Student Name: Amrutha Varshini G

Dataset: Trending videos on YouTube

YouTube Trending Video Analytics – Final Report

---

1. Introduction

With the exponential growth of video content, platforms like YouTube have become a hub for creators and consumers. Analysing trending videos provides valuable insights into user engagement, content preferences, and viewing patterns. This project explores YouTube trending video data to uncover patterns based on category, sentiment, channel popularity, and content attributes like duration and definition.

---

2. Abstract

This project aims to analyse a dataset of trending YouTube videos using Python and Power BI. The study covers data cleaning, exploratory data analysis (EDA), sentiment classification of video titles, SQL querying, and visual storytelling through dashboards. Key insights include the most engaging video categories, sentiment trends, and how video duration or HD quality affects viewership. The goal is to assist content creators, marketers, and analysts in understanding what makes a video trend.

---

3. Tools Used

- **Kaggle:** Source of the trending YouTube video dataset
- **Python:** For data preprocessing, cleaning, and sentiment analysis (NLTK + VADER)
- **SQLite (via Pandas):** To execute SQL queries for structured insights
- **Jupyter Notebook:** For writing and executing Python code
- **Power BI:** For creating interactive visual dashboards and reports

---

4. Steps Involved in Building the Project

Phase 1: Setup & Cleaning

- Loaded CSV dataset, standardized column names, handled nulls
- Converted numeric columns and parsed dates

Phase 2: Exploratory Data Analysis (EDA)

- Analysed top video categories by average views
- Explored relationships between views, likes, and comment counts
- Examined trends over time

Phase 3: Sentiment Analysis

- Used VADER from NLTK to classify titles/tags into Positive, Neutral, Negative
- Analysed sentiment impact on popularity

Phase 4: SQL Insights

- Executed SQL queries to find top-performing videos and average views by category

Phase 5: Power BI Dashboarding

- Built 3 interactive pages:
    - Overview (views by category, sentiment distribution, trend timeline)
    - Channel Analysis (top channels, engagement metrics)
    - Video Characteristics (impact of duration, HD/SD, and captions)

---

5. Conclusion

This project demonstrates how data-driven techniques can uncover trends in digital content consumption. The analysis revealed that video category, length, sentiment, and quality all influence viewer engagement. Educational and entertainment categories attract higher views, while positive sentiment titles often perform better. Power BI dashboards helped visualize these patterns in an interactive format. Such insights can guide creators and marketers in optimizing their video strategies.