

YouTube Trending Video Analytics: Final Report

Introduction

The goal of this project was to analyse YouTube trending videos across different regions (United States and India) to uncover insights into viewer preferences, genre popularity, sentiment trends, and regional differences.

Methodology

1. Data Cleaning

- Imported datasets: From Kaggle downloaded the dataset USvideos.csv and INvideos.csv
- Converted trending_date to datetime format
- Removed duplicate records within and across datasets based on video_id, title, and region
- Combined datasets using pd.concat() for comparative analysis

2. Sentiment Analysis

- Applied TextBlob sentiment analysis on video titles
- Added a title_sentiment column with polarity scores (-1 to 1)

3. SQL-style Ranking (with pandas)

- Ranked categories by average views
- Identified most popular categories using:

```
SELECT category_id, AVG(views) AS avg_views
```

```
FROM youtube_trending
```

```
GROUP BY category_id
```

```
ORDER BY avg_views DESC;
```

4. Time-Series and Visual Analysis

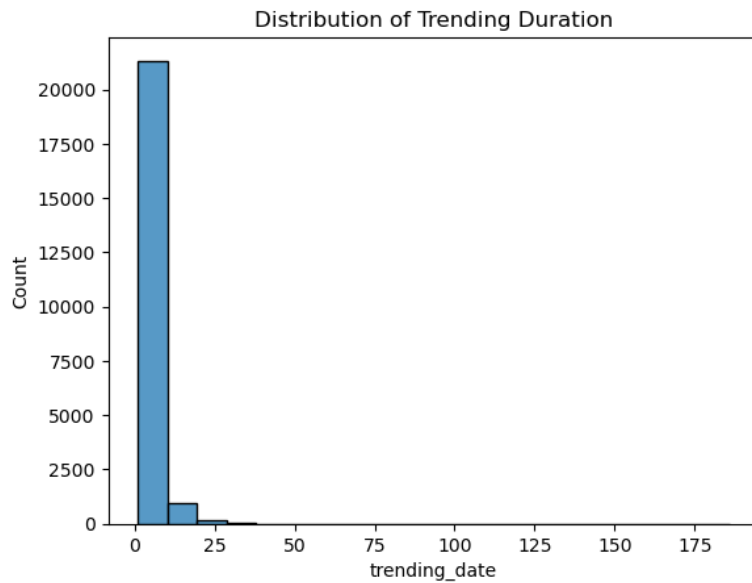
- Measured video trending duration per region
 - Created visualizations using Seaborn and Matplotlib
-

Insights

1. Time Series Visualization for Trending Duration

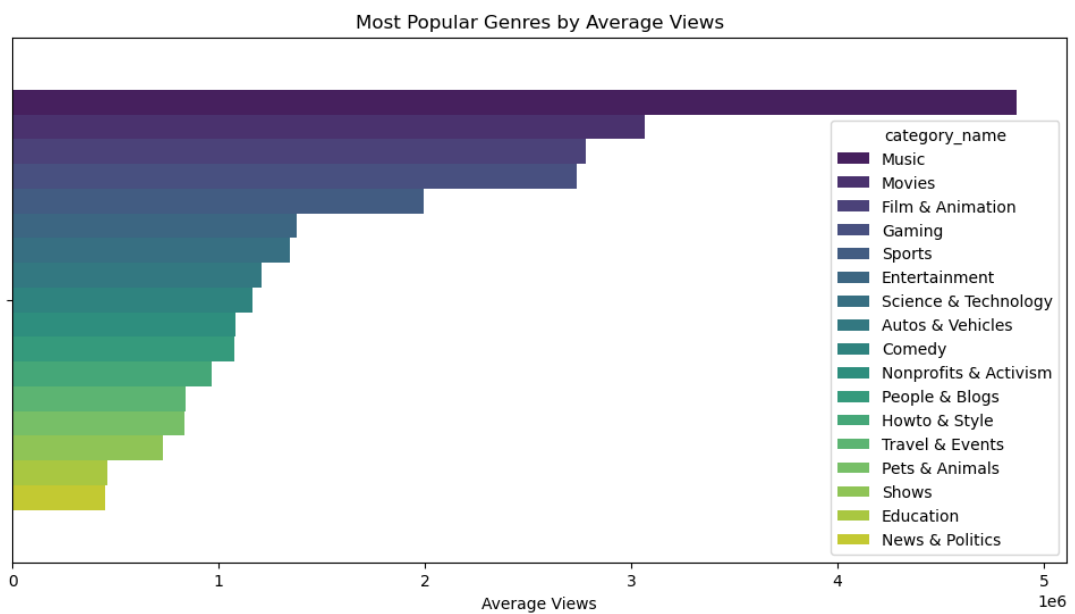
- A histogram of trending duration revealed that most videos trended for a short period — typically 1 to 3 days, with fewer videos sustaining longer visibility
- The distribution is right-skewed, indicating that only a few videos manage to stay trending for extended durations.

