YouTube Trending Video Dataset Analysis

Submitted to - Elevate labs Submitted by - Charul MANISHI

2025

**ACKNOWLEDGEMENT**

I am writing to express my profound gratitude to Elevate Labs for their invaluable contributions to the successful completion of my project titled *"YouTube trending video dataset analysis."*

I extend my heartfelt thanks to all the staff members of the Elevate Labs for their continuous guidance, advice, and encouragement. Their expertise played a crucial role in helping me navigate the complexities of analysing and interpreting crime data effectively.

**CONTENT**

1. Objective

2. Data Overview

3. Tools & Technologies Used

4. Data Preprocessing

5. Key Analyses

6. Visualizations Used

7. Outputs

8. Conclusions

**1. Objective**

The primary goal of this project is to explore the patterns and trends in the YouTube U.S. trending video dataset. We aim to identify key factors that influence a video's trending duration, engagement levels, and discover common patterns through statistical and text analysis.

**2. Dataset Overview**

* Source: Kaggle – U.S. YouTube Trending Videos dataset.
* Content: Includes video metadata such as title, views, likes, dislikes, comment counts, tags, publish time, and category.
* Additional Data: A JSON file mapping category\_id to human-readable category\_name.

**3. Tools & Technologies Used**

* Python Libraries:
  + pandas, numpy – Data manipulation
  + matplotlib, seaborn – Visualization
  + nltk, wordcloud – Text preprocessing and visual analysis
  + scipy.stats – Statistical analysis (e.g., ANOVA tests)

**4. Data Preprocessing**

* Loaded and merged video metadata with category names.
* Converted publish\_time to datetime for temporal analysis.
* Handled null values and checked variable types.
* Created new features:
  + trending\_days: Number of days a video trended
  + likes\_to\_views\_ratio, comments\_to\_views\_ratio, dislikes\_to\_views\_ratio

**5. Key Analyses**

**Trending Duration Analysis**

* Computed how long each video stayed trending.
* Found correlation between trending duration and average views.
* Insight: Videos that trend longer tend to accumulate significantly higher views.

**Engagement Metrics**

* Calculated engagement ratios (likes/views, comments/views).
* Analysed how these metrics vary across videos.
* Insight: Strong positive correlation between views, likes, and trending duration.

**Category-based Performance**

* Used ANOVA to test if video views significantly differ across categories.
* Result: Statistically significant differences found; some categories trend longer and receive more views.

**Temporal Trends**

* Analysed average views by:
  + Day of the week
  + Hour of video publishing
* Finding: Videos published during certain hours or weekdays have better average view counts.

**Textual Analysis**

* Combined video titles and descriptions.
* Pre-processed text (lowercasing, stop-word removal, etc.)
* Generated a Word Cloud to visualize most common keywords in trending videos.

