YouTube Trending Video Analytics

Objective:

The objective of this project is to uncover patterns and derive insights from YouTube trending video datasets across three countries: the United States (US), India (IN), and Great Britain (GB). The goal is to understand what factors influence a video's popularity, including timing, sentiment, and engagement metrics.

Tools Used:

Python (for data cleaning, transformation, sentiment analysis)

Tableau (for visual analytics and dashboard creation)

Google Colab (execution environment for Python)

Step 1: Dataset Collection

Datasets for US, IN, and GB trending videos were downloaded from Kaggle. Each dataset contains information such as video titles, channel names, views, likes, dislikes, comment counts, publish time, and tags.

Step 2: Data Cleaning and Standardization

Each dataset was loaded into Google Colab and cleaned by:

- Converting `publish_time` to datetime format.

- Extracting new fields like `publish_date`, `publish_hour`, and `country`.

- Handling missing values and unnecessary columns.

Step 3: Sentiment Analysis

Performed sentiment analysis using the TextBlob library:

- Added two new columns: `title_sentiment` and `tags_sentiment`.

- These fields reflect polarity scores ranging from -1 (negative) to 1 (positive).

Step 4: Data Consolidation

All three datasets (US, IN, GB) were merged into a single dataset to allow for comparative analysis.

Step 5: Exploratory Data Analysis (EDA)

Initial visual inspection of key metrics including:

- View count distributions

- Likes and comments correlation

- Sentiment distribution

Step 6: Tableau Visualizations

Multiple visualizations were created to analyze different aspects of the dataset:

Dashboard 1: Top Trending Channels and Categories

- Chart 1: Bar chart showing top 10 channels by total views

- Chart 2: Box plot comparing view counts across categories

Dashboard 2: Engagement Metrics

- Chart 3: Line chart showing average views by publishing hour

- Chart 4: Scatter plot showing correlation between likes and comments

Dashboard 3: Sentiment and Popularity Insights

- Chart 5: Bubble chart of views vs likes with dislikes as bubble size

- Chart 6: Box plot of title sentiment distribution by country

- Chart 7: Grouped bar chart showing category-wise average views per country

Step 7: Dashboard Design and Publishing

- All three dashboards were designed in Tableau using tiled layouts.

- Filters and legends were added for interactivity.

- Dashboards were saved and published via Tableau Public.

Key Insights:

- Certain hours (like 6 PM to 9 PM) see a higher average view count across all regions.

- Titles with moderately positive sentiment tend to receive more views.

- US videos generally have higher engagement (likes/comments) than those from IN and GB.

- Tech and Entertainment categories dominate across all three countries.

Conclusion:

This project successfully demonstrates how structured analysis of YouTube trending data can provide insights into content performance across different regions. The combination of Python for data preparation and Tableau for visualization created a powerful and informative analysis framework.