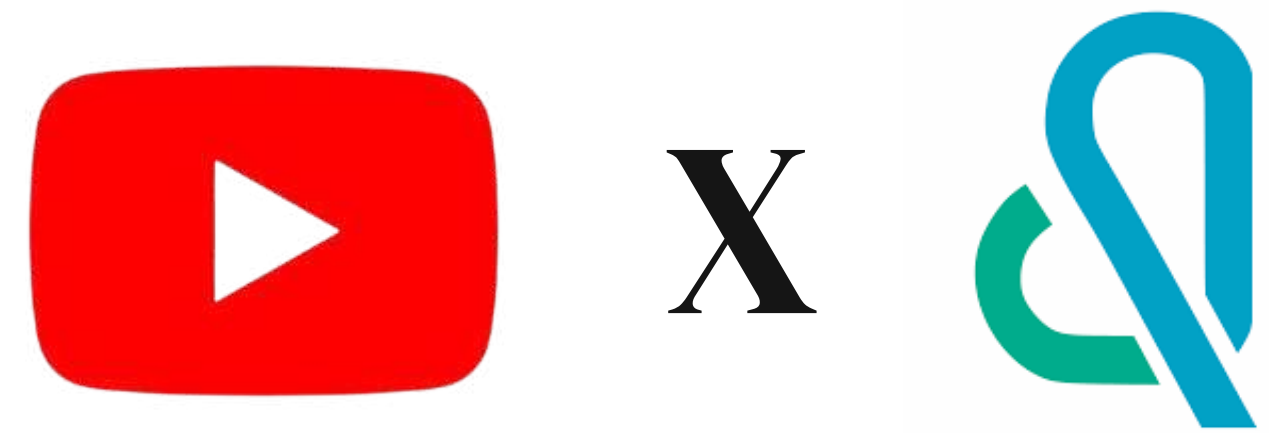


# **DATA SCIENCE & ANALYTICS PROJECT : YOUTUBE TRENDING**



**In collaboration with : Dibimbing**

**By : Melvern Amadio Hidayat**

# ABOUT MYSELF

## Melvern Amadio Hidayat



Jakarta Selatan, Jakarta



[melvern.amadio@gmail.com](mailto:melvern.amadio@gmail.com)



[melvernhidayat](https://www.linkedin.com/in/melvernhidayat)

Saya adalah mahasiswa Data Science semester 6 yang memiliki semangat tinggi di bidang IT dan Data Science. Saya memiliki pengalaman dalam mengerjakan berbagai proyek pemrograman, mengolah data, serta menerapkan model machine learning menggunakan bahasa pemrograman seperti Python dan R. Sebagai calon Data Scientist, saya sangat antusias untuk menerapkan kemampuan analisis dan pengetahuan saya dalam menyelesaikan permasalahan nyata serta menggali insight yang lebih mendalam dari data.



# TABLE OF CONTENT



Share



1

Introduction & Understanding

2

Data Understanding & Preprocessing

3

Exploratory Data Analysis

4

Data Analytics & Visualization

5

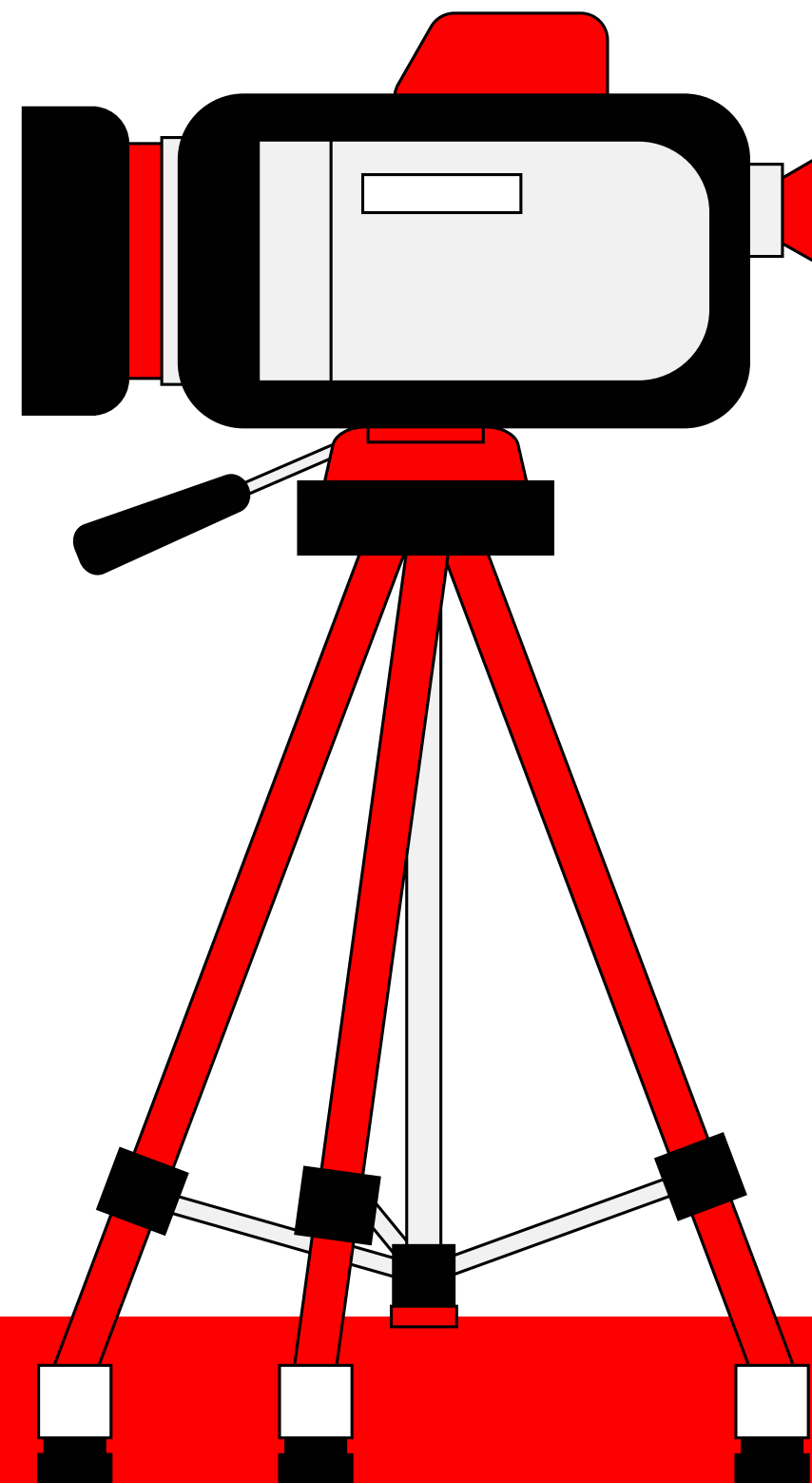
Conclusion & Solution



2.3K



# Introduction & Understanding



# About Youtube

**YouTube adalah platform berbagi video terbesar di dunia yang diluncurkan pada tahun 2005 dan kini dimiliki oleh Google. Platform ini memungkinkan pengguna untuk mengunggah, menonton, menyukai, membagikan, dan mengomentari video secara gratis. YouTube telah menjadi media utama bagi para kreator konten untuk mengekspresikan diri, menyampaikan informasi, serta membangun komunitas secara global.**

**Dengan lebih dari 2 miliar pengguna aktif setiap bulannya, YouTube menampilkan berbagai jenis konten yang terus berkembang mengikuti tren. Video trending di YouTube mencakup konten seperti vlog kehidupan sehari-hari, tantangan viral (challenge), video reaksi, podcast visual, edukasi singkat, dan YouTube Shorts berdurasi pendek. Tren ini mencerminkan minat penonton yang terus berubah dan memberikan peluang besar bagi kreator untuk tampil lebih relevan dan menjangkau audiens yang lebih luas.**





# Purpose & Understanding



## Business Understanding

Memahami tren di YouTube sangat penting karena dapat membantu kreator, brand, dan agensi dalam membuat keputusan konten yang lebih tepat sasaran. Konten trending seperti YouTube Shorts, video reaksi, dan podcast visual menunjukkan bahwa audiens kini menyukai video yang cepat, interaktif, dan mudah dikonsumsi. Dari sisi bisnis, tren ini bisa dimanfaatkan untuk meningkatkan jangkauan, pertumbuhan subscriber, serta potensi monetisasi melalui iklan, sponsor, maupun penjualan produk digital.

## Problems

1. Konten seperti apa yang paling sering muncul dalam daftar video trending YouTube?
2. Faktor apa yang paling memengaruhi sebuah video bisa trending, seperti views, likes, dan comment count?
3. Bagaimana perbedaan performa video berdasarkan waktu publish, hari dalam seminggu, dan negara?



## Purpose

1. Mengidentifikasi pola umum dalam video yang masuk trending, termasuk genre, durasi trending, dan waktu tayang.
2. Menganalisis hubungan antara metrik performa (views, likes, komentar) dengan status trending.
3. Memberikan insight bagi kreator atau brand dalam merancang konten agar lebih berpeluang masuk trending YouTube dan meningkatkan jangkauan audiens.



