JOIN (Jurnal Online Informatika)
p-ISSN: 2528-1682, e-ISSN: 2527-9165
Volume 10 Number 2 | December 2025: 372-383
DOI: 10.15575/join.v10i2.1633

A Data Science Approach to Exploring the Relationship Between TikTok Engagement and Revenue in Malaysia: A Case Study of the Beauty and Personal Care Sector

Muhammad Akmal Hakim Ahmad Asmawi[1], Pradeep Isawasan[1], K.S. Savita[2], Lalitha Shamugam[3], Khairulliza Ahmad Salleh[1]
[1]College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA, Perak Branch, Tapah Campus, Malaysia
[2]Department of Computing, Positive Computing Center, Universiti Teknologi PETRONAS, Malaysia
[3]Department of Business Analytics, Sunway Business School, Sunway University, Malaysia

Article Info

Article history:

Received May 10, 2025
Revised June 06, 2025
Accepted July 13, 2025
Published November 10, 2025

Keywords:

Business Analytics
Engagement Metrics
e-Commerce Revenue
Social Media
TikTok

ABSTRACT

TikTok has reshaped digital marketing in the beauty and personal care sector, yet the relationship between engagement metrics and revenue outcomes remains unclear. This study aims to examine how public engagement metrics (likes, comments, shares, and live interactions) relate to revenue performance among TikTok influencers. Using the Data Science Trajectories (DST) framework, data from 17 Malaysian influencers across Celebrity, Macro, Meso, and Micro categories were analyzed through descriptive statistics and machine learning models implemented in Python. The findings reveal that high engagement does not consistently lead to higher revenue. Live sessions were more effective than standard videos in driving sales due to real-time interaction. While Celebrity influencers led in revenue, Meso influencers recorded the highest engagement rates. A Random Forest regression model showed strong predictive power (R^2 = 0.94), demonstrating that public-facing metrics can be used to estimate revenue. The study also introduces category-based engagement rate benchmarks and highlights the unique value of live content in converting engagement into sales. This research contributes to the growing body of work on TikTok marketing by combining statistical and predictive techniques to link engagement behavior with commercial outcomes, offering actionable insights for both practitioners and scholars.

Corresponding Author:

Pradeep Isawasan,
College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA, Perak Branch, Tapah Campus, 35400, Tapah Road, Perak, Malaysia
Email: pradeep@uitm.edu.my

1. INTRODUCTION

TikTok has become a transformative platform in social media, particularly in the beauty and personal care industry. Its unique algorithm and interactive features enable influencers to create dynamic, engaging content that drives visibility and interaction [1]. By prioritizing short, fresh, and genuine videos, TikTok allows influencers to connect deeply with audiences and reach beyond their follower base [2]. In this industry, tutorials, product demonstrations, and personal stories resonate with viewers ([3]), and tools like live Q&A sessions and product showcases further enhance engagement in real time. In Malaysia, TikTok contributed RM2.32 billion (USD 530 million) to e-commerce revenue in 2023, underscoring its role in consumer behavior and market growth [4]. Despite its importance, little research examines how TikTok's engagement metrics, such as likes, comments, and shares that could translate into revenue in this sector.

Addressing this gap is crucial academically because existing social media research predominantly focuses on platforms such as Instagram or YouTube, often overlooking TikTok's unique algorithmic and engagement dynamics, particularly within Southeast Asia. Practically, understanding this relationship provides influencers and brands with valuable data-driven insights to optimize their marketing and revenue strategies effectively. Existing studies often focus on platforms like Instagram or YouTube, overlooking TikTok's unique features and their impact [2], [5]. This study addresses these gaps by investigating the relationship between engagement rates and revenue, analyzing content types (live sessions, pre-recorded videos), and evaluating influencer categories (Celebrity, Macro, Meso, Micro). By focusing on TikTok's "For You Page" algorithm and content amplification, the study provides insights into how these features shape user interaction and revenue growth. Data from TikTok influencers in the beauty sector was analyzed using metrics such as likes, comments, and live session sales. Standardized methods ensured reliable insights into TikTok's content ecosystem. The study contributes quantitative evidence on how TikTok drives engagement and revenue, helping influencers and brands create effective, data-driven strategies. Following this introduction, the paper outlines the research objectives, reviews relevant literature, describes the methodology, presents results, and concludes with practical implications and future research recommendations.

Related Works

Social media platforms have reshaped how brands and influencers connect with consumers, especially in the beauty and personal care industry. TikTok, in particular, stands out due to its "For You Page" (FYP) algorithm, which can spotlight content even from newer creators based on engagement metrics such as likes, shares, and video completion rates. This shift from follower-based reach to content-driven visibility offers an environment where short-form videos can achieve rapid virality. However, while TikTok's potential is widely acknowledged, literature focusing specifically on how its user engagement correlates with revenue remains underexplored [6]. Existing research across multiple social media platforms generally supports the notion that higher engagement drives revenue-related outcomes. Yoon et al., 2018 found that engagement metrics such as likes, shares, and comments, foster brand loyalty and elevate purchase intentions, though their analysis revolves around Facebook. [8] further suggest that tutorial and review formats in the beauty segment boost viewer trust, often translating into higher sales. Yet, the direct path from engagement metrics to revenue is not always mapped out with precision. [9] argue that most engagement studies still prioritize qualitative indicators (likes or shares) without consistently tying these metrics to tangible outcomes such as product sales or brand partnerships.

