YouTube Trending Video Analytics – Project Report

Introduction

YouTube is one of the world's most popular video-sharing platforms, generating vast amounts of data every day. This project focuses on analyzing trending video data from YouTube to uncover patterns and insights related to viewer engagement, regional trends, category performance, and time-based popularity. The goal is to transform raw data into meaningful visualizations and reports that help understand the dynamics of YouTube's trending content.

Abstract

The project centers around the analysis of trending YouTube videos from the US and India, spanning several years. Using a structured analytics pipeline that integrates Python, SQL, and Tableau, we processed, transformed, and visualized the data. Key insights include regional differences in content consumption, popularity across categories, and temporal viewing behaviors, enhanced by sentiment analysis on video titles and tags.

Tools Used

- **Python (Pandas, Matplotlib, Seaborn, Textblob)** – For data preprocessing, visualization and sentiment analysis.
- **SQLite** – Lightweight SQL database for efficient querying within Jupyter.
- **Jupyter Notebook** – Environment for exploratory data analysis and code execution.
- **Tableau** – Visualization platform used to create interactive dashboards and trend visualizations.

Steps Involved in Building the Project

1. **Data Collection & Cleaning**
    - Imported CSV datasets containing YouTube trending videos.
    - Cleaned data by handling null values, correcting formats, and filtering by regions (IN, US).
2. **Data Transformation**
    - Mapped category IDs to readable category names using JSON metadata.
    - Extracted temporal features (year, month) from trending dates for trend analysis.
3. **SQL Analysis in Jupyter**
    - Utilized SQLite3 within Python to execute queries including:
        - Avg views as per region and category.
        - Avg likes as per region and category.
        - Top videos as per days trending.
4. **Sentiment Analysis**
    - Applied TextBlob to analyze sentiment polarity of video titles and tags.
    - Categorized videos based on positive, negative, and neutral sentiment scores.

5. **Visualizations in Tableau**
    - Created:
        - **Bar Chart** showing avg. views by category **– India VS USA**
        - **Category Distribution Pie Chart**
        - **14 Day Average Line Chart** for video counts by region over time
        - **Stacked Bar Chart** highlighting **Viewer Sentiment** as per **Region & Category**
        - **Engagement Intensity Heatmap: Views vs Likes (Sized by Comment Count)**
        - **Interactive Dashboard**

Conclusion

This project successfully integrated data engineering and visualization techniques to derive actionable insights from YouTube's trending video data. Key findings include:

- YouTube trends vary notably between India and the USA. India sees higher average views in Pets & Animals, Music, and Film & Animation, while Entertainment, Music, and Science & Technology lead in the USA. Despite high viewership, content volume is dominated by Entertainment.
- India consistently trends more videos, with a mid-2021 spike likely due to increased digital activity.
- Sentiment-wise, Pets & Animals and How-to & Style receive the most positive responses across both regions, while Comedy and News & Politics show neutral/mixed sentiment. Slightly higher sentiment for Music in India suggests cultural influence. These insights reflect regional and emotional differences in viewer preferences.