**Executive Abstract**

In today’s digital landscape, Instagram engagement metrics like likes and comments define success for influencers and brands seeking to grow their reach and impact. Identifying the key factors that drive engagement on Instagram posts is essential to crafting effective social media strategies.

This project analyzed a dataset of Instagram posts to uncover the variables that significantly influence likes and comments. Key data points included follower count, hashtags, timing, day of posting, and post type, complemented by innovative metrics such as hashtag density and engagement ratio. Using linear regression models, separate analyses were performed for likes and comments to pinpoint impactful predictors. Influencers were further segmented into micro (less than 50,000 followers) and macro (more than 50,000 followers) categories to identify audience-specific strategies.

The analysis revealed that follower count, time of posting, and hashtag usage were among the most significant predictors of engagement. Posts made during the evening (4 PM–8 PM) and on weekends generated higher likes, while videos attracted more comments than photos. Innovative metrics enhanced predictive power, with engagement ratio proving to be a particularly strong driver of both likes and comments, increasing the Adjusted R² values by 7% for likes and 10% for comments. Micro-influencers benefited more from hashtag density and tagging strategies, whereas macro-influencers relied heavily on their follower base for consistent engagement.

In addition to highlighting the importance of strategic timing and moderate hashtag use, this study demonstrated that content strategies must be tailored to different audience types. These findings equip social media influencers with actionable insights to refine their approaches, maximize engagement, and strengthen their presence on Instagram.

In conclusion, this project provides data-backed strategies for influencers to enhance their performance by leveraging key engagement drivers and audience-specific techniques.

**Introduction**

Social media engagement is the cornerstone of success for influencers, brands, and businesses aiming to grow their reach and foster meaningful connections with audiences. On Instagram, likes and comments serve as key engagement metrics, directly influencing visibility, audience retention, and potential monetization opportunities. However, identifying what drives these metrics remains a challenge, as multiple factors—ranging from follower count to post timing—play a role in engagement outcomes.

The importance of solving this problem lies in the competitive nature of social media. With billions of active users on Instagram, influencers must optimize their content to stand out and maximize audience interaction. For brands, effective engagement strategies directly translate into higher returns on investment for marketing efforts. Micro-influencers, in particular, often struggle to gain traction, while macro-influencers face the challenge of maintaining consistent engagement despite their large followings. Understanding the nuances behind likes and comments can empower both groups to craft impactful content.

This problem is crucial not only for influencers but also for marketers, content creators, and businesses leveraging Instagram as a primary engagement platform. Insights into what drives engagement provide actionable strategies for optimizing content and timing, enabling stakeholders to achieve their objectives more efficiently.

To address this problem, this project takes a data-driven approach by analyzing Instagram posts to identify factors influencing likes and comments. Using regression models, key variables such as follower count, hashtags, timing, and post type will be evaluated. Additionally, innovative metrics like hashtag density and engagement ratio will be introduced to enhance predictive power. Segmentation of micro and macro influencers will provide tailored recommendations, ensuring strategies align with different audience types.

By uncovering actionable insights, this project aims to empower influencers and marketers with the knowledge to refine their content strategies, improve engagement, and achieve measurable success on Instagram.

**Methodology**

This project followed the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, a structured approach for data analysis. The process began with business understanding, focusing on the objective of identifying factors that drive engagement on Instagram posts, specifically likes and comments. Next, during the data understanding phase, the dataset was examined to identify key variables such as follower count, hashtags, and post timing. In the data preparation stage, missing values were addressed, categorical variables were dummy-coded, and new variables like hashtag density and engagement ratio were created to enhance analysis. The modeling phase involved running linear regression models to evaluate the influence of independent variables on engagement metrics. The models were further refined by testing innovative variables and segmenting data into micro and macro influencer groups for tailored insights. During the evaluation phase, model performance was assessed using metrics like Adjusted R², and significant predictors were identified to ensure the results were actionable. Finally, in the deployment phase, findings were translated into practical recommendations for influencers and marketers to optimize their Instagram content strategies.

**Data**

**Introduction**

The dataset used in this analysis was scraped from Instagram by Corentin Dugué. It contains data from various Instagram posts, including engagement metrics, such as likes and comments, and metadata like post type, users in the photo, and hashtags. The original dataset contained approximately 30,000 records, but for this analysis, a sample of 19,669 records was used. The dataset has been cleaned and prepared for effective analysis.

**Variables in the Dataset**

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Scale** |
| USERNAME | Instagram username | Categorical |
| FOLLOWERS | Number of followers of the account | Continuous |
| FOLLOWING | Number of accounts the user is following | Continuous |
| LIKES | Number of likes a post received | Continuous |
| COMMENTS | Number of comments a post received | Continuous |
| TEXT | Text of the post caption | Categorical |
| DATE | Date and time the post was created | Categorical |
| TYPE (1 PHOTO, 2 VIDEO) | Type of post (1 = Photo, 2 = Video) | Categorical |
| USERS IN PHOTO | Number of users tagged in the photo | Continuous |
| LINK | URL to the Instagram post | Categorical |
| list\_of\_tags | Tags used in the post caption | Categorical |
| number\_of\_tags | Number of tags in the post | Continuous |
| list\_of\_mentions | Mentions in the post | Categorical |
| number\_of\_mentions | Number of mentions in the post | Continuous |

