

```
import pandas as pd
import numpy as np
```

```
df = pd.read_csv('/content/IN_youtube_trending_data.csv')
df
```



	video_id	title	publishedAt	channelId	channelTitle	categoryId	trending_date
0	lot0eF6EoNA	Sadak 2 Official Trailer Sanjay Pooja ...	2020-08-12T04:31:41Z	UCGqvJPRcv7aVFun-eTsatcA	FoxStarHindi	24	2020-12T00:00:00Z
1	x-KbnJ9fvJc	Kya Baat Aa : Karan Aujla (Official Video) Tan...	2020-08-11T09:00:11Z	UCm9SZAI03Rev9sFwloCdZ1g	Rehaan Records	10	2020-12T00:00:00Z
2	KX06ksuS6Xo	Diljit Dosanjh: CLASH (Official) Music Video ...	2020-08-11T07:30:02Z	UCZRdNleCgW-BGUJf-bbjzQg	Diljit Dosanjh	10	2020-12T00:00:00Z
3	UsMRgnTcchY	Dil Ko Maine Di Kasam Video Amaal M Ft.Ariji...	2020-08-10T05:30:49Z	UCq-Fj5jknLsUf-MWSy4_brA	T-Series	10	2020-12T00:00:00Z
4	WNSEXJJhKTU	Baarish (Official Video) Payal Dev,Stebin Ben ...	2020-08-11T05:30:13Z	UCYe6Oz0mg46S362LwARGVcA	VYRLOriginals	10	2020-12T00:00:00Z
...
251272	VKuNHLgrQno	Powerful Hanuman Chalisa HanuMan(Telugu) T...	2024-04-09T02:30:22Z	UC2V5vzgmEmoiWqXfM2jN5_w	Tips Telugu	10	2024-15T00:00:00Z
251273	2Dxg3g2MfJg	Indian Army Admit Card 2024 Kaise Download kar...	2024-04-12T11:31:50Z	UCQ7P1lEiG0uR-G8sU_zT0wg	Super Study	27	2024-15T00:00:00Z
251274	YDzOPDxinmA	NEW Deadpool 3 Trailer Has Thor? Captain Ameri...	2024-04-12T07:11:00Z	UCKQ5Jj35sjTmJigRtlCPhVQ	ComicVerse	24	2024-15T00:00:00Z
251275	KqtXJCZMhcs	FAAD MAGICIAN-CHUSO TOH JAANE	2024-04-11T14:45:01Z	UCxyrQT5syO1KIJ9z7aUPf5Q	RJ Abhinav	24	2024-15T00:00:00Z
251276	7bKql74_9Pg	Match Highlights Hyderabad FC 1-3 Kerala Bla...	2024-04-12T17:27:51Z	UCJk-aQ7NZtqYtpcqKDt_vZg	Indian Super League	17	2024-15T00:00:00Z
251277 rows × 16 columns							

```
df.info()
```



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 251277 entries, 0 to 251276
Data columns (total 16 columns):
#   Column              Non-Null Count  Dtype
---  -
0   video_id            251277 non-null object
1   title               251277 non-null object
```

```
2 publishedAt      251277 non-null object
3 channelId        251277 non-null object
4 channelTitle     251276 non-null object
5 categoryId       251277 non-null int64
6 trending_date    251277 non-null object
7 tags            251277 non-null object
8 view_count       251277 non-null int64
9 likes           251277 non-null int64
10 dislikes        251277 non-null int64
11 comment_count   251277 non-null int64
12 thumbnail_link  251277 non-null object
13 comments_disabled 251277 non-null bool
14 ratings_disabled 251277 non-null bool
15 description     231822 non-null object
dtypes: bool(2), int64(5), object(9)
memory usage: 27.3+ MB
```

```
df.isnull().sum()
```



	0
video_id	0
title	0
publishedAt	0
channelId	0
channelTitle	1
categoryId	0
trending_date	0
tags	0
view_count	0
likes	0
dislikes	0
comment_count	0
thumbnail_link	0
comments_disabled	0
ratings_disabled	0
description	19455



```
df.duplicated().sum()
```



75

```
df = df.drop_duplicates()
```

```
df.duplicated().sum()
```



0

```
df["description"] = df["description"].fillna("Not Available")
```



<ipython-input-75-6a0dd14321b1>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df["description"] = df["description"].fillna("Not Available")

