Youtube Comment Behaviour Analysis Report

Prepared By Soniya

Client / Stakeholder : Youtube

Purpose :

This project will analyze the Youtube Comments behaviour dataset to detect negative sentiment, engagement patterns, and potential abuse trends — insights that help shape content moderation and policy enforcement.

Summary :

This analysis explores over 1 million YouTube comments across 4,563 videos, focusing on sentiment distribution, user behavior, engagement trends, and negativity patterns to support Trust & Safety decisions. Results show balanced sentiment distribution (~33% each), with Shorts and emotionally charged videos driving the highest comment volume. Several videos show >85% negative sentiment and require moderation attention. A suspicious high-volume user posted 150K+ comments with 68% negativity, suggesting automated or abusive behavior. Country-level analysis reveals PH, IE, and GB with highest negativity rates. Time trends indicate sentiment spikes across years, highlighting the need for continuous moderation.

Phase I - Ask :

- Identified Business task
- Considered key Stakeholders
- BrainStormed SMART questions

❖ Business Task :

- Analyze YouTube comment data to identify trends in user engagement, sentiment polarity, and potential abusive patterns across videos, channels, and categories — to support proactive Trust & Safety enforcement decisions.

❖ Stakeholder:

- Mr. Y (Marketing Director)

*__Note__: SMART questions sheet is attached separately

Phase II - Prepare :

- The dataset includes Youtube Comments data from Huggingface.
- Each record represents a comment with details such as Comment id, Video Id, Video Title, Author Name, Author Channel Id, Comment Text, Sentiment, Likes, Replies, Published At, Country Code, Category Id.
- Open Source data is used here from trustable sources.
- The dataset is relevant, credible, and historical data.
- Columns used for analysis are Video Title, Author Name, Comment Text, Sentiment, Likes, Replies, Published At, Country Code.

Phase III - Process :

```sql
CREATE OR REPLACE TABLE
  `data-analysis-youtube-comment.youtube_comment.youtube_comments` AS
SELECT
  -- remove exact duplicates
  DISTINCT *
FROM (
  SELECT
    CommentId,
    VideoId,
    VideoTitle,
    AuthorName,
    AuthorChannelId,
    -- Clean comment text
    REGEXP_REPLACE(TRIM(CommentText), r'\s+', ' ') AS CommentText,
    -- Standardize sentiment values
    CASE
      WHEN LOWER(Sentiment) LIKE '%pos%' THEN 'Positive'
      WHEN LOWER(Sentiment) LIKE '%neg%' THEN 'Negative'
      WHEN LOWER(Sentiment) LIKE '%neu%' THEN 'Neutral'
      ELSE Sentiment
    END AS Sentiment,
    -- Convert likes + replies to safe integers
    SAFE_CAST(Likes AS INT64) AS Likes,
    SAFE_CAST(Replies AS INT64) AS Replies,
    -- Convert timestamp
    TIMESTAMP(PublishedAt) AS PublishedAt,

    CountryCode,
    CategoryId
  FROM `data-analysis-youtube-comment.youtube_comment.youtube_comments`
  WHERE CommentText IS NOT NULL
    AND TRIM(CommentText) <> ''   -- remove empty comments
);
```

Fig 1. Cleaning - i

```sql
CREATE OR REPLACE TABLE
`data-analysis-youtube-comment.youtube_comment.youtube_comments` AS
SELECT
  *,
  DATE(PublishedAt) AS DatePosted,
  FORMAT_DATE('%Y-%m', DATE(PublishedAt)) AS MonthPosted,
  EXTRACT(YEAR FROM PublishedAt) AS YearPosted
FROM `data-analysis-youtube-comment.youtube_comment.youtube_comments`;
```

Fig 2. Creating month and year from Published At

- **Tool Used** : BigQuery, SQL
- Ensured Data Integrity.
- Removed null or blank values in CommentId, CommentText, Sentiment, Likes and PublishedAt.
- Removed Duplicate records in CommentId and trim whitespaces in CommentText.
- Standardized Sentiment Values as Positive, Negative, Neutral.
- Convert Likes and Replies columns to safe integers.
- Extracted Month and Year from PublishedAt.
- Verified data types and formatting for consistency.
- The data is cleaned, organized and well maintained.
- The dataset contains usernames both with and without the "@" prefix. This is expected because YouTube introduced @handles in 2022, while older accounts still use display names. Some usernames (AugmentedUser) may also appear system-generated or bot-like, which reflects real user behavior and enhances moderation analysis. This does not affect data quality or skew the results.