# DATA UNDERSTANDING



# About the Data

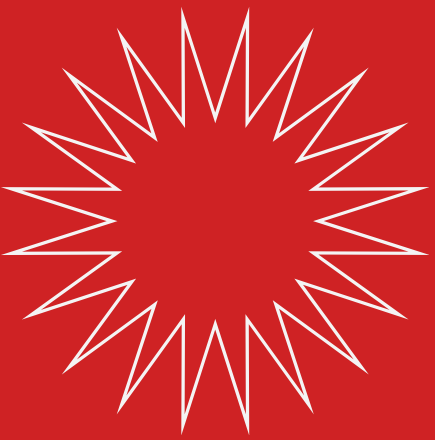
**DATA DIAMBIL DARI KAGGLE, YANG  
BERISIKAN DATA VIDEO YANG  
SEDANG TREND DI YOUTUBE PADA  
TAHUN 2017-2018**

**Source :**

**<https://www.kaggle.com/datasets/thedevastator/youtube-trending-videos-dataset>**

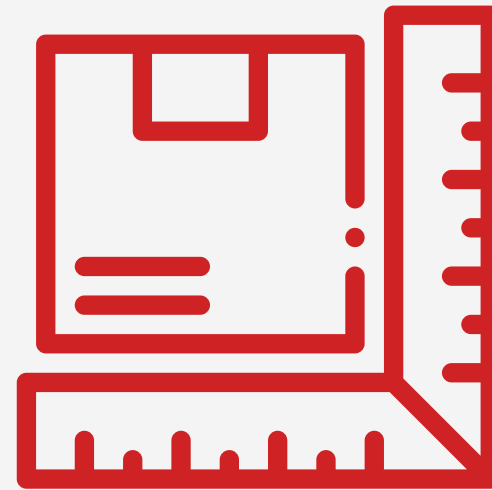


# DATASET INFORMATION



## Dataset

- **Format Dataset: CSV**
- **Nama File: youtube.csv**
- **Domain Industri:**  
**Entertainment**



## Data Structure

### Terdiri atas:

- **18 kolom**
- **161470 baris**



## Data Types

### Terdiri dari 4 tipe data

- **Kategorik (object)**
- **Numerik (integer & float)**
- **Waktu (Datetime)**
- **Booelan (True or False)**

# Data Dictionary

- **title:** The title of the video.
- **channel\_title:** The title of the YouTube channel that published the video.
- **publish\_date:** The date when the video was published on YouTube.
- **time\_frame:** The duration of time (e.g., 1 day, 6 hours) that the video has been trending on YouTube.
- **published\_day\_of\_week:** The day of week (e.g., Monday) when the video was published.
- **publish\_country:** The country where the video was published.
- **tags:** The tags or keywords associated with the video.
- **views:** The number of views received by a particular video
- **likes:** Number of likes received per each video
- **dislike:** Number of dislikes received per an individual video
- **comment\_count:** number of comments
- **comments\_disabled :** apakah komentar dinonaktifkan pada video.
- **ratings\_disabled:** apakah penilaian (likes/dislikes) dinonaktifkan untuk video.
- **video\_error\_or\_removed :** apakah video sudah dihapus dari YouTube atau terjadi kesalahan saat memuat video.

# DATA PREPROCESSING



```
df['trending_date'] = pd.to_datetime(df['trending_date'], format='%y.%d.%m')
```

```
df['publish_date'] = pd.to_datetime(df['publish_date'], format='%d/%m/%Y')
```

Perlu ada sedikit perubahan kepada 2 kolom dalam dataset, dimana kedua tipe data diatas masih dalam bentuk tipe data object, sehingga harus diganti menjadi tipe data datetime

# EXPLORATORY DATA ANALYSIS

---

# HANDLE MISSING VALUES

	0
video_id	0
trending_date	0
title	2
channel_title	0
category_id	0
publish_date	0
time_frame	0
published_day_of_week	0
publish_country	0
tags	0
views	0
likes	0
dislikes	4
comment_count	0
comments_disabled	0
ratings_disabled	0
video_error_or_removed	0

ada missing value di title dengan 2 value kosong dan dislikes dengan 4 value kosong, tapi karena sedikit saja yang missing maka cukup di drop saja



```
df = df.dropna(subset=['title', 'dislikes'])
```

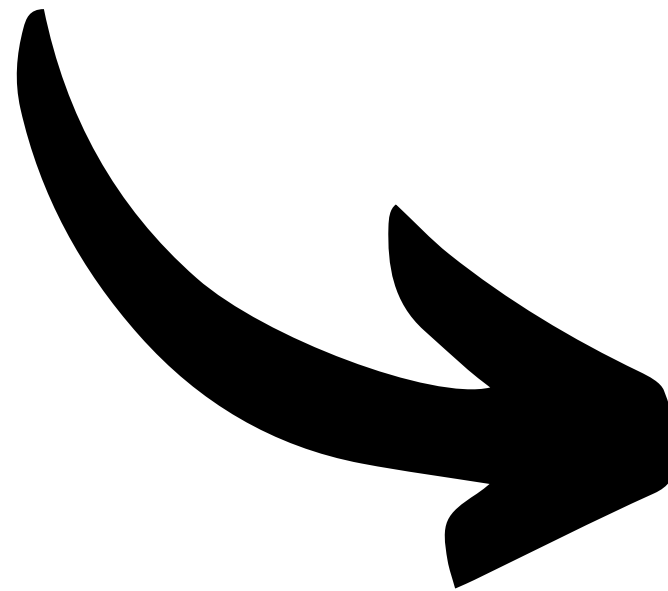


# DUPLICATED DATA

```
print("Jumlah data duplikat :",df.duplicated().sum())
```

```
Jumlah data duplikat : 220
```

Dalam Dataset terdapat 220 data duplikat



Mengatasi Data Duplikat dengan cara menghapus data yang termasuk duplikat

```
df = df.drop_duplicates()  
print("Jumlah data duplikat :",df.duplicated().sum())
```

```
Jumlah data duplikat : 0
```



# RESULT

```
<class 'pandas.core.frame.DataFrame'>
Index: 161244 entries, 0 to 161469
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   video_id                             161244 non-null object
1   trending_date                       161244 non-null datetime64[ns]
2   title                               161244 non-null object
3   channel_title                       161244 non-null object
4   category_id                         161244 non-null int64
5   publish_date                       161244 non-null datetime64[ns]
6   time_frame                         161244 non-null object
7   published_day_of_week               161244 non-null object
8   publish_country                    161244 non-null object
9   tags                               161244 non-null object
10  views                              161244 non-null int64
11  likes                              161244 non-null int64
12  dislikes                           161244 non-null float64
13  comment_count                      161244 non-null int64
14  comments_disabled                  161244 non-null bool
15  ratings_disabled                   161244 non-null bool
16  video_error_or_removed             161244 non-null bool
dtypes: bool(3), datetime64[ns](2), float64(1), int64(4), object(7)
memory usage: 18.9+ MB
```

# DATA ANALYTICS & VISUALIZATION

---

# INSIGHTS

## Video

Total Video yang menjadi trend berjumlah 161244 video

## Trend Date Range

Semua Video yang masuk ke trend berada di periode November 2017 - Juni 2018

## Country

Video yang masuk trending berasal dari 4 negara : USA, Inggris, Prancis & Kanada

## Content Creator

Sebanyak 11344 konten kreator/channel berkontribusi dalam membuat video trend selama periode tersebut

## Views

Video Youtube dengan Views Terbanyak mencapai 424538912 views

## Comment

Video Youtube dengan komen Terbanyak mencapai 1626501 Komen

## Likes

Video Youtube dengan Likes Terbanyak mencapai 5613827 Likes

## Dislikes

Video Youtube dengan Dislikes Terbanyak mencapai 1944971 Dislikes

## Problems

Terdapat 2783 video yang komen-nya disabled, 1423 video dengan rating yang di disabled, dan ada 137 yang dihapus dari Youtube

# CHANNEL / CONTENT CREATOR

5 channel dengan penyumbang video trending terbanyak selama tahun 2017-2018



1. The Late Show with Stephen Colbert dengan 651 Video

2. Late Night with Seth Meyers dengan 590 Video



3. The Ellen Show dengan 583 video

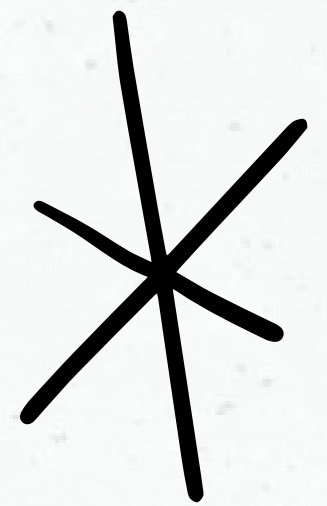
4. Tonight Show Starring Jimmy Fallon dengan 567 video



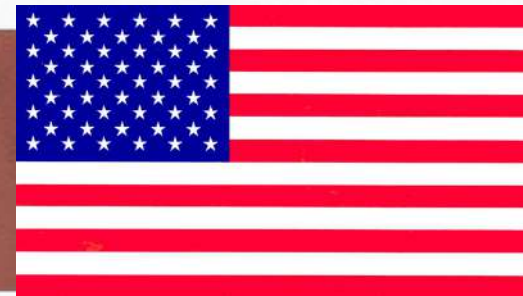
5. Jimmy Kimmel Live dengan 556



# TOP COUNTRY WITH MOST TRENDING VIDEOS



**AMERIKA SERIKAT**



Amerika Serikat menduduki peringkat pertama dengan 40.896 video trending. Hal ini menunjukkan dominasi kuat AS dalam produksi konten yang menarik perhatian global.



