

**Capstone Project**

# **YouTube Data Analytics**

**Turning Raw YouTube Data into Business Insights**

# Introduction

YouTube has become one of the world's most influential digital platforms, shaping entertainment, education, marketing, and global communication. With millions of videos uploaded daily, understanding what drives content to trend is a powerful skill for data enthusiasts. Analyzing YouTube data provides unique opportunities to explore **viewer engagement, content strategies, category dominance, and country-wise trends**.

This project is designed to give students a **hands-on, end-to-end experience** in working with real-world datasets. By exploring and analyzing YouTube trending video and channel information, students will practice skills across **data cleaning, descriptive analysis, visualization, and storytelling**, while also learning how to collaborate in structured group roles—similar to professional consulting projects.

The dataset provided is a large CSV file containing trending videos from multiple countries. Each record includes detailed **video-level metadata** (title, publish date, tags, views, likes, comments, duration, etc.) as well as **channel-level metadata** (channel name, subscribers, total views, video counts, etc.). Students will filter and extract data for **3–5 countries of their choice**, enabling comparative insights across regions and categories.

## Project Overview

The **YouTube Data Analytics Project** has been structured into well-defined roles to simulate real-world team-based analytics work. Each student will take ownership of a specific role, ensuring diverse tool usage (Excel, SQL, Python, Power BI, Tableau), while the group as a whole will deliver a unified project outcome.

## Dataset Description

This dataset provides a **comprehensive collection of YouTube video and channel metadata** curated for data analysis, visualization, and storytelling projects. It contains rich information on trending videos across multiple countries, including video performance statistics, engagement metrics, and channel-level details.

The dataset is designed to help learners and researchers explore **real-world YouTube dynamics**, such as:

- What type of content gains the highest views and engagement?
- How do categories perform across different countries?
- What role do publishing time, video duration, or tags play in driving popularity?
- Which channels dominate in terms of subscribers, views, and content consistency?

Dataset Link- <https://www.kaggle.com/datasets/keshavbansal95/youtube-trending-videos-dataset/data/data>

## Features

The dataset includes detailed **video-level fields** such as:

- Video ID, title, description, and publish time
- Trending date and country
- Tags, categories, duration, resolution, and licensed content status
- Views, likes, and comment counts

Alongside **channel-level information** including:

- Channel ID, title, and description
- Channel country, publish date, and custom URL (if available)
- Subscriber count, total views, video count, and hidden subscriber flag

With this structured dataset, students and professionals can perform **data cleaning, transformation, SQL querying, trend analysis, and dashboarding in tools such as Excel, SQL, Power BI, Tableau, and Python**. It is also suitable for advanced machine learning tasks like **predicting video performance, engagement modeling, and natural language processing on video titles and descriptions**.

## Use Cases

1. **Descriptive Analytics:** Identify top categories, channels, and countries leading the YouTube trending space.
2. **Comparative Analysis:** Compare engagement rates across different regions and content types.
3. **Visualization Projects:** Create dashboards showing performance KPIs, category trends, and time-based patterns.
4. **Storytelling:** Derive business insights and best practices for creators, marketers, and educators on YouTube.
- 5.

## Educational Value

This dataset is structured specifically for **student projects and group assignments**. It ensures every learner can take a role—whether as a data engineer, analyst, visualization specialist, or business storyteller—mirroring the structure of real-world consulting projects.

# Project Roles & Responsibilities

## Role 1: Data Engineer (Data Collector & Cleaner)

**Objective:** Prepare a clean, reliable, enriched dataset for analysis.

**Tasks:**

- Extract subset of data for 3–5 chosen countries.
- Format and standardize:
  - Convert publish time → separate date, day, and hour.
  - Clean tags (remove stop words, split multiple tags).
  - Standardize categories & handle missing values.
- Create new derived metrics:
  - Engagement Rate = (likes + comments) / views
  - Like Ratio = likes / (likes + dislikes)
- Tool Options: Excel, SQL, Python.

**Deliverable:** Clean, enriched dataset with new calculated fields.

## Role 2: Data Analyst (Descriptive & Comparative Analysis)

**Objective:** Explore and summarize key performance patterns.

**Tasks:**

- Identify top 10 channels by views, likes, and engagement rate.
- Find which categories have the highest average views.
- Compare engagement rates across selected countries.
- Analyze weekly and monthly trends in uploads and views.
- Present outputs as pivot tables, summary tables, or SQL queries.
- Tool Options: Excel, SQL, Python.

**Deliverable:** Detailed summary tables and outputs (Excel workbook, SQL views, or Python scripts).

## Role 3: Visualization Specialist (Dashboard Developer)

**Objective:** Build an interactive dashboard that brings the data to life.

**Tasks:**

- Import cleaned dataset into a visualization tool.
- Create a multi-page dashboard with:
  - **Overview Page:** KPIs (total views, likes, average engagement).
  - **Category Insights:** performance by category.
  - **Channel Leaderboard:** top-performing channels.
  - **Time Analysis:** heatmap/day-hour vs. views.

- Add interactive slicers (country, category, time).
- Tool Options: Power BI, Tableau, Excel.

**Deliverable:** Multi-page dashboard with at least 3–4 interactive pages.

## **Role 4: Business Analyst (Insights, Storytelling & Recommendations)**

**Objective:** Convert analytics into actionable insights and craft a compelling narrative.

**Tasks:**

- Review outputs from the Data Analyst and Visualization Specialist.
- Derive actionable insights such as:
  - “Best upload time for maximum engagement is evenings (6–9 PM).”
  - “Music dominates in views, but Education delivers higher engagement per view.”
- Use **storytelling techniques** to turn data into a persuasive message.
- Prepare a business insights report + presentation slides.
- Tool Options: Power BI, Tableau, Excel (for visuals), plus Word/PPT.

**Deliverable:** Business Insights Report + Storytelling Presentation (10–12 slides).

## **Role 5: Project Manager & Quality Assurance Lead**

**Objective:** Coordinate the project and ensure professional delivery.

**Tasks:**

- Define milestones and track progress.
- Perform quality checks of dataset, analysis, and dashboards.
- Consolidate all outputs into a final submission package.
- Lead the group presentation:
  - Introduce the project.
  - Ensure each student presents their role’s part.
  - Conclude with key recommendations.
- Tool Options: Open choice (management, documentation, integration).

**Deliverable:** Final consolidated project folder + leading the group presentation.

## Final Deliverables

1. **Individual Submissions (from each student):**
  - Clean dataset file / SQL scripts / Python notebook (Role 1).
  - Analysis tables, queries, or reports (Role 2).
  - Dashboard file (PBIX / TWBX / Excel) (Role 3).
  - Insights report & presentation slides (Role 4).
2. **Group Submission:**
  - Consolidated project folder with all files.
  - Final group presentation (10–12 slides).
3. **Presentation:**
  - Each student presents their part (2–3 slides).
  - Project Manager leads, introduces, and concludes.

## Why This Project?

- **Real-World Relevance:** Mimics how consulting firms analyze social media trends.
- **Collaborative Workflow:** Divides work into roles like in actual data teams.
- **End-to-End Coverage:** From raw data to actionable insights.
- **Skill Integration:** Ensures use of Excel, SQL, Python, Power BI, Tableau across groups.
- **Storytelling Focus:** Trains you to communicate insights, not just build dashboards.