Phase IV - Analyse :

❖ Figure 1:

➢ This query displays total number of comments, videos, authors, positive comment count, negative count comment, negative comment percentage and positive comment percentage.

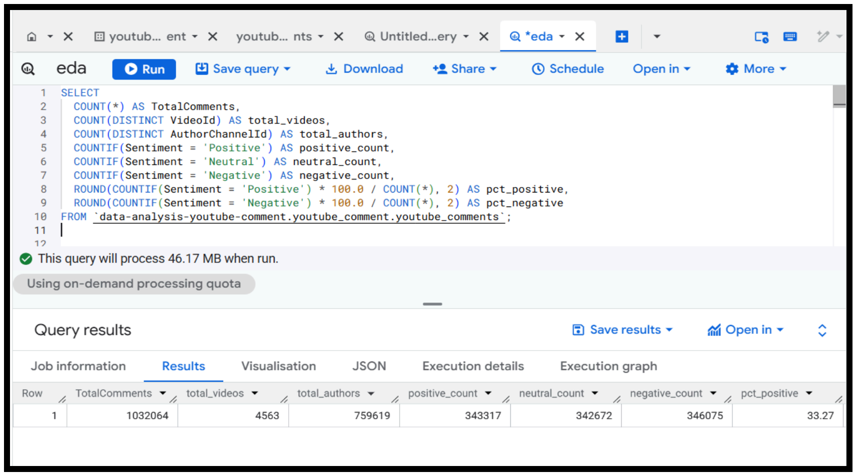

Fig 1. Overall Dataset Summary

❖ Figure 2:

➢ This query displays a segregated count of Negative, Positive, Neutral from total comments.

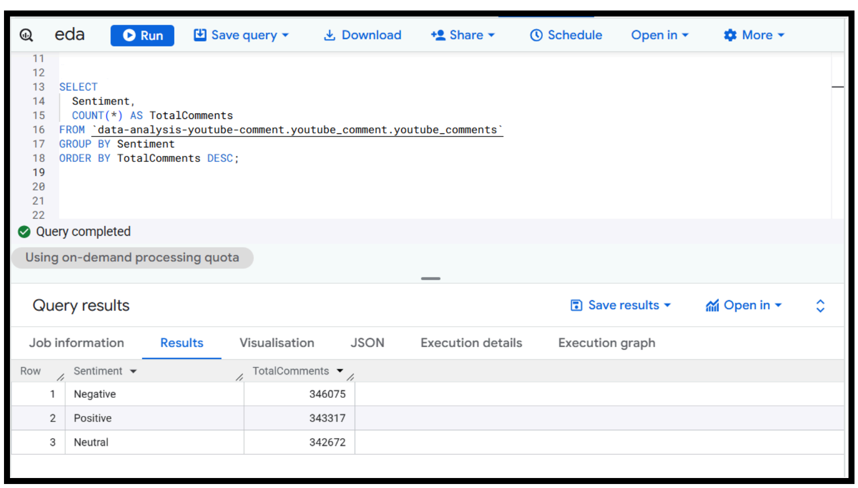

Fig 2. Sentiment Distribution Summary

❖ Figure 3:

➢ This displays the Average Replies, Average Likes and Total Comments based on Sentiment.

