**A colorful logo with hearts above it

Description automatically generated**

**Instagram Interaction Insights Report**

Table of Contents

[EXECUTIVE SUMMERY 3](#_Toc518429476)

[Key Findings: 3](#_Toc593648306)

[INTRODUCTION 4](#_Toc1052146540)

[METHODOLOGY 5](#_Toc1817413020)

[DATA 5](#_Toc561264643)

[Data cleaning 7](#_Toc1374498531)

[Date Transformation 7](#_Toc1167651348)

[Analysis Results 9](#_Toc1298369034)

[Question 1: 9](#_Toc863994899)

[Question 2: 12](#_Toc1337650228)

[Question 3: 14](#_Toc582697040)

[Question 4 : 16](#_Toc1685060385)

[Conclusion and Recommendations 19](#_Toc2145443421)

[References 20](#_Toc1970248814)

# **EXECUTIVE SUMMERY**

To give influencers useful information for maximizing likes and comments, this initiative examines Instagram post data to pinpoint the main elements influencing user interaction.

The dataset contains Instagram post data, which includes metadata such as hashtags, mentions, and posting date, and features such as likes, comments, text captions, and the type of post, whether video or photo. We employed a systematic approach to study user engagement by identifying high-impact factors of likes and comments through data translation, cleansing, and linear regression modelling. This research will seek to find patterns and trends in data that can give practical insights to influencers on how to improve engagement.

## Key Findings:

* Videos generally engage more than images, which is indicative of people finding dynamism in what they consume.
* Posts published at periods such weekends or evenings typically get more interaction than those published at other times.
* Likes and comments are greatly influenced by the quantity of hashtags and mentions, with an ideal range boosting interaction.
* In return, it is often achieved through medium caption length to give some balance between audience attention span and clear communication

By customizing Instagram posts to audience behavioural patterns and content attributes, influencers will be able to effectively achieve their social media objectives and improve engagement.

# **INTRODUCTION**

Instagram influencers aim to boost user engagement, but they frequently only look at data from their own posts, which prevents them from seeing more broad strategies and insights that could boost likes, comments, and overall involvement. Influencers face a challenge when they have to analyze their own Instagram post statistics in order to optimize user engagement. Their narrow scope prevents them from identifying more complete strategies that can significantly boost their reach and interaction rates, even though they are able to track likes, comments, and shares.

Businesses, marketers, organizations, and influencers are all interested in finding a solution to this problem. Engagement metrics are essential for influencers' expansion and revenue. Instagram interaction has measurable benefits for companies and organizations, including improved customer relationships and increased conversion rates.

We intend to use regression models and exploratory analysis to examine a comprehensive dataset of Instagram posts to identify engagement-boosting elements such as timing, content type, and hashtag usage. With the help of this method, influencers will be able to develop more impactful content strategies and get practical insights and suggestions.

This problem is important because, in addition to an influencer's exposure and audience growth, user interaction affects the profitability of businesses and organizations who utilize Instagram as a marketing platform. High interaction can lead to increased revenue and customer loyalty, as well as increased brand recognition and content reach.

# **METHODOLOGY**

Our Instagram post data research was based on the CRISP-DM approach, which is the Cross-Industry Standard Process for Data Mining. The first stage of the **business understanding** phase involved defining the goal of the research as finding out what factors influence user engagement. Then, the quality, variables, and structure of the dataset were checked in the **data understanding stage**. We cleaned the data, added more variables, and prepared it in a suitable format for analysis in the **data preparation stage**. Regression analysis was applied throughout the **modelling phase** to determine the major factors affecting likes and comments. In the **evaluation process**, we ensured relevant insights by assessing model performance and checked the results. Lastly, we converted the findings into suggestions for influencers to successfully improve engagement tactics during the **deployment phase**. A methodical and trustworthy approach to accomplishing the project's objectives was guaranteed by this organized technique.

# **DATA**

* Data set is provided by Corentin Dugué by scraping this data from Instagram.
* Total no. of records is 19,681.
* Variables

|  |  |  |
| --- | --- | --- |
| Variable | Type | Description |
| Followers | Continuous | Number of followers of the user. |
| Following | Continuous | Number of accounts followed by the user. |
| Likes | Continuous | The number of likes the post received at the time of data collection |
| Comments | Continuous | The number of comments the post received at the time of data collection |
| Number of Tags | Continuous | Number of hashtags used. |
| Number of Mentions | Continuous | Number of users mentions in the post. |
| User in photo | continuous | The number of users in the post |
| Text | Text | The text used as the post caption |
| Date | Date | The date the post was created |
| Type (Photo or Video) | Categorical | What type of post it was (1 = PHOTO, 2 = VIDEO) |

