedback.

Total Comment Likes: Combines likes and replies on user comments to capture the total level of audience interaction beyond the video itself.

Like Rate per Video: Represents the proportion of likes per view. This normalizes the like count and helps evaluate how positively a video is received relative to its exposure.

Comment Rate per Video: Measures the number of comments per view, indicating how likely viewers are to engage in discussion.

Like-to-Dislike Ratio: Calculates how many likes a video receives for each dislike. A higher value suggests that people liked the video more.

These engineered metrics provide deeper insights into influencer performance and audience engagement, enabling data-driven selection and precise campaign targeting.

3. Sentiment Analysis

3.1. Sentiment Analysis Using VADER

To analyze the emotional tone of user comments in the GB region, we first applied the VADER (Valence Aware Dictionary and sEntiment Reasoner) model, a rule-based sentiment analysis tool specifically developed for social media and user-generated content. VADER is well-suited for analyzing informal text due to its ability to capture linguistic nuances such as slang, emoticons, capitalization, and degree modifiers, which are prevalent in online communication. Unlike traditional lexicon-based methods, VADER combines a comprehensive sentiment lexicon with a set of grammatical and syntactical rules to enhance accuracy in short, casual text formats.

The model classifies textual content into three sentiment categories—positive, neutral, and negative—based on the aggregation of word-level sentiment scores adjusted by contextual intensifiers. Given the nature of YouTube comments, which are often brief, colloquial, and emotionally charged, VADER provides an efficient and interpretable approach for initial large-scale sentiment assessment.

The results show that among all comments, positive sentiments dominated with approximately 2.16 million comments (51.4%), followed by 1.24 million neutral comments (29. 5%) and 802,000 negative comments (19. 1%). This indicates that a significant portion of user engagement in trending videos in the GB region carries a generally positive tone.

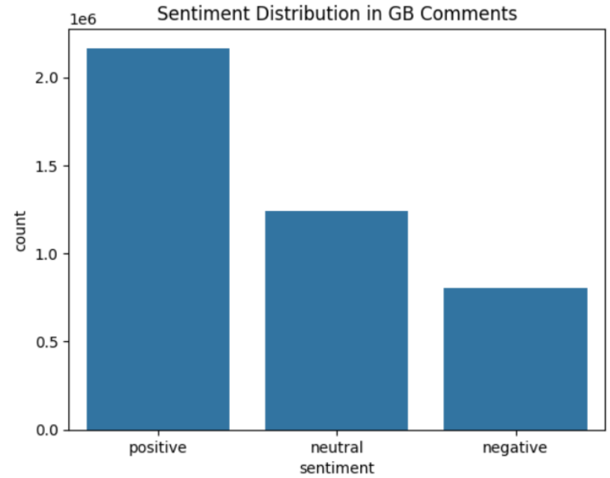

Figure 5. Sentiment distribution by VADER.

To uncover thematic variation in audience tone, we examined sentiment proportions within each video category. Categories such as *Nonprofits & Activism* (68.8%), *Howto & Style*(67.4%) and *People & Blogs* (57.0%) showed the highest proportions of positive sentiment, suggesting strong viewer approval and emotional resonance. In contrast, *News & Politics* exhibited the highest share of negative sentiment (36.1%), which may reflect the polarizing nature of political content. Most other categories showed a mix.

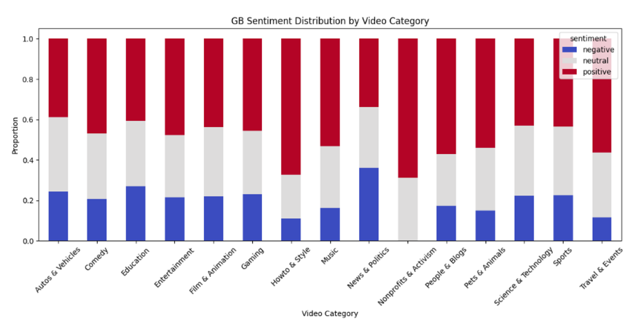

Figure 6. Proportional sentiment distribution by video category in GB(VADER).

3.2. Sentiment Analysis Using BERT

To complement the VADER analysis, incorporate machine learning techniques, and provide a deeper, context-aware interpretation of sentiment, we applied a transformer-based model—BERT, specifically the pretrained model cardiffnlp/twitter-roberta-base-sentiment.

