# Exploring Social Media Analytics with Data Science

Analyzing Video Trends and User Engagement in the US  
**Prepared by: Nikitha Gutha**

## 1. Importing Data

The dataset used in this project is a US video dataset, analyzed using Excel.

## 2. Data Cleaning Process

During the analysis, the following steps were performed to clean and prepare the data for insights:

A screenshot of a computer

Description automatically generated

Source Data Loading: Imported raw data for analysis.

Navigation: Selected relevant datasets and tables for analysis.

Promoted Headers: Converted data rows into column headers.

Removed Columns: Eliminated unnecessary columns to focus on relevant data.

Removed Errors: Excluded rows containing errors to ensure clean data.

Changed Data Types: Standardized data types for consistency and accuracy.

Grouped Rows: Consolidated data to enable grouped analysis.

## 3. Video Trends Analysis

Top 10 Most Viewed Videos  
Identified and ranked the Top 10 Most Viewed Videos in the US.



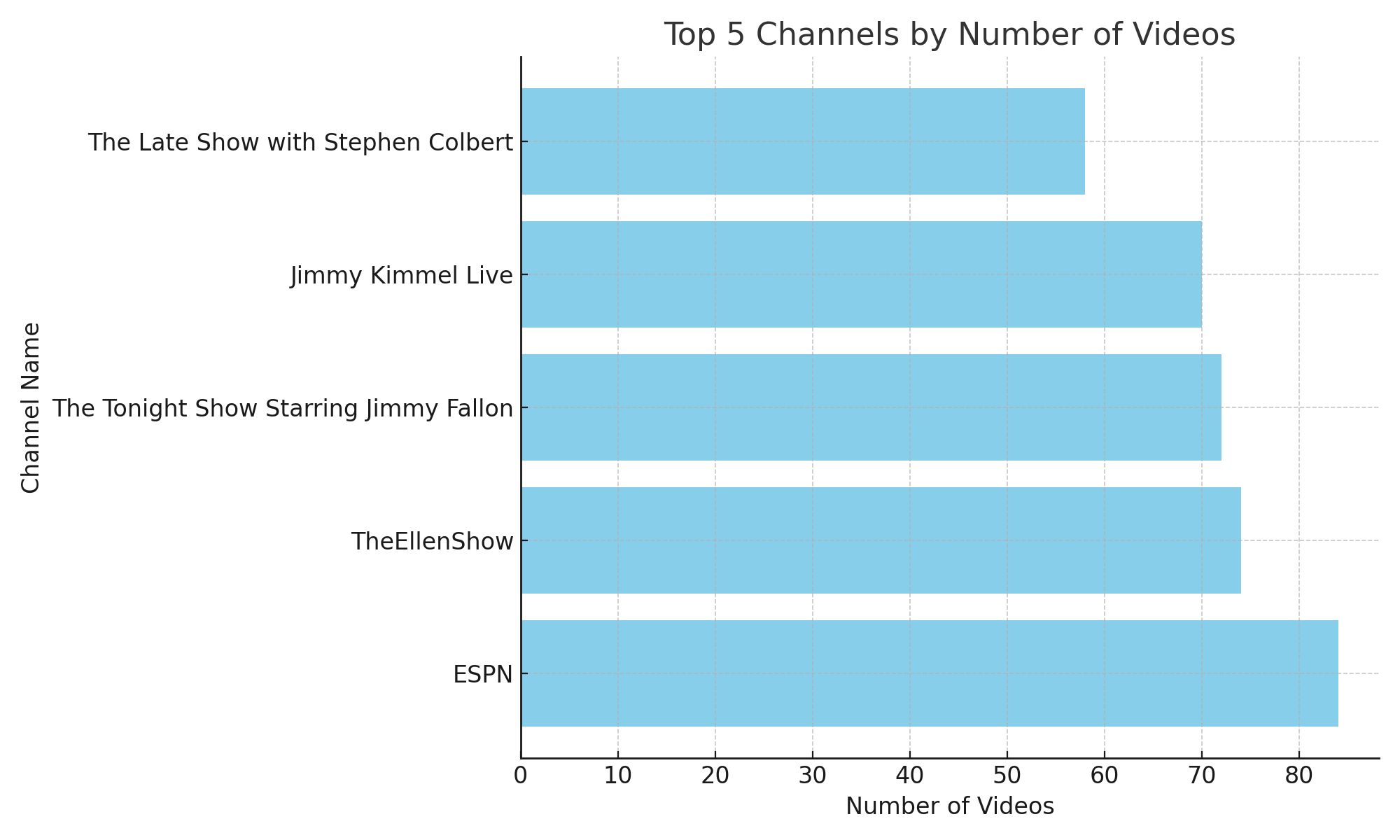
Top 5 Channels by Number of Videos  
Channels with the most video content are listed below:

ESPN with 84 videos, TheEllenShow with 74 videos, The Tonight Show Starring Jimmy Fallon  
with 72 videos , Jimmy Kimmel Live with 70 videos, The Late Show with Stephen  
Colbert with 58 videos.

A screenshot of a computer

Description automatically generated

The chart below illustrates the distribution of videos among the top 5 channels:



## 4. User Engagement Analysis

Average Like-to-Dislike Ratio

A screenshot of a computer

Description automatically generated  
The calculated average like-to-dislike ratio is 43.12 (excluding videos with missing or zero dislikes)

A close up of a white sheet

Description automatically generated with medium confidence

Videos with Extreme Like-to-Dislike Ratios  
• Highest Ratio: 1473.478

A screenshot of a computer

Description automatically generated  
• Lowest Ratios: 0 and 0.041762.