* Descriptive statistics for numerical variables:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **Std Dev** | **Min** | **Max** |
| Followers | 72571 | 52945 | 32552 | 461375 |
| Following | 886 | 1428 | 0 | 7368 |
| Likes | 3046 | 3377 | 0 | 24811 |
| Comments | 26 | 81.41 | 0 | 1814 |

|  |  |
| --- | --- |
| Variable Name | Unique Values |
| TYPE | 1 = Photo, 2 = Video |
| Day of Posting | Monday, Tuesday, Wednesday, etc. |
| Time of Day | Morning, Afternoon, Evening, Night |

* Categorical variables and their unique values
* Number of null values

1. Users in photo-8527
2. list\_of\_tags – 5819
3. list\_of\_mentions -12935

* Independent variables

Followers, Following, Date, Type, Number of Tags, Number of Mentions, Users in Photo, and Text.

* Dependent variables:

Likes and Comments.

# **Data Cleaning**

* Few variables, such as "Text," "List of Tags," and "List of Mentions," contained missing values.
* Numerical variables for "Followers," "Likes," and "Comments" have extreme levels that could distort the findings.
* Categorical variables were dummy coding for regression analysis.

1. Handle missing values in the TEXT column using. isnull() and .sum()
2. Replaced missing values with NaN with an empty string
3. Detected and handled outliers using z-scores
4. Dummy-coded.
5. Due to effecting on multi collinearity we have removed month 4 and 5.

# **Date Transformation**

* Converts the DATE column to a datetime format.
* Defining the categorize\_timing function, which is used to group the hours of the day into more general categories.
* **Morning**: 8:00 AM to 12:00 PM
* **Afternoon**: 12:00 PM to 4:00 PM
* **Evening**: 4:00 PM to 8:00 PM
* **Night**: All other hours
* Creating a new column Timing\_of\_post with categories like Morning, Afternoon.. etc
* Displaying the first rows of the DataFrame with the DATE, Month, Day, Day\_of\_week, and Timing\_of\_post columns to confirm the changes.
* Dummy coded variables for 'Month', 'Timing\_of\_post', 'Day\_of\_week', and 'Type\_of\_post'.

# **Analysis Results**

## **Question 1:**

To forecast two dependent variables the quantity of likes and comments on Instagram post we performed an OLS regression study. We ran two distinct models, one for predicting likes and another for predicting comments, because regression analysis can only handle one dependent variable at a time.

1. **Regression Models**:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Metrix | **R-squared** | **Adjusted R-squared** | **F-statistic** | **Prob(F-statistic)** |
| Model 1: Likes | 0.416 | 0.415 | 583.4 | 0.00 |
| Model 2: Comments | 0.064 | 0.063 | 55.85 | 4.06e-259 |

**Model 1:**

* 41.6% of the variation in the dependent variable LIKES is explains the linear regression model has a decent level of explanatory power.
* Considering the number of predictors in the model shows that, even after complexity is considered, the model still performs well.
* F-statistic: 583.4 with a 0.000 p-value. The overall statistical significance of the model is confirmed by the F-statistic.

**What variables had a significant influence on the dependent variables?**

* Number\_of\_followers : Controlling for all other variables, the average number of likes increases for every added follower.
* Number\_of\_people\_tagged : For every extra person tagged in a post, the likes decrease by an average.
* Number\_of\_users\_post : Increasing one user in the post is associated with a 36.81 increase in the number of likes, on average.
* Type\_of\_post\_Video :On average, videos receive less likes than photo posts.
* Month\_3will be the factors with the biggest positive influence, with Month\_3 having the greatest gain in likes compared to January.

**Model 2:**

* R-squared :This implies that the independent variables explain 6.4% of the variance in the dependent variable, COMMENTS. This performance supports a moderate fit.
* R-squared adjusted: This number confirms the model's moderate effectiveness, considering the number of predictors.

**What variables had a significant influence on the dependent variables?**

* Number\_of\_following : The number of comments rises with each new account followed.
* Number\_of\_people\_tagged : Comments increase with each additional tagged individual.
* Type\_of\_post\_Video : Compared to photo posts, video posts receive more comments.  
  LIKES : Higher comments are linked to more likes.
* Night Posts : Little more comments are left on posts made at night.  
  Evening postings : Compared to other times, evening postings get less comments.
* Most Important Factor: Month\_3 : Comments on March posts are most negatively impacted.  
  LIKES : Because of its scalability, it has a significant cumulative positive influence.
* **Which Variables Had the Greatest Influence?**

Likes Model:

Positive impact:

Number\_of\_followers (strongest effect, 𝛽 =0.667, β=0.667, p < 0.001).

COMMENTS (𝛽 = 0.485,β=0.485, p < 0.001).