Unlike rule-based methods that primarily rely on fixed lexical dictionaries and heuristics, BERT leverages deep learn-

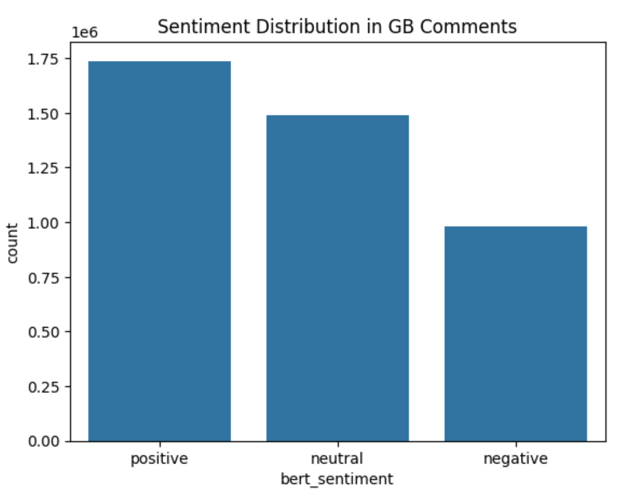

Figure 7. Sentiment distribution by BERT.

ing to model the semantic relationships between words within a sentence, allowing it to infer sentiment from context rather than isolated keywords alone. Through pretraining on large-scale social media datasets, this model is particularly adept at handling the informal, diverse, and often nuanced language commonly found in YouTube comments. Its contextual embeddings enable the model to distinguish between subtle shifts in tone, sarcasm, or complex emotional expressions, areas where traditional models such as VADER may struggle. Therefore, integrating BERT into our analysis offers a more comprehensive and precise understanding of user sentiment, particularly in scenarios involving ambiguous or multi-layered emotional content.

According to BERT's predictions, positive sentiment remains dominant, followed by neutral and negative comments. The proportion of positive comments remains high , supporting the earlier VADER results. However, BERT tends to produce a more balanced distribution, capturing subtleties that VADER may overlook, particularly in identifying neutral tones in subjective text.

BERT's category-level sentiment distribution largely confirms the trends observed in the VADER results. Categories such as *Nonprofits & Activism* and *Howto & Style* again exhibited the highest proportions of positive sentiment, reaching 66.1% and 60.8% respectively, indicating consistently strong viewer approval across both models. *Music* also received a strong positive reception (51.3%).

News & Politics had the highest proportion of negative sentiment (51.0%), followed by *Education* (36.7%). These results reaffirm that content involving current affairs or informative content often elicits stronger, sometimes polarizing emotional responses from viewers.

Furthermore, BERT identified a larger share of neutral sentiment in categories like *Science & Technology* and *Autos & Vehicles*, indicating its ability to detect more objective or informational tones that VADER might misclassify as emotional.

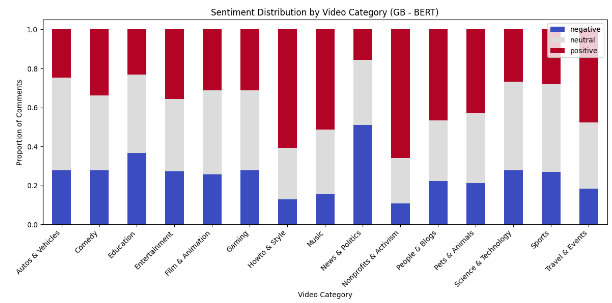

Figure 8. Proportional sentiment distribution by video category in GB(BERT).

These findings validate that sentiment varies not only across influencers, but also across video genres, underscoring the value of integrating sentiment in influencer evaluation.

4. Influence Scoring Methodology

To evaluate quantitative influence across YouTube channels, we developed a custom *Influence Score*—a composite metric that integrates four key engagement indicators: average views, likes, comments, and engagement rate. The objective was to go beyond surface-level statistics and rank creators based on normalized, aggregated interaction quality.

4.1. Metric Design

We first aggregated engagement metrics at the channel level by summing across all associated videos. For each channel, we computed:

- Average views per video: viewership represents the basic *reach* of an influencer — the number of users exposed to their content. A high average view count indicates strong baseline visibility, which is fundamental for any influencer's impact.
- Average likes per video: likes serve as a quick indicator of positive audience reception. Giving likes an equal weight as views emphasizes not just how many people see the content, but also how many react favorably to it.
- Average comments per video: comments reflect deeper audience engagement compared to likes, requiring more active participation. However, because comment activity can vary widely across content types and platforms, a slightly lower weight (20%) is assigned to avoid over-amplifying niche behaviors.