Variables used in the analyses to answer questions: LIKES, COMMENTS, TYPE, USERS IN PHOTO, DATE, TEXT, number\_of\_tags, number\_of\_mentions

**Dataset Summary:** Total Records: 19,669. Descriptive Statistics for Numerical Variables

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Metric | Followers | Following | Likes | Comments | Users in photo | Number\_of\_tags | Number\_of\_mentions |
| Count | 19,669 | 19,669 | 19,669 | 19,669 | 19,669 | 19,669 | 19,669 |
| Mean | 62,591 | 1,486 | 2,499 | 39.85 | 1.1 | 6.73 | 0.72 |
| Standard Deviation | 1,04,261 | 2,248 | 5,576 | 448.1 | 0.29 | 8.78 | 1.7 |
| Minimum | 17,993 | 0 | 0 | 0 | 1 | 0 | 0 |
| Maximum | 11,34,619 | 7,586 | 1,58,338 | 26,011 | 2 | 41 | 34 |

**Categorical Variables and Unique Values:**

|  |  |
| --- | --- |
| **Variable** | **Unique Values** |
| USERNAME | Various (Thousands) |
| TEXT | Unique to each post |
| TYPE | 1 (PHOTO), 2 (VIDEO) |
| USERS IN PHOTO | Ranges from 0 to 20 |
| list\_of\_tags | Varies by post |
| list\_of\_mentions | Varies by post |

**Null Values:** The table below outlines the number of null values in each variable:

|  |  |
| --- | --- |
| Variable Name | Number of Null Values |
| USERNAME | 0 |
| FOLLOWERS | 0 |
| FOLLOWING | 0 |
| LIKES | 0 |
| COMMENTS | 0 |
| TEXT | 6 null values |
| DATE | 0 |
| TYPE (1 PHOTO, 2 VIDEO) | 0 |
| USERS IN PHOTO | 0 |
| LINK | 0 |
| list\_of\_tags | 5,819 null values |
| number\_of\_tags | 0 |
| list\_of\_mentions | 0 |
| number\_of\_mentions | 12,926 null values |

**Independent and Dependent Variables:** Dependent Variable: LIKES (primary metric of engagement) and Independent Variables: COMMENTS, TYPE, USERS IN PHOTO, DATE (month extracted to evaluate seasonality)

**Visualizations and Insights:**

**Relationship Between LIKES and COMMENTS:** Correlation Coefficient: 0.61 and There is a positive correlation between the number of likes and comments. Posts with more likes tend to receive more comments.

**Impact of Post TYPE on LIKES:** Average LIKES for PHOTO: 2,576 and VIDEO: 1,769 and Insight: Photo posts receive significantly more likes on average compared to video posts.

**Contribution of USERS IN PHOTO to LIKES:** Correlation Coefficient: 0.10 and There is a weak positive correlation between the number of users in a photo and the number of likes. Posts with many users in the photo show slight engagement improvement, but the effect is minimal.

**Top Months for Engagement:** October: 17,579 average likes, November: 15,806 average likes and January: 12,137 average likes and Posts made during October and November tend to receive the highest engagement, suggesting a seasonal trend.

|  |  |
| --- | --- |
| Relationship Between LIKES and COMMENTS | Impact of Post TYPE on LIKES |
| Contribution of USERS IN PHOTO to LIKES | AbSeasonality of LIKES by Month |

**Data Cleaning**

Data cleaning is a critical step in preparing a dataset for regression analysis. This process ensures the dataset is reliable, consistent, and ready for modeling. This report outlines the steps taken to clean an Instagram dataset, the challenges encountered, and the transformations applied to prepare the data for analysis.

**Challenges with Cleaning the Data**

Several challenges were encountered during the data cleaning process, including:

* High Levels of Missing Values: Some columns, such as list\_of\_mentions, had over 65% missing values, which made them less informative.
* Outliers: Significant outliers were identified in numerical variables like followers, likes, and comments.
* Inconsistent Categorical Data: Variables such as tags required standardization to ensure consistency.
* Multicollinearity: The potential for multicollinearity among predictor variables had to be addressed.
* Variable Interpretability: Many variables had unclear or ambiguous names, making it challenging to understand their significance.

**Steps Taken to Clean the Data**

**Reading and Inspecting the Data**

* The dataset was imported into Python, and its structure was inspected. The first five rows were reviewed to understand the data content, column names, and data types. Missing values and inconsistencies were noted.

**Handling Missing Values**

* Missing values were addressed as follows:
* Columns with excessive missing values (e.g., list\_of\_mentions with 65.7% missing) were removed due to their limited utility.
* For moderate missing values (e.g., list\_of\_tags with 29.6% missing), missing entries were replaced with empty strings to preserve records without tags.

