**A**

Project report on

# ANALYSIS OF YOUTUBE VIDEO USING PYTHON

# INTRODUCTION

YouTube video analysis helps evaluate video performance by combining data-driven insights with content quality assessment. On the technical side, key metrics like views, watch time, and audience retention measure how engaging a video is, while click-through rates (CTR) and traffic sources indicate how effectively a video reaches its audience. SEO factors such as keywords, titles, and descriptions play a crucial role in discoverability. Additionally, audience demographics, including age, location, and returning vs. new viewers, provide insights for content optimization.

From a content perspective, storytelling and production quality significantly impact engagement. A strong hook in the first 10 seconds helps retain viewers, while well-paced videos keep them engaged. Calls to action (CTAs) encourage interaction through likes, comments, and subscriptions. Editing techniques like eye-catching thumbnails, captions, and smooth transitions enhance visual appeal. Staying updated with trends, analysing competitor strategies, and finding a unique niche further contribute to a video's success. By balancing data analysis with creative execution, creators can optimize their videos for both YouTube's algorithm and audience engagement.

## NEED FOR YOUTUBE VIDEO ANALYSIS PROJECT

1. **Optimizing Content Performance** – Helps creators understand what works best by analysing engagement metrics, retention rates, and audience behaviour.
2. **Improving Audience Targeting** – Provides insights into demographics, interests, and viewing patterns to create more relevant content.
3. **Enhancing Discoverability** – Uses SEO techniques like keyword analysis, thumbnails, and titles to improve video rankings in search and recommendations.
4. **Increasing Monetization Opportunities** – Helps maximize revenue through better ad placement, sponsorship strategies, and audience retention.
5. **Staying Competitive** – Enables creators to track trends, analyse competitors, and adapt strategies to remain relevant in the evolving YouTube landscape.

# DATASET DESCRIPTION AND DATA ANALYSIS

## DATASET DESCRIPTION:

The dataset comprises the following key columns:

- **Video_ID:** A unique identifier for each YouTube video.
- **Title:** The title of the video.☐
- **Published_Date:** The date when the video was published.
- **Views:** The total number of views the video has received.
- **Likes:** The number of likes the video has garnered.
- **Dislikes:** The number of dislikes the video has received.
- **Comments:** The count of comments posted on the video.
- **Duration:** The length of the video, typically in seconds.
- **Category:** The category under which the video is classified (e.g., Education, Entertainment).
- **Tags:** Keywords associated with the video, aiding in searchability.
- **Description:** A brief description provided by the uploader about the video's content.

## DATA ANALYSIS:

Analyzing this dataset can yield valuable insights into factors influencing video performance on YouTube:

1. **Engagement Metrics:** By examining views, likes, dislikes, and comments, one can assess audience engagement levels. For instance, a high views-to-likes ratio may indicate content that resonates well with viewers.
2. **Temporal Analysis:** Analyzing the Published_Date alongside engagement metrics can reveal trends over time, such as seasonal variations in viewer interest or the impact of publishing frequency on engagement.
3. **Content Length Impact:** Evaluating the Duration of videos in relation to views and engagement can help determine optimal video lengths that maximize viewer retention.
4. **Categorical Performance:** Assessing which Categories garner more views or engagement can guide content creators in selecting topics that align with audience preferences.
5. **Keyword Effectiveness:** Analyzing Tags and their correlation with video performance can provide insights into effective keyword strategies for enhancing video discoverability.

# METHODOLOGY

## METHODOLOGY FOR YOUTUBE VIDEO ANALYSIS :

This methodology is structured based on the Python script, which analyzes YouTube video performance using data processing, feature engineering, and visualization techniques.

### 1. Data Collection

The dataset is loaded from a GitHub repository using pandas.

**Dataset Source:**

- youtube_data_extended.csv (a CSV file containing various YouTube video attributes)

**Data Fields Considered:**

- Video Metadata: Title, Category, Upload Date
- Engagement Metrics: Views, Likes, Comments
- Additional Metrics: Watch Time, Trending Status (trend_video flag)

### 2. Data Preprocessing

To ensure a clean dataset, the following steps are performed:

- Convert upload_date to a datetime format
- Handle Missing Values:
    - Numeric columns → filled with median values
    - Categorical columns → filled with mode values
- Feature Engineering:
    - Engagement Rate: (Likes + Comments) / Views
    - Popularity Index: (Views $\times$ 0.5) + (Likes $\times$ 0.3) + (Comments $\times$ 0.2)
    - Extracting Upload Day & Month from the upload date

### 3. Data Analysis & Visualization

**a) Identifying Popular Videos**

- Top 10 Most Viewed Videos: A bar chart visualizes the most-watched videos.

**b) Audience Engagement Analysis**

- Engagement Rate Distribution: A histogram shows the distribution of engagement rates.

**c) Video Performance Evaluation**

- Watch Time vs. Views: A scatter plot compares watch time and views, distinguishing between trending and non-trending videos.
- Trending Videos by Category: A box plot shows the view count variation across different categories.

