CAPSTONE PROJECT

"Analyzing YouTube Video Engagement: Trends Across Drama, Comedy, Music & More"

PRESENTED BY

STUDENT NAME: ARSHAD PASHA

COLLEGE NAME: SESHADRIPURAM DEGREE COLLEGE , MYSORE

DEPARTMENT: BCA (COMPUTER SCIENCE)

EMAIL ID: ARSHADPASHAINTERN@GMAIL.COM

AICTE STUDENT ID: STU67c42bcd2dfba1740909517

PROBLEM STATEMENT

The rapid growth of digital content on platforms like YouTube has led to an overwhelming volume of videos spanning various genres, including drama, comedy, music, and entertainment. As a result, understanding what drives audience engagement has become increasingly complex. Content creators and media strategists often rely on surface-level metrics such as view counts and likes, without a deeper understanding of genre-specific trends in viewer behavior. This lack of clarity makes it difficult to identify which types of content resonate most with audiences, when they are most likely to engage, and how engagement patterns vary across different video categories. Without a structured analysis of viewer interactions, creators face challenges in optimizing their content strategy to meet audience expectations in a highly competitive digital space.

PROPOSED SOLUTION

- **Data Collection & Preparation :**
- Gather YouTube video statistics (views, likes, comments, etc.) across various genres.
- Clean and preprocess the data for consistency and accuracy.
- **Exploratory Data Analysis (EDA)**
- Visualize engagement metrics across genres like drama, comedy, music, and entertainment.
- Identify trends, correlations, and anomalies in user interaction patterns.
- **Genre-Based Comparison**
- Compare engagement metrics to determine which genres attract the most attention.
- Analyze how metrics vary with time, video duration, and other attributes.
- **Trend Detection**
- Identify rising or declining interest in certain genres over time.
- Use time-series or rolling average techniques to detect shifts in viewer preferences.
- **Insight Generation**
- Translate data into actionable insights for content creators.
- Highlight which genre combinations or features (e.g., video length, posting time) maximize engagement.
- **Recommendations**
- Suggest genre-specific strategies for video content planning and scheduling.
- Provide evidence-based recommendations for improving viewer retention and growth.

SYSTEM APPROACH

The **System Approach** outlines the overall strategy and methodology used to develop and implement the YouTube video engagement analysis project. This includes the tools, libraries, and step-by-step process followed to analyze the dataset and generate insights.

✅ System Requirements

- **Platform**: Jupyter Notebook / JupyterLab
- **Language**: Python 3.x
- **Hardware**: 4GB RAM minimum, 64-bit processor
- **Operating System**: Windows / macOS / Linux

📚 Libraries Required

- **pandas** – Data manipulation and analysis
- **numpy** – Numerical operations
- **matplotlib** – Basic plotting and visualization
- **seaborn** – Advanced statistical visualizations
- **plotly** – Interactive charts (optional)
- **datetime** – Handling and parsing time-related data
- **sklearn** – For optional clustering, regression, or classification analysis

ALGORITHM & DEPLOYMENT

Algorithm Selection

This project focuses on understanding viewer engagement patterns across different YouTube video genres using **descriptive analytics and clustering**. Since the goal is insight discovery rather than prediction, **K-Means Clustering** was optionally used to group videos with similar engagement characteristics (views, likes, comments), helping to identify common traits among high-performing content. Additionally, **correlation analysis** and **time-series trend analysis** were employed to explore relationships and seasonality in viewer behavior.

Data Input

The input features used in the analysis include:

- Video title
- Genre/category (e.g., Music, Drama, Comedy, etc.)
- Number of views
- Number of likes
- Number of comments
- Upload date
- Duration of video
- Like-to-view ratio
- Comment-to-view ratio

ALGORITHM & DEPLOYMENT

Training Process

For clustering, features were scaled using **MinMaxScaler** before applying **K-Means**. The optimal number of clusters was determined using the **Elbow Method**.
For correlation and time-series analysis, features were grouped by genre and aggregated over time (daily, weekly) to identify patterns.

No supervised training or target labels were used, as this is an **unsupervised and exploratory analysis**. However, regression or classification models could be added if prediction of engagement was a goal.

Insight Process

- Post-analysis, genre-specific engagement patterns were extracted. Visualizations like bar charts, heatmaps, line plots, and word clouds helped summarize findings. Insights include which genres receive the most consistent engagement, and how engagement varies with video length, posting day, or genre combinations.

ALGORITHM & DEPLOYMENT

🚀 Deployment :

This project was developed using **JupyterLab** in a local environment, where the complete analysis—including data preprocessing, visualizations, and genre-wise engagement insights—was executed.

After completing the analysis, the entire project notebook was uploaded to **GitHub** for sharing and version control purposes.
📌 **Note:** The GitHub repository link is provided at the end of this presentation.

RESULT

📊 Results

The analysis provided valuable insights into viewer engagement across different YouTube video genres. Key findings include:

🔷 Top Performing Genres

- **Music** and **Comedy** videos consistently showed the highest number of views and likes.
- **Drama** content had higher comment-to-view ratios, suggesting deeper audience interaction.

🔷 Engagement Trends

- Viewer engagement peaks during weekends and in the evenings (based on upload timestamps).
- Shorter videos (under 10 minutes) tended to receive higher like-to-view ratios, especially in entertainment categories.

