

## **Social Media Content Analysis**

Social media marketing has become an increasingly integral part of businesses. This is because the digital world with insightful data generated online has become a useful platform for businesses to reach out to their audiences and potential customers and engage them with their brand, products, and services. People spend countless hours surfing through social media posts daily, and businesses may be missing out a lot of information if they can't attract attention from the ones who use social media daily. A successful business relies on having a strong online application presence and on engaging with customers.

There are four reasons why having social media content engagement is beneficial for brands and how it can positively affect the business's relationship with its potential customers. First, with social media content engagement, businesses can have a better chance to reach a bigger audience. Social media targets many people who use social media and the Internet regularly. It gives you a better chance to reach out to many people and at the same time to increase brand awareness and exposure among them. Second, it will help people put more trust into your brand. Staying active on creating attractive and informative social media contents will help the brand focus on producing quality content and giving better customer services. Showing attractive and informative posts can build the trust of post visitors and may turn them into our potential customers. Moreover, people usually start looking for the brand when they feel they need to buy products that are relevant to the ones from the brand. Getting enough likes, shares and comments, visitors will think the content and account trustable and therefore explore the brand more.

Third, having engagement with social media content helps the brand get valuable feedback from the public, potential customers, and customers. Social media is not only important for gaining new customers but also for helping the brand to improve its services by learning its feedback from buyers. For example, following up with complaints and concerns can help the brand improve the existing product design, change the way the information is presented in front of visitors and gather valuable information to make predictions on new launching products. Fourth, it helps the brand create an interactive environment with the visitors. Social media engagement is all about interacting with people, answering concerns, and analyzing the feedback collected. People are more inclined towards the brand if they feel that their questions and feedback are responded to. This kind of interaction helps improve the relationship

with customers and creates an interactive environment.

Before running any analysis to evaluate how engagement with digital content can be predicted, we decided to first preprocess the data as we were dealing with a large dataset. The first step was to clean data by removing outliers, replacing missing values and getting rid of columns where missing values were above 80% and probably would not be relevant for our analysis. As most of the columns' content were heavy text based, we had to find a way to encode and label them for easiest manipulation later. Moreover, we had to do a data binning for date columns to observe if there is a possible trend or an impact of time/date on engagement metrics.

Aiming to get a great performance for our models, we decided to proceed for a feature engineering before fitting data into regression as the number of predictors in the dataset is over 100. Our feature engineering techniques involved correlation, feature importance (using generalized linear model), forward/backward elimination and recursive feature elimination (RFE) to see if there were any variables highly correlated that we should be aware of; and, to eliminate the weakest features from our potential predictor variables list. We went ahead with feature importance as it was giving the best combination of features for higher accuracy of the model. Below is a screenshot that displays the top 20 most important features among all our variables.

only 20 most important variables shown (out of 120)

	overall
Post.Views	100.0000
Total.Views	10.5801
Total.Views.For.All.Crossposts	7.8379
User.Name_te	6.3678
Sponsor.Name_te	6.2235
Likes.at.Posting	3.7557
`TypeNative Video`	2.6124
Sponsor.Category_te	1.7989
Page.Category_te	1.5305
hasSponsor	1.4331
interrog	1.3908
`TypeLive video complete`	1.2905
male	1.1966
see	1.0919
feel	1.0666
percept	0.9463
hear	0.9067
body	0.7801
money	0.7757
clout	0.7234

Then, we tried different combinations when fitting our models using the list of predictors from our features importance to evaluate what could be the best variables

when predicting engagement metrics. Overall, our best model gave us R-squared of 0.44. From this analysis, we were able to conclude that among all variables, those that seemed more likely to have an impact predicting engagement metrics from a post would be *post views, total views, total views for all crossposts, user name*, Note: We kept aside the positively correlated expression columns like wow, haha, anger, interactions etc..., as they were overfitting the model with a very high R-squared. These features are directly tied to the number of likes a post gets and therefore provide little value for inference.