- Engagement rate: defined as (likes + comments) divided by views. The engagement rate contextualizes interactions relative to reach. It ensures that influencers who generate a high level of audience interaction per view are appropriately recognized, regardless of their absolute view counts.

Each of these metrics was then normalized using min-max scaling to bring them onto a comparable scale. The final **Influence Score** was computed as a weighted sum:

$$\begin{aligned}\text{Influence Score} = \;& 0.3 \times \text{norm}(\text{avg views}) \\ & + 0.3 \times \text{norm}(\text{avg likes}) \\ & + 0.2 \times \text{norm}(\text{avg comments}) \\ & + 0.2 \times \text{norm}(\text{engagement rate})\end{aligned} \tag{1}$$

which reflects a carefully considered balance among different aspects of influencer quantitative performance.

4.2. Top Ranked Influencers

Based on the Influence Score, we present the top 10 influencers in the GB dataset:

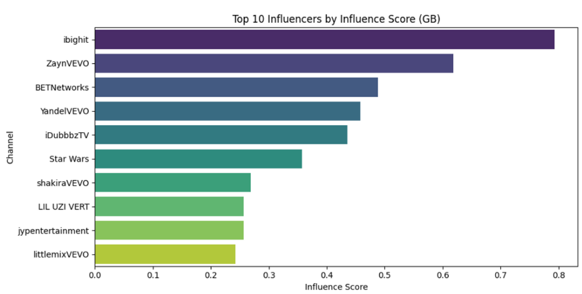

Figure 9. Top 10 YouTube influencers by Influence Score (GB).

At the top of the list is ibighit, the official YouTube channel of Big Hit Entertainment, which manages global K-pop sensations such as BTS and TXT. The channel consistently attracts massive viewership, high volumes of likes, and strong audience engagement across all content, including music videos, dance performances, and behind-the-scenes footage. Its influence is amplified by a highly loyal international fanbase, which drives both reach and interaction to exceptional levels, resulting in an outstanding Influence Score.

Following ibighit are ZaynVEVO, BETNetworks, and YandelVEVO, all of which maintain competitive influence levels, particularly within the music and entertainment domains. Other notable channels include iDubbbzTV, Star Wars, and shakiraVEVO, reflecting a mix of both personal creators and branded entertainment content.

Overall, the GB rankings show a strong presence of music-related influencers, especially VEVO-affiliated channels, indicating that the music and entertainment sectors dominate the influencer landscape in the UK. The diversity of channel types—from global entertainment franchises to independent creators—demonstrates the varied nature of influence in the UK YouTube ecosystem.

5. Performance Scoring Framework

While the Influence Score captured quantitative engagement signals such as views, likes, and comments, it did not account for the quality of audience sentiment. To bridge this gap, we introduced a second-tier metric—the **Performance Score**—which integrates both behavioral and affective signals to better rank influencer effectiveness.

5.1. Sentiment Scoring

For each channel, we computed the proportion of positive, neutral, and negative comments. A sentiment polarity score was then calculated as:

$$\text{Sentiment Score} = \frac{\text{Positive} - \text{Negative}}{\text{Total Comments}} \tag{2}$$

This metric ranges from -1 to 1 and captures the net positivity of user feedback for each channel.

5.2. Composite Performance Score

To create a unified performance metric, we normalized sentiment scores using min-max scaling and combined them with the previously computed Influence Scores. The final Performance Score was defined as:

$$\begin{aligned}\text{Performance Score} = \;& 0.6 \times \text{Influence Score} \\ & + 0.4 \times \text{Normalized Sentiment Score}\end{aligned} \tag{3}$$

This formula gives slightly greater importance to quantitative reach (60%) while still heavily considering the sentiment feedback (40%) to avoid promoting influencers with high engagement but negative reputations.

5.3. Top Performers

Figures 10 display the top 10 channels by Performance Score in the GB dataset. Notably, mainstream channels such as *ibighit* and *ZaynVEVO* rank at the top due to their strong influence and favorable audience sentiment. Meanwhile, niche creators like *booksandquills* and *Jessica Kellgren-Fozard* also appear in the top rankings, primarily because of their exceptionally high sentiment scores. This highlights the importance of audience approval in shaping overall performance, beyond raw engagement metrics alone.