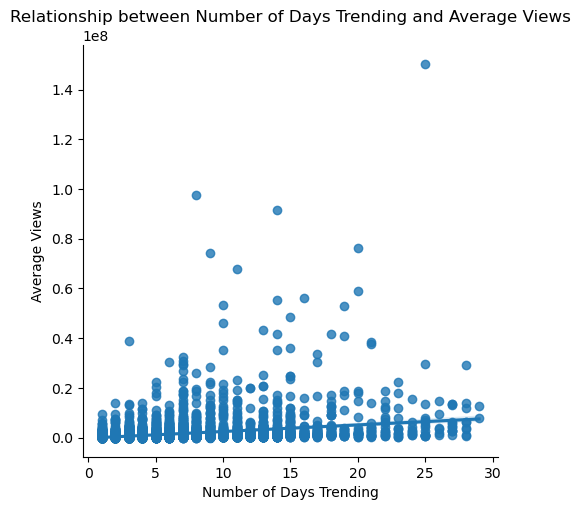
**Tags Analysis**

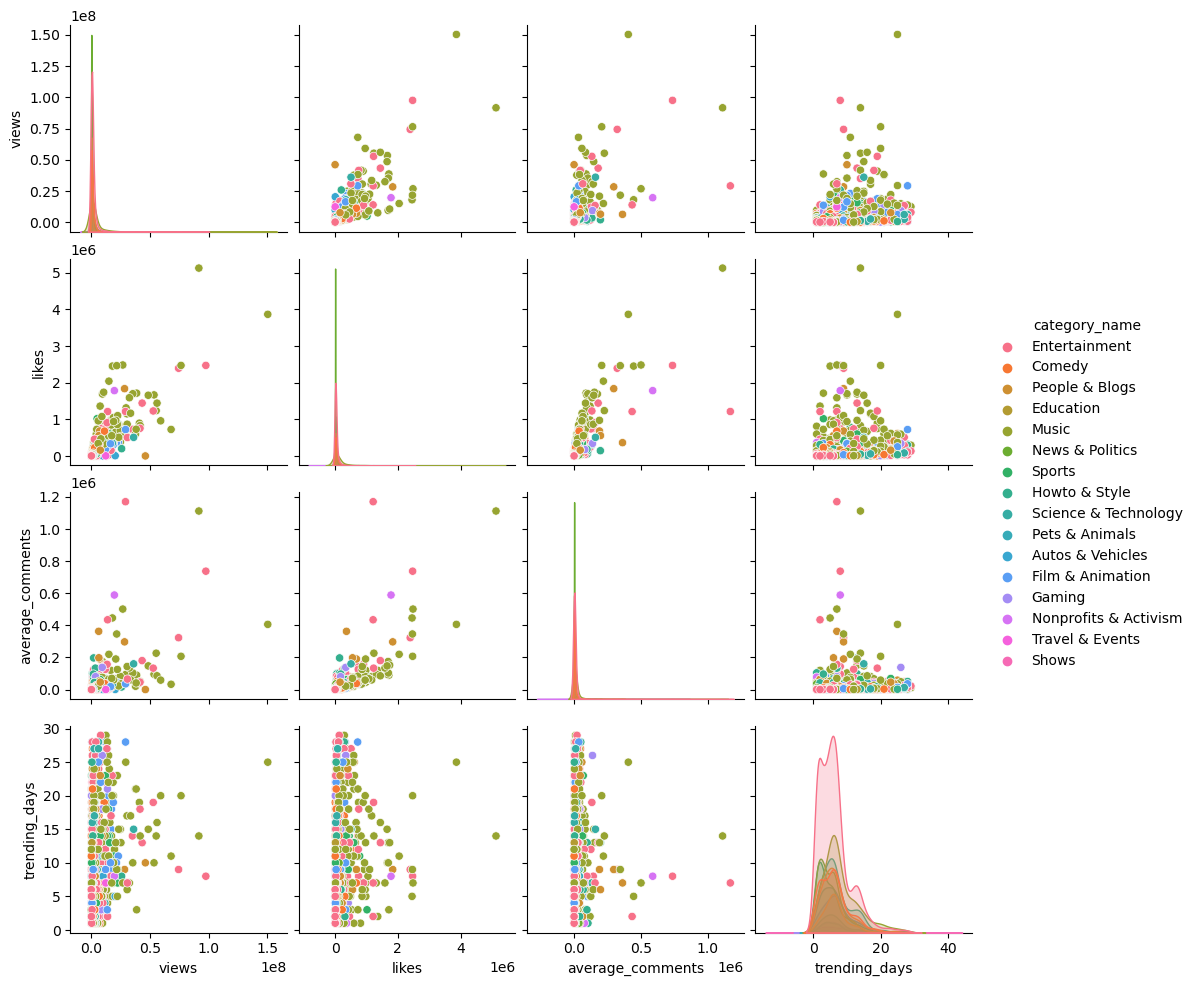
* Counted most frequent tags among top trending videos.
* Visualized via word cloud to identify hot themes.

