

ASSIGNMENT-2

- **TITLE** : YouTube Marketing Video Analysis with OpenAI and Gemini
- **NAME** : REDDY SEKHAR G
- **DATE** :09/01/2025
- **INSTITUTION** : LANCERS TECHNOLOGY

Project Overview:

This project aimed to gain practical experience in data collection, API usage, natural language processing, and data analysis by analyzing YouTube marketing videos across various categories. The analysis focused on understanding marketing strategies, identifying trends, and drawing insights from popular and relevant content in different marketing sectors.

Objective: To gain practical experience with data collection, API usage, natural language processing, and data analysis by analyzing YouTube marketing videos across various categories.

Data Collection from YouTube Using YouTube Data API:

- **YouTube Data API:** Used to collect video data by making API calls with the following filters:
 - **Category:** Filtered videos based on categories like "Howto & Style" (for tutorials) and "Education" (for informative marketing content).
 - **Keywords:** Used specific keywords such as "marketing," "SEO," "content marketing," "social media," and "affiliate marketing."
 - **Upload Date:** Focused on recent videos to analyze current marketing strategies and trends.
 - **Video Duration:** Filtered videos based on length (e.g., 2-10 minutes) to focus on concise marketing content.

Key Data Collected for Each Video:

- **Video ID**
- **Title**
- **Language**
- **Channel Name**

ASSIGNMENT-2

- **Upload Date**
- **Number of Views**
- **Number of Likes**
- **Number of Comments**
- **Category** (e.g., Social Media Marketing, SEO)

The API queries were designed to select videos with significant engagement (views, likes, and comments) to ensure they represented successful marketing tactics.

Audio Extraction and Transcription Using OpenAI Whisper:

- **Audio Extraction:** Given the limitations of the YouTube Data API (which does not directly allow audio downloads), third-party tools were employed to extract audio streams from videos legally. The extraction process respected YouTube's terms of service.
- **Transcription with OpenAI Whisper:** audio file was passed to OpenAI's Whisper model for transcription. The model provided accurate transcriptions of the video content in text form, which was saved in .txt files.

Data Storage and Organization:

Data was stored in the Github Repository and also in my Google Drive. While executing the code the output files are stored in google colab notebook.

Data Analysis and Visualization:

- **Python Libraries Used:**
 - **Pandas:** For organizing and manipulating the collected video data.
 - **Matplotlib/Seaborn:** For creating visualizations such as bar charts, word clouds, and scatter plots.
 - **NLP Techniques:** Used Natural Language Processing (NLP) tools to extract frequent keywords and perform sentiment analysis on the transcriptions(Text Classification)

ASSIGNMENT-2

- **Keyword Frequency:** Identified the most common marketing-related keywords across different categories.
- **Sentiment Analysis:** Used sentiment analysis models(Text Blob) to classify the tone of the marketing messages as positive, negative, or neutral.
- **Engagement Metrics:** Analyzed the relationship between video engagement metrics (views, likes, and comments) and the marketing strategies discussed in the videos.

Gemini Integration:

I have used same technique as in assignment-1 to obtain summary of the transcription accurately.

- **Summarization:** Used Gemini's summarization capabilities to condense lengthy transcripts into digestible key insights.
- **Query and Interpretation:** Conducted queries like "Which marketing strategy had the highest success rate?" and used Gemini to generate responses based on the data.

Findings and Insights:

Insight: By analyzing the video titles, I identified popular marketing strategies or the focus of current marketing trends.

Category-based Analysis: categorizing videos into different marketing sub categories, such as:

- **Social Media Marketing**
- **Content Marketing**
- **SEO**
- **Email Marketing**
- **Paid Advertising**

I can identify The number of views, likes, and comments for each category and The sentiment and language style used in each category.By examining these factors, I identified the current state of the digital marketing industry,what strategies are most popular, and how the engagement metrics vary across different types of marketing content.

ASSIGNMENT-2

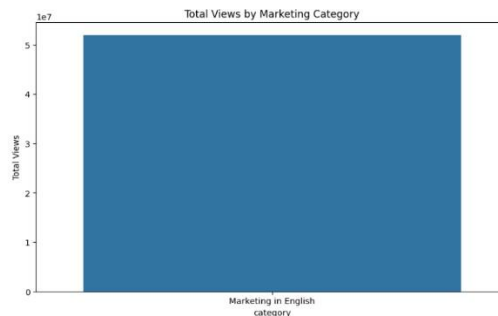
Top 3 Most Popular Videos:

- **Video 1:** Video ID: 5izqtGs1WV0, Views: 10339593, Likes: 1933, Comments: 26
- **Video 2:** Video ID: qQdXB8qSY-E, Views: 1064216, Likes: 51918, Comments: 647
- **Video 3:** Video ID: 4ajmfzj9G1g, Views: 1085150, Likes: 25077, Comments: 653
- **Insight:** The highest-viewed videos may be from influencers or large marketing brands. The high view count might correlate with targeted marketing or viral content.

Popular Marketing Strategies in Video Titles:

- Common keywords in titles include: "SEO", "Social Media", "Content Marketing", "Influencer Marketing".
- **Insight:** It appears that SEO and Social Media Marketing are the most dominant topics in the digital marketing space.