TikTok's content distribution system offers a fresh lens on engagement-driven revenue. Unlike Instagram or YouTube, where follower counts largely determine reach, TikTok's algorithm can introduce videos to broad audiences if they spark immediate engagement [10]. This mechanism provides an opportunity for both emerging and established influencers to expand their reach quickly, creating potential for faster sales conversions [11]. Although a few studies acknowledge TikTok's value within broader social media strategies Fadil Bakri & Bakri, 2023, they often focus on top-level engagement metrics such as views, likes, and comments, without isolating how these indicators might translate into actual revenue streams. Over time, influencers have been categorized by follower size, with Micro and Meso influencers often credited for deeper audience connections, while Macro and Celebrity influencers boast wider reach [13]. [14] hold that smaller influencers tend to produce higher engagement and conversion rates, whereas Ashley & Tuten, 2015 argue that larger influencers capture broader visibility. These observations generally derive from Instagram and YouTube data, raising questions about whether TikTok's specific features alter the balance between engagement and revenue across different influencer tiers.

Although several researchers propose that social media engagement should correspond to revenue growth [6], [16], TikTok-specific studies often rely on small samples or mixed-platform data that blurs platform-specific effects [12]. Most fail to incorporate direct sales figures or comprehensive analytics, making it difficult to confirm how much a TikTok engagement spike boosts tangible outcomes like product purchases or brand collaborations. This gap underscores the need for more quantitative approaches that link specific TikTok metrics, such as FYP appearances, video completion rates, and hashtag challenge participation, to actual sales or revenue data. To build a fuller picture, scholars must examine how TikTok's short-form, fast-paced environment affects user decisions to buy or recommend products. Merging engagement data (shares, comments, watch time) with robust tracking of sales or referral links may offer clearer evidence of a cause-and-effect relationship. A refined, platform-focused methodology could also address whether influencer size or type significantly moderates the engagement-revenue pathway. For the beauty and personal care sector, which relies heavily on product demonstrations and user feedback, these insights are likely to shape how brands partner with different influencer tiers [13]. While past studies validate the role of social media engagement in steering revenue, they rarely disentangle the specific effects of TikTok's distinctive features. This gap points to the need for data-driven research that correlates TikTok-centric engagement metrics with actual sales figures. By aligning influencer-tier analyses with in-depth tracking of consumer purchase behavior on TikTok, future studies may clarify how short-form videos and interactive content truly impact revenue generation in the beauty and personal care space.

Research Objectives

The research objectives are as follows:

1) To quantitatively analyze the relationship between engagement rates (likes, comments, shares, saves, and plays) and revenue generation (Live GMV and Video GMV), identifying the specific metrics that most strongly correlate with higher revenue.
2) To evaluate and compare the effectiveness of different influencer categories (Celebrity, Macro, Meso, Micro) in generating engagement and revenue, clarifying how follower size and content strategy affect their relative strengths and weaknesses.
3) To establish actionable engagement benchmarks tailored specifically for TikTok influencers in Malaysia's beauty and personal care sector, guiding brands and creators in setting realistic and effective content performance targets.
4) To develop and compare machine learning models that predict influencer revenue based on TikTok engagement metrics, and to identify key features contributing to revenue performance.

2. METHODOLOGY

This study uses a method shown in Figure 1, adapted from the Data Science Trajectories (DST) model [17], which offers a flexible plan for data science projects; we customized the model to match our research goals by including steps like Business Understanding, Data Acquisition, Data Preparation, Modeling, and Result Exploration. The Business Understanding step, discussed in the Literature Review section, involved finding research goals and combining previous studies to build a theoretical base. In Data Acquisition, we gathered relevant data from TikTok using tools like Kalodata for revenue numbers and Apify for detailed engagement data. During Data Preparation, we selected and cleaned the data to ensure it was accurate, focusing on influencers who had at least 30 recent posts. In the Modeling phase, we used Python to perform both statistical and predictive analyses. This included correlation and comparative analysis to understand the relationship between engagement metrics and revenue, as well as machine learning algorithms to predict influencer revenue based on public metrics.

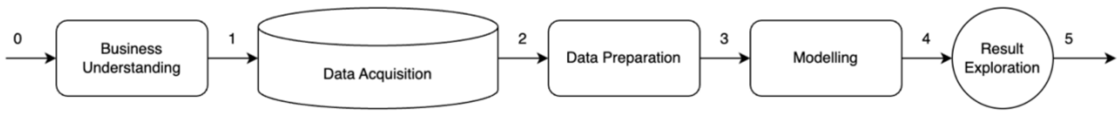

Figure 1. Research Methodology

The machine learning models applied were Random Forest Regressor, XGBoost Regressor, Gradient Boosting Regressor, and Ridge Regression. These models were chosen to evaluate predictive accuracy and identify which features were most influential in revenue estimation. Finally, Result Exploration involved interpreting the findings, spotting trends, setting engagement benchmarks, and giving recommendations for TikTok marketing strategies. The Modeling and Result Exploration steps are explained in detail in the Results and Discussion section. This structured approach ensures a thorough, data-driven analysis designed to meet the needs of the beauty and personal care industry, enhancing how practical the findings are.