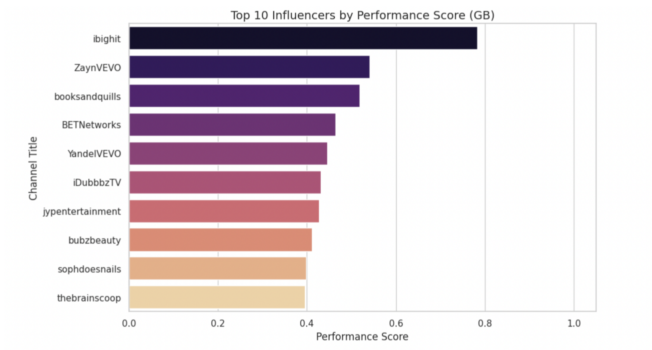

Figure 10. Top 10 YouTube influencers by Performance Score (GB).

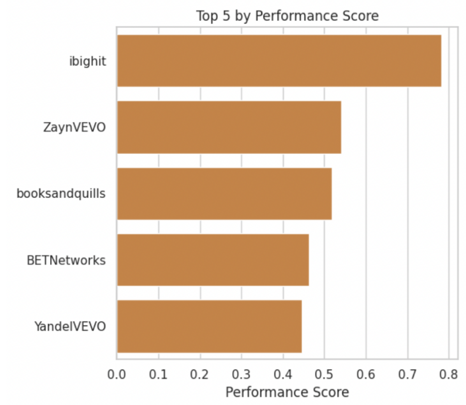

Figure 13. Top 5 Influencers by Performance Score

6. Consolidated Insights and Results

6.1. Influencer Rankings Across Multiple Metrics

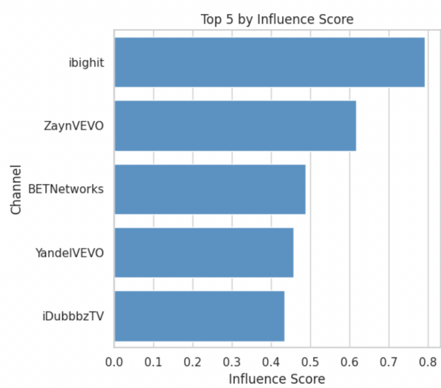

Figure 11. Top 5 Influencers by Influence Score

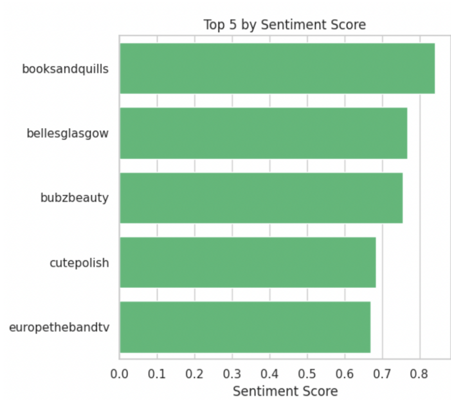

Figure 12. Top 5 Influencers by Sentiment Score

With reference to Figures 11-13, our ranking analysis reveals two key patterns. First, top influencers by influence score, such as *ibighit* and *ZaynVEVO*, are predominantly major entertainment channels with massive followers. However, when sentiment is prioritized, creators like *booksandquills* and *bellesglasgow* emerge as leaders, indicating strong emotional resonance with audiences despite smaller reach.

Interestingly, performance scores, which integrate both influence and sentiment, highlight a blend of these two extremes. Influencers with moderate reach but consistently positive sentiment, like *booksandquills*, outperform some larger channels that suffer from polarized or negative audience reactions. This underscores the importance of emotional trust and brand alignment beyond raw popularity metrics.

6.2. Dual-Metric Framework Effectiveness

The dual-metric evaluation framework, combining quantitative reach with qualitative sentiment, offers a richer perspective on influencer performance. Sentiment analysis not only shaped final rankings but also consistently emerged as a critical factor across clustering patterns and supervised model feature importance evaluations.

By balancing audience size with emotional resonance, our framework moves closer to capturing authentic influence rather than mere visibility. This approach highlights the strategic importance of weighing sentiment stability and positivity alongside traditional engagement rates when selecting influencers for marketing campaigns.

7. LLM-Powered Comment Analysis

7.1. Methodology

To complement our quantitative metrics, we developed a Large Language Model (LLM)-powered comment analysis system that extracts recurring themes and sentiment patterns from YouTube comments. This approach provides campaign managers with human-readable summaries of audience perceptions for each top-ranked influencer, offering a richer view beyond numerical scores.