- Visual: Histogram (Seaborn histplot) of TrendingDuration column, showing frequency of videos against number of days trended.



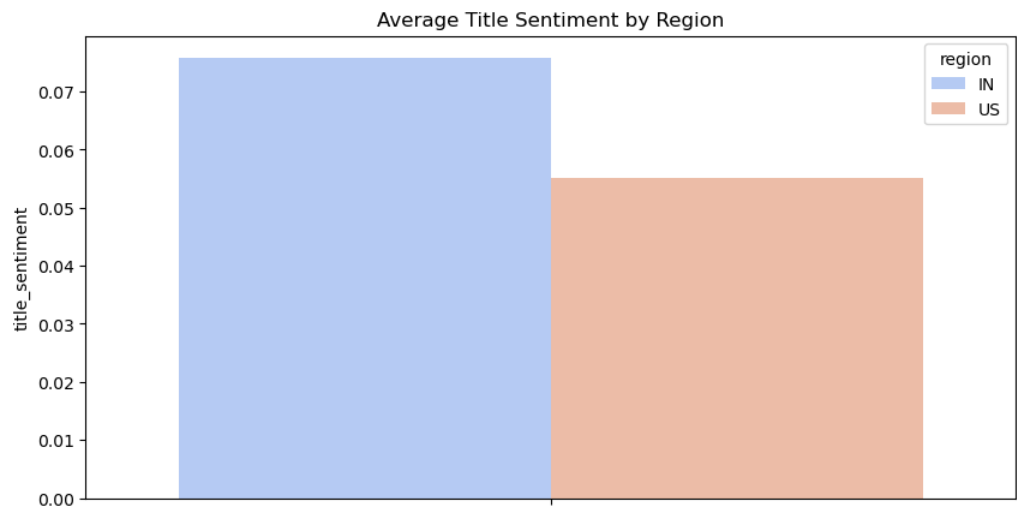
2. Most Popular Genres by Views

- Music and Entertainment** consistently had the highest average views in both US and India.
- Visual: Horizontal bar plot of average views by category_name.



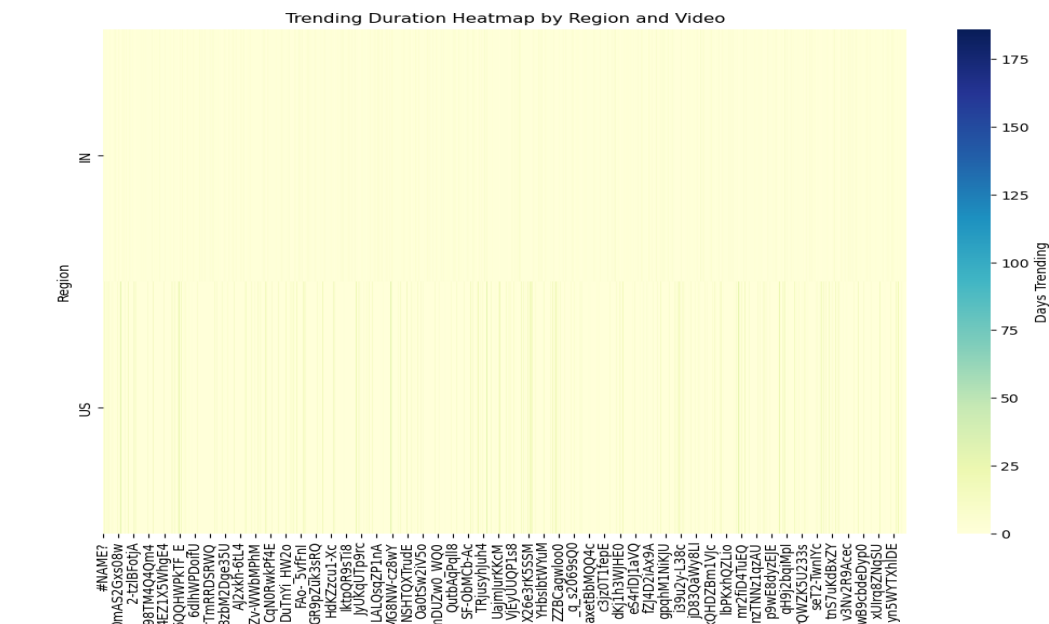
3. Sentiment Patterns Across Regions

- US video titles showed a **slightly more positive sentiment** on average.
- India had a **wider range of sentiments**, indicating more emotional diversity.
- Visual: Bar chart of title_sentiment mean by region.