2.1. Data Acquisition

This study focuses on collecting data about TikTok influencers in the beauty and personal care industry using two platforms, Kalodata and Apify. Kalodata, a platform known for providing insights into influencer revenue and business analysis [18], specializes in evaluating TikTok's business landscape. Apify, on the other hand, is a web-scraping tool that gathers detailed data directly from online sources [19]. The study found that the beauty and personal care sector was TikTok's highest revenue generator in September 2024, making it the ideal focus for this research. Influencers in this sector were carefully analyzed and divided into two main categories: "Individual/Personal" and "Organizational/Company". For this study, we selected the top 20 "Individual/Personal" influencers based on their earnings for further analysis.

Using Apify, a comprehensive dataset was created by collecting detailed information from the profiles of these selected influencers. A total of 3,913 records were compiled from posts made by these 20 influencers. To ensure the data sample was robust and relevant, the Apify scraper was configured to extract posts dated January 1, 2024, or later, with a maximum limit of 300 posts per influencer. The final dataset, titled "TikTok Influencers in Beauty and Personal Care", is publicly available on Kaggle [20], providing an open resource for further analysis and research.

2.2. Data Preparation

2.2.1. Influencer Selection Criteria

The collected data focused on follower counts, engagement per post, and posting frequency, forming the backbone of the analysis. To ensure consistency and relevance, we established strict criteria for influencer selection. Only influencers with at least 10,000 followers were considered, setting a baseline for audience reach [21]. Posting frequency, which often varied widely, also influenced the dataset's scope, ranging from 2 to 300 posts per influencer. However, only influencers with a minimum of 30 posts were included in the final selection. For those who met the threshold but had more than 30 posts, only their most recent 30 posts were analyzed. This narrowed the dataset to 17 influencers, a deliberate choice aimed at focusing on current engagement trends. By excluding outliers caused by exceptionally viral or underperforming posts, the method ensured a more balanced and accurate dataset. Concentrating on recent activity allowed the study to present a clearer view of engagement and revenue patterns, particularly among influencers with diverse posting frequencies. However, this study acknowledges methodological limitations, including the relatively small sample size of 17 influencers, potential unobserved variables such as audience demographics, and the cross-sectional nature of the dataset, which limits causal interpretations.

2.2.2. Variable Selection

To achieve the study's objectives, the analysis focused on selected variables from the dataset. Revenue-related metrics included Live GMV, representing the gross merchandise value from live sessions, and Video GMV, which captured the value generated through standard video posts. Engagement metrics, on the other hand, encompassed likes, comments, shares, saves, and plays for each post, providing a detailed view of audience interaction. Additionally, the follower count of each influencer was incorporated as a measure of their overall reach. These variables were chosen to explore and evaluate how influencer engagement connects to revenue generation, offering insights into the dynamics between audience activity and financial outcomes.

2.2.3. Influencer Categorization

To better understand how influencer reach, and engagement affected the outcomes, influencers were grouped into four categories based on their follower counts, as detailed in Table 1. This categorization is developed based on past work [22], [23], [24] to enable a comparative analysis across varying levels of reach, shedding light on the relationship between follower count, engagement, and revenue generation. By organizing influencers in this way, the study was able to identify patterns and highlight differences unique to each group, offering a clearer view of how reach impacts key metrics.

Table 1. Influencer's Categorization

Influencer Category	Follower Range
Celebrity	Over 1 million
Macro	100,000 to 1 million
Meso	50,000 to 100,000
Micro	10,000 to 50,000

2.2.4. Engagement Rate Calculation

The engagement rate is a key metric for assessing how effectively influencers connect with their audiences. It offers insights into the level of active participation their followers show in response to their content. Drawing on previous research [25], [26], [27], [28], [29], the engagement rate is calculated by dividing the total number of interactions, such as likes, comments, shares, plays, and saves, by the influencer's follower count. This measure helps evaluate the quality of audience engagement. Typically, a higher engagement rate signals a more responsive and interactive audience, suggesting that followers actively engage with the content rather than passively scrolling past it [29]. For brands, this distinction is critical, as it identifies influencers who can generate genuine interactions. Such influencers are particularly valuable for campaigns designed to influence purchasing behaviors through meaningful and authentic engagement.

2.3 Modelling

This section explains how the dataset was analyzed to find key insights and build a predictive model. It is divided into two parts: statistical analysis for descriptive insights, and machine learning for revenue prediction.

2.3.1. Statistical Modeling for Insight Extraction

The first part of the modeling process aimed to explore patterns within the dataset using descriptive analysis. Three key insights were derived from this step. The first insight involved examining the correlation between engagement metrics such as likes, comments, shares, saves, and play count with the revenue variable (LiveGmv). By checking the correlation values, it was possible to identify which engagement metrics had stronger relationships with influencer revenue. The second insight involved comparing influencers across different categories. The grouped influencer is then compared by the average revenue and engagement behavior. This helped to understand whether certain products or content categories were linked to higher monetization outcomes. The third insight focused on calculating engagement rate benchmarks. The engagement rate was computed using the formula of dividing the total number of interactions, such as likes, comments, shares, plays, and saves, by the influencer's follower count. This metric allowed for comparison of influencers based on how much interaction they generated relative to their follower base. Overall, these descriptive analyses provided a foundational understanding of the relationships and variations within the data before moving into predictive modeling.

