Social Media Analysis

GROUP 10

Lorenzo Cecconi - 10722598
Luca Panti - 10707591

Prof. Lucio Lamberti
Prof. Gloria Peggiani
Prof. Michele Di Dalmazi

Marketing Analytics
A.Y. 2023-2024

Agenda

Objectives and Expected Results
Set the stage by clearly outlining the objectives and anticipated results

New Variables
Introduce new insightful variables to fully extrapolate the key drivers empowering the social media phenomena

Data Exploration
Showcase key descriptive statistics, visualizations, and trends that provide a comprehensive view of the data landscape

Clustering
Identify relevant clusters, their characteristics, and the implications for business decision-makers.

Regression Analysis
Leverage regression analysis to evaluate models explaining the degree of engagement through significant variables

Conclusions
Summarize findings useful for companies' marketing teams and specify the limitations of the presented models

Objectives

- Categorize TikTok and Instagram content to enhance statistical analysis of the key features characterising them
- Identify creative lines and innovative patterns to distinguish the current approaches of social media platforms
- Understanding the key drivers of engagement:
  - Determine the main differences between drivers in sponsored and non-sponsored content
  - Highlight the underlying differences between TikTok and Instagram
- Provide business decision-makers with insightful suggestions to approach innovative marketing tools
- Compare the statistical results with real user perspectives

Expected Results

- Contents with high levels of information and Creativity stimulate positively engagement levels
- The Body Display and Posting frequency is driving the engagement up to a certain level
- TikTok and Instagram are saturated with positive and happy content: likely sad posts are driving engagement more effectively
- Posting on weekends provides better results
- The authenticity of the influencer has a key role when posting advertised content

New Variables

Variable	Type	Tool used	Description and Variable creation
Emotion	Categorical	AWS Rekognition	The variable can assume 8 values: Happy, Sad, Angry, Confused, Disgusted, Surprised, Calm and Fear. As in some videos, the tool recognised more emotions, the one with a higher level of confidence and that appeared the most among all the frames of the video was taken into account
Subtitles	Categorical (0/1)	AWS Rekognition	The variable was set to 1 when the tool recognised text in the video, 0 otherwise
Body Display	Numerical (1-7)	AWS Rekognition	The variable was set according on how many elements of body display the tool was able to recognise (ex. shorts, underwear, lingerie,...)
Emoji Count	Numerical	Python	Using Python it was possible to extract the number of emojis in the descriptions of the posts/videos
Followers	Numerical	Manually	The number of followers per influencer was taken manually, looking at the current number for the creators in the dataset (there might be a difference between the number of followers used (January 2024) and the number of followers when the posts/videos were published (between Sept 2022 and Nov 23))

New Variables

Variable	Type	Tool used	Description
isWeekend	Categorical (0/1)	RStudio	Converting the given dates in the JSON files into date format, it was possible to identify if the content was published during the weekend (Sat and Sun) or not
Average Posts per Week	Numerical	RStudio	Given the date of the publication of the posts/videos, it was possible to calculate the average number of publications per week
Frenetic Pace of the Video	Numerical (1-7)	Manually	A first calculation of the variable was done using the AWS Rekognition tool, but the results were not reliable. Therefore the variable was computed manually watching the videos
Gender	Categorical	AWS Rekognition	The Gender was assigned, only to TikTok videos, based on the labels provided by the image processing tool of AWS Rekognition. When both the male and female labels were present the gender was set to *Both*
Post Positivity	Numerical	Python/Google Cloud API	The Post Positivity variable was computed by translating the description of the items in English through a Google Cloud API, then all the emojis were converted into text (as they could have been relevant in measuring the positivity). A Python package was used to count the “positive” and “negative” words, finally, the Post Positivity was defined as in the * paper : $\frac{PositiveWords - NegativeWords}{PositiveWords + NegativeWords + 1}$

Leung, F. F., Gu, F. F., Li, Y., Zhang, J. Z., & Palmatier, R. W. (2022). Influencer Marketing Effectiveness Journal of Marketing, 86 6 93 115 https ::/doi org/ 10 1177 00222429221102889

Data Exploration

To gain a comprehensive understanding of the dataset's variables, an **exploratory data analysis** (EDA) was conducted. Several plots were generated, providing insights into the distribution of various factors. Notably, the target variable exhibited a concentration of observations towards the lower end of the graph. Consequently, a **logarithmic transformation** (*Figure 1*) was applied to this variable. This transformation not only helped in spreading out the data but also facilitated the detection of outliers. In total, after removing some observations due to missing and not recoverable data, **340 observations**, of which 80 were advertised, were considered for Instagram, and **488 observations**, of which 120 were advertised, for TikTok.

