YouTube Trending Video Analytics Project Report

By Ayush Katkar

1. Introduction

This project explores YouTube trending videos across different regions. We analyzed datasets from various countries using Python for preprocessing and sentiment analysis, SQL for insights extraction, and Tableau for dashboard visualization.

2. Tools & Technologies

- Python (Pandas, Matplotlib, Seaborn, TextBlob)

- SQL (SQLite)

- Tableau (Dashboards and Visualizations)

- Jupyter Notebook

3. Dataset

The dataset was sourced from Kaggle: https://www.kaggle.com/datasets/datasnaek/youtube-new. It contains information about trending videos in multiple regions like US, IN, CA, DE, etc. Files include CSVs for each country and corresponding category_id JSONs.

4. Data Cleaning and Exploration

- Checked nulls, data types, duplicates
- Parsed publish_time and trending_date into datetime
- Created Trending Duration feature
- Merged category names using category JSON files

5. Data Preprocessing

We loaded and combined datasets for multiple regions using Python. Date fields were converted to appropriate formats, and a 'region' column was added to each dataset.

6. Visualization in Python (Dash + Plotly)

- Created a Dash app showing average views by category_id
- Used Plotly Express for bar charts embedded into Dash layout

7. Sentiment Analysis

We used the TextBlob library to perform sentiment analysis on video titles and tags. The results were added as new columns: polarity and subjectivity.

8. SQL Analysis

We used SQLite to rank categories by average views and extract top-performing genres.

9. Tableau Dashboards

We built the following dashboards in Tableau:

1. **Most Popular Genres** – Bar chart showing average views by category.

2. **Video Sentiment Analysis** – Scatter plot of sentiment polarity vs views.

3. **Video Count by Region** – Pie chart of video count by region.

4. **Trending Duration** – Bar chart for number of videos by duration bin.

5. **Top 10 Channels** – Horizontal bar chart showing top 10 channels by total views.

6. **Filters Used** – Drop-downs for region, category, channel, and publish month.

10. Tableau Filters Setup

- region: Drop-down (used in all charts)

11. Conclusion

This project demonstrates the complete pipeline of data collection, preprocessing, analysis, and visualization. By combining Python, SQL, and Tableau, we uncovered insights into regional YouTube trends, sentiment patterns, and top-performing content.