2.3.2. Predictive Modeling Using Machine Learning

In the next part of the modeling process, a machine learning approach was used to estimate influencer revenue based on engagement metrics. This step expands the study beyond statistical description by introducing predictive modeling to test how well engagement data can be used to forecast revenue outcomes. The modeling phase began with several data preprocessing steps to prepare the dataset. Only relevant numerical features were selected as model inputs, which included likesCount, commentCount, shareCount, savesCount, playCount, and followers. These features were used to predict the target variable LiveGmv, representing influencer revenue from live sessions. To maintain data quality, rows with missing or invalid values in any of the selected columns were removed. Duplicate records were also eliminated to avoid skewing the results. All selected columns were converted into appropriate numerical data types to ensure compatibility with machine learning algorithms. Although tree-based models are typically robust to outliers, a basic inspection was carried out to detect any extreme values that could distort model training. These values were retained to reflect the natural variance present among influencers. For normalization, it was noted that tree-based models such as Random Forest, XGBoost, and Gradient Boosting do not require feature scaling. However, Ridge Regression does benefit from standardized inputs, so z-score normalization (transforming the features to have a mean of zero and a standard deviation of one) was applied specifically for that model. The final dataset was then split into 80% training data and 20% testing data. This split allowed the models to be evaluated fairly on unseen data. Four different machine learning models were trained and compared using the same dataset structure: Random Forest Regressor, XGBoost Regressor, Gradient Boosting Regressor, and Ridge Regression. Each model used the same engagement features and target variable, ensuring a consistent basis for performance comparison. The results of this modelling phase, including accuracy scores and feature importance, are discussed in the next section.

3. RESULT AND DISCUSSION

3.1. Insight 1

The first objective of this study examines the relationship between engagement rates and revenue in TikTok's beauty and personal care sector. As shown in the correlation matrix (Figure 2), there is a weak negative correlation between engagement and revenue, with a coefficient of -0.39. Surprisingly, this suggests that higher engagement rates do not necessarily lead to increased revenue for influencers in this sector. One possible explanation is that highly engaging content often prioritizes entertainment or community-focused interactions over direct product promotion. Influencers creating engaging content may focus on emotional or personal connections, such as humorous sketches, trending dance routines, or relatable storytelling, which captivate audiences but do not necessarily motivate purchases. However, it is crucial to acknowledge additional variables potentially influencing this relationship, including audience demographics (age, income levels, cultural background), product types, influencer content styles, pricing strategies, and seasonal promotional activities. These factors, although not explored in this study due to data constraints, could significantly shape the nuanced relationship between engagement and revenue.

Figure 2. Correlation Matrix

Conversely, influencers who focus more on driving sales may produce content that feels like an advertisement. While this type of content might lower engagement, resulting in fewer likes or comments, it can still generate significant revenue by directly targeting potential buyers. Promotional posts, though less personal or authentic, may effectively push interested viewers to make purchases. This highlights that engagement rates, while useful for measuring audience interaction, are not straightforward indicators of purchasing behavior in this context. Several other factors likely influence this complex relationship. Variables such as the type of products being promoted, the perceived authenticity of influencer endorsements, pricing strategies, and audience demographics (including spending habits and disposable income) can all impact whether engagement translates into sales.

Interestingly, revenue from live sessions (Live GMV) shows a stronger positive correlation with overall revenue, with a coefficient of 0.51. This suggests that live, interactive content may be more effective at driving immediate sales. During live sessions, influencers can engage directly with their audience, answer questions, demonstrate products, and offer time-limited deals. This real-time interaction fosters trust and creates a sense of urgency, encouraging instant purchases. In contrast, revenue from regular video posts (Video GMV) exhibits a small positive correlation with revenue (0.10) but has an almost negligible correlation with engagement rate (0.08). Notably, there is a strong negative correlation (-0.81) between Live GMV and Video GMV. This finding implies that influencers who perform strongly in live-based sales may not necessarily rely on or generate substantial revenue from pre-recorded videos, and vice versa. It may also suggest a trade-off in how influencers allocate their efforts between live streaming and video content, highlighting the importance of strategic planning to optimize overall revenue.

These findings highlight the complicated relationship between engagement and revenue. While engagement helps influencers grow their audience and foster strong connections, it does not always translate into higher sales. To succeed, influencers may need to strike a balance between creating engaging content and using effective sales strategies. Live sessions, in particular, seem to bridge the gap between interaction and revenue more effectively than other formats, and incorporating them into a content strategy could be a smart move for influencers aiming to build long-term brands and boost sales. Regularly hosting live streams not only keeps followers interested but also builds a sense of community, which can strengthen brand loyalty. By offering exclusive discounts or limited-time deals during these sessions, influencers create urgency that encourages immediate purchases. However, finding the right balance is crucial. While promotional content plays a key role in generating sales, it should be carefully integrated with authentic and engaging posts to maintain high levels of audience interaction without coming across as overly commercial. Content that resonates with the unique preferences of followers can help influencers connect emotionally while positioning themselves as relatable personalities and trusted product advocates. By understanding and adjusting to these subtle dynamics in audience behavior and content performance, influencers can refine their strategies to better align with their revenue goals.

3.2. Insight 2

The second objective examines how different types of influencers (Celebrity, Macro, Meso, and Micro) compare in terms of average engagement rates and revenue within TikTok's beauty and personal care sector. As shown in Table 2, Celebrity influencers stand out with the highest average revenue, with the amount of RM475,442.85. However, their engagement rates remain relatively moderate at 10.19. This trend suggests that although their large audience size enables substantial revenue generation, their content might not resonate with followers as strongly as that of influencers in smaller categories. It highlights a trade-off: while extensive reach contributes to financial success, it does not necessarily translate into deeper audience interaction.