```
df.info()
```



```
<class 'pandas.core.frame.DataFrame'>
Index: 251202 entries, 0 to 251276
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   video_id              251202 non-null object
1   title                 251202 non-null object
2   publishedAt           251202 non-null object
3   channelId             251202 non-null object
4   channelTitle          251201 non-null object
5   categoryId            251202 non-null int64
6   trending_date         251202 non-null object
7   tags                  251202 non-null object
8   view_count            251202 non-null int64
9   likes                 251202 non-null int64
10  dislikes              251202 non-null int64
11  comment_count         251202 non-null int64
12  thumbnail_link        251202 non-null object
13  comments_disabled     251202 non-null bool
14  ratings_disabled      251202 non-null bool
15  description           251202 non-null object
dtypes: bool(2), int64(5), object(9)
memory usage: 29.2+ MB
```

```
a = "sahil#gupta"
b = a.split("#")
b[0]
```



```
df["publish_date"] = df["publishedAt"].str.split("T").str[0]
df.head()
```

```
<ipython-input-78-bd09475dc580>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-wrapper-object

```
df["publish_date"] = df["publishedAt"].str.split("T").str[0]
```

	video_id	title	publishedAt	channelId	channelTitle	categoryId	trending_date	
0	lot0eF6EoNA	Sadak 2 Official Trailer Sanjay Pooja ...	2020-08-12T04:31:41Z	UCGqvJPRcv7aVFun-eTsatcA	FoxStarHindi	24	2020-08-12T00:00:00Z	sada 2 bhatti films
1	x-KbnJ9fvJc	Kya Baat Aa : Karan Aujla (Official Video) Tan...	2020-08-11T09:00:11Z	UCm9SZAi03Rev9sFwloCdz1g	Rehaan Records	10	2020-08-12T00:00:00Z	
2	KX06ksuS6Xo	Diljit Dosanjh: CLASH (Official) Music Video ...	2020-08-11T07:30:02Z	UCZRdNleCgW-BGUJf-bbjzQg	Diljit Dosanjh	10	2020-08-12T00:00:00Z	cl: dosa dosa
3	UsMRgnTcchY	Dil Ko Maine Di Kasam Video Amaal M Ft.Ariji...	2020-08-10T05:30:49Z	UCq-Fj5jknLsUf-MWSy4_brA	T-Series	10	2020-08-12T00:00:00Z	son: son: new sc
4	WNSEXJJhKTU	Baarish (Official Video) Payal Dev,Stebin Ben ...	2020-08-11T05:30:13Z	UCye6Oz0mg46S362LwARGVcA	VYRLOriginals	10	2020-08-12T00:00:00Z	Original Khan S Joshi

```
df["publish_date"] = pd.to_datetime(df["publish_date"])
df.info()
```

```
<ipython-input-80-6a77045b95b9>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-wrapper-object

```
df["publish_date"] = pd.to_datetime(df["publish_date"])
<class 'pandas.core.frame.DataFrame'>
Index: 251202 entries, 0 to 251276
Data columns (total 17 columns):
```

#	Column	Non-Null Count	Dtype
0	video_id	251202 non-null	object
1	title	251202 non-null	object
2	publishedAt	251202 non-null	object
3	channelId	251202 non-null	object
4	channelTitle	251201 non-null	object
5	categoryId	251202 non-null	int64
6	trending_date	251202 non-null	object
7	tags	251202 non-null	object
8	view_count	251202 non-null	int64
9	likes	251202 non-null	int64
10	dislikes	251202 non-null	int64
11	comment_count	251202 non-null	int64
12	thumbnail_link	251202 non-null	object
13	comments_disabled	251202 non-null	bool
14	ratings_disabled	251202 non-null	bool
15	description	251202 non-null	object

```
16 publish_date      251202 non-null datetime64[ns]
dtypes: bool(2), datetime64[ns](1), int64(5), object(9)
memory usage: 31.1+ MB
```

```
df.columns
```

```
Index(['video_id', 'title', 'publishedAt', 'channelId', 'channelTitle',
      'categoryId', 'trending_date', 'tags', 'view_count', 'likes',
      'dislikes', 'comment_count', 'thumbnail_link', 'comments_disabled',
      'ratings_disabled', 'description'],
      dtype='object')
```