**Regression Output(Top regression based on R2 score) - Only showing variables significant above 10% level**

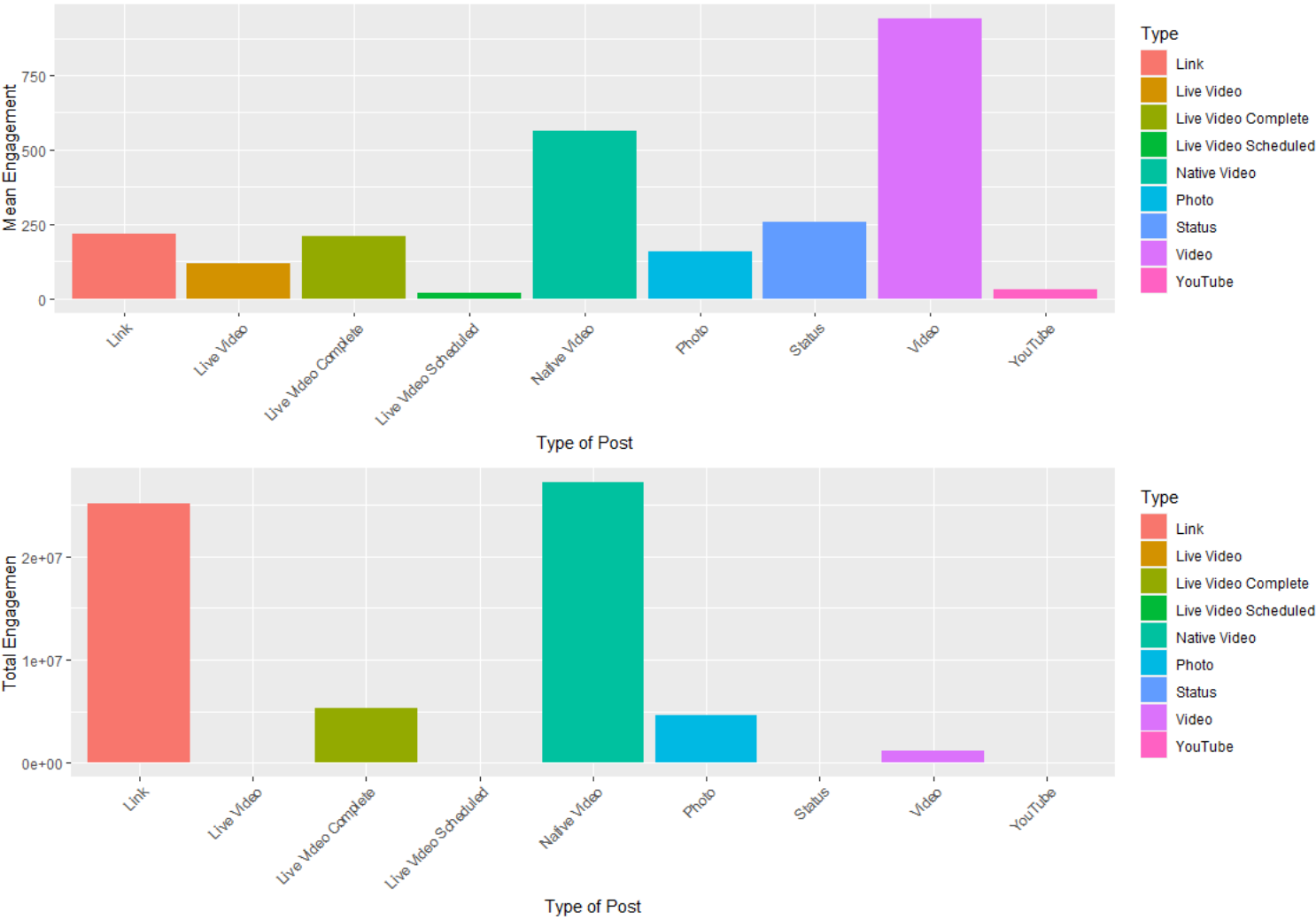
**Regression Results**

		<i>Dependent variable:</i>	
		Effects on Engagement	
Sponsor Name Encoded	0.63*** (0.05)	Anger	-20.60*** (7.18)
Post views	0.02*** (0.0001)	Nonfluencies	47.56** (23.20)
Total views	0.0003*** (0.0000)	Parentheses	16.31*** (5.56)
Total Views for all Crossposts	0.0002*** (0.0000)	Work	-5.43** (2.21)
User Name Encoded	0.41*** (0.02)	Leisure	-4.24* (2.36)
Interrogatives	-36.46*** (5.47)	Body	-10.33* (5.50)
Dashes	14.27*** (2.05)	3rd person singular	-27.09*** (7.86)
Male	38.05*** (5.72)	Female References	18.72*** (5.73)
Sponsor Category Encoded	-0.36*** (0.05)	Quantifiers	-6.45* (3.76)
2nd Person	15.15*** (3.50)	See	-6.25** (3.17)
Impersonal Pronouns	11.22*** (3.39)	Achievement	-7.90** (3.21)
Question Marks	-12.17* (6.34)	Risk	-12.73** (6.44)
Total function words	-4.89*** (1.16)	Emotional Tone	-0.69* (0.40)
Future focus	11.66*** (4.09)	Conjunctions	-6.25* (3.30)
Money	-14.06*** (3.36)	Type: Live Video Complete	-262.13** (112.19)
Post Created at Night	-48.49** (22.74)	Type: Native Video	-346.57*** (110.69)
Reward	-15.17*** (4.09)	Sponsor Name Encoded * hasSponsor	-0.21*** (0.05)
Positive emotion	13.88*** (4.01)	Intercept	475.52*** (140.23)
All Punctuation	-3.59*** (0.94)	Observations	221,127
Swear words	-76.33** (31.64)	R <sup>2</sup>	0.44
Power	9.15*** (2.53)	Adjusted R <sup>2</sup>	0.44
		Note:	*p<0.1; **p<0.05; ***p<0.01

Our analysis shows that the most important factors to drive people to engage more with social media posts are the number of views a post receives across all

platforms. This makes sense as considering that as the number of views increase, more people are susceptible to engage with the content being posted, leading to an increase in the number of likes, comments, and shares. Marketing is often a numbers game, with total impressions being an important metric to evaluate the success of a marketing campaign - or, in this case, a post.

The encoded variable for username is important for engagement. This comes without surprise, since we would expect an artist, influencer, or specialist to have higher engagement metrics for a post than a regular person. In addition, certain users have far more people seeing what they are posting due to their number of followers or friends, meaning higher number of views per post. Users that ranked highly included College Humor and Insider. (GRAPH FOR USERS AVG AND SUM ARE BELOW, reiterate social media engagement is a numbers game?)



The 'Type' variables show that the type of post matters. Post type 'Video' was used as the baseline, meaning it is included in the strongly positive intercept. Video content driving more engagement is not surprising since it tends to be easier to digest than text and keeps attention focused. The graph above shows that while links and native videos account for nearly all engagement, on average videos created more engagement than any other type of post.