**KANADA**



Kanada berada di posisi kedua dengan total 40.879 video trending, hanya terpaut sedikit dari Amerika Serikat. Ini menandakan bahwa konten dari Kanada juga memiliki daya saing tinggi di platform YouTube.



**PRANCIS**



Prancis menyumbangkan 40.724 video trending, menjadikannya negara ketiga terbanyak. Konten dari Prancis menunjukkan keberagaman dan kreativitas yang mampu menarik audiens internasional. Posisi ini menandakan pengaruh Prancis yang terus tumbuh di dunia digital.

**INGGRIS RAYA**



Inggris Raya menghasilkan 38.745 video trending, berada di peringkat keempat. Meskipun sedikit di bawah tiga besar, Inggris tetap menjadi pemain penting dalam kancah konten global. Kreator dari negara ini tetap aktif menciptakan video yang engaging dan populer.





# MOST VIEWS



**Nicky Jam x J. Balvin  
- X (EQUIS)**

424.538.912 Views



**Casper, Nio García, Darell, Nicky Jam,  
Bad Bunny, Ozuna - Te Bote Remix (Video  
Oficial)**

337.621.571 Views



**Bad Bunny -  
Amorfoda**

328.860.380 Views



**Ozuna x Romeo Santos -  
El Farsante Remix**

288.811.992 Views



**Childish Gambino - This Is  
America (Official Video)**

259.721.696 Views



# MOST LIKES



**BTS - Fake Love**

**5.613.827 Likes**



**Childish Gambino - This Is America (Official Video)**

**5.444.541 Likes**



**Drake - God's Plan (Official Video)**

**4.737.873 Likes**



**Bad Bunny - Amorfoda**

**3.823.879 Likes**



**Ariana Grande - no tears left to cry (Official Video)**

**3.394.437 Likes**



# MOST DISLIKES



**Logan Paul Vlogs -  
So Sorry**

**1.944.971 Dislikes**



**YouTube Rewind:  
The Shape of 2017**

**1.753.274 Dislikes**



**Logan Paul Vlogs - Suicide:  
Be Here Tomorrow**

**545.595 Dislikes**



**Jake Paul - It's Everyday Bro (Song) feat  
Team 10 (Official Music Video)**

**504.340 Dislikes**



**Lucas Lucco e Pablio  
Vittar - Paraíso**

**421.473 Dislikes**



# MOST DISLIKES



**Logan Paul Vlogs -  
So Sorry**

**1.626.501 Comments**



**BTS - Fake Love**

**1.228.655 Comments**



**YouTube Rewind:  
The Shape of 2017**

**845.233 Comments**



**Logan Paul Vlogs - Suicide:  
Be Here Tomorrow**

**704.824 Comments**

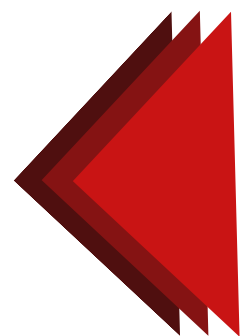


**BTS (방탄소년단) 'MIC  
Drop (Steve Aoki Remix)'**

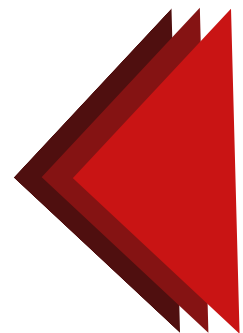
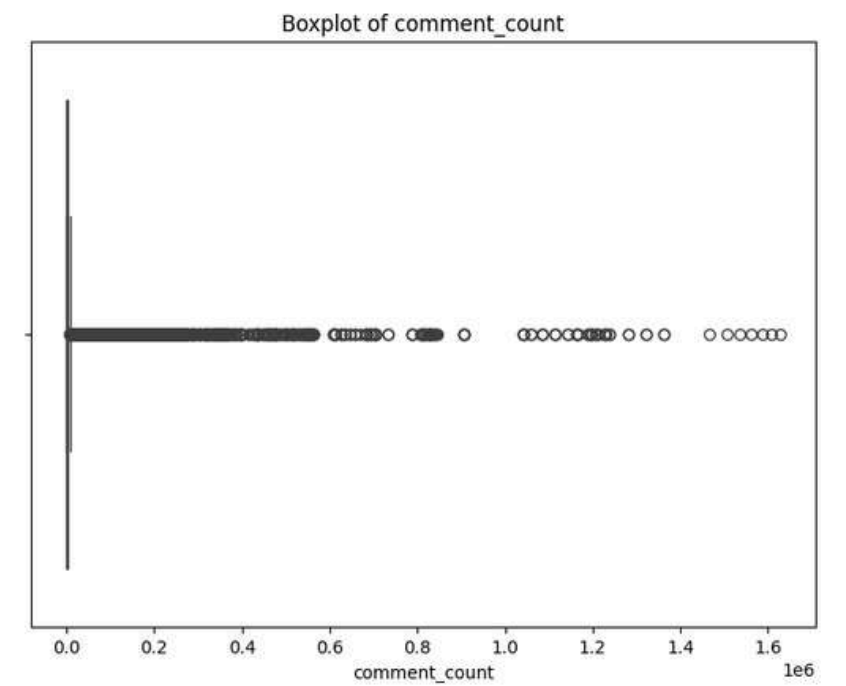
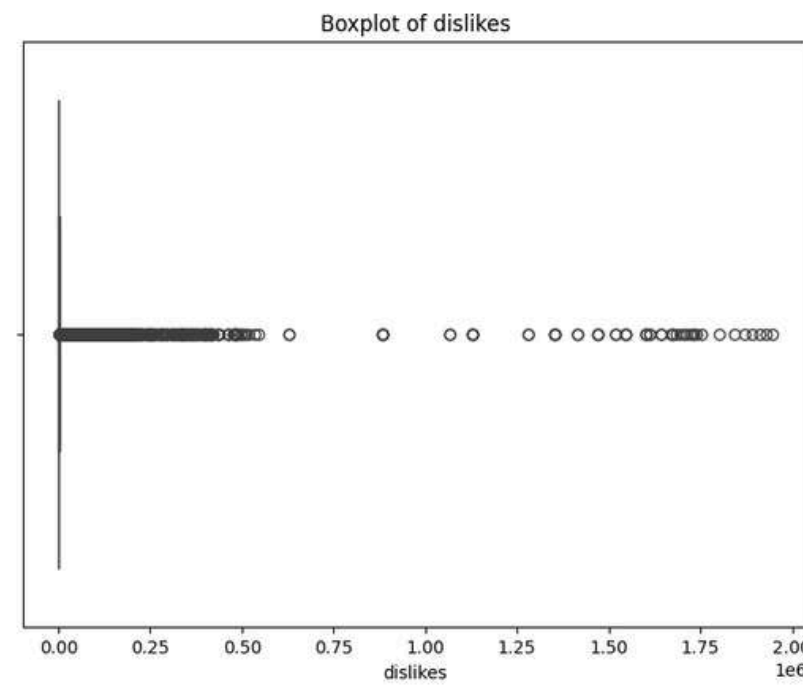
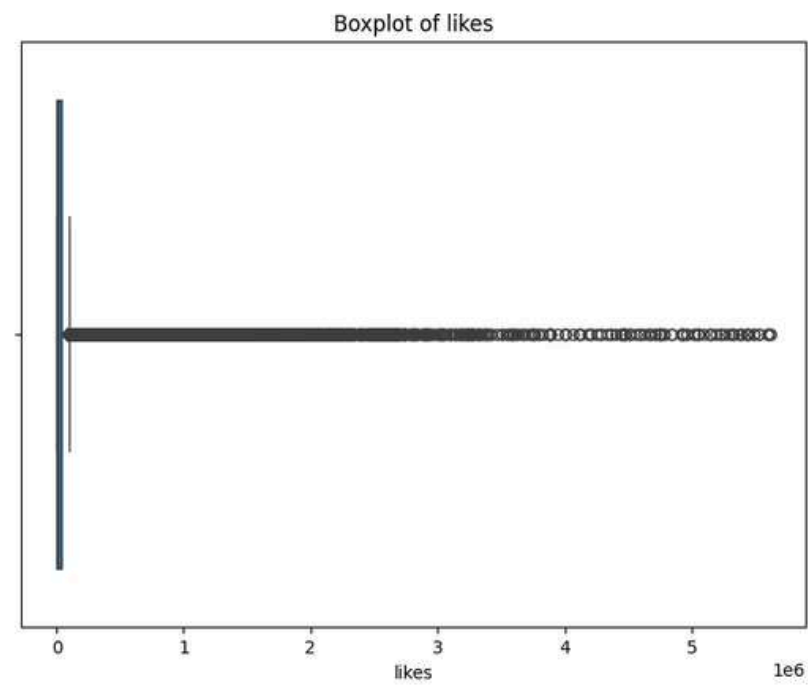
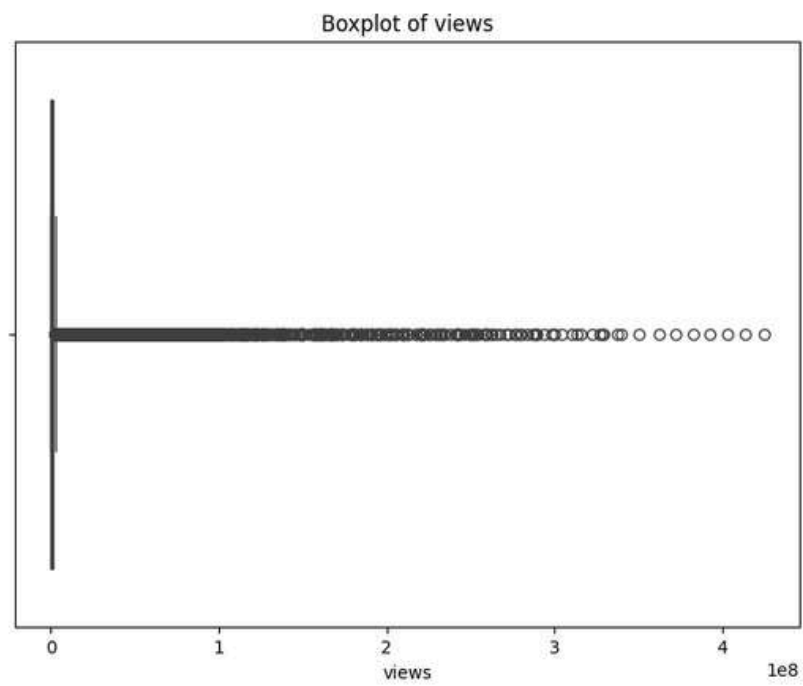
**559.197 Comments**