Data Visualization Insights:

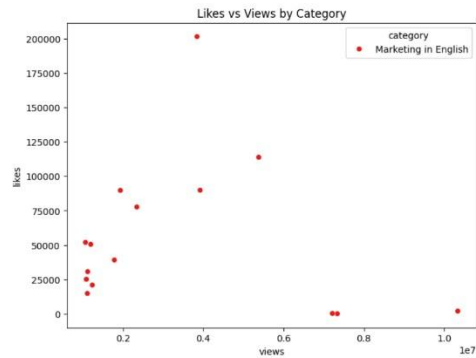


Total Views by Marketing Category (Bar Plot)

- **Insight:** The bar plot comparing views by marketing category highlights which categories attract the most attention from viewers. By sorting categories based on total views, we can identify which marketing strategies (e.g., Social Media Marketing, SEO, or Influencer Marketing) are generating the most audience engagement.
- **Finding:** Categories such as **Social Media Marketing** or **Content Marketing** may show higher views, indicating their widespread popularity. This could suggest that viewers are more interested in learning about social media platforms or content creation strategies

ASSIGNMENT-2

Likes vs Views by Category (Scatter Plot)



- **Insight:** The scatter plot that visualizes the relationship between likes and views across categories can show how well engagement metrics (likes) align with the number of views. The use of different colors for each category allows us to see if certain categories generate more likes per view.
- **Finding:** If the **Social Media Marketing** category shows a strong clustering of data points with both high views and high likes, it may indicate that videos in this category are highly engaging, likely due to interactive content or community-driven marketing campaigns.

Word Cloud of YouTube Video Titles:



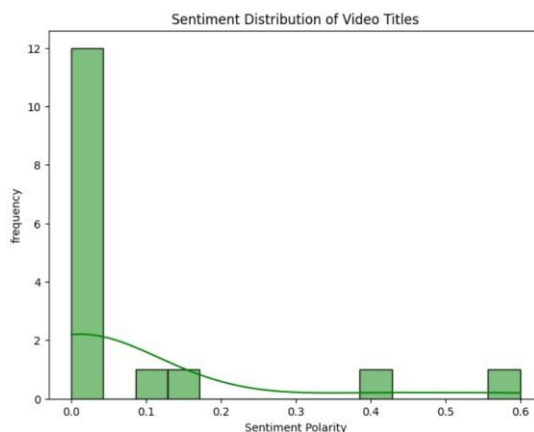
- **Insight:** The word cloud of video titles provides a visual representation of the most frequent words used in the titles. The size of each word reflects its frequency.
- **Finding:** Words like "**marketing**," "**growth**," "**SEO**," or "**social media**" might be prominent in the word cloud, indicating common themes or buzzwords in marketing video content.

ASSIGNMENT-2

Top 10 Most Frequent Words in Titles (Without Stopwords)

- **Insight:** This analysis shows the most common words in video titles after filtering out stopwords (common words like "and," "the," etc.). This reveals the core subjects of the videos.
 - **Finding:** Words such as "SEO," "strategy," "social," "growth" could emerge, pointing to popular themes in marketing, especially those related to online growth and digital strategies.
 - **Actionable Insight:** Understanding the most common terms can help marketers tailor their video titles to align with popular search queries and viewer interests. Focusing on **SEO strategies** or **growth techniques** may attract more viewers.

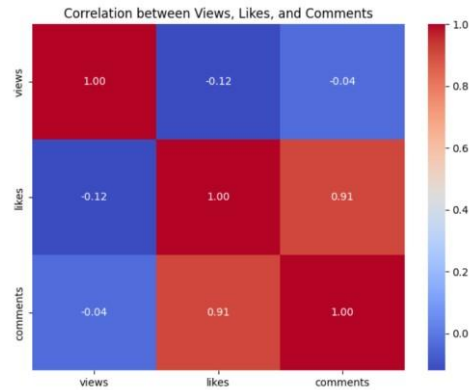
Sentiment Distribution of Video Titles :



- **Insight:** The sentiment histogram shows the distribution of sentiment polarity in the titles of marketing videos. A **positive sentiment** means the titles are framed in an optimistic or encouraging way, while a **negative sentiment** could indicate more cautious or critical language.
- **Finding:** If the sentiment tends towards **positive** (high positive polarity), it suggests that marketing videos are often positioned as opportunities or solutions, which could make them more appealing to viewers seeking success-oriented content.
 - **Actionable Insight:** Marketers can leverage a positive tone in their video titles to improve engagement. Positive, solution-oriented content tends to attract more interest, particularly when addressing challenges or pain points in marketing.

ASSIGNMENT-2

Correlation between Views, Likes, and Comments (Heatmap)



Insight: The heatmap visualizes the correlation between **views**, **likes**, and **comments**. A positive correlation between views and likes indicates that videos with more views tend to generate more likes.

- **Finding:** If the correlation is strong (close to +1), it indicates that engagement (likes) is tightly connected to views. A moderate correlation with comments could also suggest that videos with more views encourage audience interaction through comments.
- **Actionable Insight:** Marketers should focus on content that drives both views and likes. If comments are also positively correlated, engaging viewers through calls to action or interactive content (e.g., asking questions) could enhance engagement.

THANK YOU