The system leverages GPT-4o through a structured pipeline. For each influencer, 200 comments are randomly sampled (or all available comments if fewer exist). These samples are then analyzed using a prompt instructing the model to summarize audience sentiment, highlight positive and negative reactions, and identify key discussion topics and emotional tones. The output is validated against manual annotation samples to ensure consistency.

Compared to traditional sentiment analysis, this method offers three advantages: it uncovers organic discussion themes beyond pre-defined categories, captures nuanced emotional expressions such as sarcasm and emojis, and delivers executive-ready insights without requiring technical expertise.

7.2. Case Studies

Applying this analysis to top performers in the GB dataset revealed distinct audience engagement patterns.

Music Influencer: ibighit For *ibighit*, the comments reflect an overwhelmingly positive and enthusiastic fanbase centered around BTS. Audience sentiment is largely supportive and motivated, with fans actively participating in streaming and voting efforts to help BTS achieve milestones such as reaching targeted view counts or winning awards. Recurring themes include organized streaming strategies, calls for voting participation in events like the Asia Artist Awards, admiration and emotional support for BTS members, and celebrations of record-breaking achievements. Some comments also engage directly with the music video content, praising choreography and production quality. Negative reactions were rare, mainly involving isolated criticisms of the video or frustration about voting outcomes. Overall, the comments showcase a deeply committed and passionate community dedicated to BTS's success.

Niche Creator: booksandquills For *booksandquills*, feedback was overwhelmingly positive, with viewers expressing admiration and emotional responses to the video's aesthetic and message. Words such as ”beautiful,” ”amazing,” and ”lovely” appeared frequently, reflecting strong audience appreciation. Many viewers shared how the video moved them emotionally, while others related personally to the themes of cherishing books and grappling with limited time to read. The stop-motion animation style also received notable praise. Negative feedback was minimal, mostly centered around critiques of traditional literary canon choices, with some calling for broader representation. Overall, the comments reveal a deeply engaged and reflective audience that connected both emotionally and intellectually with the content.

7.3. Strategic Value

The LLM analysis surfaced three strategic insights valuable for campaign design. First, sentiment-content alignment varies by creator type: music influencers drive high-volume but more volatile engagement, while niche creators foster deeper, steadier positive sentiment. Second, recurring themes—such as BTS member focus or literary aesthetics—offer natural entry points for tailored brand messaging. Third, early detection of sensitive topics, such as health concerns or literary diversity critiques, enables preemptive risk management.

Overall, this qualitative layer enhances the influencer selection framework by revealing not just how audiences engage, but why they engage, offering a competitive edge for data-driven marketing strategies.

8. Cosmetics Industry Specific Analysis

To align our influencer evaluation framework with a real-world business objective, we conducted an industry-specific analysis tailored to the cosmetics sector. The goal was to identify influencers whose content styles were more conducive to the promotion of beauty and personal care products on YouTube. Our approach combines domain knowledge, GenAI-guided category selection, and our data-driven scoring framework to deliver actionable recommendations for campaign planning.

Category Selection via GenAI: We leveraged GPT-4o to identify the most relevant YouTube video categories for a cosmetics digital marketing campaign. Based on audience fit, content style, and influencer alignment, the following four categories were recommended:

- **Howto & Style (Category 26)**
  – Tutorials, beauty routines, and product reviews.
- **People & Blogs (Category 22)**
  – Personal endorsements and lifestyle storytelling.
- **Entertainment (Category 24)**
  – Creative, wide-reach content like challenges and skits.
- **Music (Category 10)**
  – Style-forward audiences and artist partnerships.

We then assigned a weight reflecting its strategic importance to each category respectively: 0.4 for Howto & Style, 0.3 for People & Blogs, 0.2 for Entertainment, and 0.1 for Music.

Influencer Scoring Framework: We reused the previously computed `performance_gb` dataframe, which captures each influencer's performance score based on both quantitative reach and audience sentiment. This score was joined with the relevant category data and used to compute a weighted score for each influencer by multiplying the category-specific performance by its corresponding GPT-defined weight. The results were then aggregated across all applicable categories for each influencer.