Table 2. Average Revenue and Engagement Rate

Influencer Category	Revenue (RM)	Engagement Rate (%)
Celebrity	475,442.85	10.19
Macro	421,502.50	9.63
Meso	357,672.00	22.05
Micro	322,980.00	3.83

Macro-influencers follow closely, with RM421,502.50 in revenue and an engagement rate of 9.63%, offering a balance between reach and consistent engagement. This makes them adaptable for campaigns seeking visibility while still encouraging interaction. Meso-influencers, however, stand out with the highest engagement rate at 22.05%, despite a lower average revenue of RM357,672. Their strong engagement indicates that they foster deeper connections with their audience, often building loyalty through community-focused content. Although this high engagement does not automatically translate into higher revenue, it suggests that other factors may bring about the link between follower interaction and sales. Micro-influencers, with the smallest revenue at RM322,980 and a 3.83% engagement rate, tend to have more targeted audiences, making them suitable for brands aiming at specific niches. Yet, their limited reach and engagement hint at a reduced capacity for broad revenue generation. Bar charts in Figures 3 (a) and (b) visually represent the average revenues and engagement rates across these categories, highlighting differences in influence and audience interaction.

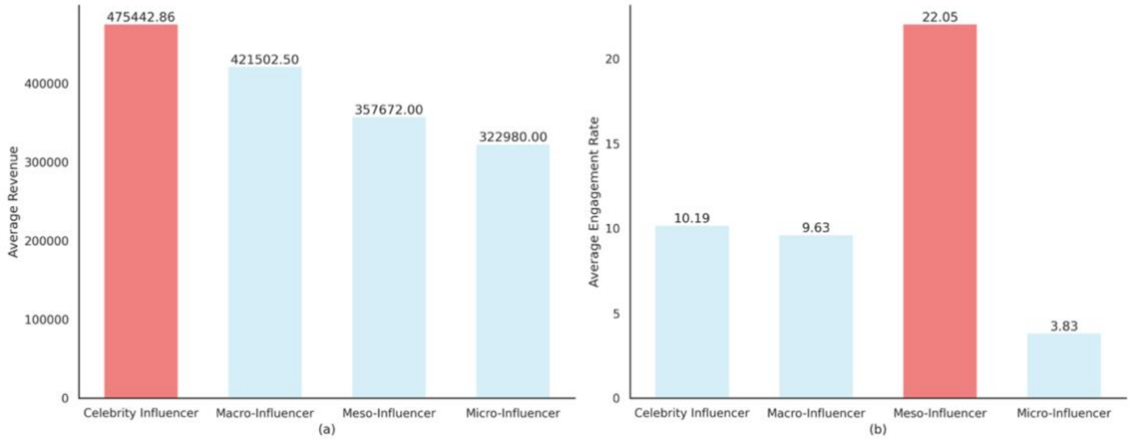

Figure 3. Average Revenue and Engagement Rate per Influencer Category

The unique impact of live sessions on revenue generation is particularly noteworthy here. While the quantitative data confirms that live sessions drive significant revenue, adding qualitative insights can provide a bigger picture. The quantitative data confirms the significant revenue potential of live sessions, but more specific strategies can enhance their effectiveness. In live sessions, influencers interact directly with their audience in real-time, creating a setting that encourages immediate engagement and influences buying behavior. By addressing questions on the spot, influencers can quickly resolve consumer doubts and provide reassurance about their products. Demonstrating products live not only builds trust but also highlights their practical benefits in a convincing and authentic way. Influencers and brands can optimize live sessions with the following actionable strategies: Regular Scheduling: Setting a consistent schedule for live sessions builds anticipation and helps retain a loyal viewership; Interactive Elements: Adding Q&A segments, polls, and direct engagement boosts viewer participation and creates a dynamic experience; Exclusive Offers: Special promotions or discounts available only during live streams to encourage immediate buying decisions; Tailored Content: Aligning live session content with audience interests and feedback enhances relevance and effectiveness.

These strategies emphasize the importance of live sessions not just for their engagement potential but for their ability to convert passive viewers into active buyers. In contrast, while standard video posts are effective for building visibility and general audience interaction, they lack the dynamic, real-time feedback loop that live sessions offer. Delayed responses to questions or comments on regular videos can dampen the immediacy needed for impulsive buying decisions, making them less effective for driving quick sales. In summary, while follower counts and engagement rates are important, the content type and quality of interaction also heavily impact revenue generation. Live sessions, with their real-time communication and ability to connect deeply with audiences, represent a powerful tool for bridging the gap between engagement and revenue.

3.3. Insight 3

The third objective focuses on setting benchmarks to help influencers establish realistic engagement targets based on their follower count and potential for interaction. As shown in Figure 5, the mean and median engagement rates for each influencer category reveal distinct patterns in audience behavior. For Celebrity influencers, the mean engagement rate stands at 10.19%, while the median drops significantly to 6.26%. This gap indicates a positive skew, where a handful of highly successful posts inflate the average, while most posts generate more moderate engagement. Such variability suggests that viral or trend-driven content plays a significant role in boosting engagement within this group, but consistency remains a challenge.