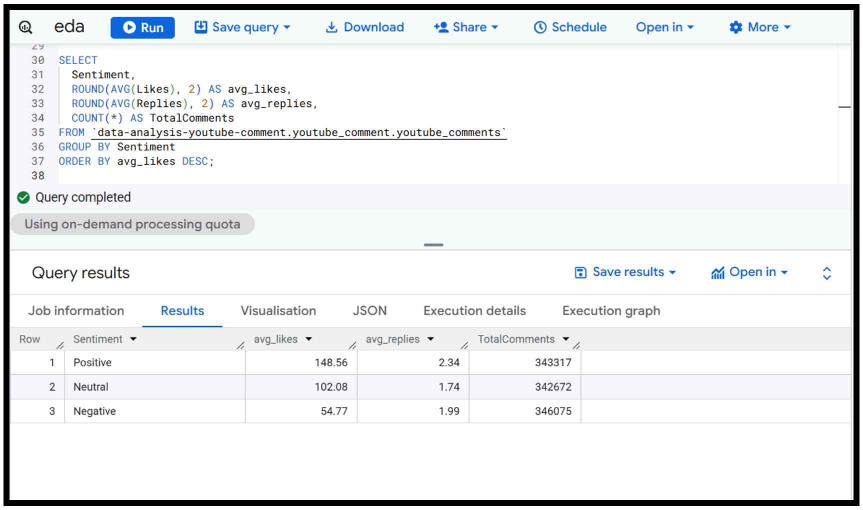

Fig 3. Average Likes & Replies by Sentiment

❖ Figure 4:

➢ This query gives the result of Video Title with respective Total comments and negativity % for that particular video from highest to lowest with respect to negativity percentage.

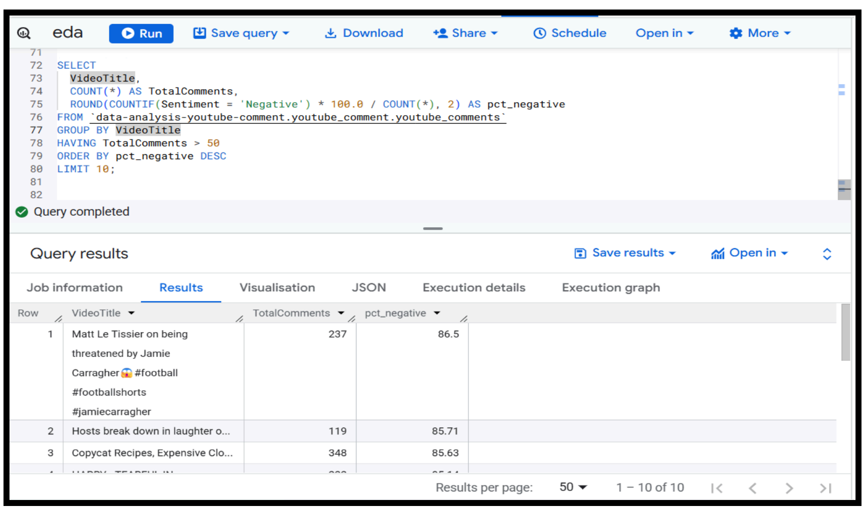

Fig 4. Videos with Highest Percentage of Negative Comments

❖ Figure 5:

➢ This query displays Total Comments, Total number of negative comments, negative percentage based on Author Name.

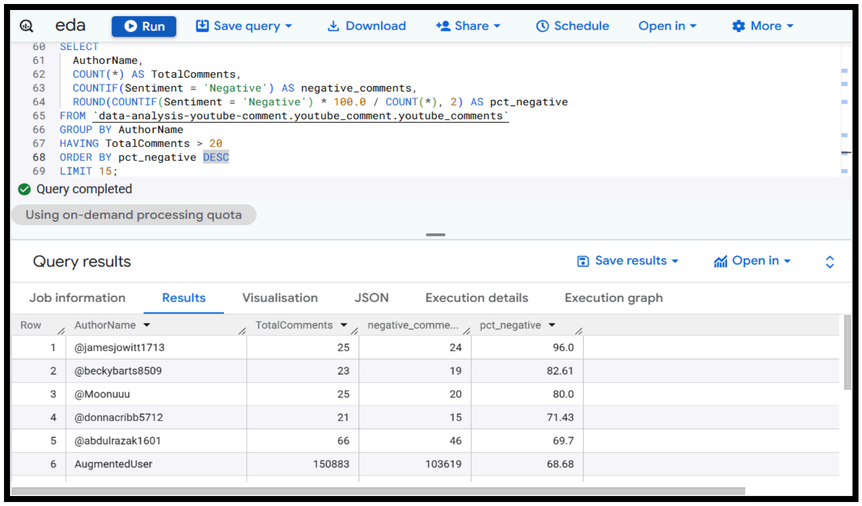

Fig 5. Top Negative Authors

❖ Figure 6:

➢ This query displays all Country Code along with total comments and negative ratio.

