YouTube Trending Video Analytics Project

Introduction

This project analyzes trending YouTube videos across regions (US and India) to uncover patterns in categories, engagement metrics, sentiment of titles, and duration of popularity. The goal is to gain insights into what drives video virality and regional differences.

Abstract

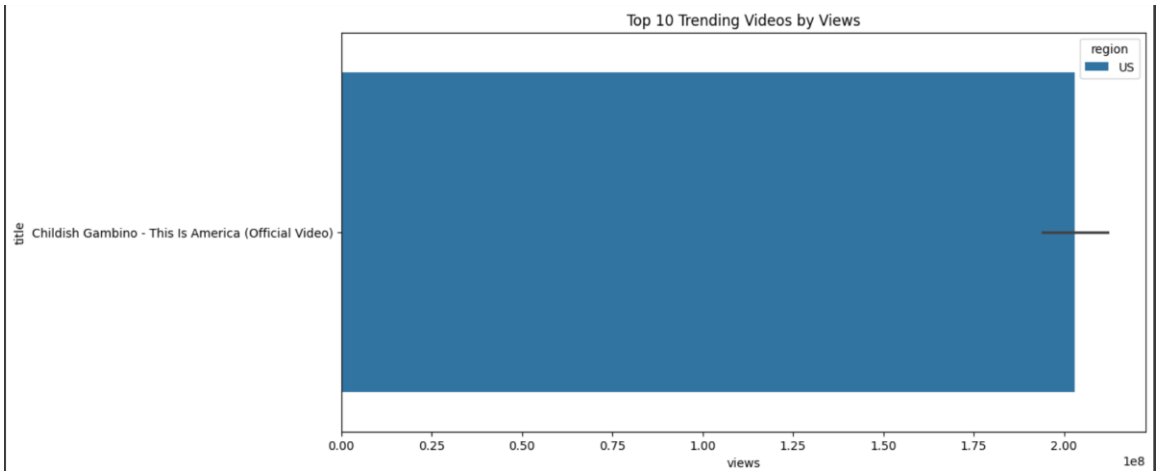
We used Kaggle's YouTube Trending dataset, combining data from US and India. The analysis covered frequency of categories, top trending videos, engagement relationships (views vs likes), sentiment analysis of titles, and WordClouds to highlight common patterns. The study highlights that Entertainment and Music categories dominate globally, while regional differences exist in engagement and trending duration.

Tools Used

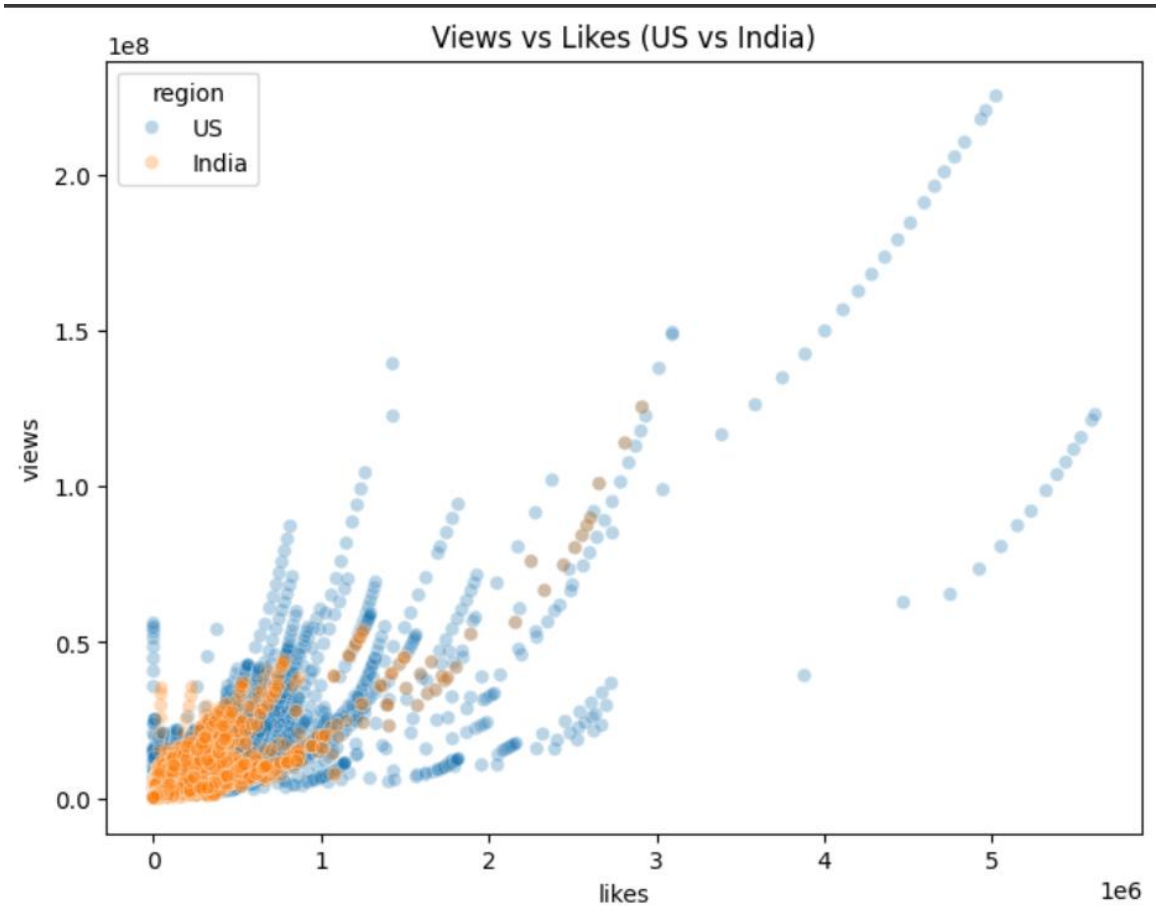
- Python (Pandas, Seaborn, Matplotlib)
- NLTK (Sentiment Analysis)
- WordCloud Library
- Google Colab

Steps Involved

1. 1. Loaded US and India YouTube trending datasets and combined them.
2. 2. Cleaned data and standardized dates.
3. 3. Analyzed top categories and top 10 trending videos.



4. 4. Visualized engagement through scatterplots of views vs likes.



5. 5. Measured trending duration of videos across regions.
6. 6. Performed sentiment analysis on video titles using VADER.
7. 7. Created WordClouds to visualize common words in trending titles.

Conclusion

Entertainment and Music dominate trending videos in both regions. Views and likes show strong correlation, suggesting engagement consistency. Most videos trend for less than a week, though some stay longer. Sentiment analysis revealed mostly neutral-to-positive titles. Regional analysis indicated small but notable differences in category dominance and engagement. This project provides actionable insights into video virality patterns.