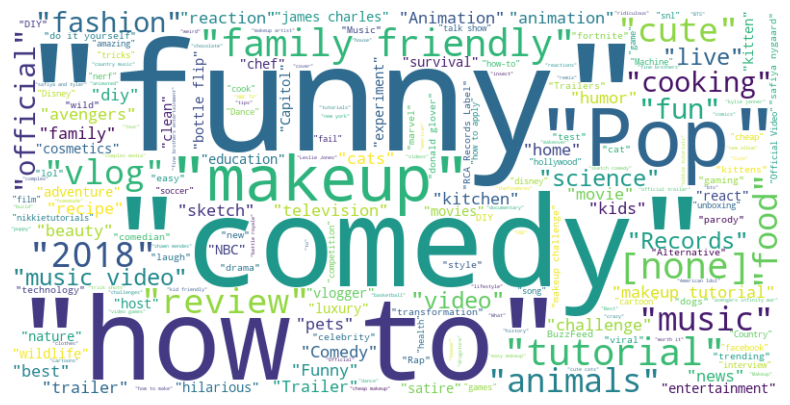
**6. Visualizations Used**

* Scatter plots (views vs. trending days)
* Pair plots (correlation of views, likes, comments)
* Word clouds (for tags and textual data)
* Donut charts (category distribution)
* Bar charts (average views by hour/day)

**7. Outputs**

****



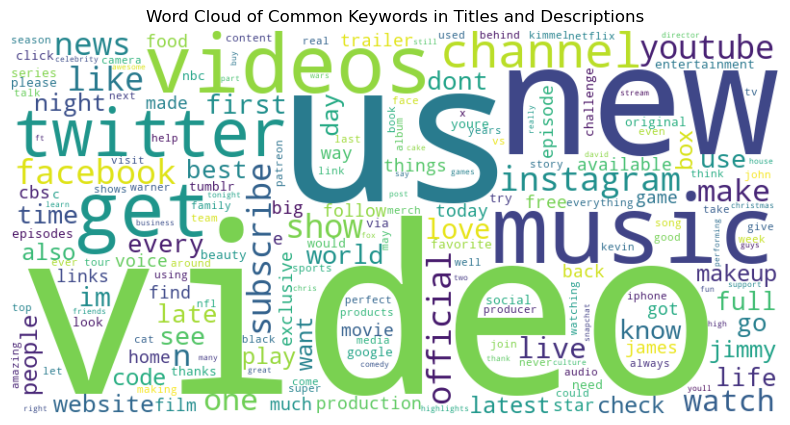


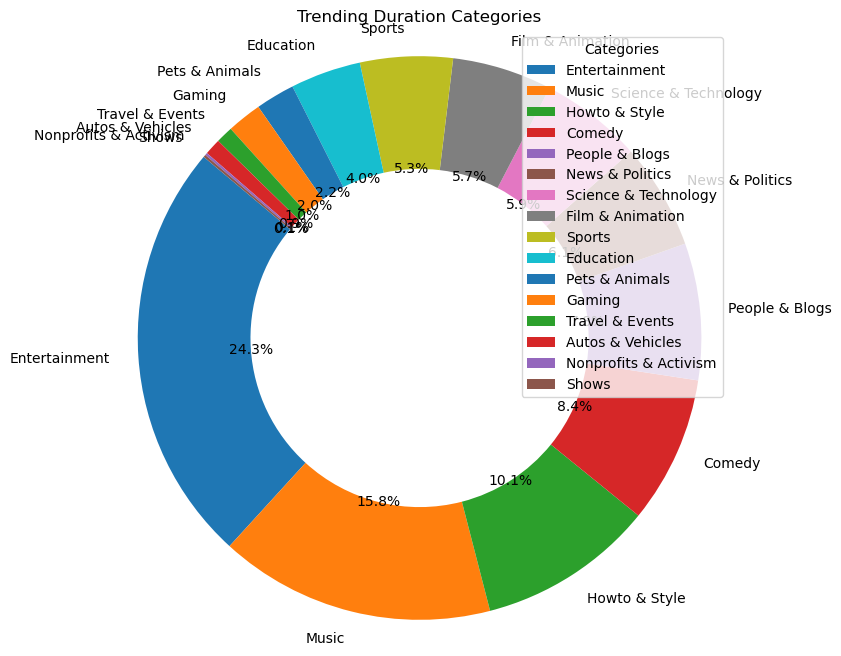
A screenshot of a graph

AI-generated content may be incorrect.

A close-up of a graph

AI-generated content may be incorrect.





**8. Conclusions**

* Engagement & Timing Matter: Videos with high engagement metrics and strategic publish times tend to trend longer.
* Content Category is Crucial: Some categories like Entertainment, Music, and News trend more often.
* Title & Tags Optimization: Keywords in titles, descriptions, and tags significantly influence video visibility and reach.