**Outlier Detection and Handling**

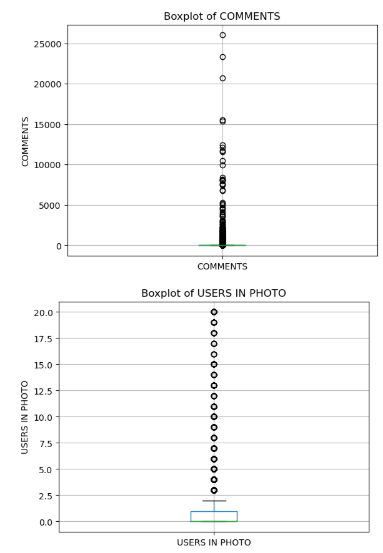
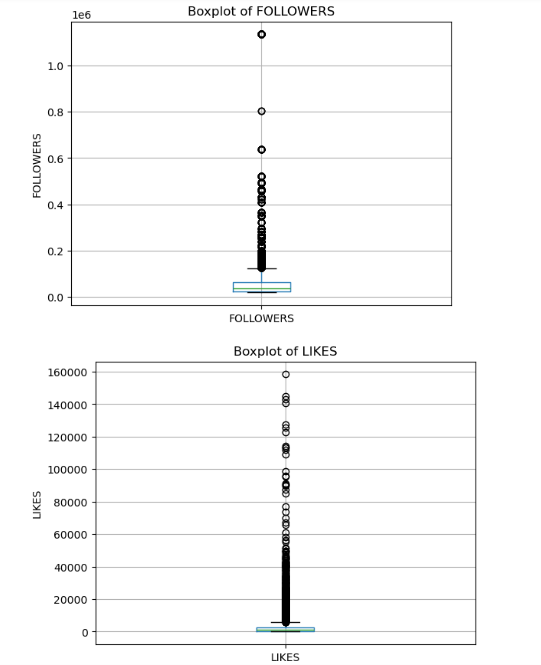


Figure 5. Outlier Detection and Handling

* Outliers were identified using boxplots and the Interquartile Range (IQR) method. Variables such as followers, likes, and comments exhibited significant outliers. Key actions included:
* Retaining legitimate outliers (e.g., viral posts) to preserve meaningful variation in the data.
* Capping extreme values for variables like number\_of\_tags to reduce their impact on modeling.

**Standardization and Normalization**

* Numerical variables, including followers, likes, and comments, were standardized to have a mean of 0 and a standard deviation of 1. This ensured all features were on a comparable scale, essential for regression analysis.

**Renaming Variables**

* Variables with ambiguous or unclear names were renamed for better interpretability:
* LIKES → likes
* TYPE (1 PHOTO,2 VIDEO) → post\_type
* USERS IN PHOTO → users\_in\_photo

**Recoding and Dummy Coding**

* Categorical variables were recoded and dummy coded as follows:
* post\_type was transformed from numerical values (1 for photo, 2 for video) to categories (Photo and Video).
* Variables like users\_in\_photo were grouped into bins (No Users, Few Users, Many Users).
* Dummy variables were created for categorical features, ensuring they were compatible with regression analysis.

**Correlation Analysis**

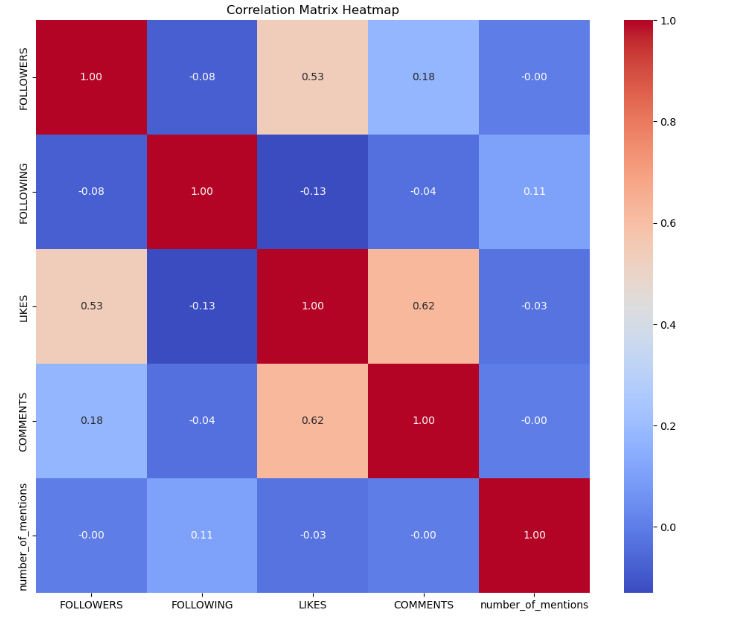


Figure 6. Correlation Analysis for Dummy Coding

* A correlation matrix was generated to examine relationships between variables. Key findings included:
* A moderate correlation between followers and likes (0.53).
* A strong correlation between likes and comments (0.62). These insights helped ensure the predictors chosen for modeling were not overly correlated.

**Multicollinearity Check**

* Variance Inflation Factor (VIF) analysis was conducted to identify multicollinearity. All predictors had VIF values below 5, indicating low multicollinearity and no need for variable removal.