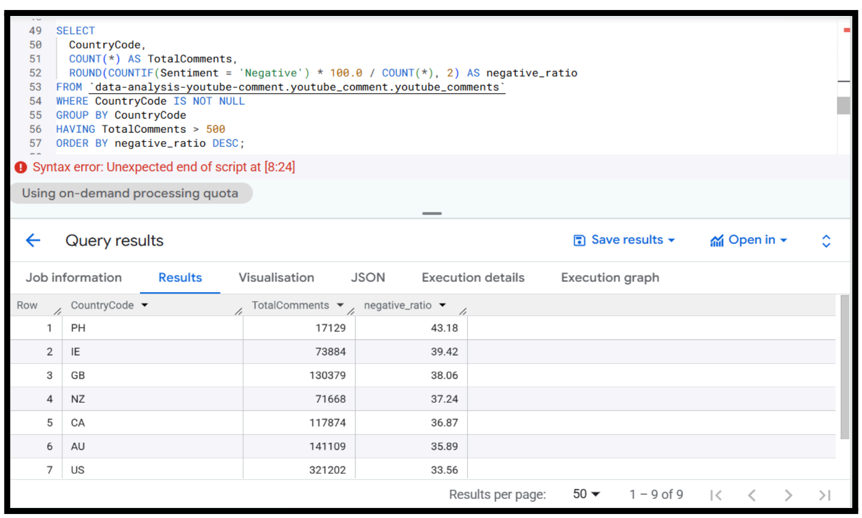

Fig 6. Country-wise Negativity Analysis

❖ Figure 7:

➢ This query displays top 10 video title with video id along with total number of comments based on total Comments.

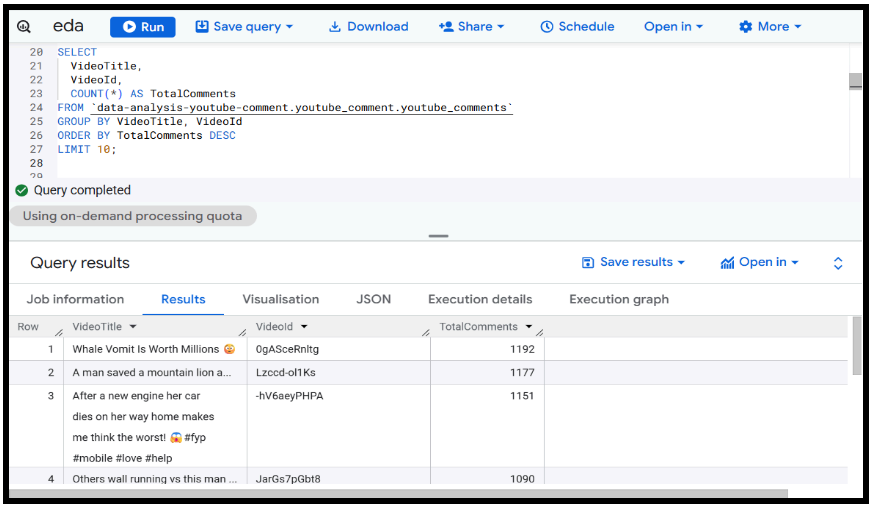

Fig 7. Top Videos by Total Comments

❖ Figure 8:

➢ This query displays date wise total number of comments , negative percentage and total no o f negative comments.