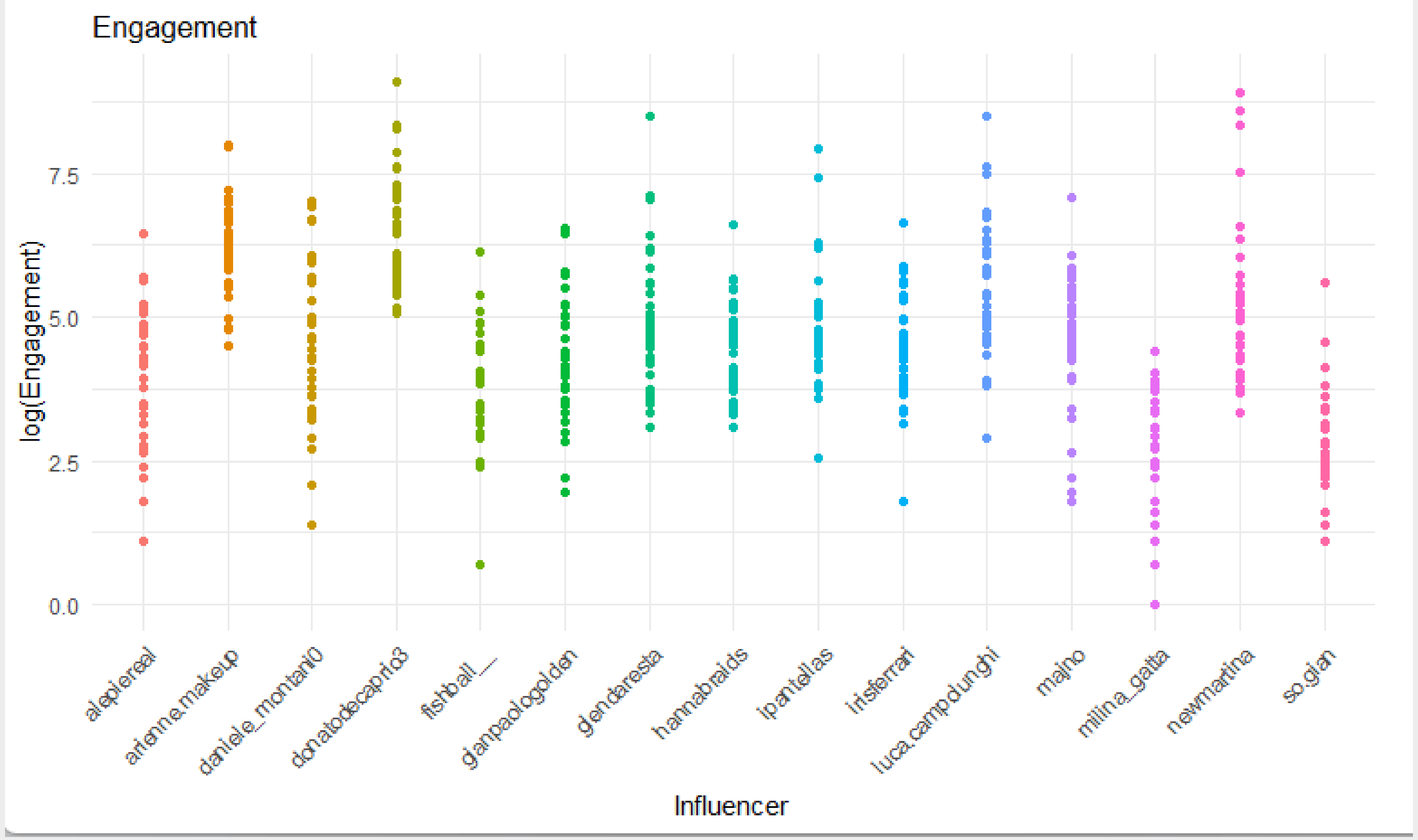


Figure 1: Comments count for TikTok content creators

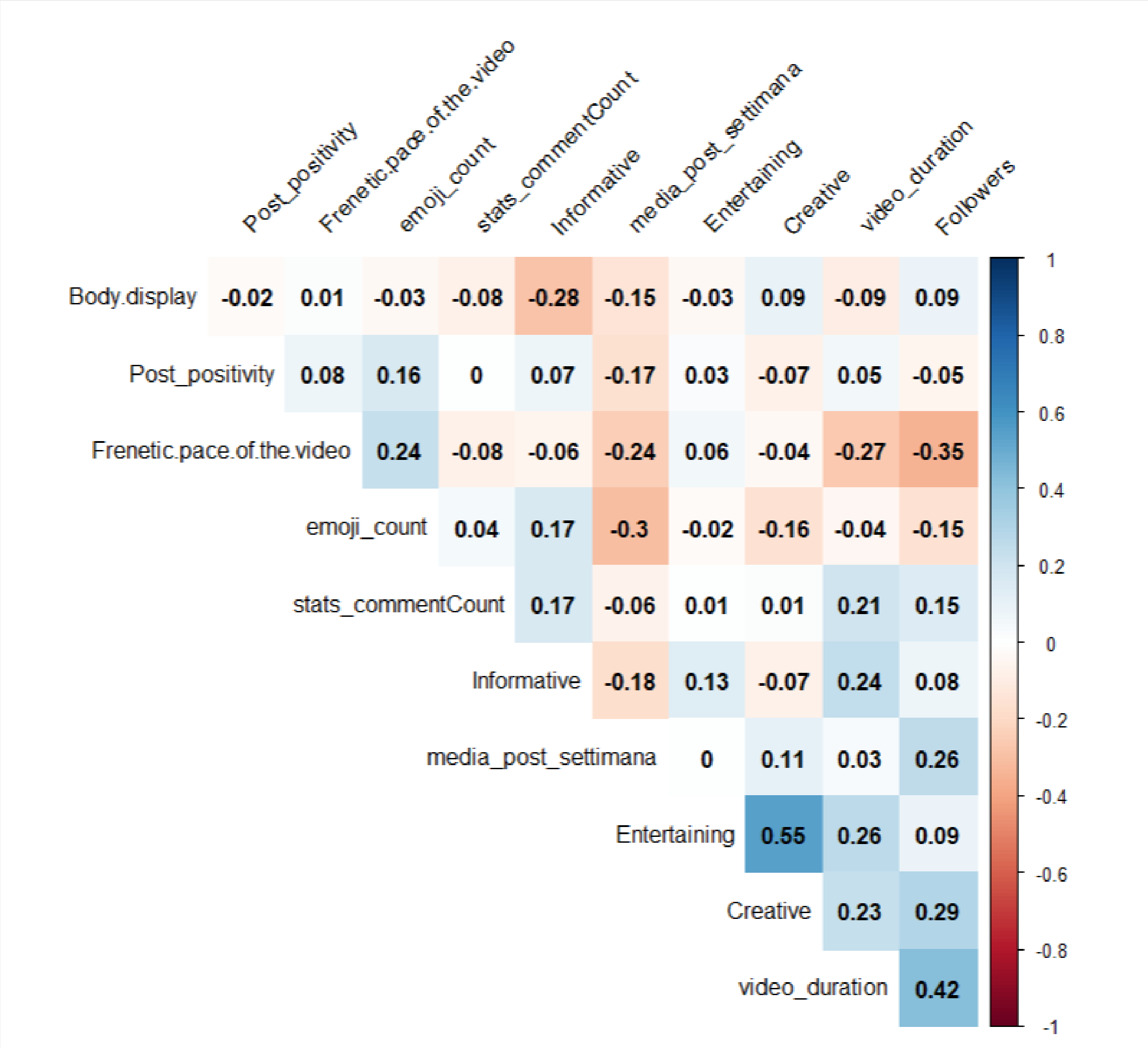


Figure 2: Correlation matrix of the numerical variables of the TikTok dataset

The dataset comprises numerous numerical variables, for which it was created a **correlation matrix** (*Figure 2*) to assess collinearity and understand the relationships between these variables. An important decision in our analysis was whether to employ **Principal Component Analysis** (PCA) for dimensionality reduction. While PCA offers benefits such as reducing the number of variables and eliminating collinearity, it also poses a risk of diminishing the interpretability of these variables. In practice, when applying PCA to the numerical variables, the principal components became less interpretable. Moreover, the reduction in the number of variables was not substantial if trying to keep a certain level of explained variability. Consequently, instead of proceeding with a PCA, irrelevant variables were **iteratively removed** from the clustering and linear regression models, to obtain significant results.

TikTok - Clustering

The **K-Prototypes Clustering** method was selected, on top of the more common K-Means, to incorporate also the categorical variables of the dataset. The first clustering on the dataset was performed considering both advertised and non-advertised items, to better capture the overall creative styles of Italian influencers on the platform. Later on a focus on the advertised content was carried out, to address the key features used in promotional activities.

To more accurately depicture the current approaches on Tiktok, the following variables were included:

- **Numerical variables**: Video_duration, Entertaining, Informative, Creative, Frenetic pace of the video, Body display, Emoji count, Post positivity, Average posts per week and Followers;
- **Categorical variables**: Verified_author, Adv/NonAdv, Creative_content, Pers_update_opinion, Subtitles, Gender, Emotion_label, isWeekend.

The optimal number of clusters was selected calculating the **Average Silhouette Width** (*Figure 3*) iteratively for different number of clusters, leading to a **K* = 5**. Also an optimal **lambda = 2.173574** was found, to optimize the algorithm clustering balancing of numerical and categorical data.

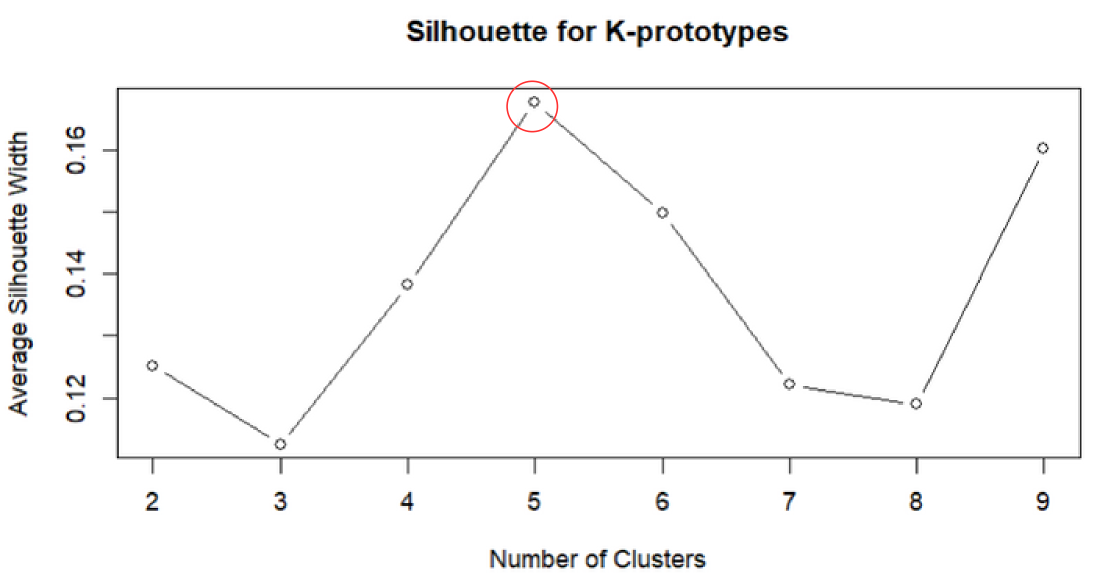


Figure 3: Avg Silhouette Width for Tiktok Clustering

Once the algorithm performed the clustering in 5 potential groups, it was noticed that some of the variables weren't significantly changing among the different clusters, such thing was also confirmed by performing a **Fisher's test**. One of them is the Emotion Lable variable which was predominantly ***Happy*** for each one of the identified clusters. The reasons behind this result may be the actual higher level of "Happy" and joyful content in the whole platform, or due to the tendency of the AWS Rekognition tool to more easily identify happy faces rather than other emotions.

Following the results of the algorithm are reported. Together with identifying the main creative lines of the videos, it was possible also to identify how **some Tiktokers are a clear representation of a specific cluster** (they are reported in the corresponding cluster), perfectly reflecting the key features of a certain creative line. Other influencers are more varying their artistic flair among different styles.

Clusters

PromoPositivity

This cluster includes all those tiktoks where mainly **males** are present and focus on **advertising** a product. The content is very **informative** of the characteristics of the product but **not much creative**. A **positive mood** is the key underpinning for this creative line. **Representative Influencers**: Donato de Caprio and GianPaoloGolden

RapidReports

Reels belonging to this creative style offer the viewers **daily updates** on the **influencers lifestyle**. Due to the very **high frequency** of publication of these videos, they don't have a high attention to details. For this reason, it's much more common to identify **sad/negative attitude** when reporting personal daily updates. **Representative Influencers**: MilinaGatta

EmpowerPhysique

These energetic contents leverage the **physical presence** as a form of expression, showcasing empowerment through body confidence. Despite a **lower emphasis on informative content**, the goal of these reels is to captivate the viewers with visually striking content centred around body display. **Representative Influencers**: Fishball, Iris Ferrari and Majno

DynamoCreations

The focus is set around high-energy and **frenetic** content creation. **Male** presence is a must in these kind of videos, delivering **rapid and engaging** content that captivates the audience. Typical contents of this cluster are jokes, making fun of weird situations and challenges among the creators. **Representative Influencers**: AlePiereal, Luca Campolunghi, iPantellas and so.gian

CreativeCharm

Despite the **high level of creativity and entertainment** value, the **pace of the videos** remains deliberately **relaxed and unhurried**, offering viewers a chance to savor the imaginative content. *Creative Charm* encapsulates a captivating concept for TikTok reels crafted mainly by **female** influencers, not only reflecting the influencers' ability to produce engaging and original material, but also emphasising their dedication to provide a more leisurely and enjoyable viewing experience on TikTok. **Representative Influencers**: NewMartina

Business Insights

This wide range of creative lines allows potential businesses to target one or more specific content types or influencers to fully exploit the marketing opportunities provided by the platform. Some of the presented creative themes, such as *CreativeCharm*, are typically adhered by **more famous** (higher followers) content creators, but this doesn't necessarily translate into higher engagement with the community, as it'll be shown in the Regression analysis. Still, a company willing to advertise its products or services on Tiktok can identify a **creator in line with its value drivers**, able to empower the features and strengths the product would already have on its own. For example, a company promoting a **daily-use product** could easily integrate its item in a *RapidReport* style reel, while a **make-up product** could be more suitable with an *EmpowerPhysique* content. Finally **further analyses** of the target audience of the influencers should be performed by the company, to confirm if the product, not only is in line with the creative style of the TikToker, but also with the characteristics of the potential customers.

TikTok Adv - Clustering

A further analysis of the TikTok dataset was done through a clustering of the **advertised content** included in the given dataset. Given the output of the first clustering, it could be noticed how the advertising or not variable was significant in describing only one of the identified clusters (*PromoPositivity*). This could have led to performing promotional activities only when the product and company were in line with the creative style of that specific cluster. However, this new clustering aims to identify whether there are **different creative lines also among the advertised content**, to **better suggest companies** when deciding which influencers entrust their promotions.

A few modifications of the selected variables was done to better capture the differences among the clusters:

- **Added variables**: Control_content and Control_product (these two variables were excluded from the previous clustering due to the high number of NAs present in the whole dataset, while they become very important in the advertising context);
- **Removed variables**: Adv (not relevant anymore as always equal to 1), Subtitles (almost all advertised content didn't include subtitles) and Average posts per week.

The optimal number of clusters was still selected calculating the Average Silhouette Width iteratively for different number of clusters, leading to a **K* = 4**. The Average Silhouette width with 4 cluster raises to 0.45, compared to the 0.165 of the clustering of the whole dataset, stating the higher reliability of this analysis.