# Cleaning Outliers

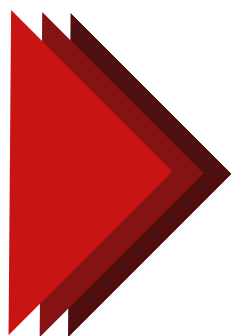
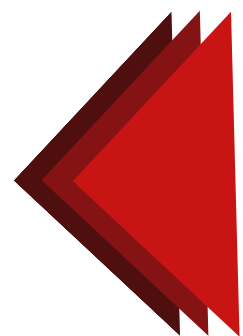
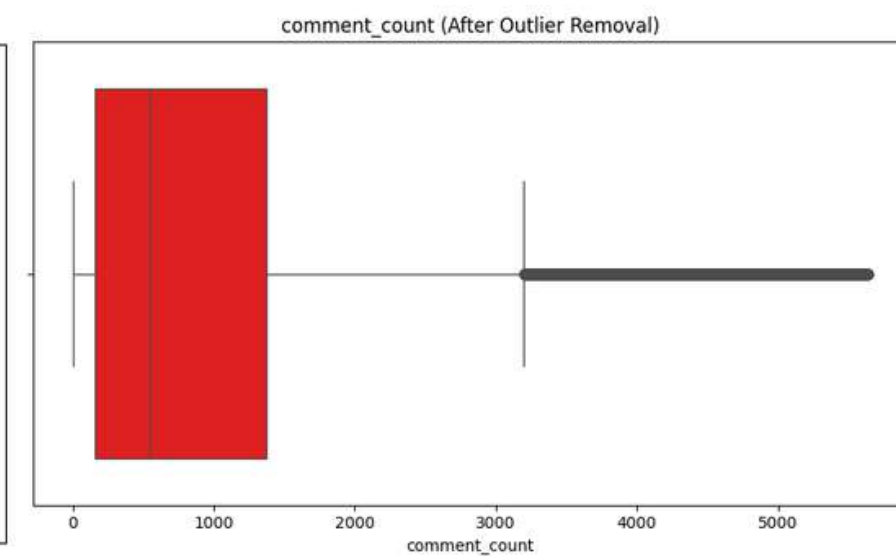
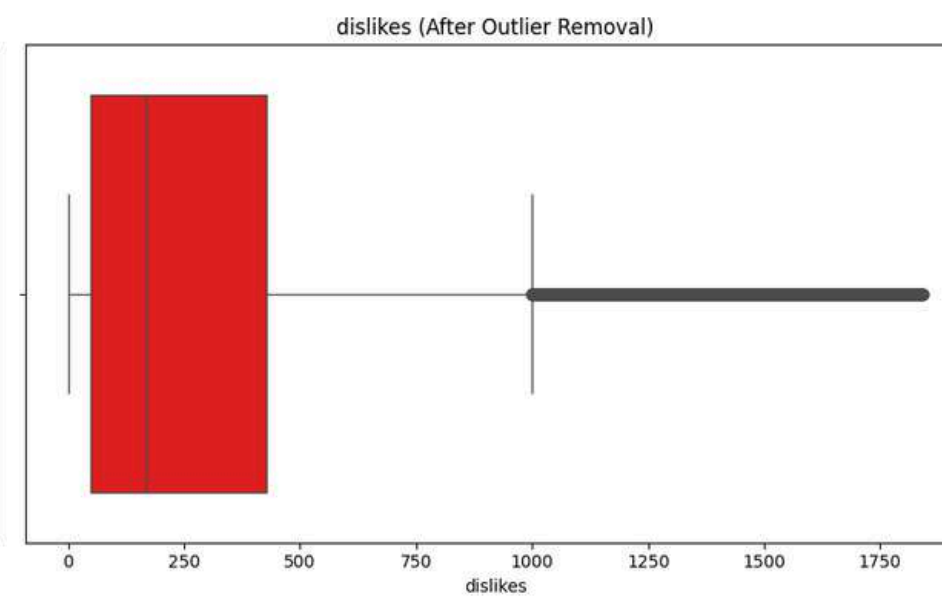
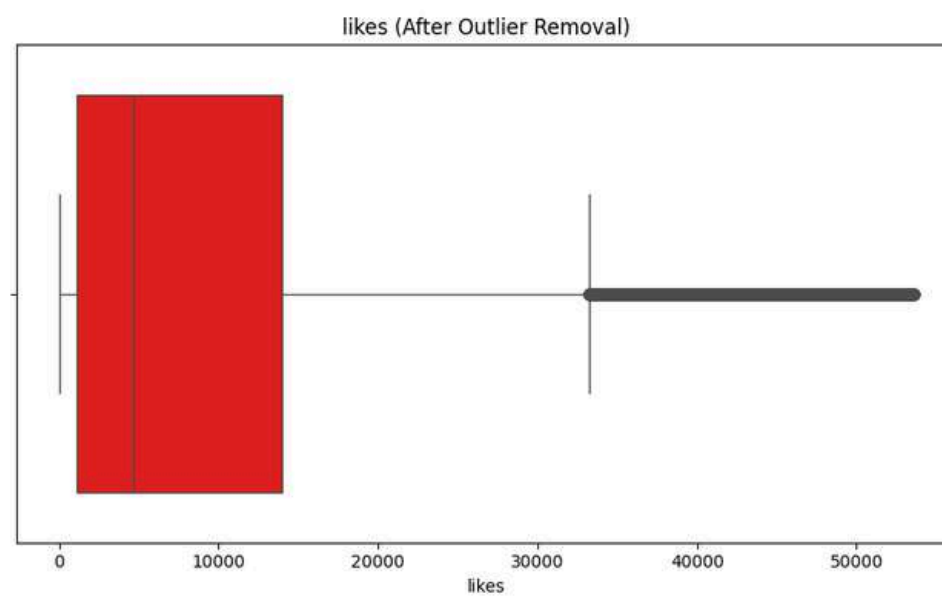
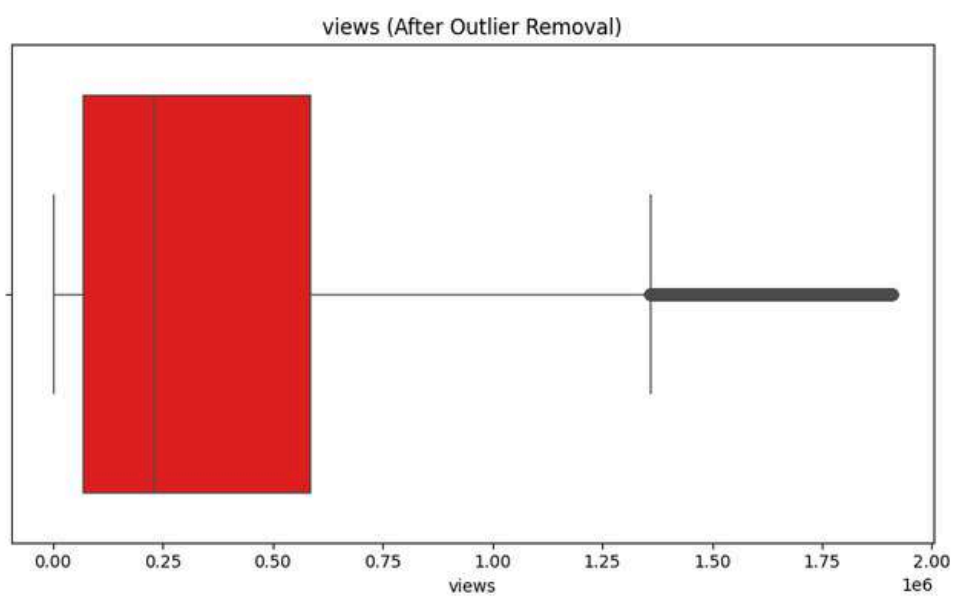
**Outliers adalah data yang memiliki nilai ekstrem atau jauh berbeda dari data lainnya, sehingga perlu ditangani. Tujuannya adalah untuk meningkatkan akurasi model analisis atau prediksi dengan mengurangi pengaruh nilai-nilai yang dapat mengganggu distribusi data. Pada kasus ini, outliers akan ditangani dengan cara menggantinya dengan nilai yang lebih wajar**



