ANP-D0449

DATA ANALYSIS USING PYTHON

YOUTUBE VIDEO PERFORMANCE METRICS: VIEWS, LIKES, AND SHARES ANALYSIS

Submitted by:

SACHIN E

ABSTRACT:

With the exponential growth of digital content, YouTube has become one of the most influential platforms for video sharing and audience engagement. Analyzing video performance metrics—such as views, likes, and shares—plays a crucial role in understanding user behavior and optimizing content strategies. This study aims to explore the interrelationship between these key engagement metrics and their impact on video visibility and audience interaction. Views indicate the reach of a video, likes reflect audience approval, and shares contribute to broader content dissemination. By employing data analytics and statistical methods, this research examines how these metrics influence video ranking, user engagement, and overall content success. Additionally, it investigates external factors such as video duration, content type, posting frequency, and social media promotions that affect these performance indicators. The findings of this study will provide valuable insights for content creators, digital marketers, and media analysts to enhance their video strategies, improve audience retention, and maximize content virality on YouTube.

PROBLEM STATEMENT

- **Unclear Video Performance Metrics** – Content creators struggle to identify the key factors contributing to video success.

- **Missing or Incomplete Data** – Some video attributes, such as tags and engagement rates, may be missing, leading to inaccurate analysis.

- **Difficulty in Summarizing Engagement** – Understanding average likes, shares, and views across different categories is challenging.

- **Limited Insights into Performance Metrics** – The lack of detailed ratio analysis (e.g., likes-to-views) makes it hard to measure audience interaction.

- **Lack of Trend Visualization** – Without clear graphical representation, it is difficult to track video popularity over time.

- **Content Strategy Optimization Challenges** – Creators need actionable insights to improve video engagement and reach but lack structured analysis.

SOLUTION APPROACH

1. **Data Collection & Preprocessing**

   - Load the dataset containing video performance metrics (views, likes, shares, and tags).

   - Handle missing values by filling in relevant placeholders or computing estimates.

2. **Statistical Summarization**

   - Compute descriptive statistics such as average likes per video.

   - Aggregate data to analyze total views per category.

3. **Performance Metric Analysis**

   - Calculate engagement rate based on likes, shares, and views.

   - Compute the ratio of likes to views to determine user engagement levels.

4. **Data Visualization**

    o   Generate separate line plots for views, likes, and shares over time.

    o   Identify trends in video popularity and engagement patterns.

5. **Insights & Optimization**

    o   Analyze which categories receive the highest engagement.

    o   Use findings to guide content strategies for better audience interaction.

IMPLEMENTATION:

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

Load dataset

df = pd.read_csv("youtube_video_metrics_large.csv")

Handling missing values

df.fillna({"Tags": "No Tags"}, inplace=True)

def engagement_rate(row):

   return (row["Likes"] + row["Shares"]) / row["Views"] if row["Views"] > 0 else 0

df["Engagement_Rate"] = df.apply(engagement_rate, axis=1)

