

# YouTube Trending Video Analytics

---

## Introduction

The YouTube Trending Video Analytics project explores the behavior of trending videos on YouTube across three countries: India, Japan, and South Korea. With YouTube being a global platform for content sharing and discovery, understanding what makes a video trend is critical for content creators and digital marketers. This project aims to identify which factors — including engagement metrics and sentiment — influence video performance across different regions. The analysis covers video titles, categories, likes, views, and comments using Python and Power BI.

---

## Abstract

We obtained daily snapshots of trending videos from Kaggle datasets, covering multiple countries. After cleaning and combining data for IN, JP, and KR, we performed Exploratory Data Analysis (EDA) to assess engagement levels. Using VADER from NLTK, we also classified video titles by sentiment. This project revealed strong correlations between engagement metrics and highlighted regional preferences in content and tone. Over 80,000 records were analyzed and presented using Power BI dashboards for better interpretability.

---

## Tools Used

- **Python** (Pandas, Seaborn, Matplotlib)
  - **NLTK** (VADER Sentiment Analysis)
  - **Power BI** (Dashboard Visualizations)
  - **Jupyter Notebook, VS Code**
  - **Data Source:** Kaggle YouTube Trending Datasets
- 

## Steps Involved

1. **Data Collection & Merging:**  
Downloaded CSVs for IN, JP, KR and added a country column.
2. **Cleaning & Transformation:**  
Parsed date fields, removed duplicates, fixed encodings.
3. **Category Mapping:**  
Converted category\_id into readable labels (e.g., Music, Comedy).
4. **Sentiment Analysis:**  
Used VADER to score titles as Positive, Neutral, or Negative.
5. **Correlation Analysis:**  
Found strong correlation between views, likes, and comment counts.
6. **Power BI Dashboarding:**  
Created KPIs, sentiment charts, and country/category filters.

# YouTube Trending Video Analytics

---

## Workflow:

1. Loaded CSVs for each country and added a country column.
2. Cleaned data: fixed encodings, parsed trending\_date, removed duplicates.
3. Mapped category\_id to human-readable names.
4. Applied **VADER sentiment analysis** to title → Positive, Neutral, Negative.
5. Ran correlation analysis for views, likes, comments, dislikes.

---

## Key Findings:

- Negative titles had the **highest average views** (~1.58M).
- India showed more sentiment variation; Japan/Korea leaned neutral.
- Engagement (likes/comments) strongly correlates with views.
- **Top categories:** India → Music & Entertainment; Korea/Japan → Education & Style.

---

## Power BI Dashboard Included:

- **KPI cards:** Total videos, views, avg likes/comments
- **Donut/Bar charts:** Sentiment, category distribution
- **Region-wise stacked bar chart**
- **Trending line graph** (optional if date supported)

---

## Conclusion

This project shows that emotional tone affects video performance. Titles with negative sentiment attracted the most views (~1.58M). India had more emotionally varied content, while Japan and Korea leaned toward neutrality. Strong correlations were found between views and other engagement metrics. Using Python for EDA and Power BI for dashboards allowed for clear, actionable insights suitable for marketers, creators, and strategists.



**Prepared by:** Babi Pepakayala



**Email:** babipepakayala162129@gmail.com



**Date:** June 2025