# Before Cleaning Outliers

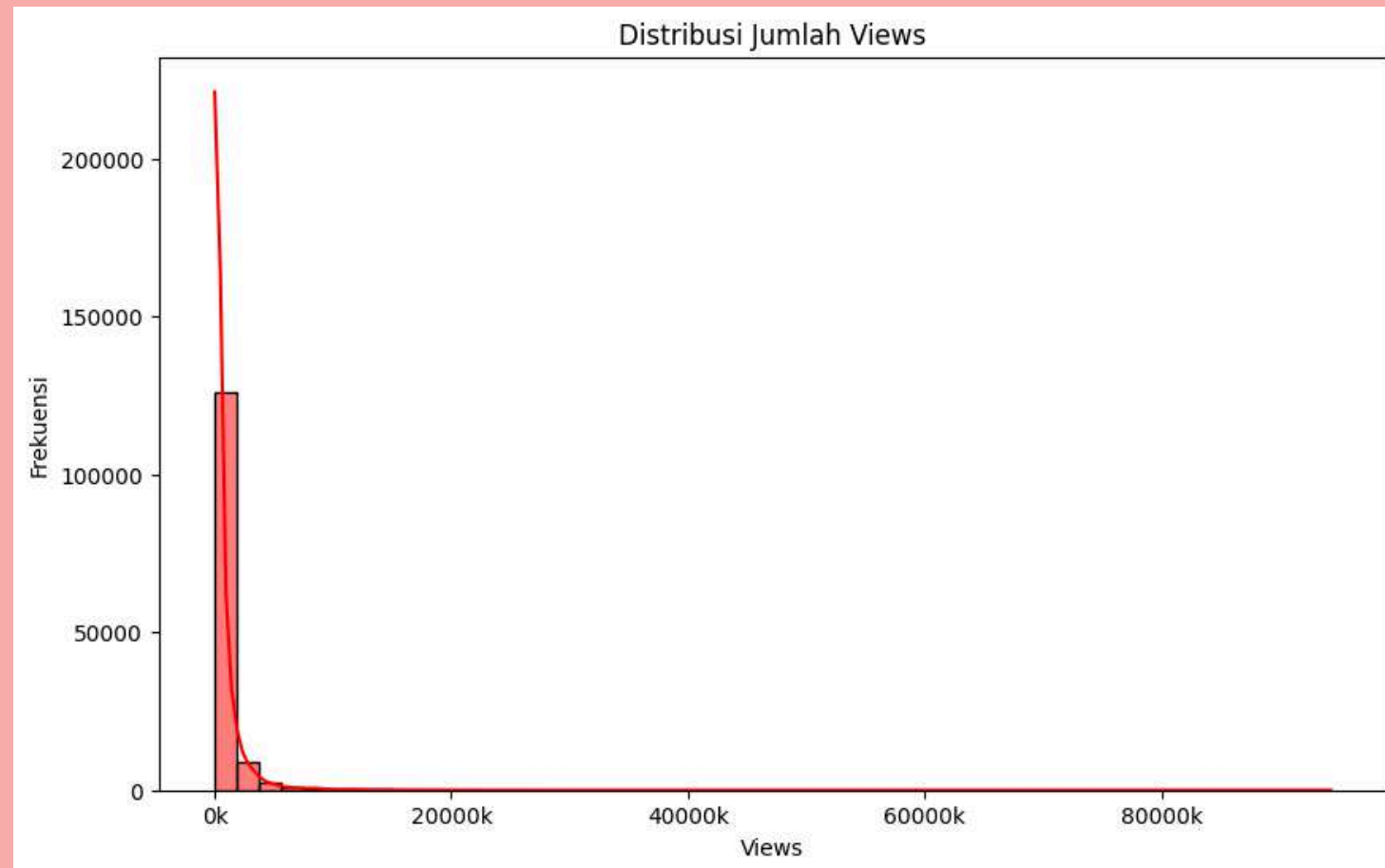


# After Cleaning Outliers

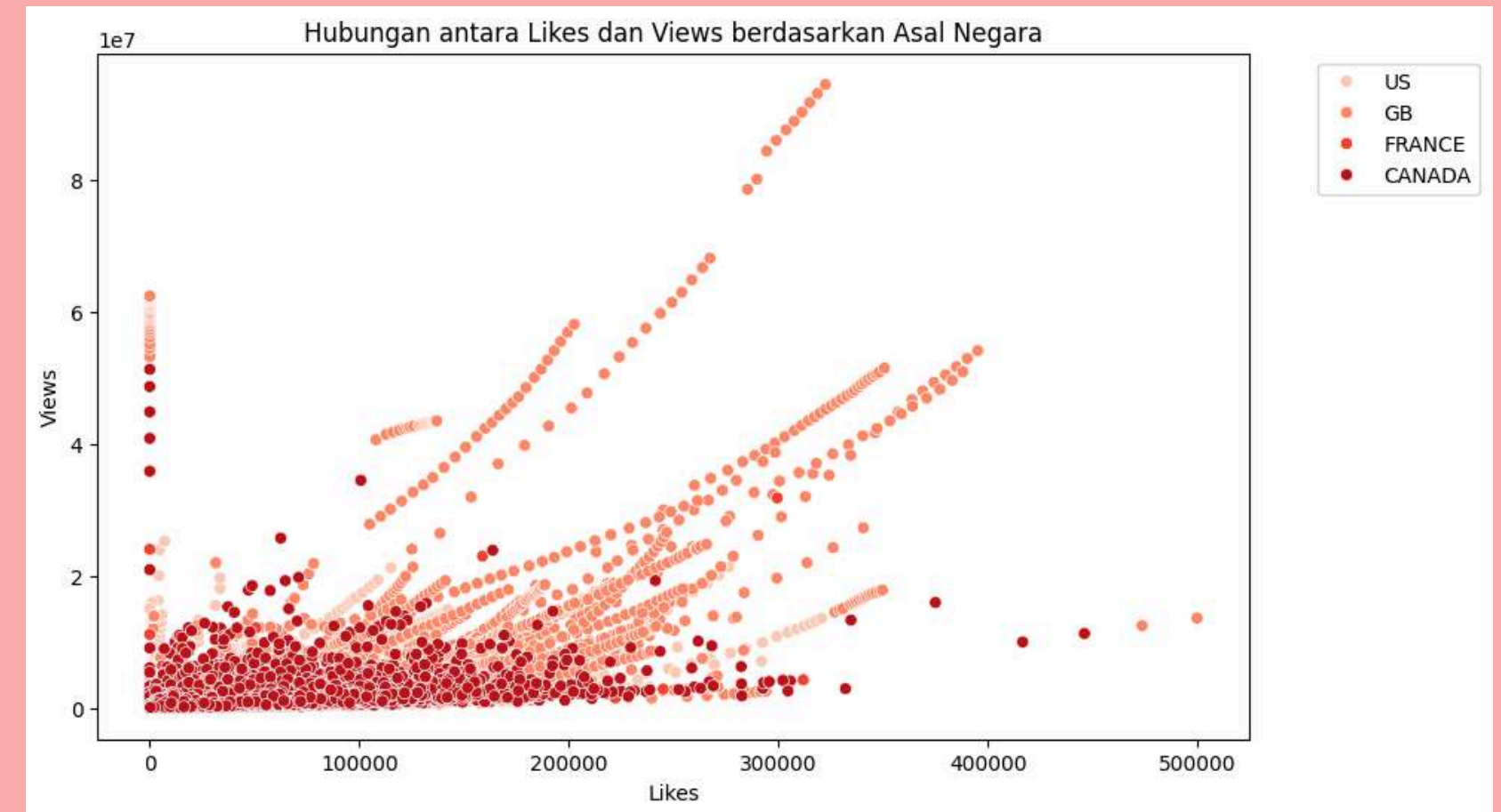




# Visualization

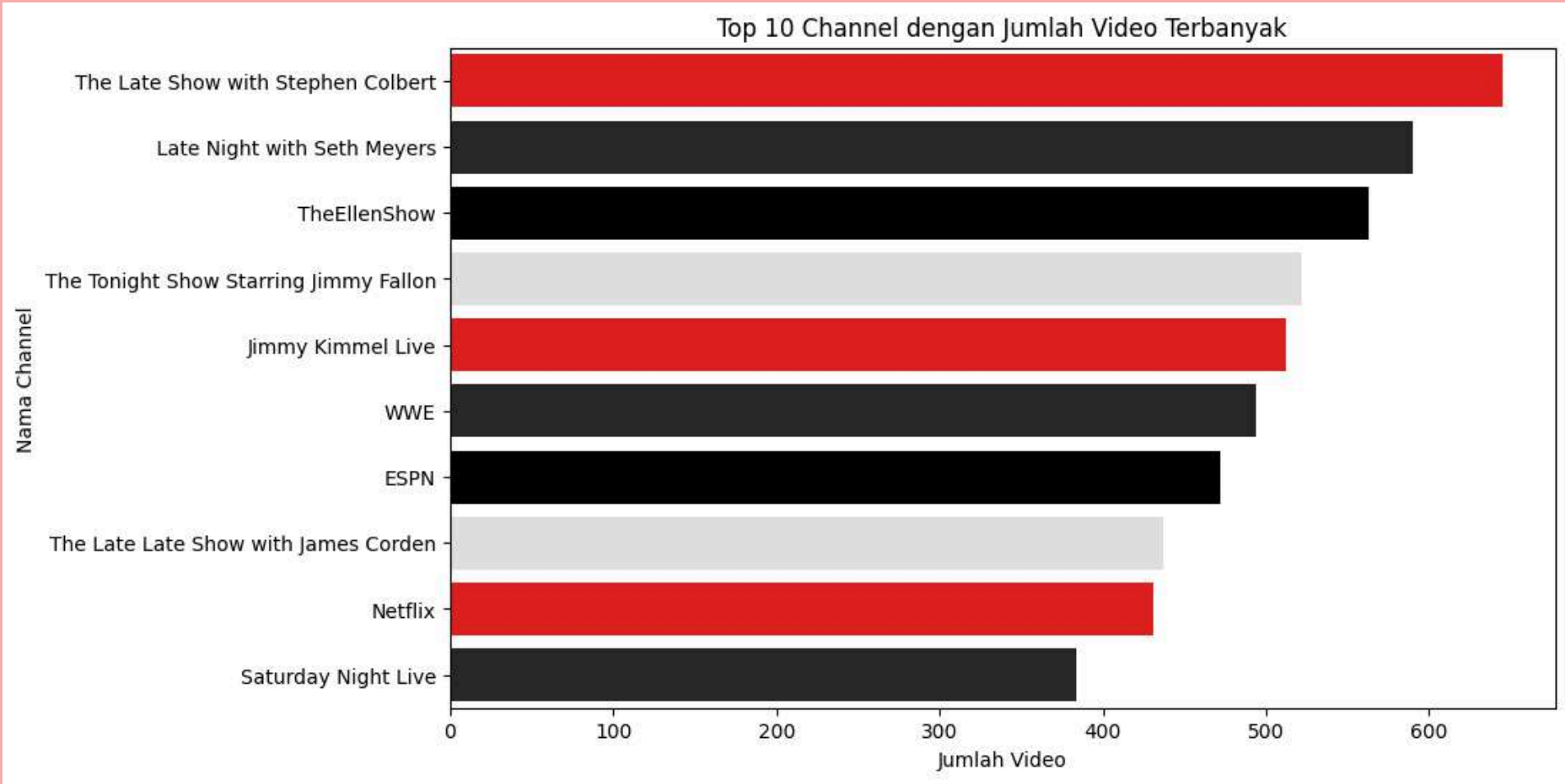


Sebagian besar video memiliki jumlah views yang rendah, sedangkan sedikit video memiliki views yang sangat tinggi. Fenomena ini mencerminkan distribusi yang tidak merata dalam konsumsi konten, di mana sebagian besar video cenderung mendapatkan sedikit perhatian, sementara sejumlah kecil video viral menarik jumlah penonton yang luar biasa besar. Hal ini sering kali dipengaruhi oleh algoritma platform yang mengedepankan konten tertentu berdasarkan preferensi pengguna.

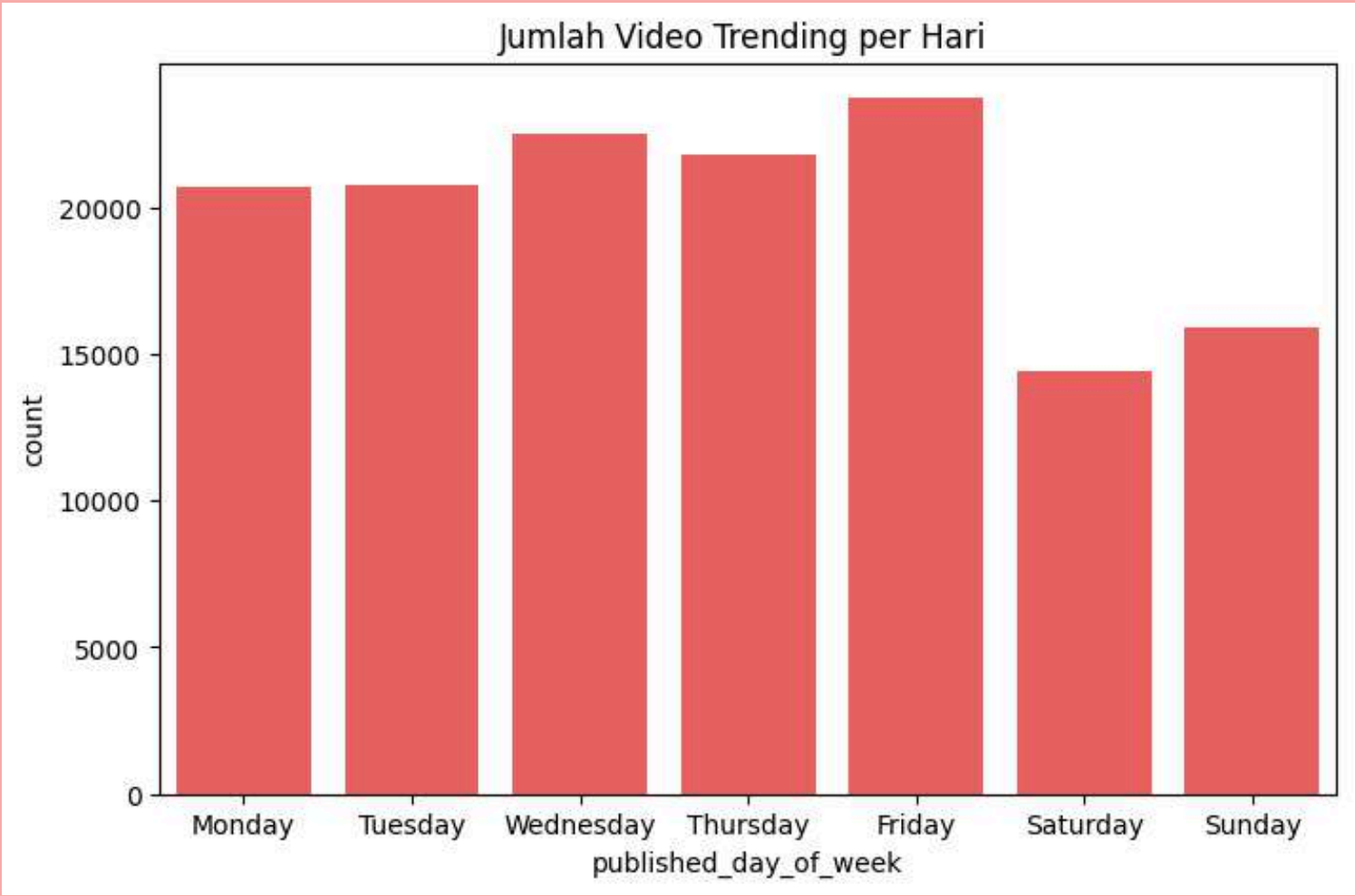