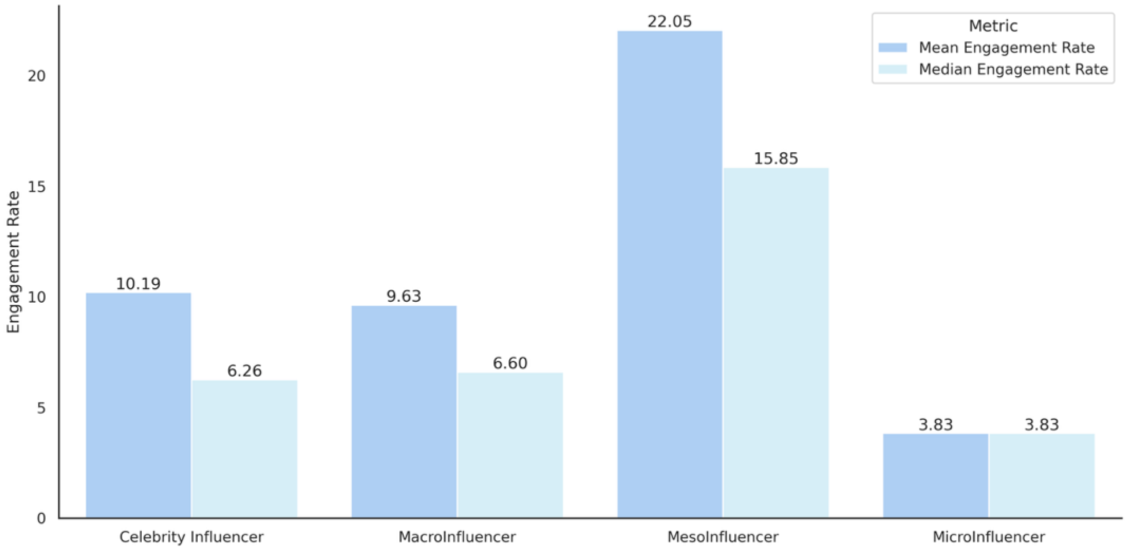

Figure 4. Mean and Median for Engagement Rate

Macro-influencers show a similar trend, with a mean engagement rate of 9.63% and a median of 6.60%. This smaller gap indicates more consistent engagement, though occasional spikes in interaction still occur. The slightly higher median compared to Celebrity influencers suggests that Macro-influencers may maintain a closer connection with their audience, striking a better balance between broad reach and follower interaction. Meso-influencers stand out with the highest engagement rates, averaging a mean of 22.05% and a median of 15.85%. While there is still a noticeable difference between the two, these higher numbers reflect their ability to generate consistently strong engagement. This suggests that Meso-influencers excel at fostering community and loyalty among their followers, making them particularly effective at sustaining audience interaction.

Micro-influencers, by contrast, show remarkable uniformity, with both the mean and median engagement rates at 3.83%. While their overall engagement levels are lower, the lack of variability highlights a dedicated audience that interacts consistently across posts. This steady interaction is particularly valuable for brands seeking predictable performance within niche markets, where reliability is often more important than scale. The benchmarks outlined in Table 3 provide influencers with practical engagement targets tailored to their category. For instance, Celebrity influencers can focus on balancing viral campaigns with consistent, relatable content like behind-the-scenes posts. Macro-influencers, with their stable engagement rates, might prioritize building authenticity through polls or "day-in-the-life" content. Meso-influencers, due to their community-driven success, should emphasize loyalty programs, personalized rewards, and interactive campaigns. Micro-influencers, given their niche appeal, can build trust by engaging directly with followers through comments and Q&A sessions. These tailored approaches can help influencers refine their strategies to align with their engagement potential.

Table 3. Engagement Rate Benchmark

Influencer Category	Description	Reason
Celebrity (6.26)	Sets a steady target for consistent engagement, focusing on regular interaction rather than occasional high spikes.	A stable engagement target helps maintain brand presence without relying on viral content, making it more predictable for long-term planning.

Influencer Category	Description	Reason
Macro (6.60)	Aims to keep engagement steady through genuine content that holds audience interest without big ups and downs.	Consistent engagement supports reliable audience connection, balancing reach with meaningful interaction.
Meso (15.85)	A strong benchmark for building loyal audience engagement, helping to create stable and trusting relationships.	High engagement indicates strong follower loyalty, making Meso-influencers ideal for community-building and trust-focused campaigns.
Micro (3.83)	Provides a realistic target for stable engagement in niche audiences, supporting growth within specialized communities.	Reliable engagement among smaller audiences allows for focused, niche-specific growth, beneficial for highly targeted brands or products.

3.4. Insight 4

This section presents the results of the machine learning models used to predict influencer revenue (LiveGmv) based on public engagement metrics. The goal was to assess whether features such as likes, comments, shares, saves, play count, and follower count could be used to estimate revenue outcomes effectively. Four different regression models were tested using the same dataset and train-test split: Random Forest Regressor, XGBoost Regressor, Gradient Boosting Regressor, and Ridge Regression. Each model was evaluated based on two performance metrics: the coefficient of determination (R^2) and Mean Absolute Error (MAE), which measures the average difference between predicted and actual revenue. Table 4 below summarizes the results.

