Youtube Shorts Performance

Within the Beauty Niche

Team Members
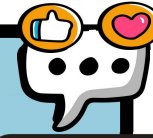
- Rachael Alvir
- Jo Beckford
- Pranav Doshetty
- Yasaman Sanatkhani
- Gabriela Terrazas

Introduction: Project Premise

Goal: Identifying video features, marketing techniques and content trends that influencers could capitalize on.

Findings will support two key audiences make data-driven decisions:

- Content Creators looking to refine their strategy to boost engagement

- Brands/Marketing Teams: identifying features that drive viewership of promotional material
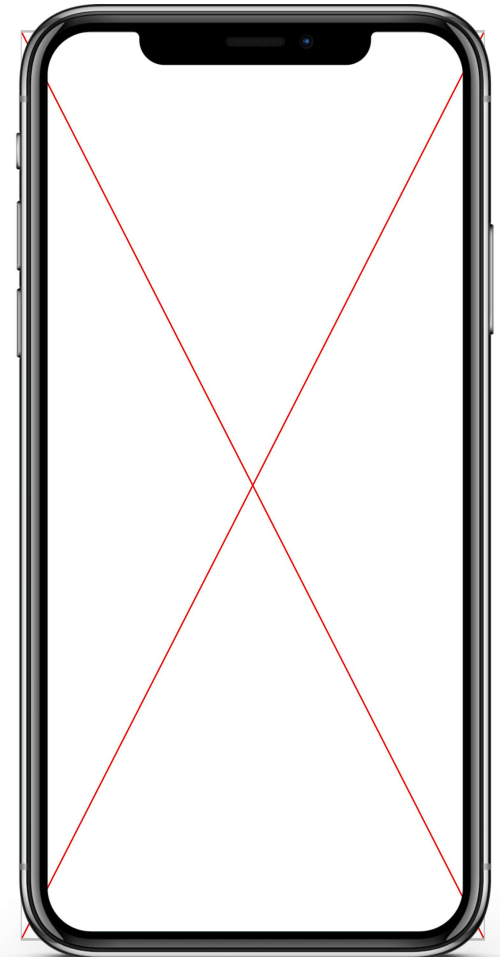
YouTube Shorts

- Aspect Ratio: portrait mode
- Length: 15 seconds – 1 minute
- Monetization: Ad Revenue
- Captures trends, teasers and ads
- Cannot see subscriber base

Data Collection

Data Source:

- List of content creators
- Scrapped YouTube Shorts data using Apify

Collection method:

- Metadata: username, likes, comments, hashtags, and posting time
- CSVs downloaded and concatenated using Python

Data Cleaning

Cleaning Process:

- Removed duplicates, handled missing values
- Dropped irrelevant columns
- Filtered out most recent posts

Tools used:

Python | Pandas | Jupyter Notebooks | APIFY

EDA Findings

- Generic hashtags in title boosts average reach (1.24 vs 0.63)
- Presence of Affiliate Links reduces average reach (0.66 vs 0.97)
- #Ad in title slashes engagement and reach both with very high statistical significance
- Medium length videos (30-60s) have highest reach and engagement both!
- Engagement peaks in February (2% higher than April which has lowest)
- Engagement peaks when video is posted at 3:00AM (EST)

```
Reach : Views/Subscriber
Engagement : (Likes+Comments)/View
```
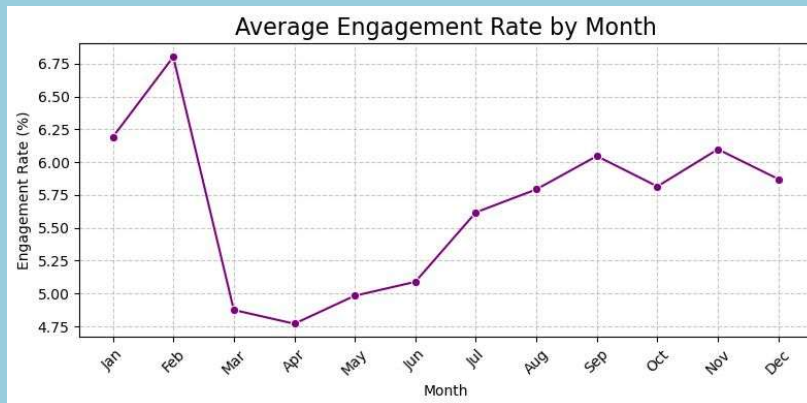
EDA Findings

- Videos from verified channels have significantly more of reach and engagement
- Mention of popular brands increases reach

Caveats :
- Video Content is never analyzed
- Limited Sample size for some features

Average Engagement Rate by Month

Reach : Views/Subscriber
Engagement : (Likes+Comments)/View

Modeling approach

Modeling Approach
- Used standard and regularized linear models, decision trees, and random forests to analyze YouTube Shorts metadata
- Linear regression with interaction terms + Lasso regularization identified as appropriate approach

Performance Metrics
- Model specification testing confirmed linear relationships predominate
- Decision trees achieved 61.7% accuracy predicting high-view videos
- Linear model with interactions + Lasso:
  - Views model: $R^2 \approx 0.099$, RMSE = 1.525
  - Engagement model: $R^2 \approx 0.069$, RMSE = 0.028

Hypothesis Testing

Hypothesis Testing Approach
- Tested if content categories significantly impact views and engagement
- Combined exploratory findings with linear model coefficients
- Used t-tests with strict significance level (α = 0.005)
- Evaluated both statistical significance (p-values) and practical significance (effect sizes)
- Views optimization: Include hashtags and Korean keywords (+), avoid affiliate links (–)
- Engagement optimization: Post during prime hours (+), avoid hashtags and brand mentions (–)

Conclusion

To increase views/subscriber:

- use hashtags
- mention popular beauty brands
- use keywords about Korean makeup
- and avoid affiliate links

To maximize engagement:

- avoid mentioning popular beauty brands
- avoid using hashtags.

Next Steps

There are many more questions our data set can answer that can be explored in the future:

- Is it better to use hashtags in the title or description?
- For reach, is it better to use keywords (such as "makeup") or hashtags (such as "#makeup"?)
- Videos that are too short or too long do not perform as well. Is there a minimum length a video should be?