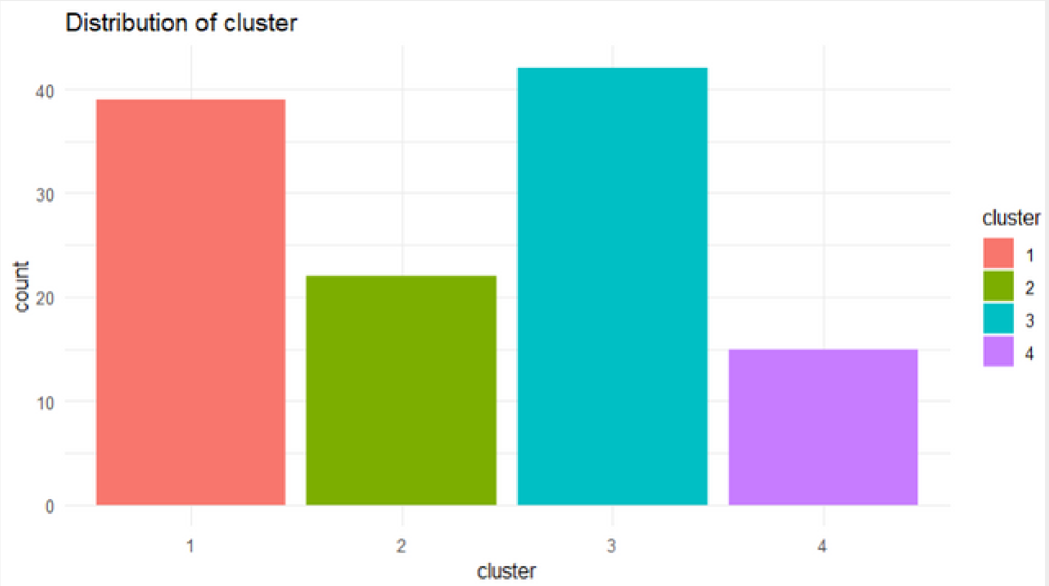


Figure 4: Number of advertised items per cluster

Figure 4 shows how two of the identified creative lines are much more common, at least among the analysed influencers. It's relevant to notice how the newly introduced variables have a **high correlation** (0.67) meaning that the more the product is co-created or co-owned by the influencer, the more he/she has control over its content and how to present it to the audience.
As for the previous clustering, a label was assigned to each of the clusters, trying to extrapolate the trends and key features of the artistic lines. It's important to consider that many of the advertised contents of the analyzed TikTokers were related to make-up and skin-care products, leading to the creation of two clusters starring make-up videos.

Clusters

SelfMade Promotion

This creative style is characterized by **male** influencers who exert a **high level of control** over the product presentation. The term *Self-made* implies a sense of independence and self-reliance, suggesting that these influencers have actively built their brand and promotional content. This style, associated with influencers with a **substantial follower count**, a **measured pace** in video delivery, and an emphasis on **informative content**, suggests a deliberate and strategic approach to promoting products.
Representative Influencers: Donato de Caprio

PositiveJournay Narration

The combination of "positive" and "journey narration" suggests a focus on **optimistic** and uplifting themes, possibly sharing **personal experiences**, stories, or journeys. A **low-creativity** indicates a straightforward and authentic approach, perhaps relying more on genuine narratives. This approach is typical of **female** influencers posting a lot of make-up videos, where they both focus on the cosmetics application and share experiences of their personal life.
Representative Influencers: GlendaResta and Hannabraids

Entertaining Show

This clustered approach is mainly brought by **men** influencers who bring **creativity, information, and entertainment**, painting a picture of engaging and enjoyable content. A key feature is the **happiness** and cheerfulness of the presentation style, driven by the reasearch of an ad-hoc creative style for the specific product.
Representative Influencers: AlePiereal and GianPaoloGolden

VibrantVogue Vlogs

This is the second cluster including a lot of make-up driven product placements. However, there are very significant differences compared to the *PositiveJourney Narration* style. This creative line presents much more **frenetic** videos, even though it's still relaying on sharing personal updates. Another distinctive trait is the **higher body display**, which could also be a reason of the **higher popularity** of the corresponding influencers.
Representative Influencers: Arienne Make-Up

Business Insights

The four clusters identified in the analysis present very **heterogeneous qualities** which can help the companies' decision-makers to target a specific set of influencers corresponding to their preferred style. While the *SelfMade Promotion* cluster may be suitable for companies willing to give the influencer full control of how to present the content, they should also consider that this creative line was very often adopted when presenting self-owned or created products. This may eventually make the promotion **perceived as unnatural**. However, this creative line might be used to enhance **mutual relationships** between influencers aiming to promote their products/services in more social accounts. The two clusters *PositiveJourney Narration* and *VibrantVogue Vlogs* can be the vehicle for companies willing to boost the target audience for **daily-use items**, allowing the marketers to decide which of the two different styles, calm and transparent vs. frenetic and material, can better fit the products value drivers. Finally, the *Entertaining Show* fully depicts the most common style on the platform in line with its user's demographics, where the average age is 7 years lower than in the other social media (*). The playful style of this last segment can, for example, enhance the promotion of **gaming or free-time products**.

(*) Osservatorio Nazionale Influencer Marketing, 2020

Instagram - Clustering

As for the TikTok dataset, the clustering was performed using the K-Prototypes clustering method, to be able to include both the numerical and categorical data in the analysis. Also both the advertised and non-advertised content was considered. It's important to notice that the **influencers of the Instagram dataset are 30** (vs. the 15 of TikTok), but fewer posts/videos are provided for each of them. Due to the distinct typologies of digital contents, some of the variables considered are different, to better evaluate the peculiarities of the platform.
The following variables were included:

- **Numerical variables**: Entertaining, Informative, Creative, Body display, Emoji count, Post positivity, Average posts per week and Followers;
- **Categorical variables**: Product_type, Verified_author, Adv/NonAdv, Creative_content, Pers_update_opinion, Subtitles, Emotion_label, isWeekend.

Among them the Frenetic pace of the video was excluded due to the low presence of videos among the Instagram items. The same had to be done for the Gender presence in the contents, as many of them didn't include any person in it.

In this case, the selection of the number of different clusters was more problematic. Still using the Average Silhouette Width, it can be seen in *Figure 5* that the ideal number of clusters would have been only 2. However performing the clustering with only the two clusters it was very difficult to identify different creative styles, as many characteristics were overlapping. Therefore, the number of clusters was set to **6**, allowing a more precise segmentation of content, but still providing a sufficient statistical level.

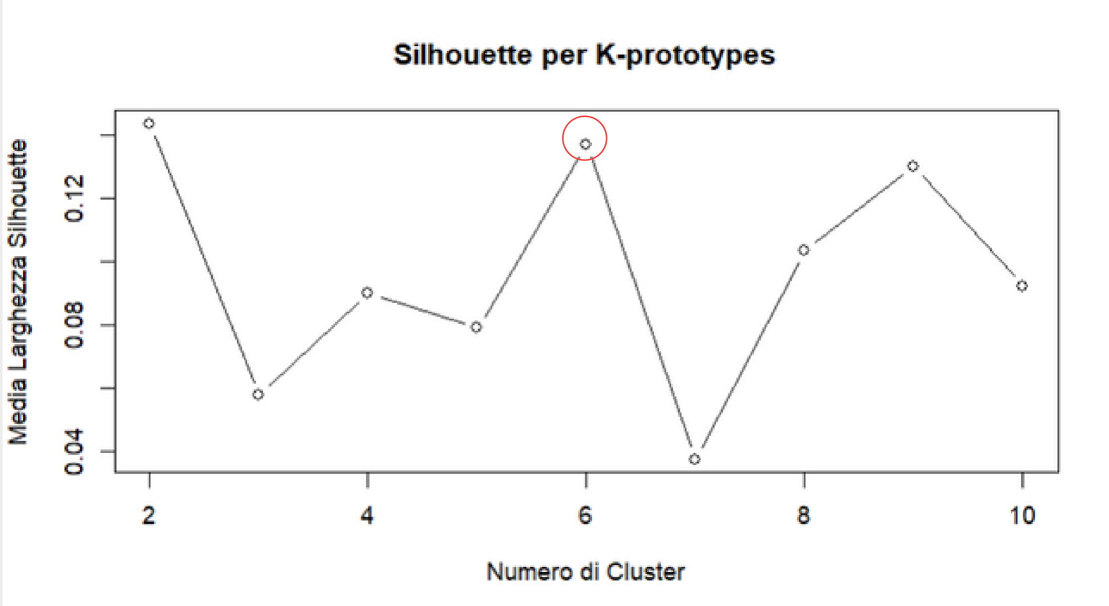


Figure 5 : Average Silhouette Width for Instagram Clustering

The higher number of influencers allowed to identify the distinctive characteristics of each of the 6 clusters. Compared to the TikTok clustering, fewer influencers belonged to only one specific creative line making it more common to **pursue different styles** when creating digital content. Still, some influencers (reported in the clusters) can be considered as a proxy of a particular way of displaying content.

Clusters

Serenity Crafting

The content is marked by a careful balance of **engaging creativity and a calming ambiance**, making it a clean cut from the often frenetic pace of social media. Visually appealing artwork, the sound of music and mindful movie narratives create a virtual oasis for followers seeking a moment of tranquility in their online experience.
Representative Influencers: Therealgue, Lazzinho and CiakClub

TranquilTones

As for the *Serenity Crafting* cluster the calm vibes of the content strongly affect the output. However loses the creative skills, encouraging followers to relax and offering a departure from the often bustling and creatively charged digital landscape.
Representative Influencers: GiuliaDeLellis, Vanessa Incontrada and Matteo Berrettini

Cheerful Realm

In this dynamic digital space, content creators infuse their posts with a **high level of positivity**, creating an uplifting and joyful atmosphere for their engaged audience. Despite the happiness and creative energy of the content, this style seems more representative of **less popular influencers**. The style also integrates a significant amount of **advertised content**, though well placed in the positive atmosphere.
Representative Influencers: CiccioGamer89 and AleCattelan

BodyElegance Chronicles

This creative style can be compared to the EpowerPhyssique identified in the TikTok clustering. They both share the **use of body as form of expression**, celebrating the beauty and diversity of the human physique through tasteful and artistic presentation. However in this case the cluster is characterized by a **very low frequency of pubblication**.
Representative Influencers: AlessiaLanza

DailyExpress

This vibrant and unfiltered creative style places a premium on real-time connection and constant updates. Although the creative aspect is not very carefully curated, probably due to the **high frequency of publication**, the approach aims at fostering a sense of relatability and immediacy, allowing followers to feel intimately connected with the daily narratives of the content creator. This style is very common among news pages.
Representative Influencers: Webboh.it, InsanityPage

EnlightenedHarmony

Also this style, as the *DailyExpress*, deals with **high frequent publication** of content. However, it is more centred on sharing **clips** (not very common for the Instagram platform as a content type), captivating a large audience with its **informative and positively charged style**.
Representative Influencers: PapuGomez, ScuolaZoo and Trash Italiano

Business Insights

The aforementioned clusters depict a wide range of creative style alternatives, common among Italian influencers. It could be important for companies to evaluate not only the style brought by creators on the platform but also their **original work**. It seems that some of the identified clusters are typical of a specific class of celebrities, such as the *Serenity Crafting* style, which is very common among **Italian singers and rappers**. Compared to TikTok, the Instagram creative lines seemed much more influenced by the number of followers: like the *Cheerful Realm* including content usually by not too famous influencers, the opposite of the *EnlightedHarmony*. This has to be taken into account by the companies' decision makers, willing to target a certain size of audience and having to consider the **related costs**.