Table 4. Summary of all Machine Learning Model results

Model	R^2 Score	MAE (RM)
Random Forest Regressor	0.94	13,095
XGBoost Regressor	0.93	15,592
Gradient Boosting	0.91	16,451
Ridge Regression	0.26	54,180

Among the four models, the Random Forest Regressor achieved the highest R^2 score and the lowest MAE, indicating it had the best predictive accuracy. Therefore, it was selected as the final model to support the study's revenue estimation component. Feature importance analysis from the Random Forest model showed that the most influential variables in predicting revenue were the number of followers, followed by comment count, like count, and play count. This finding supports earlier statistical insights and confirms that both engagement and audience size play a significant role in revenue generation on TikTok. These results demonstrate that machine learning can be used to make reliable revenue predictions using only public-facing metrics. It also shows that non-linear models like Random Forest are more suitable for capturing the complex relationships in influencer marketing data.

4. CONCLUSION

This study provides valuable insights into the relationship between engagement rates and revenue generation in TikTok's beauty and personal care sector. The results demonstrate that while engagement metrics such as likes and shares are essential indicators of audience interaction, they do not uniformly translate into higher revenue. This distinction underscores the complexity of balancing content designed to entertain and engage versus content aimed at driving direct sales. A key contribution of this study lies in highlighting the unique role of live sessions in influencing immediate purchasing behavior. The real-time interaction of live sessions offers influencers and brands a strategic platform to convert engagement into sales more effectively. Additionally, actionable strategies such as interactive Q&A sessions and time-sensitive promotions were identified as practical ways to maximize the potential of live sessions. By establishing engagement benchmarks for different influencer categories, this study provides a practical framework for influencers and brands to set realistic performance targets. The tailored benchmarks serve as a roadmap for optimizing both engagement and revenue outcomes. This study further extends its contribution by incorporating predictive modeling. Using various machine learning algorithms, including Random Forest, XGBoost, Gradient Boosting, and Ridge Regression, the study tested the ability to predict influencer revenue using engagement metrics. Among these, Random Forest achieved the best performance (R^2 = 0.94), supporting the potential of public metrics in forecasting revenue. This modeling approach enhances the study's relevance by bridging descriptive insights with predictive capabilities, offering more strategic value to both academics and practitioners.

However, this study also presents several methodological limitations. The relatively small sample size of 17 influencers restricts generalizability, and factors such as content type, audience demographics, and seasonal trends remain unexplored due to data constraints. Additionally, the cross-sectional nature of the dataset limits the ability to infer causality between engagement and revenue metrics. Future research could address these limitations by expanding the sample size to include a broader range of influencers and product categories, incorporating content-level variables and platform-specific metrics such as video completion rates, 'For You Page' visibility, and content virality indicators. Differentiating between engagement patterns in live sessions and regular videos could also provide deeper insights. Longitudinal or experimental designs could help clarify causal relationships, and exploring long-term outcomes like brand loyalty and equity would further strengthen strategic decision-making for influencers and marketers. These extensions would improve the originality and depth of analysis in understanding TikTok's algorithmic environment and commercial potential.

ACKNOWLEDGEMENTS

We would like to acknowledge the grant funder, Yayasan Universiti Teknologi PETRONAS (YUTP) [015LC0-523] for funding this research paper.

REFERENCES

[1] A. Bhandari and S. Bimo, "TIKTOK AND THE 'ALGORITHMIZED SELF': A NEW MODEL OF ONLINE INTERACTION," *AoIR Selected Papers of Internet Research*, Oct. 2020, doi: 10.5210/spir.v2020i0.11172.

[2] S.-C. Chu, T. Deng, and J. Mundel, "The impact of personalization on viral behavior intentions on TikTok: The role of perceived creativity, authenticity, and need for uniqueness," *Journal of Marketing Communications*, vol. 30, no. 1, pp. 1–20, Jan. 2024, doi: 10.1080/13527266.2022.2098364.

[3] S. V. Jin, A. Muqaddam, and E. Ryu, "Instafamous and social media influencer marketing," *Marketing Intelligence & Planning*, vol. 37, no. 5, pp. 567–579, Jul. 2019, doi: 10.1108/MIP-09-2018-0375.

[4] S. Kemp, "Digital 2024: Malaysia," Data Reportal. Accessed: Aug. 07, 2024. [Online]. Available: https://datareportal.com/reports/digital-2024-malaysia

[5] R. Tariq, P. Isawasan, L. Shamugam, M. A. H. Ahmad Asmawi, N. A. Nor Azman, and I. A. Zolkepli, "Exploring Social Media Research Trends in Malaysia using Bibliometric Analysis and Topic Modelling," *ICST Transactions on Scalable Information Systems*, vol. 12, Jan. 2025, doi: 10.4108/eetsis.7003.

[6] A. Rejeb, K. Rejeb, A. Appolloni, and H. Treiblmaier, "Foundations and knowledge clusters in TikTok (Douyin) research: evidence from bibliometric and topic modelling analyses," *Multimed Tools Appl*, vol. 83, no. 11, pp. 32213–32243, Mar. 2024, doi: 10.1007/s11042-023-16768-x.

[7] G. Yoon, C. Li, Y. (Grace) Ji, M. North, C. Hong, and J. Liu, "Attracting Comments: Digital Engagement Metrics on Facebook and Financial Performance," *J Advert*, vol. 47, no. 1, pp. 24–37, Jan. 2018, doi: 10.1080/00913367.2017.1405753.

[8] S. Gupta and R. Dutt, "Identifying consumer-based digital content marketing consumption motives: a qualitative study," *Journal of Advances in Management Research*, Sep. 2024, doi: 10.1108/JAMR-08-2023-0218.