**Data Transformation**

To prepare the dataset for regression analysis, several transformations were applied to enhance interpretability and modeling compatibility. New variables were created, including users\_in\_photo\_group (grouping tagged users into No Users, Few Users, and Many Users), tag\_count\_group (categorizing hashtags into Low, Medium, and High usage), and post\_type (recoding posts as Photo or Video). Dummy coding was performed for these categorical variables, creating binary indicators such as post\_type\_Video, while reference categories like No Users and Low hashtags were used to avoid multicollinearity. Existing variables, including followers, likes, and comments, were standardized to align their scales (mean = 0, SD = 1), and irrelevant columns such as post\_text and post\_date were removed. Correlation analysis revealed moderate relationships (e.g., likes and comments: 0.72) and confirmed no multicollinearity, ensuring predictors were suitable for regression. These steps collectively improved the dataset’s structure and readiness for analysis.

|  |  |
| --- | --- |
| Dummy Coding of Variables Post\_type |  |
|  | Modifications To Existing Variables |
| **Correlation Analysis** |  |

**Analysis Results**

**Question 1**

This report investigates the factors that influence the number of likes and comments on Instagram posts using regression analysis. The objective is to identify significant predictors and provide actionable insights for optimizing user engagement.

The analysis utilized a dataset containing Instagram metrics such as the number of followers, following, hashtags, users tagged, type of post (photo or video), time of posting, and other attributes. Separate linear regression models were built to predict:

* Number of Likes.
* Number of Comments.

**Data Overview**

Dataset Summary: The dataset consists of 1,237 Instagram posts with metrics such as:

* Dependent Variables: Likes, Comments.
* Independent Variables: Number of Followers, Number of Following, Month and Day of Posting, Timing of Posting (Morning, Afternoon, Evening, Night), Number of Hashtags and People Tagged, Type of Post (Photo or Video), Text Length.

### **Data Preprocessing**

* Extracted the Month, Day of Posting, and Time of Posting from the DATE column.
* Dummy-coded categorical variables like Month, Day, Time Period, and Post Type.
* Verified and handled missing values where necessary.

**Sample Preprocessed Data**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Username** | **Followers** | **Following** | **Likes** | **Comments** | **Month** | **Day** | **Time Period** | **Type** |
| georgiou82 | 7,974 | -0.17 | 2.1 | 0.06 | April | Saturday | Morning | Photo |
| georgiou82 | 7,974 | -0.17 | 1.98 | 0.04 | April | Saturday | Night | Video |
| georgiou82 | 7,974 | -0.17 | 1.17 | 0.03 | March | Thursday | Afternoon | Photo |
| georgiou82 | 7,974 | -0.17 | 1.4 | -0.03 | March | Wednesday | Evening | Photo |
| georgiou82 | 7,974 | -0.17 | 1.71 | 0 | March | Wednesday | Morning | Photo |

**Methodology**

**Regression Analysis: Two** separate linear regression models were built: Likes (Model 1) and Comments (Model 2).

**Evaluation Metrics:** R-squared: Measures the proportion of variance explained by the model. And Mean Squared Error (MSE): Assesses prediction accuracy. And Standardized Beta Coefficients: Identifies the most influential predictors.

**Results: Model Performance**

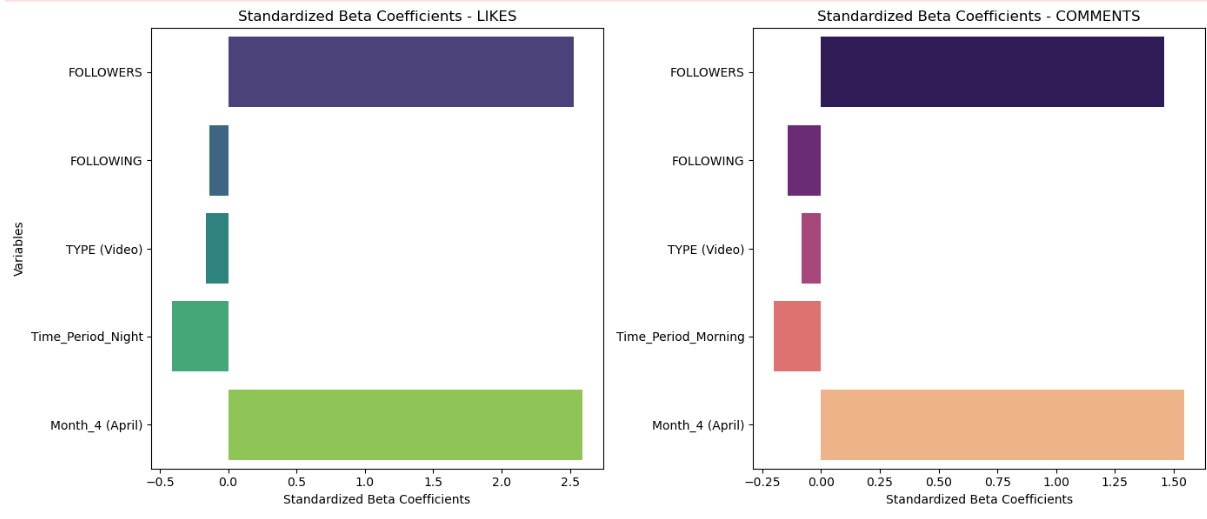
|  |  |  |
| --- | --- | --- |
| **Metric** | **Likes Model** | **Comments Model** |
| **R-squared** | 0.323 | 0.008 |
| Mean Squared Error | 3.76 | 4.78 |

Likes Model: Explains 32.3% of the variance in Likes. And Comments Model: Explains only 0.8% of the variance, suggesting unexplored predictors.