Berdasarkan Grafik diatas, dapat ditunjukan bahwa semakin banyak likes, jumlah views juga cenderung meningkat, meskipun tidak selalu sebanding. Beberapa video memiliki views tinggi meski likes-nya rendah, kemungkinan karena faktor viral atau algoritma platform. Perbedaan warna pada grafik juga mengindikasikan bahwa tren ini bisa bervariasi tergantung pada negara asal video.

# Visualization

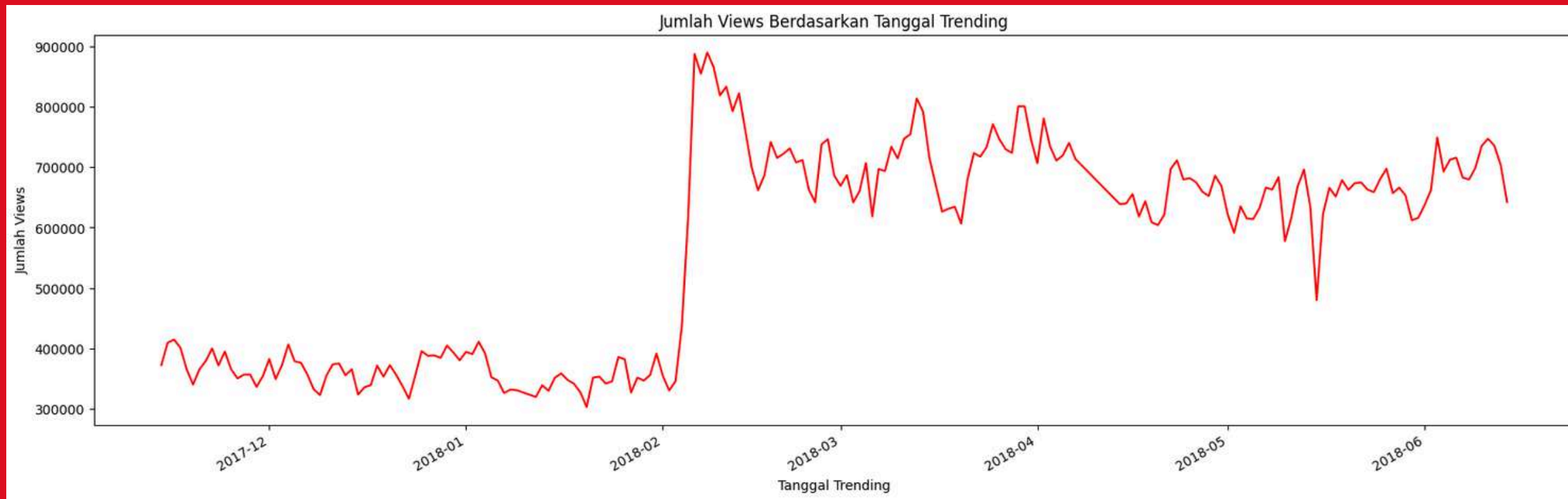


Channel "The Late Show with Stephen Colbert" menempati posisi pertama dengan lebih dari 650 video, mengungguli channel lain seperti "Late Night with Seth Meyers" dan "TheEllenShow". Sebagian besar channel dalam daftar ini merupakan acara talk show atau hiburan populer yang rutin mengunggah konten. Hal ini menunjukkan konsistensi dan produktivitas dalam produksi konten untuk menjaga engagement penonton.