**d) Upload Timing Analysis**

- Total Views by Upload Day: A bar chart identifies which days generate the most views.
- Trending Videos by Month: A count plot displays the frequency of trending videos per month.

**e) Trending vs. Non-Trending Video Comparisons**

- **Views, Likes, and Comments:** Box plots compare these metrics between trending and non-trending videos.

**f) Feature Correlation Analysis**

- **Correlation Heatmap:** A heatmap highlights relationships between views, likes, comments, watch time, engagement rate, and popularity index.

**4. Insights & Recommendations**

Based on the analysis, key takeaways could include:

- **Ideal Upload Time:** Which day of the week/month gets the most views?
- **Content Strategy:** What makes a video more likely to trend?
- **Engagement Optimization:** How to boost likes/comments for better engagement?
- **SEO & Visibility:** Are certain categories performing better in terms of views and popularity?

# DATA INSIGHTS

The dataset from the provided CSV file comprises 76 YouTube videos and includes the following columns:
- **video_id**: Unique identifier for each video.
- **title**: Title of the video.
- **category**: Genre or category of the video (e.g., Music, Gaming, Vlogs).
- **views**: Number of times the video has been viewed.
- **likes**: Number of likes the video has received.
- **comments**: Number of comments on the video.
- **upload_date**: Date when the video was uploaded.
- **watch_time**: Total watch time in minutes.
- **trend_video**: Boolean indicating if the video is trending (True or False).

**Data Insights:**
1. **Category Distribution**:
    o The dataset encompasses various categories, including Music, Gaming, Vlogs, Education, and Tech.
2. **View Counts**:
    o View counts range from approximately 12,350 to 959,378 views.
3. **Engagement Metrics**:
    o Likes range from 320 to 49,040.
    o Comments range from 11 to 5,953.
4. **Upload Dates**:
    o Videos were uploaded between February 13, 2022, and February 24, 2024.
5. **Watch Time**:
    o Watch times vary from approximately 13.3 minutes to 498.84 minutes.
6. **Trending Videos**:
    o The dataset includes both trending and non-trending videos, with the 'trend_video' column indicating their status.

This dataset provides a comprehensive overview of video performance metrics across different categories, enabling analysis of factors contributing to video popularity and engagement.

**EXAMPLE:**

| video_id | title | category | views | likes | comments | upload_date | watch_time | trend_video |
|----------|--------|-----------|--------|-------|----------|-------------|------------|-------------|
| vid_1 | Video 1 | Music | 336513 | 16284 | 188 | 15-04-2022 | 397.71 | FALSE |
| vid_2 | Video 2 | Music | 789325 | 40834 | 906 | 11-02-2024 | 82.51 | FALSE |
| vid_3 | Video 3 | Gaming | 293717 | 2680 | 1561 | 01-03-2023 | 53.05 | FALSE |
| vid_4 | Video 4 | Vlogs | 622195 | 40125 | 501 | 12-09-2022 | 73.61 | FALSE |
| vid_5 | Video 5 | Music | 959378 | 29067 | 70 | 05-03-2023 | 438.69 | FALSE |
| vid_6 | Video 6 | Music | 733722 | 2259 | 1087 | 10-03-2023 | 264.57 | TRUE |
| vid_7 | Video 7 | Vlogs | 887388 | 49040 | 2905 | 26-07-2022 | 234.16 | FALSE |
| vid_8 | Video 8 | Education | 743820 | 14000 | 4170 | 28-01-2022 | 54.25 | FALSE |
| vid_9 | Video 9 | Vlogs | 783381 | 11540 | 3693 | 13-11-2023 | 325.13 | TRUE |
| vid_10 | Video 10 | Vlogs | 906012 | 14778 | 595 | 02-12-2022 | 434.13 | FALSE |
| vid_11 | Video 11 | Education | 427583 | 47304 | 3970 | 11-08-2023 | 414.26 | TRUE |
| vid_12 | Video 12 | Vlogs | 153271 | 12121 | 3780 | 20-09-2023 | 313.4 | FALSE |

# CONCLUSION

The YouTube video analysis conducted using Python provided valuable insights into video performance, audience engagement, and trends. By leveraging data preprocessing techniques, feature engineering, and various visualization methods, we identified key factors contributing to a video's success on the platform.

Key findings from the analysis include:

- **Engagement and Popularity:** Videos with higher engagement rates (likes and comments relative to views) tend to have greater visibility and audience retention.
- **Trending Video Characteristics:** Trending videos generally accumulate significantly more views, likes, and comments than non-trending ones, highlighting the importance of user interaction.
- **Upload Timing:** Certain days of the week and months show higher video viewership, indicating the optimal time for uploading content to maximize reach.
- **Category Performance:** Different categories exhibit varying levels of engagement, with some genres consistently performing better in terms of views and interaction.
- **Watch Time vs. Views:** A strong correlation between watch time and views suggests that longer audience retention positively impacts video performance.

These insights can help content creators optimize their video strategies by improving content quality, selecting ideal upload times, enhancing engagement through calls-to-action, and leveraging SEO techniques. Continuous monitoring and A/B testing can further refine strategies to maximize video reach and effectiveness on YouTube.