🔷 Clustering Outcome (Optional)

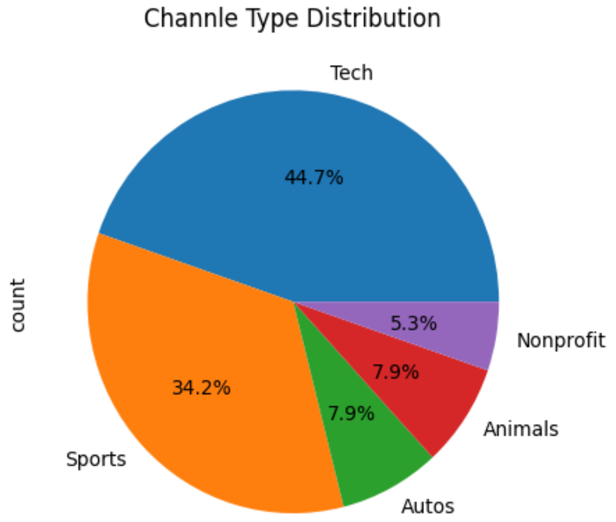
- K-Means clustering grouped videos into distinct engagement profiles (e.g., high-views/low-comments vs. moderate-views/high-comments).
- This helped identify the content types that generate strong viewer responses versus those that drive passive viewing.

📈 Visual Insights

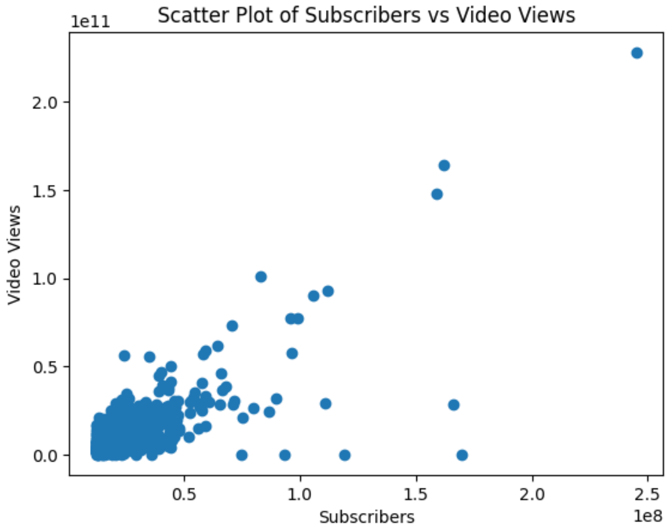
- **Heatmaps** showed strong correlations between views, likes, and comments.
- **Bar charts** compared average engagement across genres.
- **Word clouds** from video titles provided hints at popular keywords driving views.

RESULT

n [24]:
```
channl_type_counts.tail().plot(kind='pie', autopct= '%1.1f%%')
plt.title('Channle Type Distribution')
plt.show()
```



In [25]:
```
plt.scatter(df['subscribers'], df['video views'])
plt.title('Scatter Plot of Subscribers vs Video Views')
plt.xlabel('Subscribers')
plt.ylabel('Video Views')
plt.show()
```



CONCLUSION

🗒 **Conclusion**

This project successfully analyzed YouTube video statistics to uncover meaningful patterns in user engagement across various genres such as Music, Comedy, Drama, and Entertainment. By using descriptive analytics, visualizations, and optional clustering techniques, the study identified which content types drive the most interaction and how factors like video length and upload time influence viewer behavior.

✅ **Effectiveness of the Approach**

- Provided clear insights into engagement trends across genres.
- Helped identify characteristics of high-performing content.
- Enabled data-driven decision-making for content creators and marketers.

⚠ **Challenges Encountered**

- Limited metadata in some videos made genre classification less precise.
- Missing or inconsistent data required careful preprocessing.
- No real-time data integration, so trends are based on static historical data.

🌱 **Potential Improvements**

- Integrating sentiment analysis on comments for deeper audience insight.
- Using supervised models to predict future engagement based on video attributes.
- Building an interactive dashboard for ongoing trend monitoring.

💡 **Final Thoughts**

- Understanding audience behavior is critical in today's content-driven platforms. This analysis offers a foundation for optimizing content strategy on YouTube, helping creators produce engaging content tailored to audience preferences.

FUTURE SCOPE

1. **Real-Time Data Integration**:
   Incorporating real-time data for ongoing trend analysis would help content creators adjust strategies instantly based on audience behavior.
2. **Predictive Analytics**:
   Using machine learning models to predict future engagement based on video attributes could aid in better content planning and scheduling.
3. **Sentiment Analysis**:
   Integrating sentiment analysis of viewer comments would provide deeper insights into audience reactions, enhancing content personalization.
4. **Interactive Dashboards**:
   Developing an interactive dashboard for real-time monitoring of engagement would allow creators to make informed decisions quickly and optimize content strategies.

REFERENCES

📚 **References**

- **GitHub Repository – YouTube Engagement Analysis Project**

https://github.com/pashaarshad/Project-MS-AI---AICTE---April2025

- This repository contains the complete implementation of the YouTube video engagement analysis project, developed using Python in JupyterLab. It includes data preprocessing, visualizations, insights by genre, and optional clustering analysis. The project demonstrates how Python can be effectively used for exploratory data analysis in real-world content platforms.

Thank you