

NAME: PAREENITA A. SHIRSATH B.E.A.I.&D.S. ROLL.NO:49

SMA EXPERIMENT NO : 03

```

!pip install google-api-python-client pandas
from googleapiclient.discovery import build
import pandas as pd

API_KEY = "AIzaSyDFZViTcfxxjFMr67zLyr82Z7ylElnipKg"

youtube = build("youtube", "v3", developerKey=API_KEY)

def fetch_youtube_data(keyword, max_results=10):
    request = youtube.search().list(
        part="snippet",
        q=keyword,
        type="video",
        maxResults=max_results
    )
    response = request.execute()

    video_ids = [item["id"]["videoId"] for item in response["items"]]

    video_request = youtube.videos().list(
        part="snippet,statistics",
        id=",".join(video_ids)
    )
    video_response = video_request.execute()

    data = []
    for video in video_response["items"]:
        data.append({
            "Video_ID": video["id"],
            "Title": video["snippet"]["title"],
            "Published_Date": video["snippet"]["publishedAt"],
            "Views": int(video["statistics"].get("viewCount", 0)),
            "Likes": int(video["statistics"].get("likeCount", 0)),
            "Comments": int(video["statistics"].get("commentCount", 0))
        })

    return pd.DataFrame(data)

```

Requirement already satisfied: google-api-python-client in /usr/local/lib/python3.12/dist-packages (2.188.0)

Requirement already satisfied: pandas in /usr/local/lib/python3.12/dist-packages (2.2.2)

Requirement already satisfied: httplib2<1.0.0,>=0.19.0 in /usr/local/lib/python3.12/dist-packages (from google-api-python-client)

Requirement already satisfied: google-auth!=2.24.0,!>=2.25.0,<3.0.0,>=1.32.0 in /usr/local/lib/python3.12/dist-packages (from google-api-python-client)

Requirement already satisfied: google-auth-http<1.0.0,>=0.2.0 in /usr/local/lib/python3.12/dist-packages (from google-api-python-client)

Requirement already satisfied: google-api-core!=2.0.*,!=2.1.*,!=2.2.*,!=2.3.0,<3.0.0,>=1.31.5 in /usr/local/lib/python3.12/dist-packages (from google-api-python-client)

Requirement already satisfied: uritemplate<5,>=3.0.1 in /usr/local/lib/python3.12/dist-packages (from google-api-python-client)

Requirement already satisfied: numpy>=1.26.0 in /usr/local/lib/python3.12/dist-packages (from pandas) (2.0.2)

Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas) (2.9.0.post0)

Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.2)

Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.3)

Requirement already satisfied: googleapis-common-protos<2.0.0,>=1.56.2 in /usr/local/lib/python3.12/dist-packages (from google-api-python-client)

Requirement already satisfied: protobuf!=3.20.0,!>=3.20.1,!>=4.21.0,!>=4.21.1,!>=4.21.2,!>=4.21.3,!>=4.21.4,!>=4.21.5,<7.0.0,>=3.19.5 in /usr/local/lib/python3.12/dist-packages (from google-api-core!=2.0.*,!=2.1.*,!=2.2.*,!=2.3.0,<3.0.0,>=1.31.5 in /usr/local/lib/python3.12/dist-packages)

Requirement already satisfied: proto-plus<2.0.0,>=1.22.3 in /usr/local/lib/python3.12/dist-packages (from google-api-core!=2.0.*,!=2.1.*,!=2.2.*,!=2.3.0,<3.0.0,>=1.31.5 in /usr/local/lib/python3.12/dist-packages)

Requirement already satisfied: requests<3.0.0,>=2.18.0 in /usr/local/lib/python3.12/dist-packages (from google-api-core!=2.0.*,!=2.1.*,!=2.2.*,!=2.3.0,<3.0.0,>=1.31.5 in /usr/local/lib/python3.12/dist-packages)

Requirement already satisfied: cachetools<7.0,>=2.0.0 in /usr/local/lib/python3.12/dist-packages (from google-auth!=2.24.0,!>=2.25.0,<3.0.0,>=1.32.0 in /usr/local/lib/python3.12/dist-packages)

Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.12/dist-packages (from google-auth!=2.24.0,!>=2.25.0,<3.0.0,>=1.32.0 in /usr/local/lib/python3.12/dist-packages)

Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.12/dist-packages (from google-auth!=2.24.0,!>=2.25.0,<3.0.0,>=1.32.0 in /usr/local/lib/python3.12/dist-packages)

Requirement already satisfied: pyparsing<4,>=3.0.4 in /usr/local/lib/python3.12/dist-packages (from httplib2<1.0.0,>=0.19.0 in /usr/local/lib/python3.12/dist-packages)

Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages)

Requirement already satisfied: pyasn1<0.7.0,>=0.6.1 in /usr/local/lib/python3.12/dist-packages (from pyasn1-modules>=0.2.1 in /usr/local/lib/python3.12/dist-packages)

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.18.0 in /usr/local/lib/python3.12/dist-packages)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.18.0 in /usr/local/lib/python3.12/dist-packages)

Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.18.0 in /usr/local/lib/python3.12/dist-packages)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.18.0 in /usr/local/lib/python3.12/dist-packages)

```

df_search_videos = fetch_youtube_data("Data Analytics",20)
df_search_videos

