Data-Driven Insights from YouTube Trending Videos: A Cross-Category Analysis

1. Objective

The goal of this project is to uncover **patterns in trending YouTube videos** by analyzing datasets across regions. We focused on:

- Most popular categories and genres

- Sentiment analysis of titles

- Engagement trends (likes vs dislikes)

- Time-series analysis of views

- Regional comparisons (optional if multiple datasets available)

2. Tools Used

- **Python (Google Colab)** → Data cleaning, preprocessing, sentiment analysis, visualizations

- **Power BI** → Interactive dashboards and storytelling

- **Libraries** → Pandas, Matplotlib, Seaborn, TextBlob

3. Methodology

1. **Data Collection & Cleaning**

    o Loaded YouTube trending dataset (.csv) into Colab

    o Removed duplicates and missing values

    o Standardized column types (view_count, likes, dislikes)

    o Added category names using categoryId mapping

2. **Sentiment Analysis**

    o Applied **TextBlob** to video titles

    o Classified into *Positive, Neutral, Negative*

3. **Data Export**

    o Cleaned data exported as .csv

    o Loaded into **Power BI** for visualization

4. Key Visualizations

◆ **Bar Chart: Top Categories by Views**

- Music, Entertainment, and Comedy dominate trending charts

- Shows clear preference for entertainment-related content

◆ **Donut Chart: Sentiment Distribution**

- Majority of titles are **Neutral** in sentiment

- Positive sentiments appear in music and lifestyle categories

- Negative sentiment is minimal

◆ **Line Chart: Views Over Time**

- Views increase during weekends and festive seasons

- Some videos sustain trending for longer durations

◆ **Scatter Plot: Likes vs Dislikes (Engagement)**

- Most videos have high likes compared to dislikes

- Controversial videos stand out above diagonal line

- Bigger bubbles (higher views) tend to receive more likes

5. Insights

- **Entertainment & Music** are the most engaging categories globally

- **Neutral sentiment titles** dominate trending videos → suggests titles are descriptive rather than opinionated

- **Engagement (likes/dislikes)** shows audience reactions strongly aligned with category type

- Trending duration is **short-lived** for most videos, except viral hits

6. Conclusion

The analysis shows that YouTube trending videos are largely driven by **entertainment value** and **mass appeal**. While content sentiment in titles doesn't strongly affect trending, **views and engagement** are heavily linked to video category and timing.

7. Deliverables

- ✅ Cleaned dataset (.csv)

- ✅ Python scripts (Colab notebook)

- ✅ Power BI Dashboard (interactive)

- ✅ Final Report (this document)