A screenshot of a computer

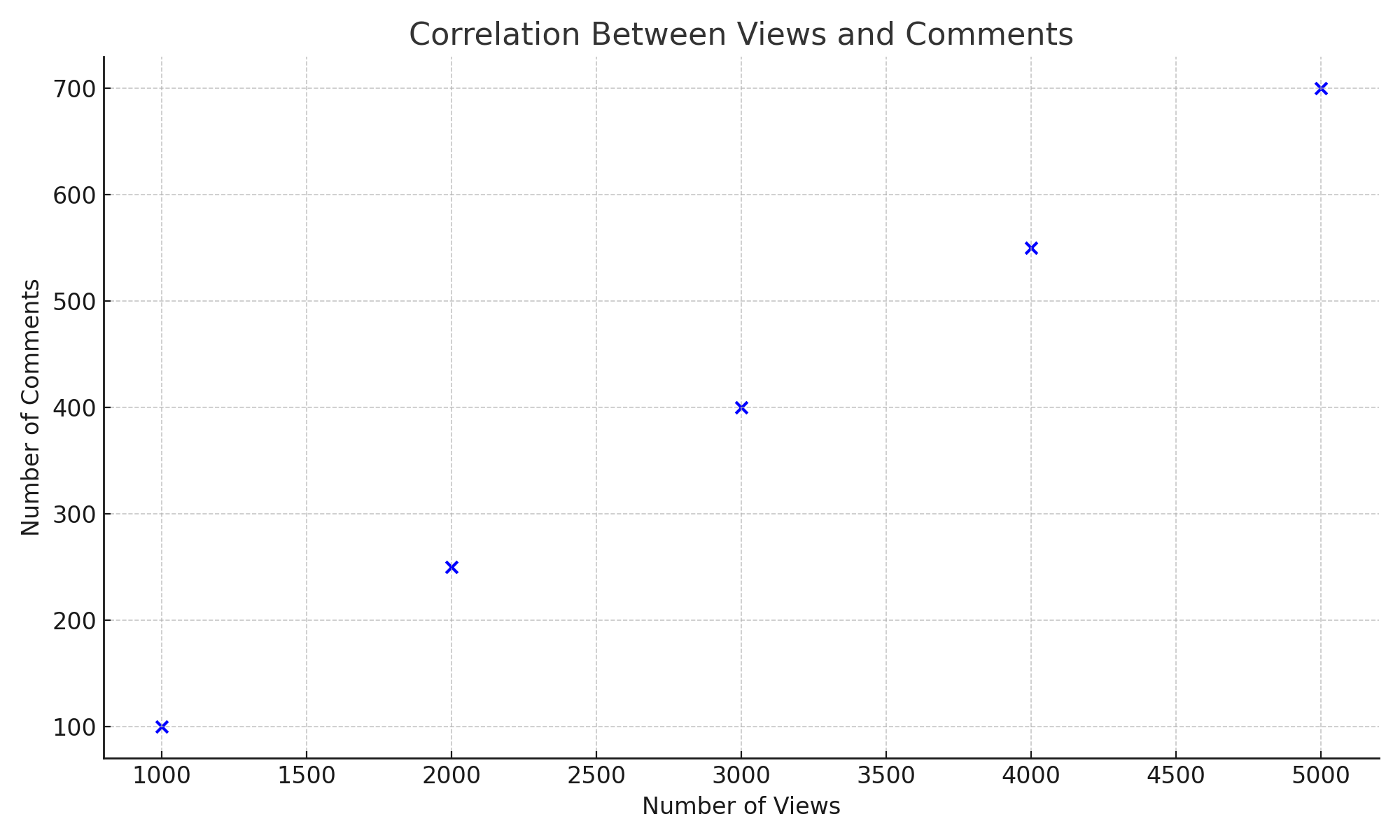
Description automatically generated

Correlation Between Views and Comments  
A positive correlation of 0.692 was found between views and comments, indicating that higher views are associated with more comments.

A screenshot of a computer

Description automatically generated

The scatter plot below visualizes the correlation between views and comments:



## 5. Key Takeaways

The analysis revealed the following insights:  
• ESPN and major talk show channels dominate content creation and viewership.  
• The overall sentiment, as measured by the like-to-dislike ratio, is positive.  
• The correlation between views and comments suggests that popular videos engage audiences more effectively.

## 6.Tools Used

Excel for statistical computations and data visualization

## 7.Conclusion

This social media analytics project provided a comprehensive exploration of the US video dataset, uncovering valuable insights into video trends and user engagement. By employing statistical tools and Excel, I successfully analyzed key metrics such as the most viewed videos, dominant channels, and user engagement patterns, including the like-to-dislike ratios and the correlation between views and comments.

The findings revealed that major entertainment channels, such as ESPN and popular talk shows, lead in video content creation and viewership, showcasing their influence in the social media landscape. Additionally, the positive correlation between views and comments underscores the importance of fostering engagement for content success. The like-to-dislike ratio analysis highlighted user sentiment, which can guide content creators in tailoring their strategies to resonate with audiences.

Through this analysis, I not only derived actionable insights for improving content performance and audience engagement but also enhanced my expertise in data preparation, visualization, and interpretation. The structured data cleaning process, combined with advanced analysis, emphasizes the critical role of clean and accurate data in generating meaningful conclusions.

In summary, this project reinforces the potential of data-driven strategies in optimizing social media performance, providing a robust foundation for decision-making in the dynamic digital content space.