```
print(df.columns)
```

```
Index(['video_id', 'title', 'publishedAt', 'channelId', 'channelTitle',
      'categoryId', 'trending_date', 'tags', 'view_count', 'likes',
      'dislikes', 'comment_count', 'thumbnail_link', 'comments_disabled',
      'ratings_disabled', 'description', 'publish_date'],
      dtype='object')
```

Start visualization

Question_1: 1. Unique Videos

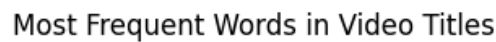
```
unique_videos = df['video_id'].nunique()
print("Number of unique videos:", unique_videos)
```

```
Number of unique videos: 78847
```

Question_2: Most Frequently Used Words in Titles.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud
```

```
wordcloud = WordCloud(width=800, height=400, background_color='white').generate(' '.join(df['title'].astype(str)))
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title("Most Frequent Words in Video Titles")
plt.show()
```



```
df['publishedAt'] = pd.to_datetime(df['publishedAt'])
plt.figure(figsize=(10, 5))
df['publishedAt'].dt.date.value_counts().sort_index().plot(kind='line')
plt.title("Videos Published Over Time")
plt.xlabel("Publication Date")
plt.ylabel("Number of Videos")
plt.show()
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
df['publishedAt'] = pd.to_datetime(df['publishedAt'])
```

Questions_4: Channels with Highest Number of Trending Videos

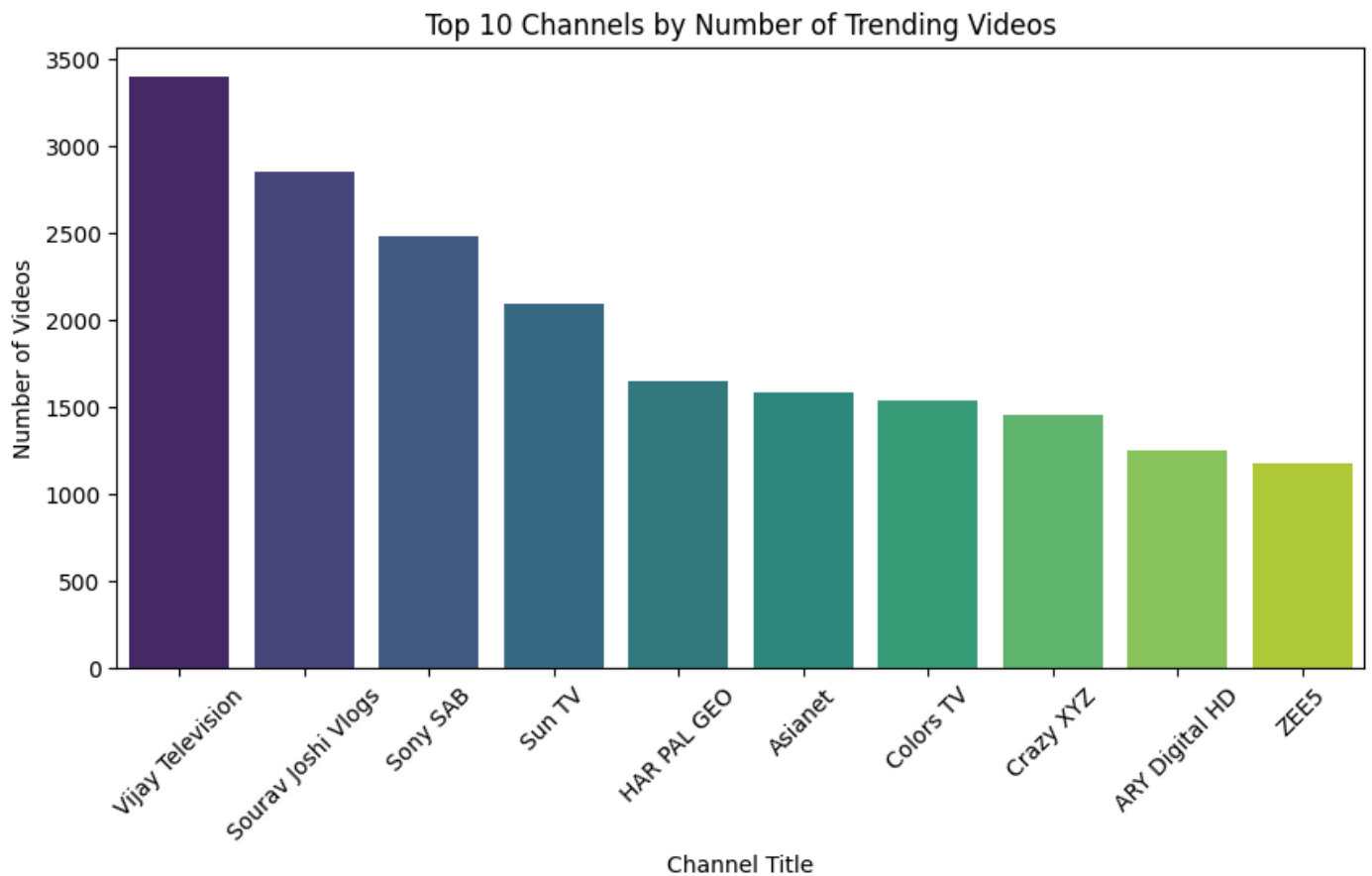
```
channel_counts = df['channelTitle'].value_counts().head(10)
plt.figure(figsize=(10, 5))
sns.barplot(y=channel_counts.values, x=channel_counts.index, palette='viridis')
plt.title("Top 10 Channels by Number of Trending Videos")
plt.xticks(rotation = 45)
plt.ylabel("Number of Videos")
plt.xlabel("Channel Title")
plt.show()
```