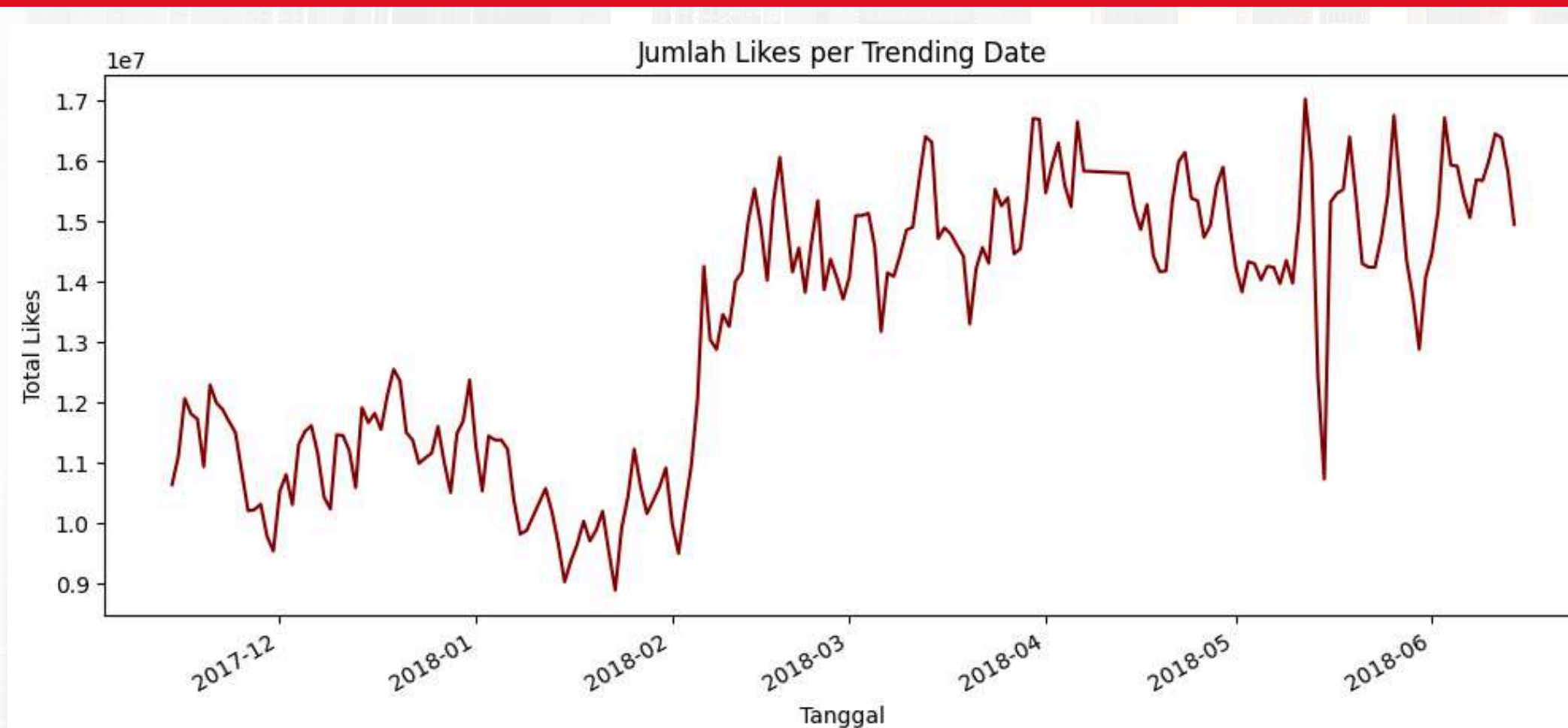


Terlihat bahwa hari Jumat menjadi puncak tertinggi dengan jumlah video trending terbanyak, disusul oleh Rabu dan Kamis. Sementara itu, akhir pekan seperti Sabtu dan Minggu justru menunjukkan jumlah yang lebih rendah. Hal ini bisa mengindikasikan bahwa lebih banyak konten viral dirilis atau mendapat perhatian di hari kerja.



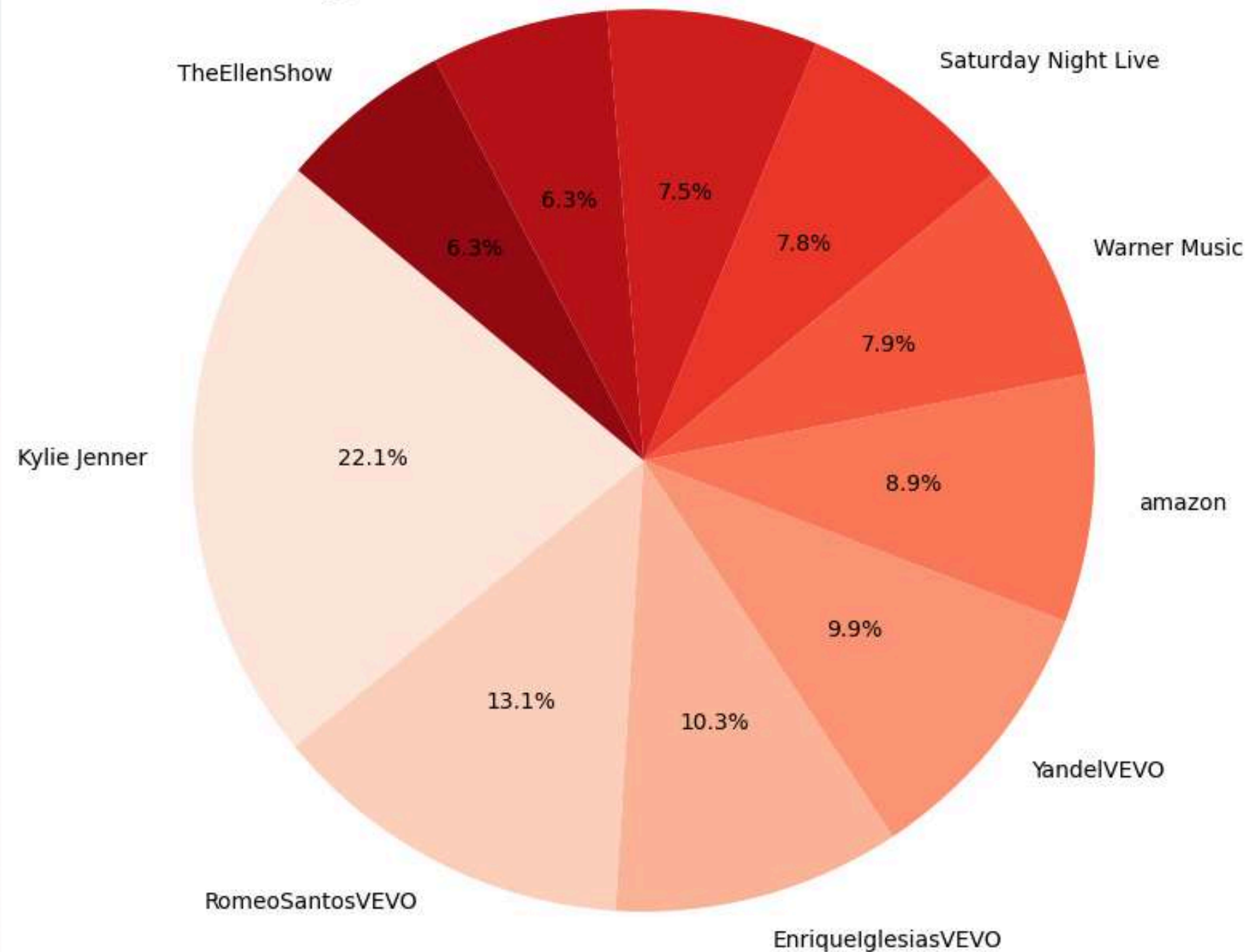


Terdapat lonjakan tajam pada awal Februari 2018, yang bisa menandakan adanya video viral atau peristiwa besar yang menarik perhatian publik. Setelah itu, jumlah views cenderung stabil di level yang lebih tinggi dibandingkan periode sebelumnya. Hal ini menunjukkan adanya peningkatan minat penonton secara signifikan setelah lonjakan tersebut.