Number\_of\_users\_post (𝛽= 0.017,β=0.017, p < 0.001).

Negative impact:

Number\_of\_following (𝛽 = − 0.063,β= 0.063, p < 0.001).

Length\_of\_Post (𝛽 = 0.0098,β=0.0098, p < 0.001).

Specific months

August 𝛽 = 0.103 , September 𝛽 = 0.166

Day of the Week and Timing of Post:

Minimal or no significant effects for most days or times, suggesting other factors might outweigh timing.

Comments Model:

The R-squared (0.064) indicates that 6.4% of the variation in COMMENTS is explained by the model.

The model is statistically significant overall, with a significant F-statistic.

Key Predictors:

Positive impact:

Certain months : August 𝛽=0.073, p < 0.001; September 𝛽= 0.126

LIKES (𝛽=0.779,β=0.669, p < 0.001).

Negative impact:

Timing of posting

night 𝛽= 0.042, p < 0.001

Day of the Week:

Most days show insignificant coefficients, with only marginal significance for Saturday, 𝑝=0.024

## **Question 2:**

**New variable which we have created:**

emoji\_count:

Post\_Activity\_Intensity

Has\_URL

Tag\_Usage

Hashtags\_Count:

**What analysis did you run for this question?**

**Analysis of Regression:**  
To predict likes and comments, an Ordinary Least Squares (OLS) regression model was constructed utilizing the existing predictors and the new variables (emoji\_count,, Post\_Activity\_Intensity, Tag\_Usage, and Hashtags\_Count).

**The correlation matrix**  
To comprehend the relationships, the correlation between the independent factors and the variables that are dependent (likes and comments) was examined.

**How well did the model perform**

R-squared (for likes): 0.417

In comparison to the original model , the model explains 41.7% of the variance in likes.  
R-squared adjusted: 0.416   
The model's explanatory power is considerably increased by the new variables without overfitting, as indicated by the adjusted R-squared, which is still robust.  
P-value < 0.001 indicates that the F-statistic is significant.  
shows that the model is statistically significant overall.

**variables which has significant influence on the dependent variables:**

Number of followers : Positive influence is a significant variable for likes.  
Avg\_Likes\_Per\_Follower: Strongly favorable impact.  
COMMENTS: Positive effect   
emoji\_count: Adverse impact   
Important Factors for Remarks:  
Favorite :Very favorable impact.  
The coefficient for emoji\_count is, which indicates a slight negative influence.

**Interpreting the coefficients of significant variables:**

For Likes:

1. Number\_of\_followers Each additional follower increases likes on average.
2. Avg\_Likes\_Per\_Follower :A one-unit increase in Avg\_Likes\_Per\_Follower increases likes on average.
3. COMMENTS :Each additional comment increases likes by on average.
4. emoji\_count :Each additional emoji decreases likes by on average.

For Comments:

1. LIKES : Each additional like increases comments by on average.
2. emoji\_count : Each additional emoji slightly decreases comments by on average.

**variables added to the model improved the performance**

Yes, the added variables improved the model:

For Likes: The R-squared increased from 0.416 (Q1) to 0.417, which means the new variables explain more variance in the likes.

For Comments: Additional variables such as LIKES and emoji\_count provided meaningful insights and improved engagement predictions.

## 

## **Question 3:**

**Linear Regression Analysis:**

Using independent factors, such as IsWeekend as a new variable, and the dependent variable (LIKES or COMMENTS), Ordinary Least Squares (OLS) regression was performed.  
To examine the independent variables' predictive ability, both standardized and unstandardized models were created.

**How well did the model perform**

LIKES is the model's dependent variable.

R-squared: 0.415: The model explains 41.5% of the variation in likes.

Adjusted R-squared: 0.415: This demonstrates that, after the number of components is controlled for, the model has a reasonable explanatory power.

COMMENTS is the model's dependent variable.

R-squared: 0.064: The model can account for 6.4% of the variance in comments.

Adjusted R-squared: 0.063

Comparatively speaking, its predictive value is lower than that of the LIKES model.

**The dependent variables were significantly impacted by variables**  
LIKE model  
These variables have a p-value of less than 0.05:  
Positive Influence: Number of Followers: Likes rise with each new follower.  
Time\_of\_post\_Night (Favorable Impact): Posts made at night get more likes.  
Month Variables: Positive impact Months 2, 12.  
Months seven, nine, and eleven: Negative impact.  
COMMENT model  
These variables have a p-value of less than 0.05:  
LIKES (Significant Impact): More likes result in more comments.  
Negative impact: Posts with a higher number of user interactions tend to receive fewer comments.  
Month-specific variables:  
Month 2 Month 3: Negative impact.  
Month 12: Positive impact.