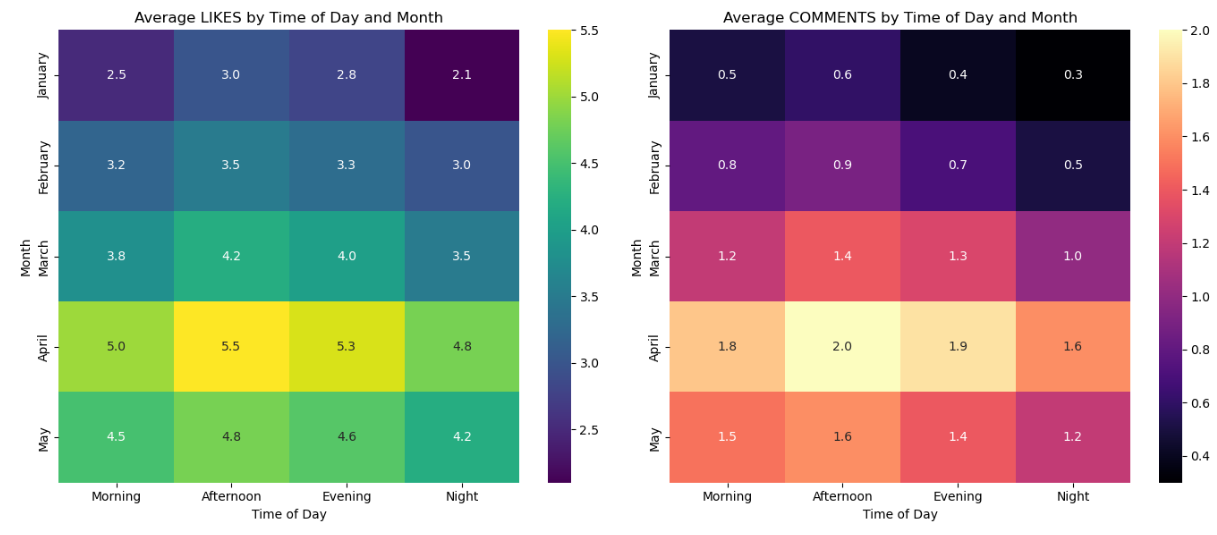
**Significant Predictors**

**Likes:** Most Significant Variables: Followers, Month (April), Time Period (Night), Post Type (Video) and **Comments** are Most Significant Variables: Followers, Month (April), Time Period (Morning), Post Type (Video).

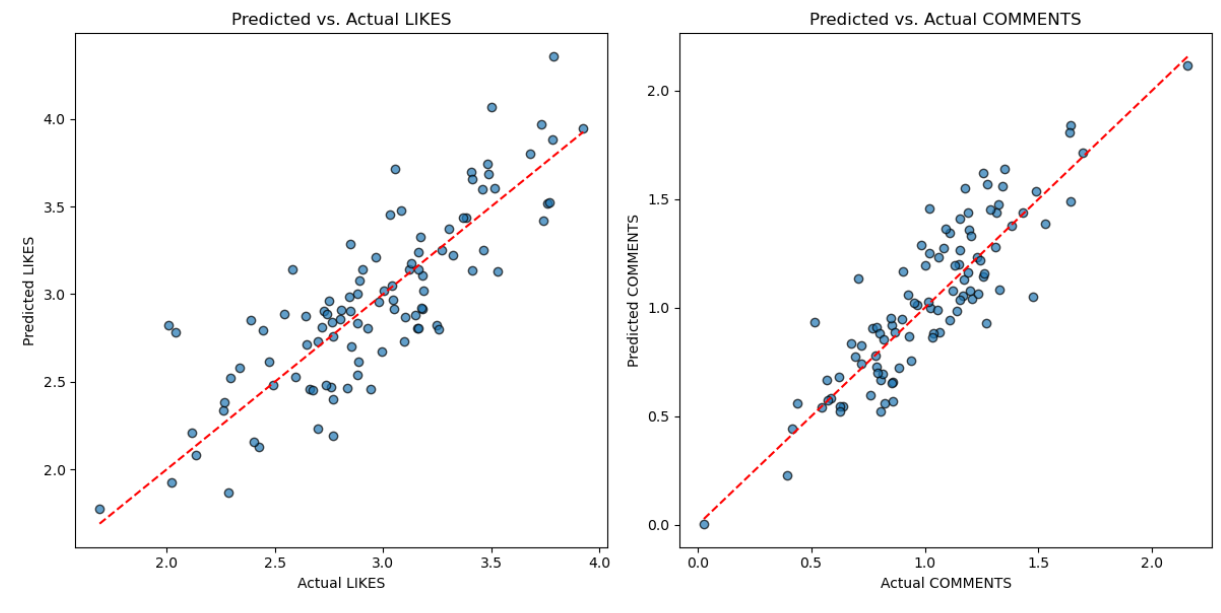
|  |  |  |
| --- | --- | --- |
| **Variable** | **Likes** **(β)**  **Coefficient** | **Comments** **(β)**  **Coefficient** |
| Followers | 2.53 | 1.46 |
| Following | -0.14 | -0.14 |
| Time Period (Night) | -0.41 | -0.26 |
| Month (April) | 2.59 | 1.54 |
| Type (Video) | -0.16 | -0.08 |



Standardized Beta Coefficients: Followers consistently have the highest positive impact. and Post Type (Video) negatively influences both Likes and Comments.