Statistical summarization

average_likes = df["Likes"].mean()

```python
print(f"Average Likes per Video: {average_likes:.2f}")


# Data Aggregation: Views per category

views_per_category = df.groupby("Category")["Views"].sum()

print("\nTotal Views per Category:\n", views_per_category)


# Performance Metric Analysis: Likes to Views ratio

df["Likes_to_Views"] = df["Likes"] / df["Views"]


# Individual Graphs for Views, Likes, and Shares

plt.figure(figsize=(12, 4))

sns.lineplot(x=df.index, y=df["Views"], marker="o", color="blue")

plt.xlabel("Video Index (Ordered by Time)")

plt.ylabel("Views")

plt.title("YouTube Video Views Over Time")

plt.show()


plt.figure(figsize=(12, 4))

sns.lineplot(x=df.index, y=df["Likes"], marker="o", color="green")

plt.xlabel("Video Index (Ordered by Time)")

plt.ylabel("Likes")

plt.title("YouTube Video Likes Over Time")

plt.show()
```

```
plt.figure(figsize=(12, 4))

sns.lineplot(x=df.index, y=df["Shares"], marker="o", color="red")

plt.xlabel("Video Index (Ordered by Time)")

plt.ylabel("Shares")

plt.title("YouTube Video Shares Over Time")

plt.show()
```
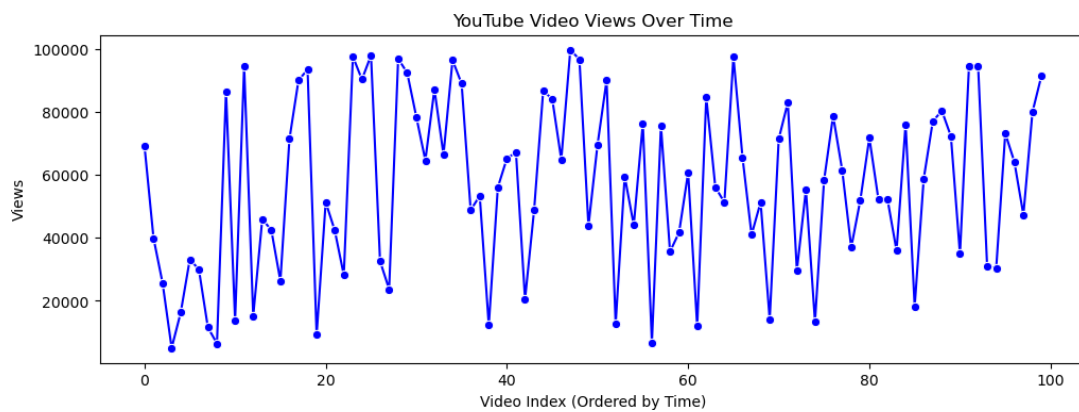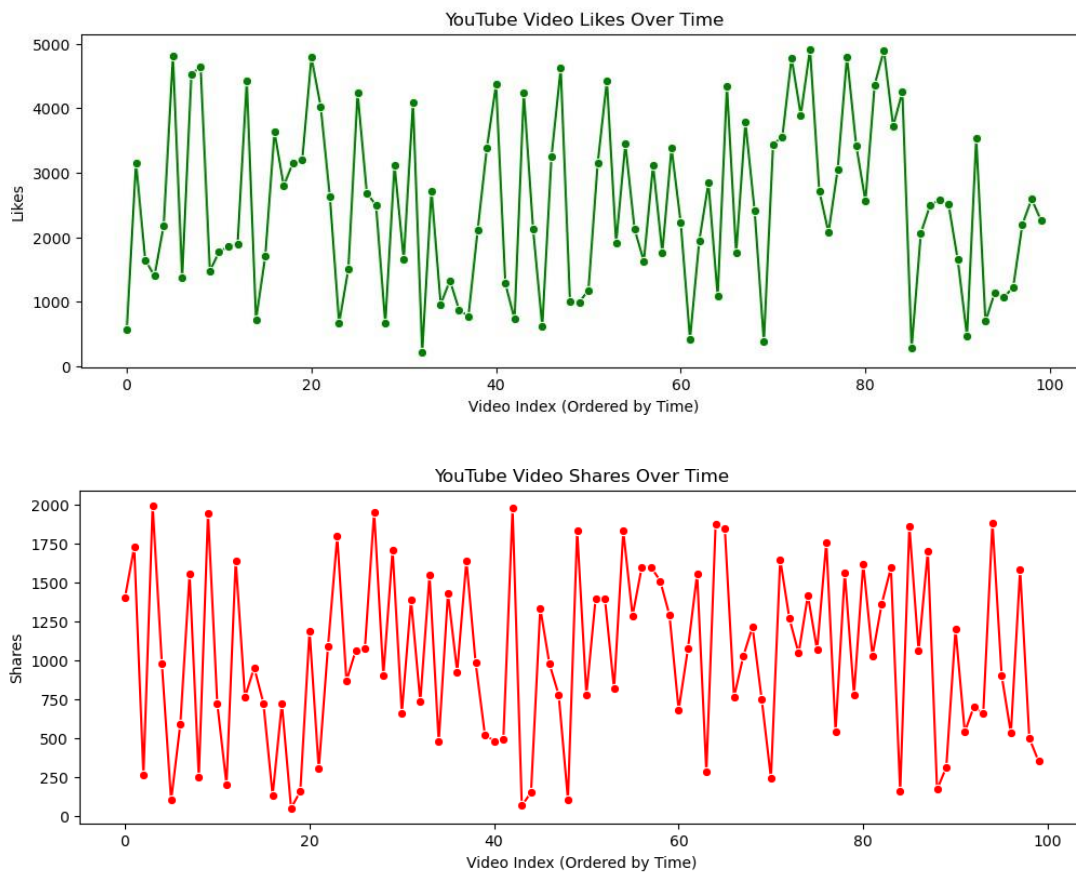
OUTPUT VISUALIZATION:

```
Average Likes per Video: 2495.69

Total Views per Category:
 Category
Education        1173817
Entertainment     498597
Gaming           1441849
Lifestyle        1440303
Tech             1082972
Name: Views, dtype: int64
```

YouTube Video Likes Over Time

YouTube Video Shares Over Time

CONCLUSION:

This project provides a structured approach to analyzing YouTube video performance using key metrics such as views, likes, and shares. By handling missing data, summarizing engagement statistics, and visualizing trends, it helps content creators and marketers understand audience engagement patterns and optimize content strategies. Performance metric analysis, including the likes-to-views ratio and engagement rate, offers valuable insights into video success factors. The trend visualizations further aid in recognizing audience preferences over time, enabling data-driven decision-making to maximize video reach and impact. Ultimately, this analysis supports improved content planning, audience targeting, and strategic growth in digital marketing.