

YouTube Comment and Channel Data Analysis

1. Project Overview

This project aimed to perform data analysis on YouTube video and comment datasets to uncover patterns in user engagement, sentiment, and video performance. The analysis provides insights into how audience behavior correlates with video popularity, comment activity, and viewer sentiment.

2. Objectives

- Analyze the relationship between video metrics (views, likes, comments) and content performance.
 - Understand audience engagement through comment analysis.
 - Perform sentiment analysis on comments to determine viewer perception.
-

3. Technical Approach

Data Collection & Preprocessing:

- Utilized datasets containing video performance metrics and user comments.
 - Cleaned the data by handling missing values, correcting data types, and removing duplicates using **pandas** and **numpy**.
 - Standardized text data for analysis, including **lowercasing**, **punctuation removal**, and **tokenization** where required.
-

Exploratory Data Analysis (EDA):

- Computed descriptive statistics to understand distributions of views, likes, and comment counts.
- Created **correlation matrices** to analyze relationships between video metrics.
- Visualized data using **matplotlib** and **seaborn** for clearer insights into trends.

- Videos with higher engagement (likes, comments) tended to have better viewership metrics.
 - Viewer sentiment skewed positive on videos with high interaction, while controversial or lower-engagement videos showed more neutral or negative sentiments.
 - Certain keywords in comments correlated with spikes in engagement, suggesting topics that drive audience interest.
-

5. Tools & Technologies Used

- **Python:** pandas, numpy, matplotlib, seaborn, TextBlob
 - **Jupyter Notebook** for interactive data exploration and visualization
 - **Visualization Techniques:** Heatmaps, Bar Charts, Word Clouds, Correlation Plots
-

6. Conclusion

The project successfully highlighted how data analysis of YouTube metrics and comments can reveal audience behavior patterns and inform content strategy decisions. Future work could expand this analysis with more advanced NLP techniques or machine learning models for predictive insights.