Results Overview:

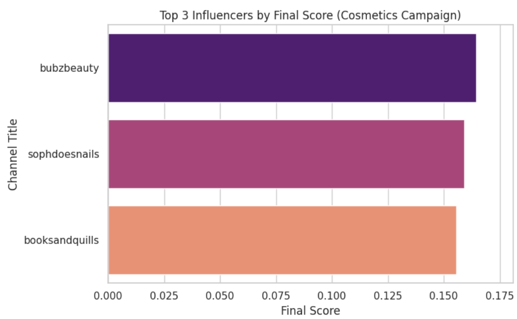

Figure 14. Top 3 Influencers by Final Score

Figure 14 visualizes the final scores of the top three influencers in our cosmetic digital marketing campaign. The highest ranked was **bubzbeauty** with a final score of 0.1645, followed by **sophdoesnails** (0.1593), and **booksandquills** (0.1557). Notably, each of these influencers derived their score primarily from a single dominant category: bubzbeauty and sophdoesnails from Howto & Style, and booksandquills from People & Blogs.

Multi-Category Contributor Insight: Interestingly, only one influencer–`danisnotinteresting`–contributed significantly across two categories. He achieved scores of 0.075 from People & Blogs and 0.05 from Entertainment, resulting in a combined final score of 0.125.

While multi-category presence can be beneficial by offering diversification, our findings suggest that dominance in a single highly relevant and weighted category may yield greater impact for targeted campaigns.

Business Implications: This industry-specific analysis framework enables cosmetics brands to systematically align influencer selection with the relevance of content genre and sentiment quality, ensuring stronger message resonance and better audience targeting. By integrating both performance metrics and business domain relevance, our approach presents a more informed and effective campaign design methodology.

9. Conclusion

This study demonstrates the value of integrating quantitative engagement metrics with qualitative sentiment analysis to evaluate YouTube influencers' effectiveness for digital marketing campaigns. By developing the Influencer Performance Score (IPS) and combining it with large language model (LLM)-assisted analysis, we go beyond traditional surface-level metrics and capture a more holistic view of influencer impact based on real community engagement.

9.1. Key Insights and Findings

Our dual-metric evaluation framework revealed that the top influencers in terms of raw visibility (views, likes) are not always the most valuable contributor considering sentiment quality and real community participation. Influencers with moderate reach but consistently high positive sentiment often outperform larger channels in overall performance rankings. In addition, industry-specific adaptations, such as our cosmetics campaign case study, showed that category alignment further sharpens influencer selection.

9.2. Key Challenges and Limitations

We've encountered several challenges during the study. First, reliance on English-language sentiment models like VADER and BERT can introduce bias in multicultural or multilingual datasets. Second, publicly available YouTube datasets can suffer from incomplete metadata (e.g., missing category information), which limits full-scale industry targeting. Finally, while LLM-powered comment analysis offers qualitative depth, it also introduces potential subjectivity and dependency on GPT prompt design.

9.3. Future Directions

As the next step, we could extend this framework to short-form video platforms such as TikTok and Instagram Reels, where influencer dynamics differ significantly. By incorporating data across multiple platforms, including video thumbnails and audio sentiment, we could further refine IPS. In addition, integrating time series engagement patterns would enable us to predict the emerging influencer trends rather than static evaluations.

9.4. Final Conclusion

By stacking traditional engagement data, advanced sentiment modeling, and LLM-based analysis, we provide a robust and scalable framework for data-driven influencer marketing. This approach empowers brands to prioritize high-impact partnerships. It also improves resource efficiency in campaign management and lays the foundation for more sophisticated and responsible digital marketing strategies.

References

[1] C.J. Hutto and Eric Gilbert. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2014.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT*, 2019.

[3] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, et al. HuggingFace's Transformers: State-of-the-art Natural Language Processing. *arXiv preprint arXiv:1910.03771*, 2019.

[4] OpenAI. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774*, 2024.

[5] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Impact Statement

This study provides a systematic and scalable framework for influencer evaluation by integrating engagement-based performance metrics, sentiment analysis, and industry-specific category relevance. By combining data-driven analytics with GenAI-assisted insights, our approach offers a holistic view of influencers' effectiveness beyond raw popularity measures. The performance framework empowers marketing teams to make more strategic, data-driven decisions when selecting influencer partners. It also ensures better audience alignment and greater authenticity of the campaign. Beyond its immediate application to cosmetics marketing, this methodology is adaptable across industries and platforms, supporting future innovations in personalized and ROI-optimized influencer marketing strategies.