```

	Video_ID	Title	Published_Date	Views	Likes	Comments
0	yZvFH7B6gKI	What Is Data Analytics? - An Introduction (Ful...	2021-03-25T13:13:45Z	2009570	47442	924
1	pCJ15nGFgVg	Excel for Data Analytics - Full Course for Beg...	2024-10-14T12:01:06Z	2116149	54461	2755
2	tqmExNhqGxw	Stop Learning Everything! Do This Instead in D...	2025-04-13T09:16:47Z	6223	553	21
3	dcXqhMqhZUo	Data Analytics vs Data Science	2023-11-08T12:00:38Z	725512	14538	236
4	7mz73uXD9DA	SQL for Data Analytics - Learn SQL in 4 Hours	2024-03-11T10:00:11Z	1233574	31305	1239
5	YRJbhFLLPyE	The Complete Data Analyst Roadmap	2024-08-06T13:00:37Z	623940	16670	347
6	IgCNTuLBMK4	A Beginners Guide To The Data Analysis Process	2021-09-30T10:00:32Z	988783	17453	196
7	6MbN0MVaQ5Y	How I'd Become Data Analyst in 2026 From Zero ...	2026-01-27T12:50:18Z	1523	268	48
8	Hk4rRvmZPQ8	Data Analyst Roadmap 2026 The NEW Way To Bec...	2025-10-07T18:51:03Z	70808	3014	152
9	dJA7k58zIA8	How I'd become a data analyst (if i had to sta...	2024-01-12T14:00:27Z	1038097	48981	1264
10	NOJKAzIH8hA	How to tell if a career in Data Analytics is r...	2022-09-16T12:00:26Z	727788	26259	407
11	YsRDM6eoRq8	Don't Become a Data Analyst if	2024-08-13T14:08:48Z	2683671	69486	1091
12	kghcAk7l6eA	Introduction to Data Analysis with Excel: 2-Ho...	2023-03-07T16:00:09Z	729129	12344	293
13	SfE3aO3LWi0	I started my data analyst career taking these ...	2021-05-11T19:28:57Z	552961	25408	1384
14	JFA7H1uJ8wU	How I Would Learn to be a Data Analyst	2025-01-17T13:00:22Z	202924	9342	198
15	GbL-42kv5LI	Data Analytics for Beginners Google Data Ana...	2021-02-25T17:40:58Z	1858234	14611	418
16	481IC6AeLlg	Your Ultimate Roadmap to Becoming a Data Analy...	2025-04-12T08:54:27Z	1228114	0	36
17	hGCKdMsfU7g	How I'd Become a Data Analyst in 2026 (the ful...	2026-01-02T14:01:10Z	59134	3576	186
18	8chbARSxT2o	Analyzing data as a business analyst in tech	2024-10-24T22:35:37Z	630199	29391	249
19	F25knulvz-A	What do data analysts DO?! Here's a project wa...	2023-07-02T00:36:35Z	1067212	36289	390



Next steps: [Generate code with df_search_videos](#) [New interactive sheet](#)

```
# 1. Inspect & Copy Data
# Always work on a copy (good practice for business pipelines).

import pandas as pd
import re
import sqlite3

df = df_search_videos.copy()
# 2. Preprocessing (Cleaning & Formatting)
# Clean video titles
def clean_text(text):
    if pd.isna(text):
        return ""
    text = text.lower()
    text = re.sub(r"^[^w\s]", "", text) # remove punctuation
    text = re.sub(r"\s+", " ", text).strip()
    return text

df["Clean_Title"] = df["Title"].apply(clean_text)
# Convert datatypes
df["Published_Date"] = pd.to_datetime(df["Published_Date"])

numeric_cols = ["Views", "Likes", "Comments"]
for col in numeric_cols:
    df[col] = pd.to_numeric(df[col], errors="coerce").fillna(0)
# 3. Feature Engineering (Business Metrics)
# These are very important for analytics & decision-making.

df["Engagement"] = df["Likes"] + df["Comments"]

df["Engagement_Rate"] = (
    df["Engagement"] / (df["Views"] + 1)
).round(4)
# 4. Filtering (Business Logic)
# Example filtering rules:

# Must be Data Science related
```

```

# At least 50,000 views

# Engagement rate above a threshold

keywords = ["data science", "data scientist", "machine learning"]

pattern = "|".join(keywords)

df_filtered = df[
    df["Clean_Title"].str.contains(pattern, regex=True) &
    (df["Views"] >= 50000) &
    (df["Engagement_Rate"] >= 0.01)
]
# You can adjust thresholds depending on your business goal (marketing, trend analysis, education, etc.).

# 5. Sort for Business Insights
# Example: Top performing videos

df_filtered = df_filtered.sort_values(
    by=["Engagement_Rate", "Views"],
    ascending=False
)

# 6. Store the Data
# Store as CSV (reports, Excel, BI tools)
df_filtered.to_csv("youtube_data_science_videos.csv", index=False)
# Store in SQLite (dashboards, analytics, ML pipelines)
conn = sqlite3.connect("youtube_business_data.db")



df_filtered.to_sql(
    name="data_science_videos",
    con=conn,
    if_exists="replace",
    index=False
)

conn.close()

# 7. Final Output Check
print("Total videos fetched:", len(df))
print("Videos after filtering:", len(df_filtered))
df_filtered.head()

```

Total videos fetched: 20
Videos after filtering: 1

1 entry Filter  

index	Video_ID	Title	Published_Date	Views	Likes	Comments	Clean_Title	Engagement	Engagement_Rate
3	dcXqhMqhZUo	Data Analytics vs Data Science	2023-11-08 12:00:38+00:00	725512	14538	236	data analytics vs data science	14774	0.0204

Show 25 per page



Like what you see? Visit the [data table notebook](#) to learn more about interactive tables.

Time series

