What Drives Engagement on YouTube?

A Data Analysis of Influencer Metrics

Athina Agiakatsikas & Stella Christou

2025-12-10

Introduction

Research question

What factors most strongly influence engagement on YouTube?

Examples of factors

- Number of views
- How long people watch
- Day of the week posted
- Number of likes and comments

Why this matters

Understanding what drives engagement helps us:

- Make better content
- Post at the right time
- Grow our audience faster

Why is our topic economically relevant?

- Influencer marketing is a multi-billion-euro industry
- Engagement influences income
- Better allocation of marketing budgets
- Content performance = business performance
- Platform algorithms reward engagement

Data Overview

What data did we analyze?

- Time period: September 2019 - November 2024 (5 years)
- Number of videos: About 300 videos
- Daily records: 234,889 observations
- Source: YouTube Analytics (official data)

What we measured

We looked at:

- Views: How many people watched
- Likes: Approval
- Comments: Active engagement
- Watch time: How long people watched
- Subscribers gained: New followers

What is Engagement?

Simple definition

Engagement = How much people interact with your videos

$$\text{Engagement Rate} = \frac{\text{Likes} + \text{Comments}}{\text{Views}} \times 100$$

How do we measure if the engagement rate is good or bad?

- There is no universal standard for engagement across all youtube channels, so we used our own data

Our Findings

First steps

Status of our data set: Messy, has gaps, duplicates, mixed-up data types

Example of what “messy” looks like:

```
date          | views | likes | comments | watch_time | notes
2019-09-01    | 1500  | 45    | (blank)  | 12500      | (incomplete)
2019-09-01    | 1500  | 45    | (blank)  | 12500      | (duplicate row!)
2019-09-02    | NULL  | 50    | 3        | (text!)    | (broken data)
```

What we did: Clean up this mess into usable files

Cleaning the Data

Here’s what we fixed:

- **Fixed data types**: Converted text numbers to real numbers
- So we can do math on them
- **Converted dates**: Turned text dates into proper date format
- Extracted year, month, day of week for analysis
- **Removed incomplete rows**: Deleted rows with zero views
- Can’t calculate engagement rates without views
- **Fixed missing values**: Replaced blanks with 0

Result: Ended up with 4 clean data sets

Created Engagement Metrics

We created new columns to measure engagement:

- **Engagement Rate** = (Likes + Comments) ÷ Views × 100
- Example: 50 likes + 5 comments ÷ 1000 views = **5.5%**
- **Like Rate** = Likes ÷ Views × 100
- **Comment Rate** = Comments ÷ Views × 100
- **Engagement Score** = (Comments × 2 + likes + shares × 3) / views × 100

  Why? With raw numbers it’s difficult to compare videos. With rates it isn’t.

Example:

The raw CSV has:

- **views** = 1000
- **likes** = 50
- **comments** = 5

But you can’t directly answer questions like:

- “Is 50 likes good or bad for 1000 views?”

Removed Duplicates & Aggregated

- **Removed duplicate rows**: Deleted identical records that appeared twice
- **Aggregated by video**: Collapsed 234,889 daily records into **~300 videos**
- **Before:** Each row is one day for one video. So if a video has been up for 100 days, it had 100 rows.
- **Now:** 1 row per video with:
- Calculated totals: total views, total likes, total comments
- Calculated averages: average engagement per video
- Found: first publish date, last recorded date, days active

Average Rates

Average & Median Engagement Rate

Avg Engagement	Median Engagement
3.32%	1.89%

- **We look at them to see what is ‘good’ or ‘bad’**

Like and Comment Rates

Like Rate	Comment Rate
3.12%	0.21%

Key statistics

- Average engagement: 3.32% (meaning on average, 3.32 out of 100 viewers like or comment)
- Median engagement: 1.89% (middle value - many videos have lower engagement)
- Likes are more common: About 1.76% average like rate
- Comments are rare: Only 0.09% average comment rate (comments are 15-20x less common than likes!)

What this means

Anything above **3.32%** is above our channel baseline (good); anything below is below average.