<ipython-input-86-b669e5f9b2fa>:3: FutureWarning:


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue`

```
sns.barplot(y=channel_counts.values, x=channel_counts.index, palette='viridis')
```



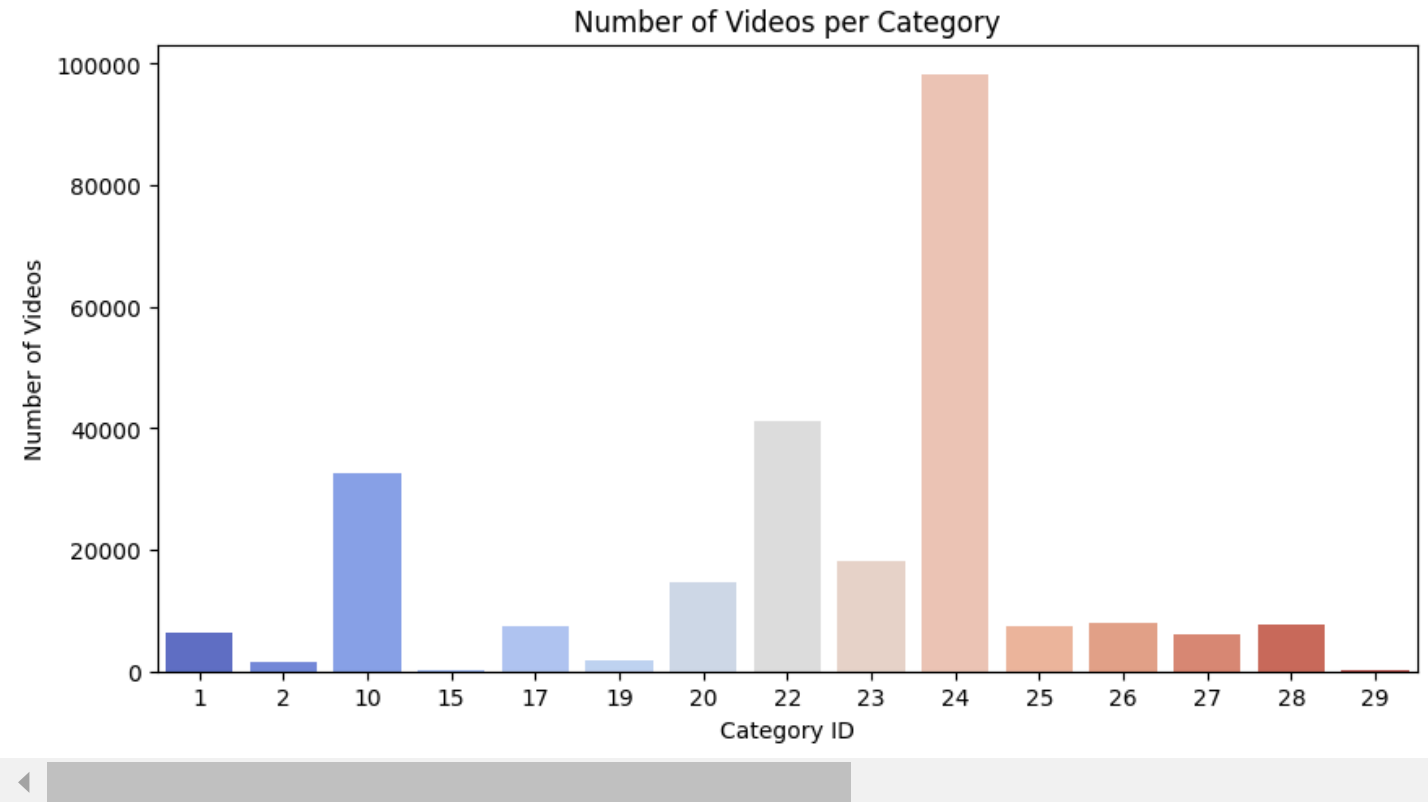
Question_5:Video Categories Distribution

```
category_counts = df['categoryId'].value_counts()
plt.figure(figsize=(10, 5))
sns.barplot(x=category_counts.index, y=category_counts.values, palette='coolwarm')
plt.title("Number of Videos per Category")
plt.xlabel("Category ID")
plt.ylabel("Number of Videos")
plt.show()
```

