Executive Summary: Youtube Short Performance Within the Beauty Niche

Team: Rachael Alvir, Jo Beckford, Pranav Doshetty, Yasaman Sanatkhani, and Gabriela Terrazas

Introduction

We explored which factors of a social media post contribute to performance. Specifically, we examined a dataset of YouTube Shorts videos in the beauty niche to uncover how elements such as post timing, hashtags, and keyword usage relate to metrics such as views, likes, and comments. Often, creators tailor their content for better performance within their niche. For example, beauty content creators may hop on current beauty trends, including hashtags relevant to those trends. Gaming creators may choose to feature newly launched video games or gaming systems. We chose to analyze data taken strictly from youtubers in the beauty niche in order to identify niche-specific posting strategies.

Defining the Dataset

We scraped public Youtube Shorts data using the Apify platform. All posts from 3/1/2024 to 3/1/2025 were scraped from a list of 155 beauty influencers. This allowed us to gather metadata for thousands of posts from beauty influencers, including creator's username, total likes, total shares, total comments, post description, hashtags, and posting time. Data cleaning was conducted in multiple stages:

- Initial cleaning: We removed duplicate entries, handled missing values, and normalized text fields.
- Temporal filtering: We identified that posts made in the final month of our data collection window (close to March 1, 2025) had significantly fewer views, likely due to limited time to accumulate engagement. Using t-tests at a significance level of $\alpha = 0.005$, we confirmed that only the most recent month had statistically lower view counts. Therefore, we excluded this final month of data when analysing views to avoid skewing results.
- Feature engineering: In preparation for modeling, we engineered features that better captured patterns in post performance. This included extracting hour-of-day and day-of-week indicators, creating categorical variables that identified when certain keywords were used in the title or description, and log-transforming engagement metrics like views and likes to normalize skewed distributions.

Exploratory Data Analysis (EDA)

Data was split into training and test sets. EDA was performed using only the training dataset (50% of the full dataset). The training set and EDA were to be used to generate hypotheses about which hypothesis tests to run on the testing set for confirmation.

During the EDA stage, we explored the effect of hashtags, mentions of popular brands, time of posting, the use of affiliate links, whether or not the post was sponsored, video length, channel verification, and groups of keywords with the same theme. Visualizations such as histograms, boxplots, scatterplots were used to identify patterns and outliers in the data.

Modeling

Modeling was conducted using only the training dataset. We focused on predicting two targets: log-transformed views per subscriber and a composite engagement score ((Likes+Comments)/(Views+1)). We log-transformed the views/subscriber metric to reduce the impact of viral videos, which are extreme outliers. We began by checking the distribution of our target variables and found that they were normally distributed.

We trained and compared several models on each target: (1) a baseline that always predicted the average value, (2) a basic linear regression model, (3) a linear regression model fit with lasso regression, (4) a regression model that included all pairwise interaction terms, and (5) a regression model that included all interaction terms fit with lasso regression. The lasso regression was used to zero out any features or interaction terms that were not meaningful. The model with the lowest RMSE in cross-validation in both cases was Model 5.

We also trained a decision tree to predict when a video has at least 10,000 views.

For our views model, the R^2 score was 0.099 and for our engagement model the R^2 score was 0.069, indicating model fit is poor. Because of this, we performed model specification testing by comparing our best-performing model against a random forest regressor on both our actual data and on simulated linear data. Improvement was similar on both real and simulated data, indicating that our best-performing model was a good choice. The poor fit of our model is likely due to to the fact that the video of the youtube short, the largest contributor to post performance, is not a part of our data set.

Despite the fact that our models should not be used for accurately predicting views or engagement, we can learn something about our data set from them. For example, our R^2 scores tell us to what extent views and engagement can be explained with our features. We also looked at which categorical variables had the largest coefficients in our best performing model to see which categories of posts might have the most significant effects on views and engagement.

Hypothesis Testing and Results

We conducted targeted follow-up tests on the testing set, based on the hypotheses generated by our EDA and modelling, to confirm if certain posts had higher views and engagements. We used a strict significance level of $\alpha = 0.005$ to account for multiple hypothesis testing.

We found:

1) Posts that contain at least one hashtag in the title or description have higher views/subscriber with a desired effect size
2) Posts that contain an affiliate link have lower views/subscriber with a desired effect size
3) Posts from users that are verified have lower views/subscriber with a desired effect size
4) Posts that mention a popular brand have higher views/subscriber with a desired effect size
5) Posts that contain a keyword from the "korean" keyword group have higher views/subscriber with a desired effect size
6) Posts from users that are verified have higher engagement with a small effect size
7) Posts posted during a prime hour have higher engagement but without a significant effect size
8) Posts that mention a popular brand have lower engagement with a small effect size
9) Posts that contain at least one hashtag in the title or description have lower engagement with a small effect size
10) Posts that contain at least one keyword from the "comparing_products" keyword group have lower engagement but without a significant effect size
11) Posts that contain at least one keyword from the "product" keyword group have lower engagement but without a significant effect size