Instagram Adv - Clustering

Also for the Instagram dataset, a focused clustering was performed on the advertised content only. This dataset contained 80 advertised items, against the 120 of TikTok, therefore it was necessary to **remove some variables**, in order to avoid problems of high dimensionality. This further step aims to provide companies and business actors with insightful elements of Instagram influencers when aiming to promote their products through the platform.

Following are reported the added and removed variables, compared to the total Instagram dataset clustering:

- **Added variables**: Control_product and Control_content;
- **Removed variables**: Adv/Non-Adv, Subtitles, Average Posts per week, Body display and isWeekend

The Average Silhouette Width analysis suggested an optimal number of **3 clusters**, which was also confirmed by the differential characteristics of each of the creative styles.
It's important to notice that in this case, differently from TikTok, the observations are more or less equally distributed among the identified clusters.

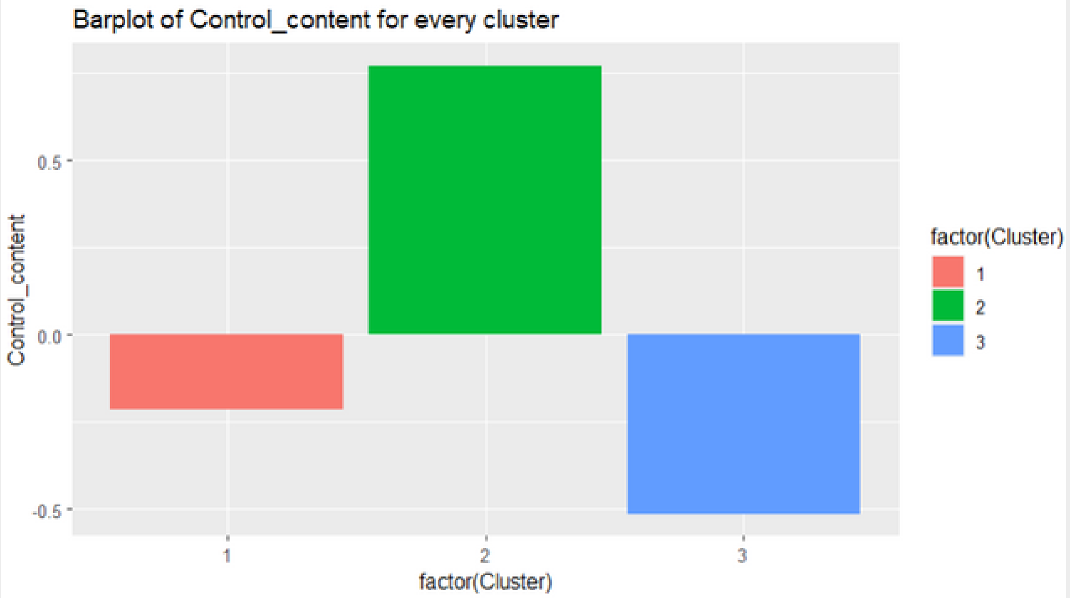


Figure 6 : Control_content variable for the 3 identified clusters

As observed in the previous advertised clustering, the *control_product* and *control_content* variables have a strong correlation, highly impacting the outcome of the analysis. Moreover, some of the variables resulted low differentiated among the different clusters: one of them, is the *personal_updates_and_opinions* which has not a big relevance on the advertised content, such as the *product_type*, which is probably not perceived by the digital audience as a distinctive feature.

Clusters

CreativeFiesta
CreativeFiesta is a digital wonderland where **creativity** takes center stage in a whirlwind of **entertainment** and **emojis**. Despite a **low positivity tone**, the creative energy makes each post a burst of imaginative expression. The high entertainment level ensures a lively and engaging experience for the select audience, even with a **modest following**. A vibrant usage of emojis adds a touch of fun to every content posted.
Representative Influencers: CiccioGamer89

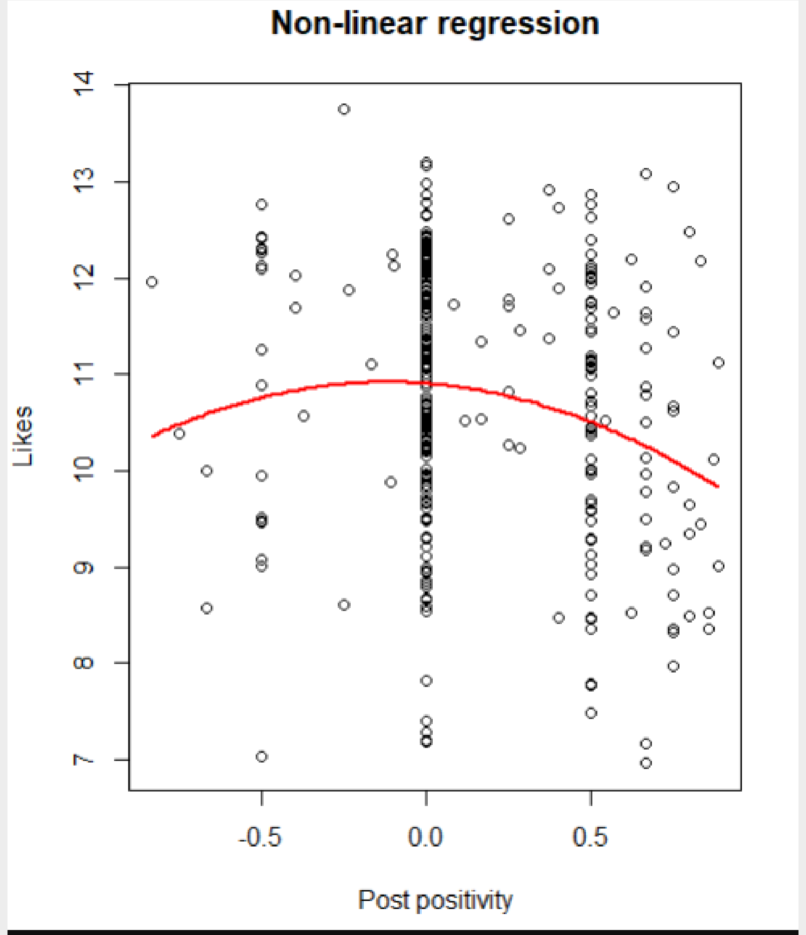
ProductPinnacle Pro
In this digital space, the emphasis is on delivering **comprehensive information** with a professional touch. The creative style is characterized by the mastery of product presentation. With a substantial and engaged following, *ProductPinnacle Pro* applies also in providing **personal opinions**. While not overly creative , the **calm** and calculated approach to creativity ensures that the audience experiences a refined and thoughtful digital environment.
Representative Influencers: Lazzinho, Therealgue and MecontroTe

Joyful Oasis
Joyful Oasis is a burst of **happiness and creativity** in a digital realm that thrives on positivity. Although the **advertised content seems highly influenced by the product owner's** company, a vivacious and celebratory creative style radiates with positivity and a festive spirit
Representative Influencers: CiccioGamer89 and Amedeusonoio

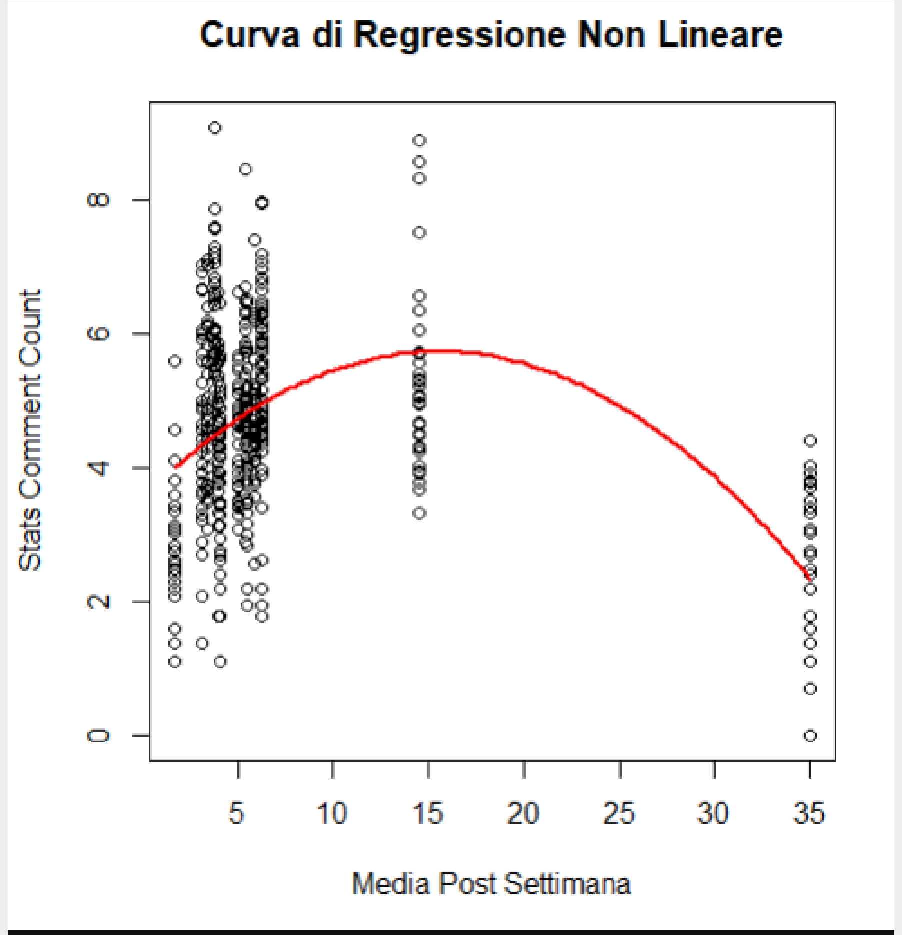
Business Insights

In general, the advertised content of the platform seems **very creative and with a positive mood**. This makes the range of creative style approaches less wide than in the TikTok platform, but this result may be biased by the different compositions and limited abundance of the dataset. Anyway, the *ProductPinnacle Pro* segment seems to have a more distinctive approach than the other two clusters. As shown in *Figure 6* this difference is led by how the products are displayed by the influencers, and this is probably due to, not only the freedom that the content creator has in displaying the product, but also to the **high control_product allowing them to be more autonomous**. In fact among the best representative influencers of this cluster, two rappers can be identified, Lazza and GuePequeno, mainly promoting their musical creations. The main difference among the other two clusters, *CreativeFiesta* and *Joyful Oasis*, is more related to the post description rather than the post itself, and this may influence how a company is perceived, for example **tagging** its Instagram page or how providing specific information on the product in the **post's description.**

Non-linearity




```
Parameters:
  Estimate Std. Error t value Pr(>|t|)
a -1.31991    0.49138  -2.686  0.00762 **
b -0.25226    0.27967  -0.902  0.36776
c 11.00181    0.09505 115.748  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```




```
Parameters:
    Estimate Std. Error t value Pr(>|t|)
a -0.0090256  0.0009684  -9.320  < 2e-16 ***
b  0.2812763  0.0367399   7.656 1.17e-13 ***
c  3.5485308  0.1719782  20.634  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

A **non-linear pattern** was oserved in certain variables, consistent with our expectations and supported by findings in the paper titled 'Influencer Marketing Effectiveness' by Leung*. Specifically, two variables, namely '**Post Positivity**' and '**Media Posts Per Week**,' exhibit such non-linear behavior.