 <ipython-input-87-7352826068e2>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue`

```
sns.barplot(x=category_counts.index, y=category_counts.values, palette='coolwarm')
```



Question_6 channel title find corresponding categoryId.

```
import pandas as pd

# Function to get channel name by channel ID
def get_channel_name_by_id(categoryId):
    channel_name = df.loc[df['categoryId'] == categoryId, 'title']
    if not channel_name.empty:
        return channel_name.iloc[0]
    else:
        return "Channel ID not found"

# Example usage
categoryId= 24
print(f"Channel Name for ID {categoryId}: {get_channel_name_by_id(categoryId)}")
```


 Channel Name for ID 24: Sadak 2 | Official Trailer | Sanjay | Pooja | Alia | Aditya | Jisshu | Mahesh Bhatt | 28 Aug

Double-click (or enter) to edit

```
false_count = (df['comments_disabled'] == False).sum()

print("Number of video in comments_disabled:", false_count)
true_count = (df['comments_disabled'] != False).sum()

print("Number of video in comments is on:", true_count)
total=false_count+true_count
print("total",total)
```

 Number of video in comments_disabled: 249505
Number of video in comments is on: 1697
total 251202

Question_8:Trending Videos by Date

```
# Get the value counts for 'trending_date' and sort them
rending_date_counts = df['trending_date'].value_counts()

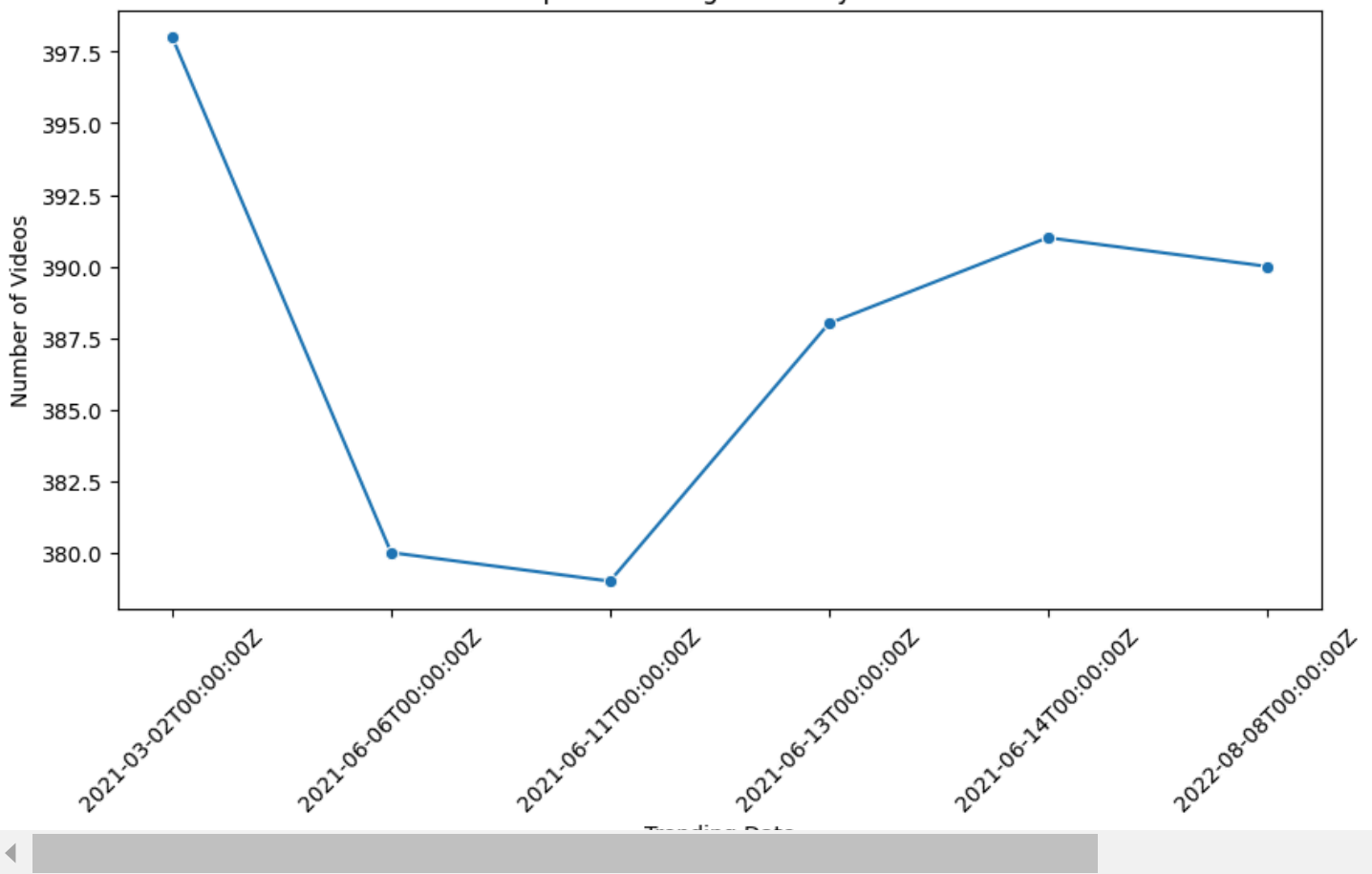