Lonjakan signifikan pada Likes juga terjadi pada pertengahan Februari 2018, menandakan peningkatan popularitas konten yang cukup besar saat itu. Setelah lonjakan tersebut, jumlah likes cenderung stabil di angka tinggi, meskipun ada sedikit fluktuasi. Ini menunjukkan bahwa setelah titik tersebut, konten-konten trending mendapat perhatian dan keterlibatan yang lebih konsisten dari penonton.

Proporsi Total Views dari 10 Channel Teratas  
KygoOfficialVEVO WWE



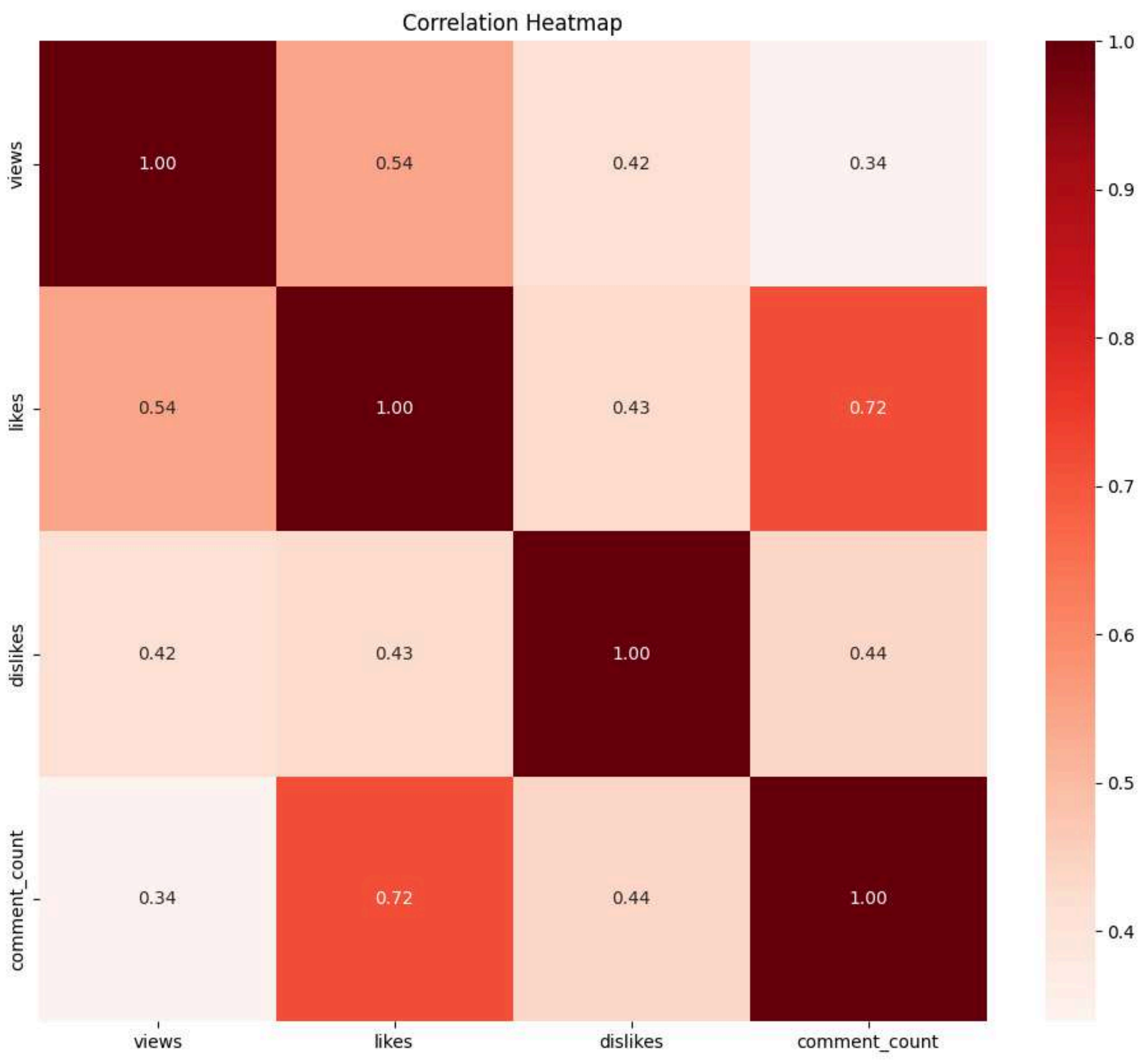
Untuk Jumlah views pada Youtube, channel Kylie Jenner mendominasi dengan porsi sekitar 22,1%, jauh di atas channel lain. Channel seperti RomeoSantosVEVO dan EnriqueIglesiasVEVO juga memiliki kontribusi besar terhadap total views. Sementara itu, TheEllenShow dan KygoOfficialVEVO berada di posisi terbawah, masing-masing hanya menyumbang 6,3%.

Hal ini mengindikasikan bahwa selebriti individu seperti Kylie Jenner bisa menarik audiens yang jauh lebih besar dibandingkan dengan brand atau channel hiburan lainnya. Perbedaan ini bisa dipengaruhi oleh popularitas global, frekuensi unggahan, serta daya tarik konten yang ditawarkan.



Likes memiliki korelasi cukup kuat dengan comment count (0.72) dan views (0.54), yang menandakan bahwa video yang disukai cenderung banyak ditonton dan dikomentari. Dislikes ternyata memiliki korelasi paling lemah dengan views (0.42), namun masih memiliki hubungan moderat dengan komentar. Secara keseluruhan, metrik-metrik ini saling berkaitan, tetapi tidak ada yang benar-benar berkorelasi sempurna, menunjukkan bahwa banyak faktor memengaruhi interaksi penonton terhadap video.

Hal ini mengindikasikan bahwa meskipun like, comment, dan view saling terkait, interaksi penonton terhadap video tidak hanya dipengaruhi oleh satu metrik, tetapi oleh berbagai faktor seperti konten, waktu tayang, dan audiens target. Oleh karena itu, strategi untuk meningkatkan keterlibatan harus mempertimbangkan kombinasi dari berbagai elemen ini.



# CONCLUTION & SOLUTION

---



# CONCLUTION

**BERDASARKAN ANALISIS DATA PERFORMA KONTEN YOUTUBE, DAPAT DISIMPULKAN BAHWA ELEMEN-ELEMEN SEPERTI JUDUL VIDEO, DURASI, THUMBNAIL, SERTA TINGKAT INTERAKSI AUDIENS MEMILIKI PENGARUH YANG SIGNIFIKAN TERHADAP JANGKAUAN DAN EFEKTIVITAS PENYEBARAN KONTEN. VIDEO DENGAN JUDUL YANG INFORMATIF DAN MENARIK CENDERUNG MEMPEROLEH JUMLAH TAYANGAN DAN WAKTU TONTON YANG LEBIH TINGGI. NAMUN, DURASI VIDEO YANG TERLALU PANJANG TANPA SEGMENTASI YANG JELAS DAPAT MENURUNKAN TINGKAT RETENSI PENONTON.**

**SELAIN ITU, CTR (CLICK-THROUGH RATE) YANG TINGGI TERBUKTI BERKORELASI POSITIF DENGAN DESAIN THUMBNAIL YANG MENARIK DAN RELEVAN. MAYORITAS TRAFFIC BERASAL DARI FITUR REKOMENDASI DAN HASIL PENCARIAN YOUTUBE, YANG MENUNJUKKAN PENTINGNYA STRATEGI PENGOPTIMALAN UNTUK MENDUKUNG ALGORITMA PLATFORM. INTERAKSI PENGGUNA SEPERTI LIKE, KOMENTAR, DAN SHARE JUGA MEMAINKAN PERAN PENTING DALAM MENINGKATKAN VISIBILITAS VIDEO DI EKOSISTEM YOUTUBE.**

# SOLUTION

**UNTUK MENINGKATKAN PERFORMA KONTEN YOUTUBE, DIPERLUKAN LANGKAH-LANGKAH STRATEGIS YANG SEDERHANA NAMUN BERDAMPAK BESAR. FOKUS UTAMA ADALAH MENARIK PERHATIAN PENONTON SEJAK AWAL, MENJAGA KUALITAS ISI, SERTA MENDORONG INTERAKSI YANG AKTIF. BERIKUT BEBERAPA SOLUSI YANG DAPAT DITERAPKAN:**

- **GUNAKAN JUDUL YANG JELAS DAN MENARIK.**
- **BUAT THUMBNAIL YANG SIMPEL TAPI MENCOLOK.**
- **JAGA DURASI VIDEO AGAR TIDAK TERLALU PANJANG.**
- **AJAK PENONTON UNTUK LIKE, KOMENTAR, DAN SUBSCRIBE.**
- **UPLOAD VIDEO SECARA RUTIN DAN KONSISTEN.**
- **CEK PERFORMA VIDEO LEWAT YOUTUBE ANALYTICS.**

# THANK FOR YOUR ATTENTION

My Contact :



**+62 856 9291 3741**



**melvern.amadio@gmail.com**



**<https://just-club-880.notion.site/Melvern-Amadio-Hidayat-10b0cb7c9cad8022a6c1d97dd5ed7f9a>**



**[melvernhidayat](#)**