Regarding 'Post Positivity,' an excessively high level of positivity in posts can actually lead to a decrease in engagement. This occurs due to a perceived **loss of authenticity** and, in the case of sponsored content, a potential **'hard sell' effect**. Moderate levels of post positivity, on the other hand, tend to establish a more genuine connection with followers, fostering trust in the influencer and indirectly, the products they endorse. This suggests an **inverted U-shaped** relationship for 'Post Positivity,' which can be effectively modeled using a quadratic equation. Analytically, this is substantiated by the **non-linear regression** analysis, which underscores the significance of the coefficient of the **quadratic term** ('a' : p << 0.05 and negative coefficient).

When an influencer posts content much more frequently than followers check their Instagram feeds, there is initially a positive impact on engagement as followers have more opportunities to engage. However, as the influencer's posting frequency keeps increasing, it eventually leads to diminishing returns. This is because the time gap between posts and followers' visits widens, causing some content to be **missed,** resulting in lower overall engagement despite the higher posting frequency.

Business Insights

- **Find the Sweet Spot**: Engage influencers who understand the importance of balancing positivity in their posts. It's crucial to maintain authenticity and avoid overly positive content that may be perceived as insincere or a hard sell.
- **Emphasize Authenticity**: Encourage influencers to maintain a level of sincerity and authenticity in their content. This builds trust and fosters a stronger connection with their audience, which can lead to higher engagement and conversion rates.
- **Avoid Content Saturation**: Be mindful of the diminishing returns associated with excessive posting frequency. Strive for a balanced posting schedule that keeps the audience engaged without overwhelming them or causing content to be missed.

*Leung, F. F., Gu, F. F., Li, Y., Zhang, J. Z., & Palmatier, R. W. (2022). Influencer Marketing Effectiveness Journal of Marketing, 86 6 93 115 https ::/doi org/ 10 1177 00222429221102889

Instagram - Regression

What do we mean for engagement ?

To comprehensively comprehend the factors exerting the most significant influence on engagement, we conducted a regression analysis, employing the number of "**Likes**" as the primary proxy for engagement, particularly within the context of the Instagram platform. The rationale behind this selection stems from the inherent **controllability exercised** by influencers **over the "Comments"** feature. Influencers frequently possess the ability to limit or even disable comments on specific posts, thereby rendering the "Comments" variable less reliable and less representative of true engagement levels. Secondly, we rejected the use of "Views" as the primary proxy for engagement. This decision was primarily driven by the observation that "Views" alone provide a relatively weak signal of engagement. Views do not necessarily reflect active user interaction with the content, as a mere passive consumption of the content can lead to a view count without a deeper engagement.

Model

$$
\begin{aligned}
\ln(\text{target\_like\_count}_{ij}) = {} & \beta_{0j} + \beta_{1j}\text{product\_type}_{ij} + \beta_{2j}\text{is\_paid\_partnership}_{ij} \\
& + \beta_{3j}\text{Adv}_{ij} + \beta_{4j}\text{Pers\_updates\_opinions}_{ij} + \beta_{5j}\text{Informative}_{ij} \\
& + \beta_{6j}\text{Subtitles}_{ij} + \beta_{7j}\text{BodyDisplay}_{ij} + \beta_{8j}\text{Gender}_{ij} \\
& + \beta_{9j}\text{user\_is\_verified}_{ij} + \beta_{10j}\text{emoji\_count}_{ij} + \beta_{11j}\text{Post\_positivity}^2_{ij} \\
& + \beta_{12j}\text{media\_post\_settimana}^2_{ij} + \beta_{13j}\text{Followers}_{ij} + \gamma_j\text{df\_like}_{ij} + \varepsilon_{ij}
\end{aligned}
$$

We decided to adopt a **logaritmic model** as visually suggested from the scatter plot of the variables in the exploratory analysis and confirmed with the **Box-Cox model.** Starting gradually the analysis from lower-order terms we end up with this model using a stepwise regression to select only relevant variables.
The model satisfy the homoscedasticity and the normal distribution of the residuals proved by a Shapiro test >0.05.

Results

```
Coefficients:
                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                 10.740788   0.381541  28.151  < 2e-16 ***
product_typeclips           -1.121964   0.153345  -7.317 2.44e-12 ***
product_typefeed            -0.215226   0.153943  -1.398 0.163141    
is_paid_partnershipTRUE     -0.870820   0.544275  -1.600 0.110681    
Adv1                        -0.741355   0.162731  -4.556 7.66e-06 ***
Pers_updates_opinions1       0.665539   0.130412   5.103 6.00e-07 ***
Informative                  0.077995   0.033991   2.295 0.022460 *  
Subtitles1                   0.271129   0.166991   1.624 0.105530    
BodyDisplay                  0.223825   0.063272   3.538 0.000469 ***
GenderFemale                 0.280142   0.189039   1.482 0.139431    
GenderMale                   0.525037   0.166632   3.151 0.001796 ** 
user_is_verifiedTRUE        -0.491708   0.311515  -1.578 0.115540    
emoji_count                  0.089869   0.046639   1.927 0.054952 .  
I(Post_positivity^2)        -1.319197   0.316061  -4.174 3.95e-05 ***
I(media_post_settimana^2)   -0.026209   0.004321  -6.066 4.04e-09 ***
Followers                    0.093821   0.047572   1.972 0.049525 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.064 on 294 degrees of freedom
Multiple R-squared:  0.4317,    Adjusted R-squared:  0.4027 
F-statistic: 14.89 on 15 and 294 DF,  p-value: < 2.2e-16
```

Business Insights

- **Follower Count and Personal Opinions**: Influencers with more followers and who openly express their personal opinions tend to generate higher engagement. Businesses should partner with such influencers to boost engagement.
- **Emoji Usage**: Incorporating emojis in content descriptions has a positive impact on engagement, particularly likes. Brands can encourage influencers to use emojis creatively to make content more appealing.
- **Content Body Significance**: The quality and relevance of content's body play a pivotal role in increasing likes. Captivating and well-crafted content resonates better with the audience.
- **Inverted U-shaped Relationships**: Some factors exhibit an inverted U-shaped relationship, implying an optimal range for those variables. Businesses should aim to stay within this range to maximize engagement.
- **Adverse Impact of Advertisements**: Advertisements can have a negative impact on engagement, as followers may find them bothersome. Brands should strive for subtlety and authenticity in their advertising efforts.
- **Reevaluation of Creativity:** Creativity, while valuable, may not always be the primary driver of engagement. Brands should balance creativity with other engagement-focused factors.

Instagram - Robustness

Model Roboustness

To evaluate the robustness of our model, as elucidated in the paper *Influencer Marketing Effectiveness**, we decided to incorporate the **comment variable** into our analysis. This choice is motivated by the observation that comments are the second most important engagement metric, following closely behind likes in terms of significance.

By employing a linear model with the comment count as the target variable, we aim to scrutinize the model's performance and **generalizability** beyond likes. This additional analysis is crucial for assessing whether the variables that we identified as relevant in the likes model also exhibit significance when applied to the comment model. It serves as a valuable check to confirm whether the factors driving engagement, as identified with likes as the target variable, maintain their relevance and consistency when we shift our focus to comments. This validation process enhances the overall **robustness and reliability** of our findings.

