# A Project Report on YouTube Trending Video Analytics

## Introduction

YouTube is one of the largest video-sharing platforms, where content trends are constantly evolving. Understanding the dynamics of trending videos can provide valuable insights into audience behaviour, content preferences, and factors that contribute to a video's success. This project aims to analyse patterns in YouTube trending videos, utilizing data science and SQL techniques to extract meaningful insights regarding content performance, popularity metrics, and user engagement. By leveraging structured queries and analytical tools, this study investigates the patterns that make certain videos trend and how engagement metrics influence visibility.

## Abstract

This analysis utilizes the US YouTube Trending Dataset, sourced from Kaggle, which contains 268,788 rows and 16 columns of trending video data. The dataset includes information on views, likes, comments, video category, trending duration, sentiment metrics, and publishing details. To process and analyse this data effectively, tools such as VS Code, PostgreSQL, and Python libraries (Seaborn, Matplotlib, Pandas) are employed. The project focuses on multiple key aspects, including data standardization, handling null values, sentiment analysis, ranking categories based on engagement, and evaluating trends in daily video counts. These insights help decode the viral nature of YouTube videos and determine which factors contribute to their sustained visibility.

## Tools Used

The project employs the following tools for data processing and visualization:

- **Programming:** Python, utilizing libraries like Seaborn, Matplotlib, and Pandas for data visualization and exploratory analysis.

- **Data Processing:** PostgreSQL for SQL queries related to standardization, formatting, cleaning null values, sentiment analysis, ranking categories, and counting trending video durations.

- **Development Environment:** VS Code serves as the primary interface for writing and executing scripts.

- **Data Source:** Kaggle provides the US YouTube Trending Dataset, ensuring a structured collection of real-world trending video data for comprehensive analysis.

## Steps Involved in Building the Project

### 1. Data Cleaning & Preprocessing

Before performing meaningful analysis, it is essential to clean and prepare the dataset. This phase involves:

- **Handling missing/null values:** Videos with incomplete data are addressed by either imputing missing values or filtering irrelevant entries.

- **Standardization & Formatting:** Data is standardized to maintain consistency across variables, such as transforming text-based attributes into lowercase, aligning date formats, and structuring categorical attributes for ease of analysis.

## 2. Exploratory Data Analysis (EDA)

The next step is exploring trends within the dataset by applying EDA techniques to derive initial observations. Key analysis includes:

- **Sentiment analysis of video titles & descriptions:** Evaluating positive, neutral, and negative sentiments using NLP techniques to assess their impact on engagement.

- **Ranking categories based on popularity & engagement:** Determining which video categories garner higher views and engagement, helping content creators understand trending genres.

## 3. Trend Analysis & Visualization

This phase focuses on visualizing patterns in trending video behaviour using SQL queries and Python-based graphs. Significant aspects include:

- **Daily count of trending videos:** Measuring how many new videos trend each day and identifying peak periods of video virality.

- **Duration analysis of video trends:** Assessing how long a video remains in the trending section, examining whether certain factors contribute to sustained visibility.

- **Identifying factors influencing trend longevity:** Analysing likes, comments, viewership spikes, and their correlation with a video's ability to trend over multiple days.

# Conclusion & Insights

Through this analysis, several key patterns emerge regarding video popularity, audience engagement, and content longevity. Notable findings include:

- **Patterns in video popularity across different categories:** Certain content types such as entertainment, gaming, and music videos tend to dominate the trending section, indicating consistent audience preferences.

- **The role of engagement metrics** (likes, comments, views): Videos with high interaction rates (likes and comments) tend to trend longer, reinforcing the importance of audience participation.

- **Sentiment trends influencing audience interaction:** Positive sentiments in titles and descriptions correlate with higher engagement, whereas controversial or negatively perceived content often gains traction due to polarizing discussions.