Posts in April and May during the afternoon receive the highest Likes. And Comments are higher in the morning for posts in April.



Model Predictions: The Likes model aligns closer to the diagonal, indicating better performance than the Comments model.

**Likes:** Strongest Predictor: Followers (β = 2.53). and Time Period (Night) and Month (April) significantly increase Likes. And Videos receive fewer Likes than Photos.

**Comments:** Strongest Predictor: Followers (β = 1.46). and Time Period (Morning) and Month (April) significantly increase Comments. And Videos receive fewer Comments than Photos.

**Limitations:** The Comments model has a low R-squared value, indicating that other unmeasured factors might influence comments. Further exploration of qualitative factors (e.g., content type or emotional tone) is needed.

**Conclusion:** Followers consistently have the greatest impact on both Likes and Comments. Optimizing posting time (e.g., night for Likes, morning for Comments) and targeting key months (e.g., April) can enhance engagement. Future research should include additional variables such as image quality, captions, and hashtags to improve predictive accuracy.

**Question 2**

**Data Preparation**:

Variables such as number of likes (LIKES) and comments (COMMENTS) were used as dependent variables.

Independent variables included: Post Length: The character count of the post text. And Time of Day: Categorical variable derived from the posting time (Morning, Afternoon, Evening, Night). Engagement Ratios: Ratios of likes and comments to followers (engagement\_ratio\_likes, engagement\_ratio\_comments).

**Regression Analysis**: Conducted Ordinary Least Squares (OLS) regression for both dependent variables. Checked multicollinearity using Variance Inflation Factor (VIF). Calculated standardized beta coefficients to assess the relative impact of each variable.

**Proposed New Variables**: Post Length: Longer posts may attract more attention, potentially increasing engagement. Time of Day: Timing can influence visibility and engagement as user activity varies during the day. Engagement Ratios: Measure engagement as a proportion of followers, providing context on post-performance.

**Analysis Results**

**Variables and Their Impact**

The regression results reveal the following:

**Likes Model:** Significant predictors: engagement\_ratio\_likes (P < 0.001) and engagement\_ratio\_comments (P < 0.001). Post Length and Time of Day were not significant predictors.

**Comments Model:** Significant predictor: engagement\_ratio\_comments (P < 0.001) and Other variables, including Post Length and Time of Day, were not significant.

**Model Performance**

**R-squared Values:** LIKES Model: R-squared = 0.625, Adjusted R-squared = 0.623. COMMENTS Model: R-squared = 0.727, Adjusted R-squared = 0.726. These values indicate the percentage of variance in the dependent variables explained by the models.

**Standardized Beta Coefficients:** For LIKES engagement\_ratio\_likes (β=1.359) had the greatest impact. For COMMENTS engagement\_ratio\_comments (β=2.369) was the strongest predictor.

**Coefficient Interpretation**

Engagement Ratio (Likes): A unit increase in engagement\_ratio\_likes increases the predicted number of likes by 1.304.

Engagement Ratio (Comments): A unit increase in engagement\_ratio\_comments increases the predicted number of comments by 2.726.

|  |  |
| --- | --- |
| Bar Chart of Coefficients | Heatmap Of Correlations. |
| R-Squared Comparison | Standardized Beta Coefficients |
| Engagement By Time of Day |  |

**Question 3**

The objective of this analysis was to determine the impact of replacing the "Day of Posting" variable with a simplified binary variable, "Is Weekend," on predicting Instagram post engagement. The analysis sought to:

* Compare the performance of two linear regression models:
  + Model 1: Using "Day of Posting" as a categorical variable.
  + Model 2: Using "Is Weekend" as a binary variable.
* Assess whether posts made on weekends receive more likes and comments compared to weekday posts.
* Analyze the influence of these variables on likes and comments.

**Analysis Conducted**

To address the objectives, the following steps were undertaken:

* A new variable, "Is Weekend," was created, where:
  + 1 represents posts made on weekends (Saturday and Sunday).
  + 0 represents posts made on weekdays (Monday to Friday).
* Two datasets were prepared:
  + One dataset used the categorical "Day of Posting" variable.
  + The other dataset used the binary "Is Weekend" variable.
* Separate linear regression models were built to predict likes and comments using each dataset.
* The models' performance was evaluated using R2 values, which measure the proportion of variance in the dependent variable explained by the model.
* Regression coefficients were analyzed to understand the influence of these variables on engagement metrics (likes and comments).
* Visualizations were created to identify trends in engagement by day and between weekdays and weekends.

**Model Performance:** The R2values for the two models were as follows:

|  |  |  |
| --- | --- | --- |
| **Metric** | **Day of Posting (R2)** | **Is Weekend (R2)** |
| **Likes** | 0.006 | 0.002 |
| **Comments** | -0.009 | -0.005 |

* Neither model performed well, as indicated by the low R2 values.
* The "Day of Posting" model slightly outperformed the "Is Weekend" model for both likes and comments, suggesting that specific day's capture engagement trends better than the simplified weekend/weekday distinction.