**Replacing this variable improve the predictive performance of the model?**

Replacing IsWeekend: The inclusion of IsWeekend did not significantly improve the model's performance:

In the LIKES model, R-squared remained at 0.438.

In the COMMENTS model, R-squared remained at 0.064

The coefficient for IsWeekend was not statistically significant (p-value > 0.05), indicating that weekends do not have a significant influence on likes or comments.

## **Question 4 :**

The analysis involved separate OLS regression models for micro-influencers and macro-influencers to understand the relationships between the dependent variable and independent variables such as follower count, post timing, hashtags, and user interactions.

**How well did the models perform?**

Micro-influencers: The R-squared of likes is 0.083, meaning that the model explains 8.3% of the variance in the likes.

The R-squared for comments is 0.007, explaining 0.7% of the variance.

Macro-Influencers: The R-squared for likes is 0.381, accounting for 38.1% of the variance.

R squared for comments is 0.072, meaning that 7.2% of the variance can be explained.

It is also observed that the macro-influencer model performs a lot better compared to the micro-influencer model since the former can explain more variance in engagement.

**What variables had a significant influence on the dependent variables for micro and macro influencers**

Micro-Influencers:

Likes: Significant variables include Number\_of\_followers, Number\_of\_people\_tagged, and Length\_of\_Post.

Comments: Few variables showed significance; notable ones included Timing\_of\_post\_Evening.

Macro-Influencers:

Likes: Significant variables include Number\_of\_followers, Number\_of\_people\_tagged, Timing\_of\_post\_Night, and Length\_of\_Post.

Comments: Significant variables include Number\_of\_followers and Timing\_of\_post\_Night.

**Did the coefficients for the input variables differ between the two models?**

Yes, the coefficients varied significantly:

This would have a more pronounced effect in the case of Number\_of\_followers for the macro-influencer model since, with more followers, the engagement grows proportionally.

It appears that timing effects were stronger, such as Timing\_of\_post\_Night in the macro model, indicating that timing has a more significant effect on larger audiences.

Other variables such as Number\_of\_people\_tagged and Length\_of\_Post were different in magnitude from micro to macro models, since audiences behave quite differently.

**Recommendations about increasing engagement to micro and macro influencers**

A good way for the micro-influencers to increase engagement would be to focus on tagging more people in their posts because more tagged people mean more likes. This means that scheduling posts preferably during evening hours can really boost interaction; similarly, keeping posts short and direct increases their potential to connect with their audiences deeper and promote conversations.

The secret to successfully generating likes and comments for macro-influencers is by leveraging their enormous following. Posting at night is very beneficial, mainly because it increases audience engagement. Another tactic to increase likes is to write interesting and thorough captions, as longer and more educational postings appear to have a greater effect. Moreover, tagging other accounts might amplify this even further, and such a method could be useful for increasing interactions with a larger audience.

# **Conclusion and Recommendations**

This research offers valuable insights for influencers by emphasizing key factors that influence Instagram engagement. Due to their dynamic and engaging nature, videos consistently attract more attention than images. Moreover, timing plays a crucial role: micro-influencers with lesser followers experience higher engagement during the evening, while macro-influencers with large audiences post at night for improved results.

When used in the correct amounts, hashtags and mentions serve as powerful tools for increasing likes and comments. Moreover, posts shared on weekends or during the evenings generally perform better, emphasizing the importance of aligning posting times with audience engagement patterns.The data indicates that mentions and follower counts have a considerable impact on engagement, representing approximately 42% of the differences in likes. Although predicting comments was more challenging, factors such as hashtag frequency and emoji usage provided valuable insights.

Micro-influencers can enhance their reach by tagging people, scheduling evening posts, and crafting brief, impactful captions. Macro-influencers should focus on maximizing their large audiences by posting during the evening, crafting engaging and detailed captions, and tagging strategically. By implementing these strategies and understanding audience preferences, influencers can significantly boost engagement and achieve their social media goals.

# **References**

Dessler, Wendy. “Social Media: Videos vs. Photos - Which One Is More Engaging.” *Promo.com*, 23 Nov. 2020, promo.com/blog/social-media-videos-vs-photos-which-one-is-more-engaging.

“The Ultimate Guide to Instagram Content Management for Growth and Engagement | OnlySocial.” *OnlySocial*, 28 Nov. 2023, onlysocial.io/instagram-content-management/.

Thomas, Monique. “The Best Types of Instagram Posts Ranked: Carousel, Image, or Video? | Later.” *Later.com*, Later Blog, 8 Apr. 2022, later.com/blog/best-types-of-instagram-posts/?#what-type-of-instagram-feed-post-gets-the-most-engagement/. Accessed 12 Dec. 2024.