[9] F. Ellis-Chadwick and D. Chaffey, *Digital marketing: strategy, implementation and practice*. Pearson, 2012.

[10] D. Delmonaco, S. Mayworm, H. Thach, J. Guberman, A. Augusta, and O. L. Haimson, "'What are you doing, TikTok?': How Marginalized Social Media Users Perceive, Theorize, and 'Prove' Shadowbanning," *Proc ACM Hum Comput Interact*, vol. 8, no. CSCW1, Apr. 2024, doi: 10.1145/3637431.

[11] V. Rasmussen and L. Grure, "# Beautytok going viral," University of Agder, 2023.

[12] Z. Fadil Bakri and Z. F. Bakri, "Analyzing the Influence of Digital Marketing Strategies on Business Performance in the Beauty Industry: A Comprehensive Analysis of Social Media Engagement and Influencer Collaborations," 2023. [Online]. Available: www.plus62.isha.or.id/index.php/JEMBUT

[13] B. E. Duffy, "Social Media Influencers," in *The International Encyclopedia of Gender, Media, and Communication*, Wiley, 2020, pp. 1–4. doi: 10.1002/9781119429128.iegmc219.

[14] C. Lou and S. Yuan, "Influencer Marketing: How Message Value and Credibility Affect Consumer Trust of Branded Content on Social Media," *Journal of Interactive Advertising*, vol. 19, no. 1, pp. 58–73, Jan. 2019, doi: 10.1080/15252019.2018.1533501.

[15] C. Ashley and T. Tuten, "Creative Strategies in Social Media Marketing: An Exploratory Study of Branded Social Content and Consumer Engagement," *Psychol Mark*, vol. 32, no. 1, pp. 15–27, Jan. 2015, doi: 10.1002/mar.20761.

[16] L. Theodorakopoulos and A. Theodoropoulou, "Leveraging Big Data Analytics for Understanding Consumer Behavior in Digital Marketing: A Systematic Review," *Hum Behav Emerg Technol*, vol. 2024, no. 1, Jan. 2024, doi: 10.1155/2024/3641502.

[17] F. Martinez-Plumed *et al.*, "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," *IEEE Trans Knowl Data Eng*, vol. 33, no. 8, pp. 3048–3061, Aug. 2021, doi: 10.1109/TKDE.2019.2962680.

[18] "Kalodata." Accessed: Sep. 28, 2024. [Online]. Available: https://www.kalodata.com/category

[19] "Apify." Accessed: Nov. 05, 2024. [Online]. Available: https://apify.com/about

[20] M. A. H. Ahmad Asmawi and P. Isawasan, "Top 20 TikTok Beauty & Personal Care Influencers," 2024, *Kaggle*. doi: 10.34740/KAGGLE/DS/6111672.
[21] E. A. van Reijmersdal, T. D. Aguiar, and G. van Noort, "How Is Influencer Follower Size Related to Brand Responses?," *J Advert Res*, vol. 64, no. 4, pp. 451–464, Dec. 2024, doi: 10.2501/JAR-2024-024.
[22] E. Djafarova and C. Rushworth, "Exploring the credibility of online celebrities' Instagram profiles in influencing the purchase decisions of young female users," *Comput Human Behav*, vol. 68, pp. 1–7, Mar. 2017, doi: 10.1016/j.chb.2016.11.009.
[23] J. Marti-Ochoa, E. Martin-Fuentes, and B. Ferrer-Rosell, "Airbnb on TikTok: Brand Perception Through User Engagement and Sentiment Trends," *Soc Sci Comput Rev*, 2024, doi: 10.1177/08944393241260242.
[24] S. C. Boerman, "The effects of the standardized instagram disclosure for micro- and meso-influencers," *Comput Human Behav*, vol. 103, pp. 199–207, Feb. 2020, doi: 10.1016/j.chb.2019.09.015.
[25] A. A. Arman and A. Pahrul Sidik, "Measurement of Engagement Rate in Instagram (Case Study: Instagram Indonesian Government Ministry and Institutions)," in *2019 International Conference on ICT for Smart Society (ICISS)*, IEEE, Nov. 2019, pp. 1–6. doi: 10.1109/ICISS48059.2019.8969826.
[26] C. Chartier, A. A. Chandawarkar, D. J. Gould, and W. G. Stevens, "Insta-Grated Plastic Surgery Residencies: 2020 Update," *Aesthet Surg J*, vol. 41, no. 3, pp. 372–379, Mar. 2021, doi: 10.1093/asj/sjaa172.
[27] J. Vassey *et al.*, "E-cigarette brands and social media influencers on Instagram: a social network analysis," *Tob Control*, vol. 32, no. 2 e, pp. E184–E191, Aug. 2023, doi: 10.1136/tobaccocontrol-2021-057053.
[28] R. L. H. Yew, S. B. Suhaidi, P. Seewoochurn, and V. K. Sevamalai, "Social Network Influencers' Engagement Rate Algorithm Using Instagram Data," in *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)*, IEEE, Oct. 2018, pp. 1–8. doi: 10.1109/ICACCAF.2018.8776755.
[29] E. Yost, T. Zhang, and R. Qi, "The power of engagement: Understanding active social media engagement and the impact on sales in the hospitality industry," *Journal of Hospitality and Tourism Management*, vol. 46, pp. 83–95, Mar. 2021, doi: 10.1016/j.jhtm.2020.10.008.