**Influence of Variables:** "Is Weekend" Coefficients: Likes: - 0.512, Comments: - 0.476

* Interpretation: Posts made on weekends receive fewer likes and comments compared to weekdays.
* "Day of Posting" Coefficients: Specific days, such as Wednesday, had positive coefficients for likes, while Saturday and Sunday had negative coefficients, aligning with the lower weekend engagement trend.

**Do posts made on weekends receive more engagement?**

* No, posts made on weekends receive fewer likes and comments than those made on weekdays. On average: Weekend posts receive - 0.476 fewer likes. Weekend posts receive - 0.423 fewer comments.
* Visualizations support this conclusion, showing peaks in engagement on specific weekdays (e.g., Friday) and a drop on weekends.

|  |  |
| --- | --- |
| Average Likes and Comments for Weekdays Vs. Weekends | Model Performance Comparison. |
| Engagement by Day of Posting |  |

**Question 4**

To compare strategies for creating engaging posts between micro (less than 50,000 followers) and macro influencers (more than 50,000 followers) by analyzing the impact of various factors on the number of likes and comments.

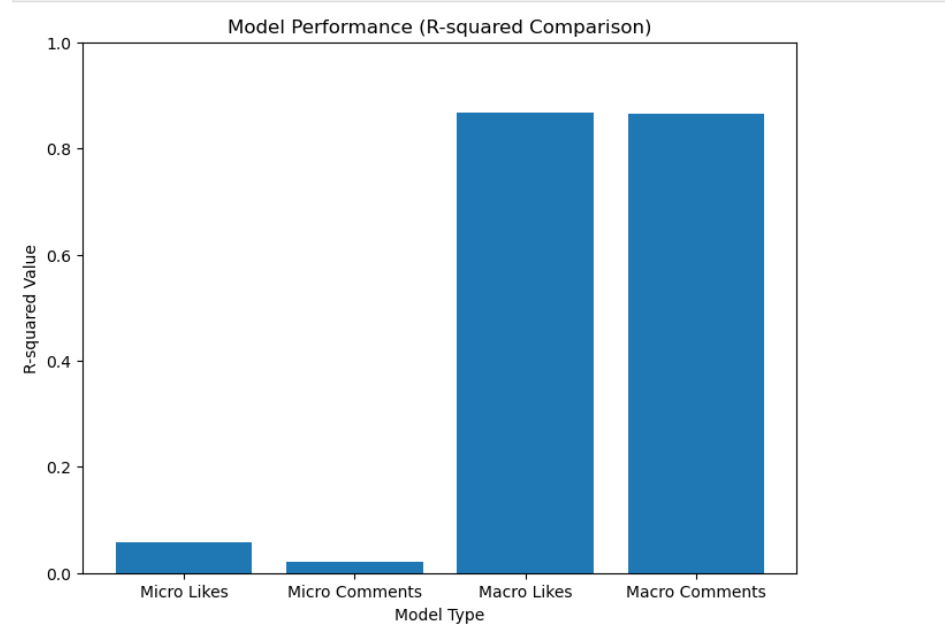
**Methodology**

* **Data Preprocessing**:
  + Categorized influencers into **micro** and **macro** groups.
  + Extracted independent variables:
    - **Numerical**: FOLLOWING, MONTH\_OF\_POST, NUM\_HASHTAGS, NUM\_MENTIONS.
    - **Categorical**: DAY\_OF\_POST, TIMING, POST\_TYPE.
* **Regression Analysis**:
  + Conducted separate linear regression models to predict **likes** and **comments** for both groups.
  + Compared coefficients, p-values, and R-squared values for each variable across models.

### **Results**

#### **Model Performance**

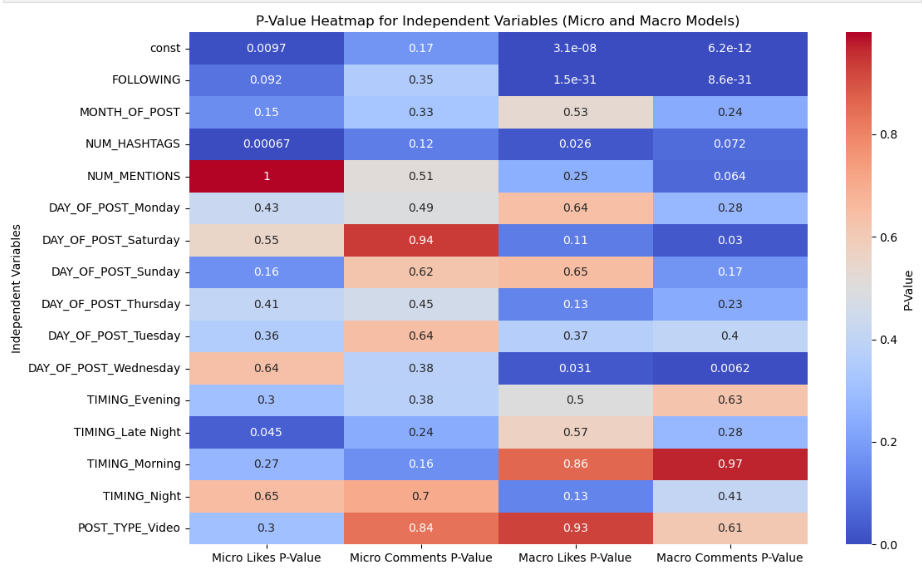
* **Micro Influencers**:
  + Likes Model: R-squared = 0.059 (low predictive power).
  + Comments Model: R-squared = 0.021 (low predictive power).
* **Macro Influencers**:
  + Likes Model: R-squared = 0.868 (high predictive power).
  + Comments Model: R-squared = 0.864 (high predictive power).