# Select the top 10 dates
top_10_trending_dates = rending_date_counts.nlargest(6).sort_index()

# Reset the index to prepare for plotting
top_10_trending_dates = top_10_trending_dates.reset_index()
top_10_trending_dates.columns = ['trending_date', 'video_count']

# Plot using seaborn for a line graph
plt.figure(figsize=(10, 5))
sns.lineplot(x=top_10_trending_dates['trending_date'], y=top_10_trending_dates['video_count'], marker='o')
plt.xticks(rotation=45)
plt.title("Top 10 Trending Videos by Date")
plt.xlabel("Trending Date")
plt.ylabel("Number of Videos")
plt.show()
```



Top 10 Trending Videos by Date



Question_9:Channel Title corresponding Total view.

```
gb = df.groupby("channelTitle").agg({"view_count": "sum"})
gb
```



view_count

channelTitle

NIPUN Bharat Uttar Pradesh	8094787
Veena Singer	5731524
#RohiL Fitness Yug	2741126
0300 Brotherhood Records	11566577
1 Million Views Tamil	107724641
...	...
తెలుగు రైతుబడి	10870342
ಉದ್ಯೋಗ ಮಿತ್ರ	60956
ആവൂസ് avoosss family	55219860
ജോഷ് Talks	34027288
이지금 [IU Official]	164796273



df.columns



```
Index(['video_id', 'title', 'publishedAt', 'channelId', 'channelTitle',  
      'categoryId', 'trending_date', 'tags', 'view_count', 'likes',  
      'dislikes', 'comment_count', 'thumbnail_link', 'comments_disabled',  
      'ratings_disabled', 'description', 'publish_date'],  
      dtype='object')
```