Results

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)               5.727247   0.225645  25.382  < 2e-16 ***
is_paid_partnershipTRUE  -1.219949   0.607224  -2.009 0.045464 *  
is_dash_eligible1        -0.356726   0.160608  -2.221 0.027122 *  
Adv1                     -0.621135   0.182725  -3.399 0.000771 ***
Pers_updates_opinions1    0.268698   0.144867   1.855 0.064648 .  
Entertaining             -0.066027   0.041684  -1.584 0.114295    
Informative               0.143304   0.038485   3.724 0.000236 ***
BodyDisplay               0.237705   0.064466   3.687 0.000271 ***
emoji_count               0.079023   0.050860   1.554 0.121342    
I(Post_positivity^2)     -0.857504   0.349999  -2.450 0.014880 *  
I(media_post_settimana^2) -0.019562   0.004282  -4.569 7.28e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.18 on 288 degrees of freedom
Multiple R-squared:  0.217,	Adjusted R-squared:  0.1898 
F-statistic: 7.982 on 10 and 288 DF,  p-value: 2.364e-11
```

Model

$$
\begin{aligned}
\ln(\text{target\_comment\_count}_{ij}) = {} & \beta_{0j} + \beta_{1j}\text{is\_paid\_partnership}_{ij} + \beta_{2j}\text{is\_dash\_eligible}_{ij} \\
& + \beta_{3j}\text{Adv}_{ij} + \beta_{4j}\text{Pers\_updates\_opinions}_{ij} + \beta_{5j}\text{Entertaining}_{ij} \\
& + \beta_{6j}\text{Informative}_{ij} + \beta_{7j}\text{BodyDisplay}_{ij} + \beta_{8j}\text{emoji\_count}_{ij} \\
& + \beta_{9j}\text{Post\_positivity}_{ij}^2 + \beta_{10j}\text{media\_post\_settimana}_{ij}^2 + \gamma_j\text{df\_comment}_{ij} + \varepsilon_{ij}
\end{aligned}
$$

The quadratic variables, as evidenced by their continued relevance in the model, affirm the existence of an inverted U-shaped relationship between post positivity and posting frequency. This suggests that both very high and very low levels of post positivity and posting frequency may not be optimal for engagement, emphasizing the importance of finding the right balance.

Furthermore, the negative coefficient associated with the Advertisement variable underscores the notion that advertisements in influencer content tend to be perceived negatively by followers, impacting engagement adversely.

However, it's worth noting that the robustness of the Emoji Count variable is weakened by its relatively high p-value ($p > 0.05$). This suggests that Emoji Count may not significantly influence engagement in the same way as other variables, indicating a lower level of relevance for this particular factor in the model.

Leung, F. F., Gu, F. F., Li, Y., Zhang, J. Z., & Palmatier, R. W. (2022). Influencer Marketing Effectiveness Journal of Marketing, 86 6 93 115 https ::/doi org/ 10 1177 00222429221102889

Instagram - Regression - Adv only

In this specific analysis, we are focusing exclusively on the observations related to advertisements. Consequently, the dataset no longer includes the "Adv" variable, as all the posts being studied are advertisements by default. However, in this context, we have introduced two additional variables, "**Control Content**" and "**Control Product**" which are only present in the case of advertised posts. These variables are introduced to explore the impact of controlling content and products within influencer marketing campaigns.

$$\begin{aligned}\ln(1+\text{target\_like\_count}_{ij}) = {} & \beta_{0j} + \beta_{1j}\text{product\_type}_{ij} + \beta_{2j}\text{Pers\_updates\_opinions}_{ij} \\ & + \beta_{3j}\text{Entertaining}_{ij} + \beta_{4j}\text{Informative}_{ij} + \beta_{5j}\text{BodyDisplay}_{ij} \\ & + \beta_{6j}\text{Control\_product}_{ij} + \beta_{7j}\text{Control\_content}_{ij} + \beta_{8j}\text{Post\_positivity}^2_{ij} \\ & + \beta_{9j}\text{media\_post\_settimana}^2_{ij} + \gamma_j\text{df\_like\_adv}_{ij} + \varepsilon_{ij}\end{aligned}$$

It's important to highlight that, in this particular case, the model's target variable is the natural logarithm of (1 + likes). This transformation is applied to ensure that the residuals adhere to a normal distribution, thus preserving the validity of the results and upholding the underlying assumptions of the analysis.

Business Insights

- **Co-Branding Partnerships**: Consider collaborating with influencers who are willing to promote both their own products/services and your company's offerings. This co-branding strategy can boost engagement by leveraging the authenticity and trust that influencers have with their followers. This dual endorsement can create a mutually beneficial relationship, enhancing both brand awareness and product credibility.
- **Guidelines and Brand Consistency**: Provide influencers with clear guidelines and brand consistency standards to maintain a cohesive brand image. This ensures that the content produced by influencers aligns with your brand's values and messaging while still allowing them creative freedom.
- **Carousels**: Companies looking to maximize their influencer marketing impact should consider prioritizing "Carousel" posts over "Clips" and "Feeds."

Results

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)              10.960747   0.402692  27.219  < 2e-16 ***
product_typeclips        -0.876372   0.225516  -3.886 0.000306 ***
product_typefeed         -0.423186   0.268849  -1.574 0.121907    
Pers_updates_opinions1    0.672735   0.216042   3.114 0.003081 ** 
Entertaining             -0.205896   0.062163  -3.312 0.001744 ** 
Informative               0.122156   0.060988   2.003 0.050731 .  
BodyDisplay               0.226248   0.107051   2.113 0.039677 *  
Control_product           0.248182   0.074314   3.340 0.001610 ** 
Control_content          -0.118466   0.067566  -1.753 0.085798 .  
I(Post_positivity^2)     -1.444907   0.434892  -3.322 0.001693 ** 
I(media_post_settimana^2) -0.059604   0.008024  -7.428 1.44e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7243 on 49 degrees of freedom
Multiple R-squared:  0.7637,	Adjusted R-squared:  0.7155 
F-statistic: 15.84 on 10 and 49 DF,  p-value: 3.513e-12
```

Remarkably, both the Control on Product and Control on Content variables are pertinent, but they exhibit opposite signs in their coefficients. Notably, Control on Product exerts a highly positive influence on engagement, implying that when an influencer sponsors their own brand, it leads to enhanced engagement. Therefore, the influencer marketing company should consider promoting its own products through collaborations with influencers' brand, possibly through **co-branding partnerships.**

Conversely, the Control variable associated with content creation shows a negative correlation with engagement. This suggests that structuring the advertisement content, rather than giving influencers full creative control, might be more effective in terms of engagement. However, it's important to note that this unexpected result is not highly significant, as indicated by the Control content variable's p-value being very close to 0.05. Further investigation may be needed to determine the robustness of this relationship.

Indeed, it's worth noting another crucial finding from the analysis, which pertains to the **type of post**. Specifically, both "Clips" and "Feeds" exhibit a negative correlation with engagement. This implies that when it comes to the type of post, "Carousel" appears to be the most effective in terms of engagement.

Instagram - Regression - No Adv

In this specific analysis, our focus was exclusively on **non-advertised content**. Consequently, the Control variables and the Adv variable were not considered in this particular model.

Model No Adv

$$
\begin{aligned}
\ln(0.01 + \text{target\_like\_count}_{ij}) = {} & \beta_{0j} + \beta_{1j}\text{product\_type}_{ij} + \beta_{2j}\text{emoji\_count}_{ij} \\
& + \beta_{3j}\text{Pers\_updates\_opinions}_{ij} + \beta_{4j}\text{Informative}_{ij} + \beta_{5j}\text{BodyDisplay}_{ij} \\
& + \beta_{6j}\text{Gender}_{ij} + \beta_{7j}\text{user\_is\_verified}_{ij} + \beta_{8j}\text{Emotion}_{ij} \\
& + \beta_{9j}\text{Post\_positivity}^2_{ij} + \beta_{10j}\text{media\_post\_settimana}^2_{ij} + \beta_{11j}\text{Followers}_{ij} \\
& + \gamma_j\text{df\_comment\_no\_adv}_{ij} + \varepsilon_{ij}
\end{aligned}
$$

The model's target variable is the natural logarithm of (0.01 + likes). This particular transformation is employed to ensure that the residuals closely approximate a **normal distribution.** By doing so, we maintain the validity of the results and uphold the fundamental assumptions underpinning the analysis, as previously discussed in the provided text.

Adv vs No Adv

The results of this regression analysis underscore the significance of certain factors, aligning with the patterns observed in both sponsored (Adv) and non-sponsored (No-Adv) content. Specifically, factors such as **body display, product type**, and the **number of followers** are consistently relevant for engagement, irrespective of whether the content is sponsored or not. This **reaffirms** the earlier reasoning and conclusions drawn.
Additionally, the inclusion of quadratic terms further **reinforces the notion of U-shaped** relationships in variables, which hold true for both Adv and No-Adv posts. This highlights the importance of finding the **optimal balance** in variables like post positivity and posting frequency to maximize engagement.
On a somewhat **surprising note**, the emotional variables feature in the final model, although only "happy" and "sad" emotions exhibit **slight relevance**, and intriguingly, they have opposite effects. Happy content seems to decrease engagement, while sad posts have the opposite effect. This may initially appear counterintuitive, but our analysis, along with findings from a survey, suggests that people tend to be more drawn to and engaged by melancholic and reflective topics, even more so than by happy content.

Results Adv

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)             10.960747   0.402692  27.219  < 2e-16 ***
product_typeclips       -0.876372   0.225516  -3.886 0.000306 ***
product_typefeed        -0.423186   0.268849  -1.574 0.121907    
Pers_updates_opinions1   0.672735   0.216042   3.114 0.003081 ** 
Entertaining            -0.205896   0.062163  -3.312 0.001744 ** 
Informative              0.122156   0.060988   2.003 0.050731 .  
BodyDisplay              0.226248   0.107051   2.113 0.039677 *  
Control_product          0.248182   0.074314   3.340 0.001610 ** 
Control_content         -0.118466   0.067566  -1.753 0.085798 .  
I(Post_positivity^2)    -1.444907   0.434892  -3.322 0.001693 ** 
I(media_post_settimana^2) -0.059604   0.008024  -7.428 1.44e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7243 on 49 degrees of freedom
Multiple R-squared:  0.7637,	Adjusted R-squared:  0.7155 
F-statistic: 15.84 on 10 and 49 DF,  p-value: 3.513e-12
```

Results No Adv

```
Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)               11.603097   0.647556  17.918  < 2e-16 ***
product_typeclips         -1.020773   0.173621  -5.879 1.45e-08 ***
product_typefeed          -0.193643   0.174526  -1.110 0.268366    
emoji_count                0.142796   0.051080   2.796 0.005623 ** 
Pers_updates_opinions1     0.550473   0.154576   3.561 0.000449 ***
Informative                0.075614   0.038768   1.950 0.052352 .  
BodyDisplay                0.207318   0.068960   3.006 0.002940 ** 
GenderFemale               0.151101   0.210489   0.718 0.473581    
GenderMale                 0.377726   0.189324   1.995 0.047219 *  
user_is_verifiedTRUE      -0.787910   0.364638  -2.161 0.031753 *  
EmotionCALM               -0.451082   0.493353  -0.914 0.361515    
EmotionCONFUSED           -0.067281   0.634527  -0.106 0.915649    
EmotionDISGUSTED          -0.673478   1.159263  -0.581 0.561845    
EmotionFEAR               -0.310075   0.824531  -0.376 0.707220    
EmotionHAPPY              -0.934707   0.501922  -1.862 0.063853 .  
EmotionSAD                 0.367656   0.565653   0.650 0.516367    
EmotionSURPRISED          -0.877748   0.604201  -1.453 0.147670    
I(Post_positivity^2)      -1.054765   0.385510  -2.736 0.006708 ** 
I(media_post_settimana^2) -0.023803   0.004481  -5.311 2.58e-07 ***
Followers                  0.122339   0.051717   2.366 0.018842 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.055 on 228 degrees of freedom
Multiple R-squared:  0.4341,	Adjusted R-squared:  0.387 
F-statistic: 9.206 on 19 and 228 DF,  p-value: < 2.2e-16
```

TikTok- Regression

Results