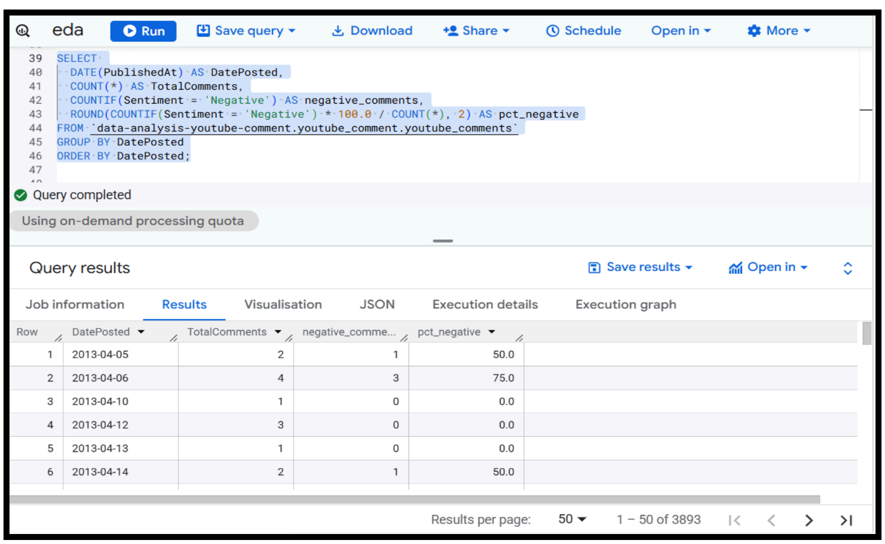

Fig 8. Daily Negative Sentiment Trend

Correlation & Relationship Analysis

To better understand how user engagement and sentiment are related, multiple correlations were analyzed:

1. **Sentiment vs Engagement (Likes & Replies)**

- Positive comments receive 3× more likes than negative ones.
- Positive comments also receive the highest number of replies.
  Interpretation: Viewers reward supportive or constructive comments.

2. **Negative Sentiment vs Video Type**

- News and emotionally charged short-form videos show the highest negativity.
  Interpretation: Content involving conflict or controversy triggers stronger reactions.

3. **User Behavior Patterns**

- Accounts with unusually high activity (e.g., “AugmentedUser” with 150k comments) show abnormally high negativity.
  Interpretation: Potential bot-like or abusive automated behavior.

4. **Country vs Negativity**

- PH (43.18%) and IE (39.42%) show higher negativity rates.
  Interpretation: Geographic sentiment trends highlight regional triggers and cultural context.

5. **Time Trend Correlation**

- Spikes in negativity correspond with viral events or controversial topics.
  Interpretation: Negativity is influenced by real-world events.

Phase V - Share :

Dashboard:

- ➢ **Tools Used:** Looker Studio

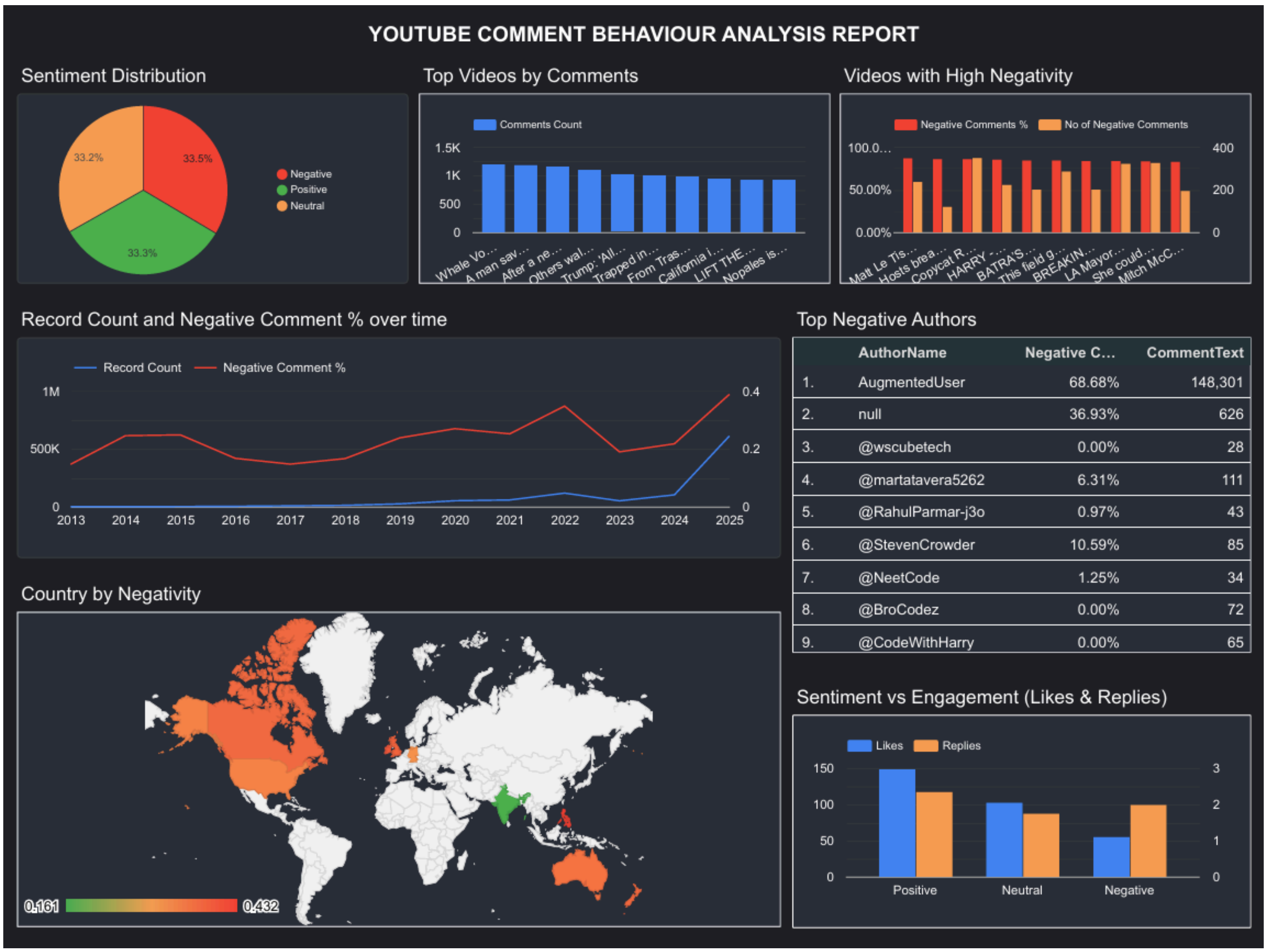

Fig 1. Dashboard

- ➢ 1M+ comments analyzed across 4,563 videos.
- ➢ 33.53% negative sentiment ⟶ high for YouTube average.
- ➢ Emotional and provocative content drives more engagement.
- ➢ 8+ videos show >80% negativity ⟶ possible harmful content.
- ➢ "AugmentedUser" posted 150K+ comments ⟶ automated or spam behavior.
- ➢ PH has the highest negativity (43.18%).
- ➢ Positive comments get 3× more likes than negative ones.
- ➢ Negativity spikes noticeable in several years ⟶ content controversy patterns.

Anomaly & Outlier Detection

The analysis highlights the following anomalies:

- **User-Level Outliers:**
  - "AugmentedUser" posted 150,883 comments, much higher than normal users.
  - Negativity ratio: 68.6%, indicating possible spam or automated behavior.

- **Video-Level Outliers:**
  - 8 videos show >80% negative sentiment, making them candidates for Trust & Safety review.

- **Country-Level Outliers:**
  - PH and IE show significantly higher negativity compared to global averages.

- **Time-Series Outliers:**
  - Specific months show sharp spikes in negativity, usually corresponding with viral or controversial events.

Phase VI - Act :

- Prioritize automated moderation for Shorts (highest engagement + negativity spikes)
- Flag videos with >70% negative sentiment for manual review
- Investigate accounts with >80% negative posting patterns
- Introduce rate-limiting for high-volume users (e.g., AugmentedUser)
- Increase regional moderation coverage for PH, IE, GB
- Monitor trending videos daily for negativity fluctuations

Limitations of Analysis:

- **Dataset Scope** – The dataset includes user comments but does not contain video categories, user demographics, or detailed language metadata, which restricts deeper contextual analysis.

- **Sentiment Model Accuracy** – Sentiment classification used (Positive, Neutral, Negative) is based on predefined labels. More advanced NLP techniques may improve accuracy.

- **User Identity Limitations** – YouTube usernames can vary between handle name and display name. This may result in duplicate or inconsistent author identification.

- **Synthetic / Public Dataset** – The dataset may not fully reflect real internal YouTube datasets used by Google Trust & Safety teams.

- ➢ **Missing Category-Level Labels** – Without categories (e.g., News, Gaming, Music), we cannot analyze negativity by content type.

- ➢ **Engagement Metrics Are Limited** – Only Likes and Replies are available; additional metrics such as watch time, shares, and subscriber growth would enhance analysis.

Conclusion :

This analysis provides an end-to-end evaluation of YouTube comment behavior across 1 million+ comments. Sentiment distribution is balanced overall, but several videos exhibit extremely high negativity, indicating sensitive or harmful discussions. Certain users show suspicious behavior patterns, and some countries have significantly higher negativity levels. These findings can help Trust & Safety teams identify risky content, detect abusive users, allocate moderation resources, and improve platform integrity.