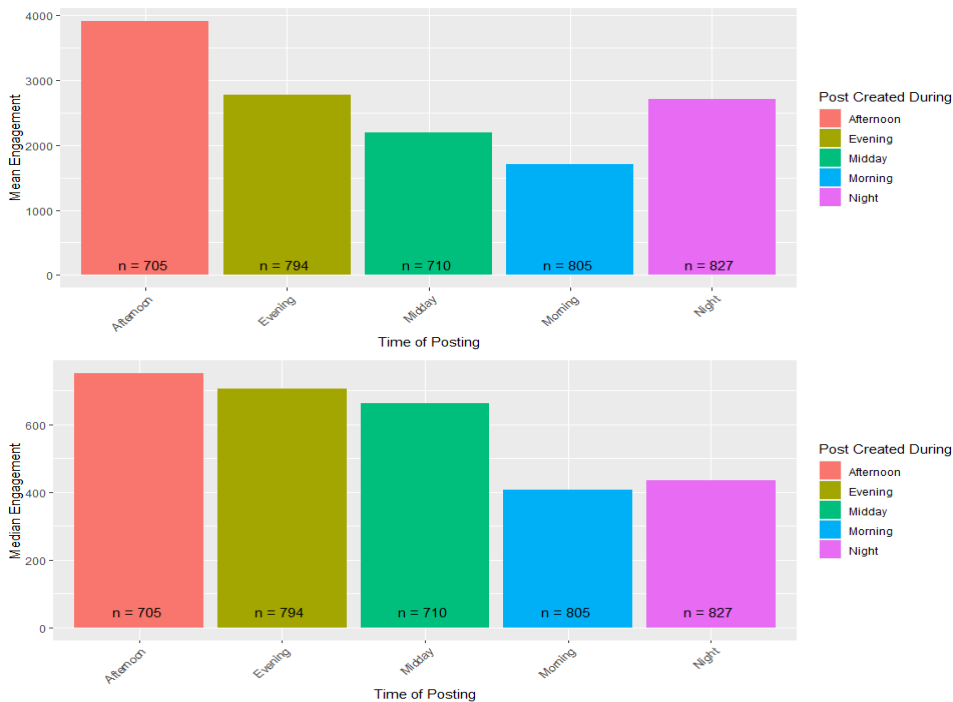
While most date and time features were not found to be important for driving engagement, posts made at night performed worse on average. This may be due to a common effect on social media: posts that do not find early success get buried quickly. Since most users are not active at night, posts made then will struggle with engagement.

Language analysis showed that the kind of language used matters as well. Swear words have a negative effect on engagement, which can be attributed to major pages that see a lot of traffic avoiding the use of swears. Second person pronouns have a positive effect, probably because they directly implicate a reader or viewer. References to sexes (i.e words like boy, girl, man, woman) had a positive effect. This may be because users react more strongly to content, they feel concerns them, like how brands will use gender-specific targeting for uni-gender products.

Our analysis shows that while both extrinsic and intrinsic factors matter for engagement, the former are the most crucial. Feature importance and regression models showed that engagement depended heavily on the who, when and what of a post as well as who sponsored it. While social media theoretically allows users to browse any kind of content, they want no matter who posted it, users prefer to stick to popular pages.

There are two types of implications: those that remain constant for all managers and those that vary depending on conditions. A general implication would be to avoid posting anything during the night, limit the use of swear words (an exception would be for more specialized users - certain communities may look favorably on swearing and have higher activity at night), use second person pronouns and prioritize videos instead of text posts. Short videos will be likely to drive more engagement than longer ones according to the success of apps like TikTok short videos.

The implications can differ depending on the resources available to social media managers. We have seen that the total views a post gets is the main factor driving engagement. A manager with access to sizable funds will prefer to post an ad on a popular page, for example popular videos, social media accounts and topics of significant events, with many viewers, however the bidding process will drive out those



lacking sufficient resources. Those with less funds should avoid sponsored posts, which often drive less engagement than non-sponsored ones, perhaps due to an adverse reaction by viewers to the presence of an explicit advertisement. If the manager is attempting to drive engagement

on their own page it will be worthwhile to attempt to cross post their page to another popular page, thereby capturing traffic. Also, those with less funds can try to reach out to social media influencers who have specific target audiences and have a decent number of followers.

We will give recommendations to the page “People”, which focuses on celebrity news. The graph below shows the mean and median engagement of posts made at certain times of the day. We can see that this page makes 42% of its posts in the morning or at night. As it shows, the average engagement of posts made in the morning is low, and while it is high for those at night the medians suggest that this average is pushed high by outliers. It would be better to increase the number of posts made for this page during the afternoon and the evening at the expense of those made in the morning and at night. They should still post during those hours to maintain a continuous stream of content, however. Analysis of the relationship between engagement and sponsors shows that sponsored posts have around 74% lower average engagement (and 86% lower median engagement) than non-sponsored posts. While sponsored posts are a way for the page to generate revenue, managers should be careful not to sponsor posts that they expect could drive larger revenue than average (unless contractually obliged).