```
Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                2.2206811  0.7700451   2.884  0.00412 **
Emotion_LabelConfused      3.4211502  1.5003826   2.280  0.02307 *
Emotion_LabelHappy         1.7471223  0.7533561   2.319  0.02084 *
Emotion_LabelSad           2.4792913  0.7712858   3.214  0.00140 **
Emotion_LabelSurprised     1.9654236  0.9061991   2.169  0.03062 *
is_weekendTRUE            -0.2334821  0.1523305  -1.533  0.12606
emoji_count                0.1539371  0.0666763   2.309  0.02142 *
I(media_post_settimana^2) -0.0018581  0.0002156  -8.619  < 2e-16 ***
video_duration             0.0015106  0.0009317   1.621  0.10565
Pers_updates_opinions1     0.5031386  0.1547627   3.251  0.00124 **
Informative                0.1168634  0.0436078   2.680  0.00764 **
Adv1                      -0.2715905  0.1778405  -1.527  0.12744
Frenetic.pace.of.the.video -0.1427832 0.0542476  -2.632  0.00878 **
Followers                  0.1465038  0.0504318   2.905  0.00386 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.287 on 443 degrees of freedom
Multiple R-squared:  0.3058,    Adjusted R-squared:  0.2855
F-statistic: 15.01 on 13 and 443 DF,  p-value: < 2.2e-16
```

The findings from our analysis on TikTok **differ** notably from those observed on Instagram. Specifically, we've identified the significant influence of **emotion** labels, particularly the "sad" label, which exhibits a strong positive impact on engagement. This highlights a trend that was not evident in our Instagram analysis, indicating that emotional content plays a more substantial role in TikTok videos.

Additionally, the relevance of the "Adv" variable, which was significant in the Instagram context, is notably absent here in TikTok. This suggests that advertising content may have a different impact on engagement across these two platforms.

A noteworthy observation is the negative influence of a frenetic pace in videos on engagement. This includes factors such as a high number of cuts in the video and excessive camera movement. It appears that followers on TikTok do not appreciate overly busy or frenetic video styles.

Furthermore, we've found that videos posted on **weekends drive less engagement**. This observation can be attributed to the reduced smartphone usage on weekends, leading to decreased platform activity. Consequently, content posted during this time may garner less attention and engagement from users.

What do we mean for engagement ?

Previously, we conducted a comprehensive regression analysis to understand the primary factors influencing engagement within the context of the Instagram platform. In that analysis, we employed the number of '**Likes**' as our primary proxy for measuring engagement.

Now, as we shift our focus to the **TikTok** platform, we find ourselves reevaluating our choice of **engagement proxy**. In the context of TikTok, 'Likes' can be easily assigned with a simple tap and are more effortless, making them a **weaker** indicator of engagement. Therefore, we are now considering '**Comments**' as our primary proxy for measuring engagement on TikTok. 'Comments' offer a more nuanced and deliberate form of interaction, providing valuable insights into the audience's active engagement with the content.

Model

$$\begin{aligned}
\ln(\text{stats_commentCount}_{ij}) = {} & \beta_{0j} + \beta_{1j}\text{Emotion_Label}_{ij} + \beta_{2j}\text{is_weekend}_{ij} \\
& + \beta_{3j}\text{emoji_count}_{ij} + \beta_{4j}\text{media_post_settimana}_{ij}^2 \\
& + \beta_{5j}\text{video_duration}_{ij} + \beta_{6j}\text{Pers_updates_opinions}_{ij} \\
& + \beta_{7j}\text{Informative}_{ij} + \beta_{8j}\text{Adv}_{ij} \\
& + \beta_{9j}\text{Frenetic.pace.of.the.video}_{ij} + \beta_{10j}\text{Followers}_{ij} \\
& + \gamma_j\text{df_comment}_{ij} + \varepsilon_{ij}
\end{aligned}$$

Regarding the quadratic terms, it's interesting to note that in this TikTok analysis, only the "Media Posts Per Week" variable appears to be relevant, while the "Post Positivity" variable does not exhibit significance. This is in contrast to what we observed in the Instagram analysis, where "Post Positivity" demonstrated relevance.

Business insights

- **Embrace Emotional Content**: TikTok users appear to respond positively to emotional content, particularly content labeled as "sad." Consider incorporating emotionally resonant storytelling in your influencer campaigns to enhance engagement.
- **Platform-Specific Strategies**: TikTok and Instagram have distinct engagement dynamics. Tailor your content strategies accordingly, recognizing that what works on one platform may not yield the same results on the other.
- **Optimal Posting Times**: Recognize that engagement tends to dip on weekends. Adjust your posting schedule to align with the days and times when TikTok users are most active on the platform.

TikTok- Robustness

Results

```
Coefficients:
                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                 2.0053199  0.9889406   2.028 0.043157 *  
GenderFemale               -0.8835795  0.2509137  -3.521 0.000472 ***
GenderMale                 -0.4319741  0.2491020  -1.734 0.083560 .  
Emotion_LabelConfused       3.7359966  1.5269681   2.447 0.014788 *  
Emotion_LabelHappy          2.6453384  0.9709227   2.725 0.006682 ** 
Emotion_LabelSad            2.9368508  0.9906638   2.965 0.003188 ** 
Emotion_LabelSurprised      3.1592966  1.1674309   2.706 0.007057 ** 
is_weekendTRUE             -0.4257839  0.1918367  -2.220 0.026935 *  
emoji_count                 0.1180351  0.0836584   1.411 0.158939    
I(media_post_settimana^2)  -0.0027430  0.0003068  -8.942  < 2e-16 ***
Informative                 0.0652289  0.0441724   1.477 0.140438    
I(Frenetic.pace.of.the.video^2) -0.0433806  0.0113488  -3.822 0.000150 ***
Followers                   0.2065181  0.0608417   3.394 0.000747 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.663 on 464 degrees of freedom
Multiple R-squared:  0.2127,	Adjusted R-squared:  0.1924 
F-statistic: 10.45 on 12 and 464 DF,  p-value: < 2.2e-16
```

To assess the robustness of our model within the TikTok context, as outlined in the paper "Influencer Marketing Effectiveness," we opted to incorporate the comment variable into our analysis. This decision was driven by the recognition that comments rank as the second most significant engagement metric on TikTok, closely trailing likes in terms of importance.
By employing a linear model with the comment count as the target variable, our objective is to scrutinize the model's performance and its applicability beyond likes. This supplementary analysis is essential for evaluating whether the variables previously identified as significant in the likes model **maintain their relevance** and **consistency** when applied to the **comment model**. This validation process strengthens the overall robustness and reliability of our conclusions, confirming that the factors influencing engagement, as identified when likes were the target variable, continue to hold true when our focus shifts to comments on TikTok.

Model Roboustness

In this specific case, we observe a notable shift in our findings, particularly regarding the variable related to the frenetic pace of the video. While the initial analysis indicated that the frenetic pace was not highly relevant in the general model, it now emerges as significant, especially when considering its **non-linear term**.
Specifically, we sought to test whether the inverted U-shaped relationship, as observed in other variables, **also holds true for the frenetic pace**. Interestingly, it does appear to follow this pattern in this particular context.
However, it's essential to highlight that all the other variables maintain their consistency with the previous model's findings.

$$\begin{aligned}\ln(0.01 + \text{stats\_shareCount}_{ij}) = {} & \beta_{0j} + \beta_{1j}\text{Gender}_{ij} + \beta_{2j}\text{Emotion\_Label}_{ij} \\ & + \beta_{3j}\text{is\_weekend}_{ij} + \beta_{4j}\text{emoji\_count}_{ij} + \beta_{5j}\text{media\_post\_settimana}^2_{ij} \\ & + \beta_{6j}\text{Informative}_{ij} + \beta_{7j}\text{Frenetic.pace.of.the.video}^2_{ij} + \beta_{8j}\text{Followers}_{ij} \\ & + \gamma_j\text{df\_share}_{ij} + \varepsilon_{ij}\end{aligned}$$

Business insights

- **Consider Video Pace Carefully**: Recognize the importance of video pacing in TikTok content. While it may not be a significant factor in all contexts, it appears to have relevance in certain situations, following an inverted U-shaped pattern. Experiment with different video pacing styles to find the optimal balance that resonates with your target audience.

TikTok - Regression - Adv only

Business insights

- **Moderate Body Display:** Recognize the importance of moderation in body display within sponsored TikTok content. Excessive body display may lead to lower engagement, so work with influencers to strike the right balance, ensuring that the content aligns with your brand message without appearing overly self-promotional.

In our analysis of sponsored videos on TikTok, we introduced a new variable, "Body.Display squared," as part of a stepwise process that involved **gradually** including higher-order terms. Ultimately, the best-fitting model revealed an interesting trend: the linear term of "Body Display" **disappeared** from the final model, leaving **only the quadratic term.** This suggests that, for sponsored content, an excessive display of the body may lead to lower engagement, while a moderate level appears to be the optimal choice.

It's noteworthy that we incorporated the **Control variables** in this analysis, as we focused exclusively on advertised content. However, it's interesting to observe that "Control Content" did not exhibit relevance in this context. This finding indicates a substantial difference in advertising dynamics between the two platforms. Nonetheless, in alignment with our findings from Instagram, "Control Product" demonstrates a positive relationship with engagement in TikTok. Therefore, the insights and conclusions drawn previously remain applicable in this TikTok context.

Model

$$\begin{aligned}\ln(\text{stats\_commentCount}_{ij}) = {} & \beta_{0j} + \beta_{1j}\text{Body.display}_{ij}^2 + \beta_{2j}\text{emoji\_count}_{ij} \\ & + \beta_{3j}\text{Post\_positivity}_{ij}^2 + \beta_{4j}\text{media\_post\_settimana}_{ij} + \beta_{5j}\text{media\_post\_settimana}_{ij}^2 \\ & + \beta_{6j}\text{video\_duration}_{ij} + \beta_{7j}\text{Entertaining}_{ij} + \beta_{8j}\text{Creative}_{ij} \\ & + \beta_{9j}\text{Control\_product}_{ij} + \beta_{10j}\text{Frenetic.pace.of.the.video}_{ij} + \beta_{11j}\text{Frenetic.pace.of.the.video}_{ij}^2 \\ & + \gamma_j\text{df\_comment}_{ij} + \varepsilon_{ij}\end{aligned}$$

Results

```
Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   2.117842   1.047858   2.021  0.04581 *
I(Body.display^2)            -0.090314   0.033481  -2.697  0.00814 **
emoji_count                   0.168930   0.125391   1.347  0.18081
I(Post_positivity^2)         -1.060129   0.710797  -1.491  0.13884
media_post_settimana          0.613159   0.254588   2.408  0.01776 *
I(media_post_settimana^2)    -0.033549   0.014343  -2.339  0.02122 *
video_duration                0.005248   0.001757   2.987  0.00351 **
Entertaining                 -0.215816   0.120266  -1.794  0.07561 .
Creative                     -0.288434   0.098150  -2.939  0.00405 **
Control_product               0.459516   0.075208   6.110 1.72e-08 ***
Frenetic.pace.of.the.video    0.876916   0.462739   1.895  0.06084 .
I(Frenetic.pace.of.the.video^2) -0.184309 0.079047 -2.332  0.02163 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.167 on 105 degrees of freedom
  (2 osservazioni eliminate a causa di valori mancanti)
Multiple R-squared:  0.4989,    Adjusted R-squared:  0.4465
F-statistic: 9.505 on 11 and 105 DF,  p-value: 1.106e-11
```

TikTok - Regression - No Adv

Results Adv