Keep in mind this benchmark is specific to our data. Factors we did not account for include:

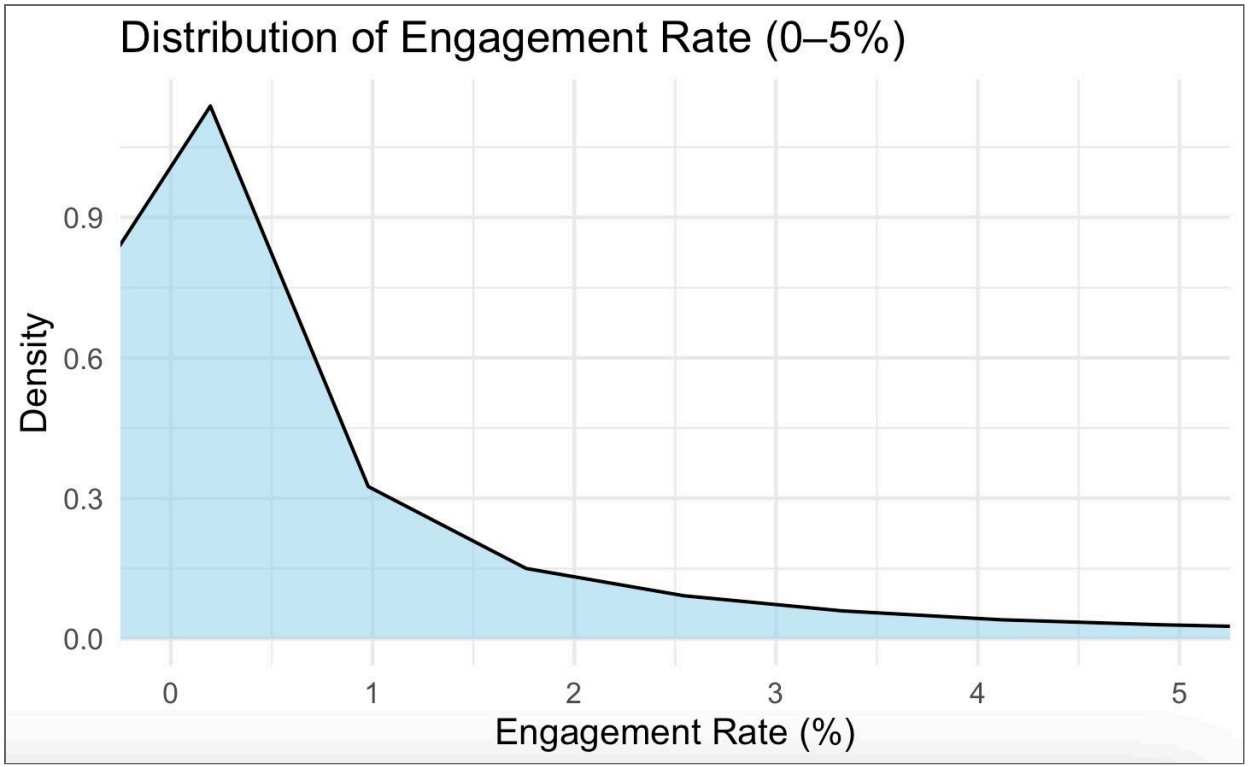
- Channel size
- Content type
- Audience loyalty
- Upload frequency

Review

Difficulties

- Data cleaning took longer than expected
- Messy & complicated at first
- Packages didn’t load correctly (many errors)
- Only defined engagement rate and not all additional metrics early on
- Sharing the R project and reproducibility steps remain
- R & data analysis tools were new

Failed Diagram



- **Used the uncleaned data**
- **Wrong rates**

What Comes Next

Now we dig deeper to find what **actually drives** engagement:

- **Correlation Analysis**
- Which metrics are connected?
- Do longer videos get more likes?
- Do videos posted on weekends get more engagement?
- Does watch time predict comments?

- **Visualizations & Graphs**
- Show trends visually
- Line charts: Engagement over time
- Bar charts: Performance by day of week
- Scatter plots: Relationships between variables

- **Key Findings & Recommendations**
- Which factors matter most for engagement
- Actionable recommendations for better content

Thank you so much for listening!

Speaker notes