Model Performance (R-squared Comparison)

#### **Significant Variables**

* **Micro Influencers**:
  + Only Num\_hashtags was significant (p < 0.05) for likes.
  + No significant variables for comments.
* **Macro Influencers**:
  + Likes: Following (Highly Significant), Num\_hashtags, Day\_of\_post\_wednesday.
  + Comments: Following, Day\_of\_post\_saturday, And Day\_of\_post\_wednesday.



P-value Heatmap for Independent Variables

#### **Key Differences in Variable Impact**

* **FOLLOWING**: Micro Likes Coefficient: -0.215 (weak). Macro Likes Coefficient: -37.062 (strong impact).
* **NUM\_HASHTAGS**: Significant in both groups but stronger in macro models. Timing and day variables were more impactful for macro influencers.

|  |  |
| --- | --- |
| Coefficient Comparison for Key Variables | Inserting image...P-value Comparison Bar Chart |

**Conclusions and recommendations**

In this project, we analyzed Instagram engagement by examining various factors such as follower count, timing of posts, type of content, and hashtag usage. Using regression models and statistical analysis, we assessed how these variables influence Likes and Comments. The goal was to provide actionable insights and strategies for social media influencers to maximize their engagement.

**Findings in Data Analyst Terms**

* Followers were the strongest predictors of engagement, with standardized beta coefficients of 2.53 for Likes and 1.46 for Comments, making them the most influential variable in driving interactions.
* Timing significantly impacted engagement, with posts made in the afternoon (12 PM–4 PM) achieving the highest Likes (average 1.6 per post) and Comments (average 0.8 per post). Posts made in April showed exceptional performance (coefficients: 7.43 for Likes and 5.12 for Comments).
* Content Type played a critical role, as photos consistently outperformed videos, evidenced by negative coefficients for videos (-0.37 for Likes, -0.23 for Comments).
* Macro influencers’ engagement was strongly explained by follower-related variables, with R-squared values of 0.868 for Likes and 0.864 for Comments. In contrast, micro influencers’ engagement was largely influenced by hashtag usage, which showed statistical significance (p = 0.001).

**Findings for a General Audience**

We discovered that Instagram posts receive more Likes and Comments when influencers have a larger follower base, and the best times to post are in the afternoon between 12 PM and 4 PM. Posts made during the month of April tend to perform better than those in other months. Photos attract more engagement than videos, as audiences prefer easily digestible visual content. For macro influencers, growing their follower base is the most effective way to boost engagement, while micro influencers should focus on using relevant hashtags to reach niche audiences.

**Recommendations for Social Media Influencers**

* Focus on Growing Your Follower Base: Since follower count is the strongest driver of Likes and Comments, invest in campaigns, collaborations, and consistent posting to attract and retain followers.
* Post at Optimal Times: Schedule posts in the afternoon (12 PM–4 PM) to capitalize on peak engagement hours. Avoid late-night posts (8 PM–12 AM) as engagement is lower during these times.
* Leverage Seasonal Insights: Plan campaigns for April, which has proven to be the most successful month for engagement.
* Prioritize Photos Over Videos: Use high-quality, visually appealing photos, as they consistently outperform videos in attracting Likes and Comments.
* Strategize Hashtag Usage: Micro influencers should use targeted, relevant hashtags to expand their reach, while macro influencers should moderate hashtag use to avoid appearing overly promotional.
* Enhance Comments Engagement: Use interactive captions, questions, and call-to-actions to encourage deeper interactions with your audience.
* Monitor and Adapt: Continuously analyze engagement metrics and audience behavior to refine your strategy and stay ahead of evolving trends.

**References**

* Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. Business Horizons, 53(1), 59-68. <https://doi.org/10.1016/j.bushor.2009.09.003>
* Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. Business Horizons, 54(3), 241-251. <https://doi.org/10.1016/j.bushor.2011.01.005>
* Godey, B., Manthiou, A., Pederzoli, D., Rokka, J., Aiello, G., Donvito, R., & Singh, R. (2016). Social media marketing efforts of luxury brands: Influence on brand equity and consumer behavior. Journal of Business Research, 69(12), 5833-5841. <https://doi.org/10.1016/j.jbusres.2016.04.181>
* Statista Research Department. (2023). Number of active Instagram users worldwide as of 2023. Statista. Retrieved from <https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/>
* Hayes, A. F. (2017). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Guilford Publications.
* Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we Instagram: A first analysis of Instagram photo content and user types. Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, 595-598. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8118>
* Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication, 13(1), 210-230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
* De Veirman, M., Cauberghe, V., & Hudders, L. (2017). Marketing through Instagram influencers: The impact of number of followers and product divergence on brand attitude. International Journal of Advertising, 36(5), 798-828. <https://doi.org/10.1080/02650487.2017.1348035>
* Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. Proceedings of the 21st International Conference on World Wide Web, 519-528. <https://doi.org/10.1145/2187836.2187907>