```
Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                 2.117842   1.047858   2.021  0.04581 *
I(Body.display^2)          -0.090314   0.033481  -2.697  0.00814 **
emoji_count                 0.168930   0.125391   1.347  0.18081
I(Post_positivity^2)       -1.060129   0.710797  -1.491  0.13884
media_post_settimana        0.613159   0.254588   2.408  0.01776 *
I(media_post_settimana^2)  -0.033549   0.014343  -2.339  0.02122 *
video_duration              0.005248   0.001757   2.987  0.00351 **
Entertaining               -0.215816   0.120266  -1.794  0.07561 .
Creative                   -0.288434   0.098150  -2.939  0.00405 **
Control_product             0.459516   0.075208   6.110 1.72e-08 ***
Frenetic.pace.of.the.video  0.876916   0.462739   1.895  0.06084 .
I(Frenetic.pace.of.the.video^2) -0.184309   0.079047  -2.332  0.02163 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.167 on 105 degrees of freedom
  (2 osservazioni eliminate a causa di valori mancanti)
Multiple R-squared:  0.4989,	Adjusted R-squared:  0.4465
F-statistic: 9.505 on 11 and 105 DF,  p-value: 1.106e-11
```

Results No Adv

```
Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               1.8172464  0.6939307   2.619  0.00921 **
GenderFemale              0.2745157  0.1805692   1.520  0.12935
GenderMale               -0.2194439  0.1811565  -1.211  0.22658
Emotion_LabelConfused     1.9595627  1.0512731   1.864  0.06316 .
Emotion_LabelHappy        1.4246952  0.6698437   2.127  0.03413 *
Emotion_LabelSad          1.9226210  0.6865383   2.800  0.00539 **
Emotion_LabelSurprised    1.5426733  0.8933206   1.727  0.08507 .
emoji_count               0.2711061  0.0679229   3.991 8.01e-05 ***
I(media_post_settimana^2) -0.0014570  0.0002043  -7.130 5.84e-12 ***
video_duration            0.0044952  0.0010169   4.420 1.32e-05 ***
Pers_updates_opinions1    0.6780586  0.1423016   4.765 2.78e-06 ***
Frenetic.pace.of.the.video 0.0779242  0.0548706   1.420  0.15646
Followers                 0.1316747  0.0537044   2.452  0.01470 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.143 on 348 degrees of freedom
Multiple R-squared:  0.4044,	Adjusted R-squared:  0.3839
F-statistic: 19.69 on 12 and 348 DF,  p-value: < 2.2e-16
```

A noteworthy observation is that the quadratic form of the "Body.Display" variable and the "Post Positivity" variable significantly influence engagement only in the context of advertised content. Interestingly, these variables do not exhibit the same relevance in non-advertised content.
Furthermore, it's worth highlighting that the **emotional** variables come into **play exclusively in non-advertised content**. Among these emotions, posts that evoke "surprise" show the highest positive correlation with engagement (estimate > 0). This finding emphasizes the importance of crafting content that elicits surprise and resonates emotionally with the audience in non-advertised TikTok posts.
These distinctions between advertised and non-advertised content underscore the need for tailored influencer marketing strategies that align with the specific dynamics and audience expectations on TikTok.

Limitations

A significant limitation that should be acknowledged is the noticeable **imbalance** between the number of advertised (adv) and non-advertised (non-adv) content within the dataset. The dataset contains a significantly higher proportion of non-advertised content compared to advertised content.
This imbalance raises concerns of **potential bias** in the results, and it is advisable to conduct further analysis on a larger dataset to **confirm the robustness** and generalizability of the findings. A more balanced dataset would provide a more accurate representation of influencer marketing dynamics on TikTok and mitigate potential biases stemming from the current dataset's imbalance.

Model No Adv

$$\begin{aligned}\ln(\text{stats\_commentCount}_{ij}) = {} & \beta_{0j} + \beta_{1j}\text{Gender}_{ij} + \beta_{2j}\text{Emotion\_Label}_{ij} \\ & + \beta_{3j}\text{emoji\_count}_{ij} + \beta_{4j}\text{media\_post\_settimana}^2_{ij} \\ & + \beta_{5j}\text{video\_duration}_{ij} + \beta_{6j}\text{Pers\_updates\_opinions}_{ij} \\ & + \beta_{7j}\text{Frenetic.pace.of.the.video}_{ij} + \beta_{8j}\text{Followers}_{ij} \\ & + \gamma_j\text{df\_comment}_{ij} + \varepsilon_{ij}\end{aligned}$$

Survey

Methodology

The survey methodology employed was **quantitative**, utilizing a **structured approach** with **multiple closed questions** presented through a **Google Form**. This method facilitated the collection of numerical data and allowed for a systematic analysis of respondents' opinions and attitudes toward various aspects of social media and advertising.

Measurement scale | Likert scale

The survey employed a **Likert scale** for measurement, where respondents were asked to indicate their level of agreement using a scale ranging from 1 to 5, with 1 indicating "Strongly Disagree" and 5 indicating "Strongly Agree." The intermediary options allowed for a nuanced level of agreement.

Structure

Two initial questions were included to ascertain respondents' age and gender, helping to ensure that the survey effectively reached its intended target segment, which, in this case, consists of adolescents and students.

No screening questions were included, as it was assumed that all individuals within the target segment either use or have at least used social media, and this approach helped to keep the survey concise.

The questionnaire was divided into **two distinct parts**. The first part addressed **general topics** related to social media usage and perception. This section aimed to gather insights into respondents' overall experiences with social media platforms.

The second part of the questionnaire specifically concentrated on the topic of **advertising** within the realm of social media. It delved into respondents' attitudes, behaviors, and preferences related to advertisements encountered on these platforms. This separation allowed for a more focused exploration of advertising-related factors while still considering the broader context of social media.

***Scan to open the form**

Business insights

Higher marketing **investments** for popular influencers is not necessarily translated in higher **returns**

Results

The survey collected 63 participant opinions* among which the following aspects can be highighted:

Confirming results

- The frequency of posting (*Figure 7*) resulted in driving down the engagement when the value was too high, confirming relevance of the negative coefficient of the quadratic variable

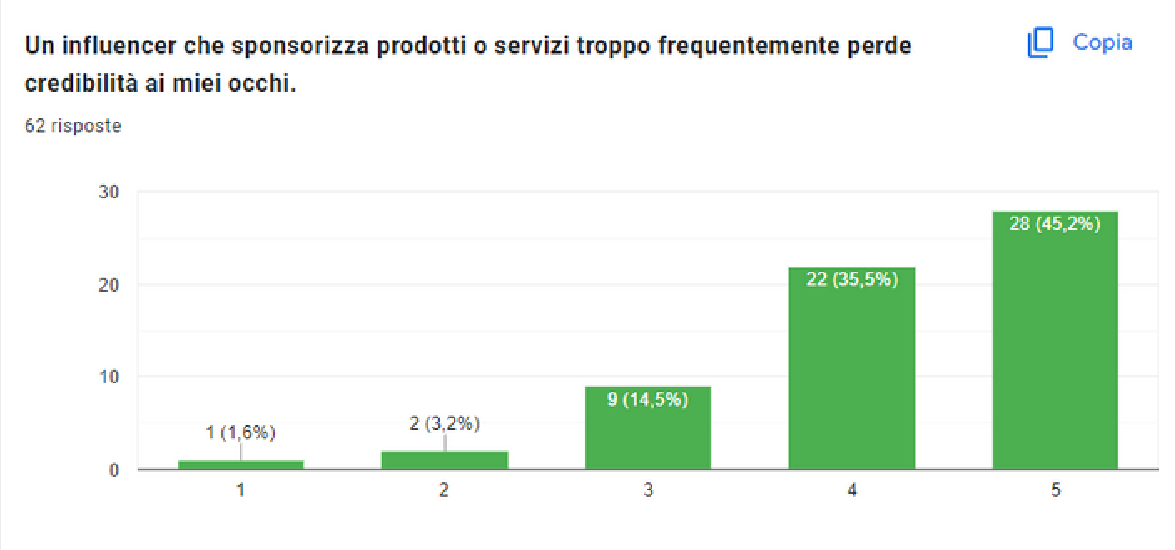


Figure 7: Survey question

- The relevance of the **comment_count** selection of the engagement variable for TikTok was confirmed by the respondent’s tendency to comment the reel when interested in the viewed product
- **Informative** contents are generally appreciated by social media users, in fact, 85.5% of the participants agreed with the statement “*I find valuable informative and educative content*”
- Almost 70% of the interviewed are bothered by the **sponsored content**, and this is reinforced by the fact that almost 50% of participants strongly sustained that they lose credibility in the influencer in case of frequent repetitive advertising
- Also the **body display** attractiveness is proved by an increasing engagement, only when it is not too explicit (over 75% prefers a discrete doby dispay level)

Non-Confirming results

- The survey collected a popular tendency of not considering sponsored content by **popular influencers** more reliable. This could be a **useful insight for businesses** when considering more popular influencers more engaging, as resulted in the regression analysis. A further analysis should be carried on about the relationship between the popularity of the influencer and the return as brand image of the company

Conclusions

Target cluster identification

To conclude the analysis and provide a general suggestion to businesses, a **combination of the Clustering** results and the **Regression** outcomes was performed.
For Instagram, the ProductPinnacle Pro cluster, identified when considering the advertised content, reflected many of the features driving the most engagement in terms of likes. Among them, there is a limited frequency of posting (negatively affecting engagement), a high sharing of personal opinions and a high follower base. Influencers corresponding to this creative style should be highly considered by companies' marketing divisions as possible vehicles of a social media marketing campaign.

On the other hand, TikTok engagement seemed still driven by high personal updates and opinions, though using low frenetic and long videos. Also, the general positive atmosphere of the platform seems to enhance the comments and strengthen the community. The TikTok cluster better reflecting these characteristics is the *PositiveJourney Narration,* where promotions of the company could be well integrated into the straightforward and authentic approach used by these content creators.

Expectations satisfaction

To conclude, an overall comparison of the results with the initial expectations is provided.

- Informative content resulted driving higher engagement in both platforms. On the other hand, creative content emerged as not significant for the Instagram platform, while negatively impacting the engagement on TikTok
- Body display and Frequency of posting happened to be inverted U-shaped as expected, meaning a moderate value of these variables is preferable
- Especially TikTok, which is more saturated with happy and positive content, happens to tease more the community with sad content
- The weekend variable, which was expected to be very significant in driving the engagement of the platforms, actually resulted in being irrelevant. It could be that the algorithm steering the views works to prevent content from being hidden when published in a low-activity time frame
- Moreover the audience positively welcomes when the influencer is co-creator or co-owner of the advertised product

Limitations

- **Limited Data**: When you have a small amount of data, regression models may suffer from limited representativeness and generalizability. The model may not capture the true underlying relationships due to the lack of data points.

- **High Variance**: Limited data can result in higher variance in the estimates of the regression coefficients, making the model less stable and reliable.

- **Subjective Variable Assignment**: When variables are subjectively assigned to observations, there is a risk of introducing bias or errors in the clustering process. Different individuals may assign variables differently, leading to inconsistent results.

- **Social Desirability Bias**: Respondents may provide answers they think are socially acceptable or expected, rather than their true opinions. This can lead to a bias in the survey results, as people may not always express their genuine thoughts or feelings.