

# YouTube Trending Video Analytics – Final Report

---

## 1. Introduction

With over 2 billion monthly active users, YouTube is one of the most popular platforms globally. Trending videos reflect public interest, sentiment, and regional engagement patterns. This project aims to uncover insights from YouTube trending video data from the United States and India, using Python for data processing and visualization, VADER for sentiment analysis, SQL for aggregating insights, and Power BI for storytelling through dashboards.

## 2. Abstract

This project involved analyzing YouTube trending datasets to identify content trends, sentiment patterns, and regional behavior. We cleaned and merged datasets from India and the US, analyzed video title sentiment using VADER, visualized patterns in Python, and ranked content using SQL. Finally, a dynamic dashboard was created in Power BI to allow interactive data storytelling.

## 3. Tools Used

- Python: Data cleaning, transformation, and visualization
- Pandas, Matplotlib, Seaborn: EDA and plotting
- VADER (NLTK): Sentiment analysis on video titles
- MySQL: SQL-based insights (ranking, aggregation)
- Power BI: Interactive dashboard and visuals
- Jupyter Notebook: Execution environment for Python code

## 4. Steps Involved in Building the Project

### Step 1: Data Collection & Merging

- Downloaded datasets for US and IN from Kaggle.
- Extracted CSVs and JSON files, added country labels.
- Merged datasets into one DataFrame.

### Step 2: Data Cleaning (Python)

- Dropped nulls, duplicates, and rows with 0 views.
- Converted 'trending\_date' and 'publish\_time' to datetime format.
- Mapped 'category\_id' using JSON to meaningful category names.

### Step 3: Sentiment Analysis (VADER)

- Applied SentimentIntensityAnalyzer on 'title' field.
- Extracted sentiment scores and labeled as positive, neutral, or negative.
- Added columns: 'sentiment\_score' and 'sentiment\_label'.

### Step 4: Exploratory Data Analysis (Python)

- Visualized top categories by count and average views.
- Compared sentiment distribution using bar plots.
- Created region-wise and sentiment-wise plots using Seaborn and Matplotlib.

### Step 5: SQL-Based Insights (MySQL)

- Imported cleaned CSV to MySQL.
- Executed queries to:
  - Rank categories by average views
  - Count sentiments per country
  - Identify top 10 most viewed videos

### Step 6: Dashboard Creation (Power BI)

- Built visuals including:
  - Sentiment-wise likes/comments (bar chart)
  - Comments by country (donut chart)
  - Likes over time (line chart)
  - KPI card for total comments
  - Filter/slicer for year and sentiment

## 5. Insights & Data Storytelling

The project provided a range of key insights supported by data:

- Entertainment and Music were the most viewed categories across regions.
- Positive sentiment titles generated more likes and comments, highlighting emotional impact.
- The US audience showed higher engagement levels than India overall.
- Videos with earlier publish dates (2007–08) still performed well, indicating algorithmic resurfacing.
- The Power BI dashboard enabled interactive filtering by sentiment and time, supporting dynamic exploration.

## 6. Conclusion

This project showcases a complete data analytics pipeline—from data cleaning and sentiment analysis to SQL insights and dashboard storytelling. It successfully meets all deliverables: identifying trending categories, regional comparison, sentiment patterns, and building an

interactive dashboard. These findings can help creators and marketers understand what content resonates across regions and emotional tones.