**Data Curation Project Outline**

**1. Question of Interest:** What is the relationship between the popularity of trending music videos on YouTube and the various engagement metrics (views, likes, comments) over time? Specifically, how do these metrics evolve in the initial weeks following the video release?

**2. Website and/or API Used:** For this project, I utilized the YouTube Data API to gather data on trending music videos. The API allows access to video details, including view counts, likes, comments, and more.

**3. Final Sample Size and Features:**

* **Final Sample Size:** 500 trending music videos
* **Features Collected:**
  + Video Title
  + Channel Name
  + Views
  + Likes
  + Comments
  + Publish Date
  + Trending Date

**4. Outline of Planned Blog Post:**

**Introduction**

* **Motivation:** Explain the significance of understanding engagement metrics for trending music videos. Highlight the importance for content creators, marketers, and data enthusiasts.
* **Question of Interest:** State the primary question and its relevance.

**Data Collection**

* **Source:** Detail the YouTube Data API as the primary data source.
* **Ethics:** Discuss the ethical considerations and how the data collection complies with YouTube's terms of service.
* **Method:** Briefly explain the steps taken to gather the data, emphasizing the use of APIs and Python scripts.

**Data Exploration and Cleaning**

* **Initial Dataset:** Describe the raw data and its initial form.
* **Cleaning Process:** Explain any cleaning steps taken, such as handling missing values, formatting dates, and removing duplicates.

**Exploratory Data Analysis (EDA)**

* **Summary Statistics:** Present numerical summaries for each feature (e.g., average views, likes, comments).
* **Visualizations:**
  + Time series plots showing the evolution of views, likes, and comments over the first few weeks.
  + Distribution plots for views, likes, and comments.
  + Correlation matrix showing relationships between different metrics.
* **Findings:** Highlight key insights from the EDA, such as trends, patterns, and any surprising observations.

**Conclusion**

* **Summary:** Recap the main findings from the analysis.
* **Implications:** Discuss the potential implications for content creators and marketers.
* **Future Work:** Suggest possible extensions or further analysis that could be done with the dataset.

**Resources and Links**

* **GitHub Repo:** Provide a link to the GitHub repository containing the code and dataset.
* **Additional Resources:** Include any additional resources or articles that were helpful during the project.

**CODE:**

import requests

import pandas as pd

import json

# Define the API key and base URL

api\_key = 'YOUR\_API\_KEY'

base\_url = 'https://www.googleapis.com/youtube/v3/videos'

# Parameters for the API request

params = {

'part': 'snippet,statistics',

'chart': 'mostPopular',

'regionCode': 'US',

'videoCategoryId': '10', # Music category

'maxResults': 50,

'key': api\_key

}

# Make the API request

response = requests.get(base\_url, params=params)

data = response.json()

# Extract relevant information and store it in a DataFrame

video\_data = []

for video in data['items']:

video\_info = {

'title': video['snippet']['title'],

'channel': video['snippet']['channelTitle'],

'views': int(video['statistics']['viewCount']),

'likes': int(video['statistics'].get('likeCount', 0)),

'comments': int(video['statistics'].get('commentCount', 0)),

'publish\_date': video['snippet']['publishedAt'],

'trending\_date': pd.Timestamp.now()

}

video\_data.append(video\_info)

df = pd.DataFrame(video\_data)

df.to\_csv('youtube\_trending\_videos.csv', index=False)