4. Trending Duration Patterns

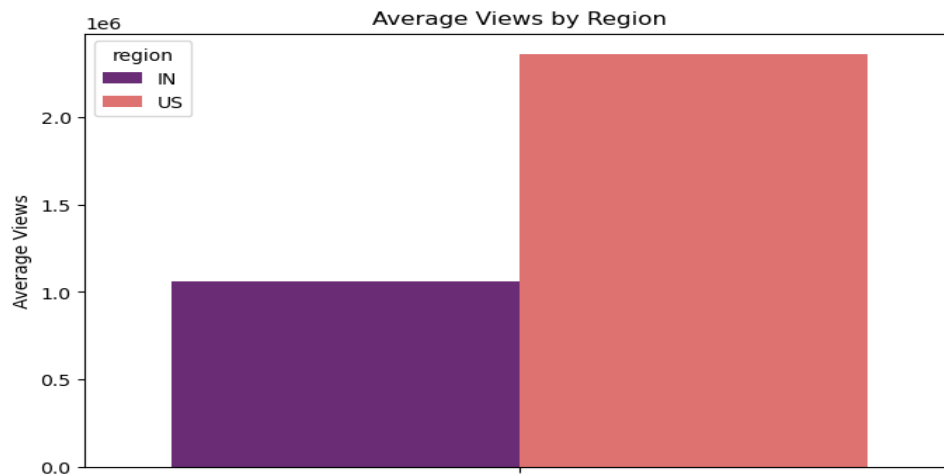
- Videos in the US tend to trend **longer**, possibly due to broader audience engagement.
- In India, trending cycles are **shorter but more frequent**, especially for music content.



5. Region-Wise Comparison

- US had **higher average views**, while India had **higher video publishing frequency**.

- Heatmaps showed regional time-of-day and weekday preference patterns.
- Some genres trend more in one region (e.g., tech in the US, music in India).



Visualizations Included (PowerBI)

Most Popular Genres by Views

- **Visual Type:** Clustered Bar Chart
- **Description:** Displays the total number of views grouped by video category (using category_id).

Trending Duration by Category

- **Visual Type:** Line Chart
- **Description:** Shows which categories tend to stay trending the longest using a sum of days between publish and trending dates.

Views Over Time (Daily Trend)

- **Visual Type:** Line Chart
- **Description:** A time-series chart that illustrates how view counts have fluctuated over daily trending dates.

Regional Distribution – Title Count

- **Visual Type:** Donut Chart
- **Description:** Compares the total number of trending video titles between regions (IN vs US).

KPI Cards – Summary Metrics

- **Visual Type:** Card / KPI
 - **Description:** High-level numeric indicators that provide overall statistics for the dataset.
-

Views by Year

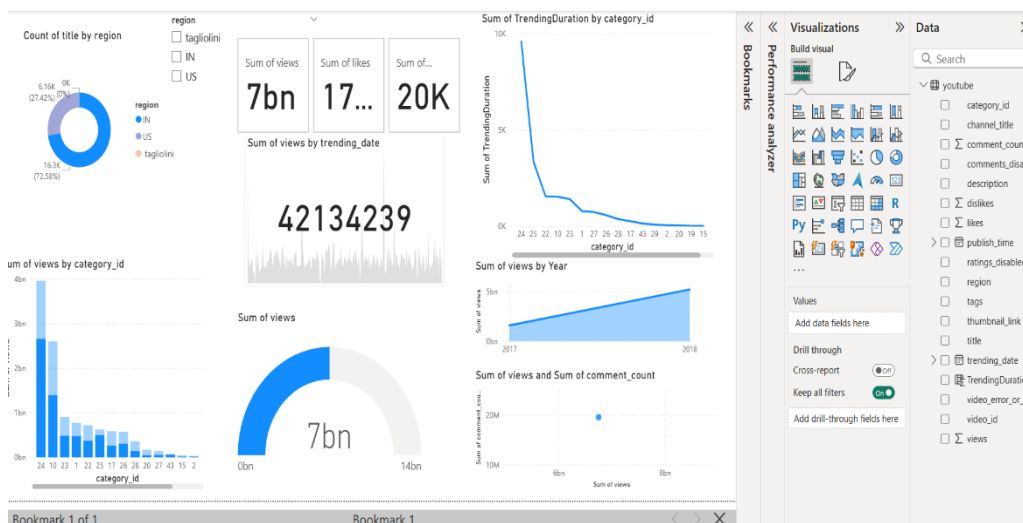
- **Visual Type:** Area Chart
 - **Description:** Displays how total view counts have grown between 2017 and 2018.
-

Views vs Comment Count

- **Visual Type:** Scatter Plot
 - **Description:** A single-point scatter comparing total views to total comment count (consider using more granular data for richer insights).
-

Gauge – Total Views Visualization

- **Visual Type:** Gauge Chart
- **Description:** Visual representation of total views as part of a possible benchmark (e.g., 7bn out of 14bn max).



Conclusion

Key Findings

- Music and Entertainment dominate trending categories.
- US content shows more stable and positive sentiment.
- India trends are faster, more dynamic, and emotionally varied.
- Clear regional differences in genre preference and engagement metrics.

Recommendations

- **Creators** should tailor content to regional sentiment trends.
- **Marketers** should leverage genre-specific insights for targeted campaigns.
- **YouTube** can enhance discovery by promoting regionally underrepresented categories.