Question_10:Top 6 Liked youtube chanel

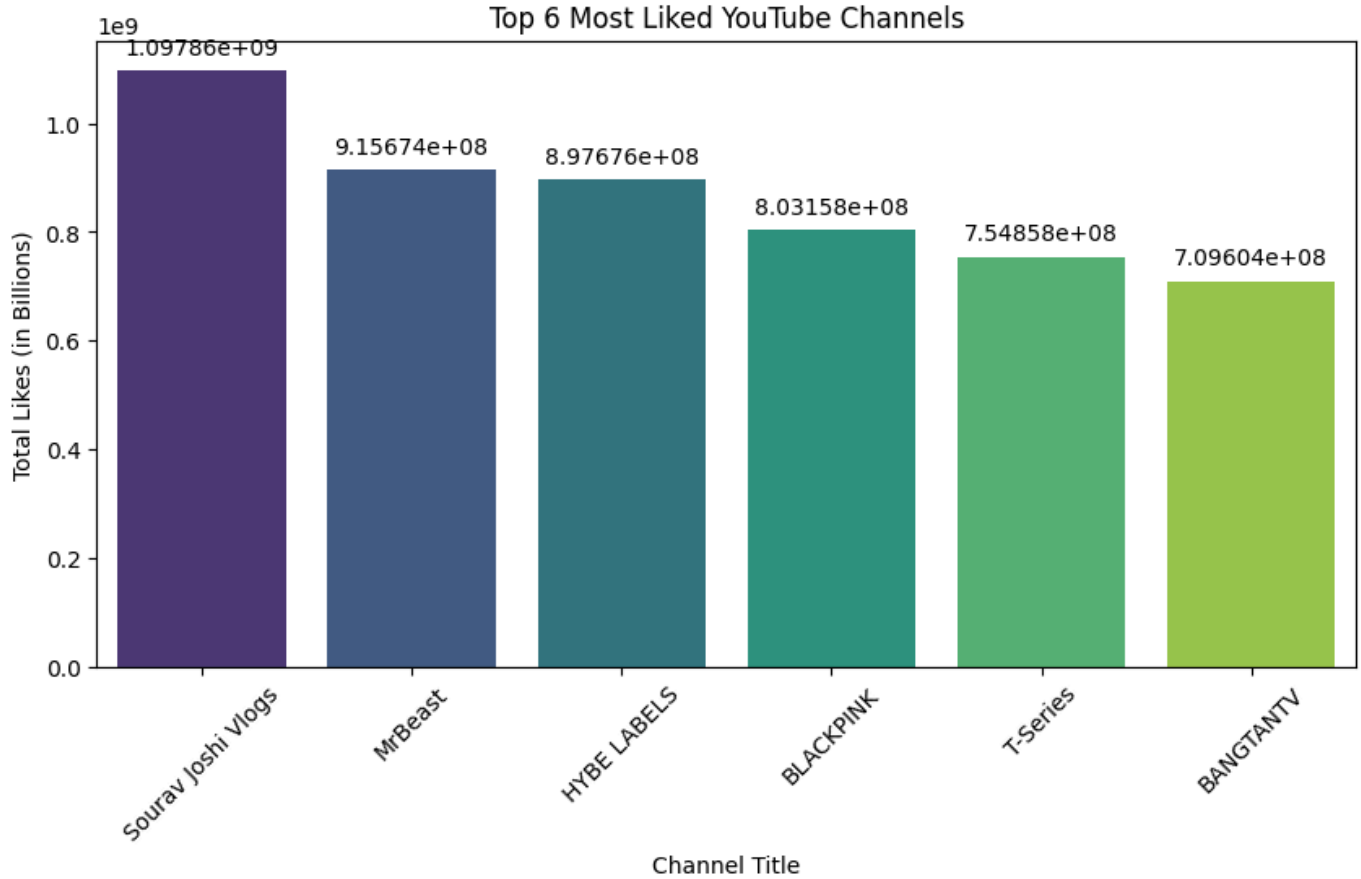
```
top_6_channels_likes = df.groupby('channelTitle')['likes'].sum().sort_values(ascending=False).head(6)
```

```
# Plot  
plt.figure(figsize=(10, 5))  
ax=sns.barplot(x=top_6_channels_likes.index, y=top_6_channels_likes.values, palette='viridis')  
  
for container in ax.containers:  
    ax.bar_label(container, padding=5)  
plt.title("Top 6 Most Liked YouTube Channels")  
plt.xticks(rotation=45)  
plt.xlabel("Channel Title")  
plt.ylabel("Total Likes (in Billions)")  
plt.show()
```

<ipython-input-128-4002dfee5df8>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue`

```
ax=sns.barplot(x=top_6_channels_likes.index, y=top_6_channels_likes.values, palette='viridis')
```



```
# Group by 'channelTitle' and sum the 'dislikes' for each channel
top_6_channels_likes = df.groupby('channelTitle')['dislikes'].sum().sort_values(ascending=False).head(6)

# Scale the values to billions
top_6_channels_likes_in_billions = top_6_channels_likes / 1e9

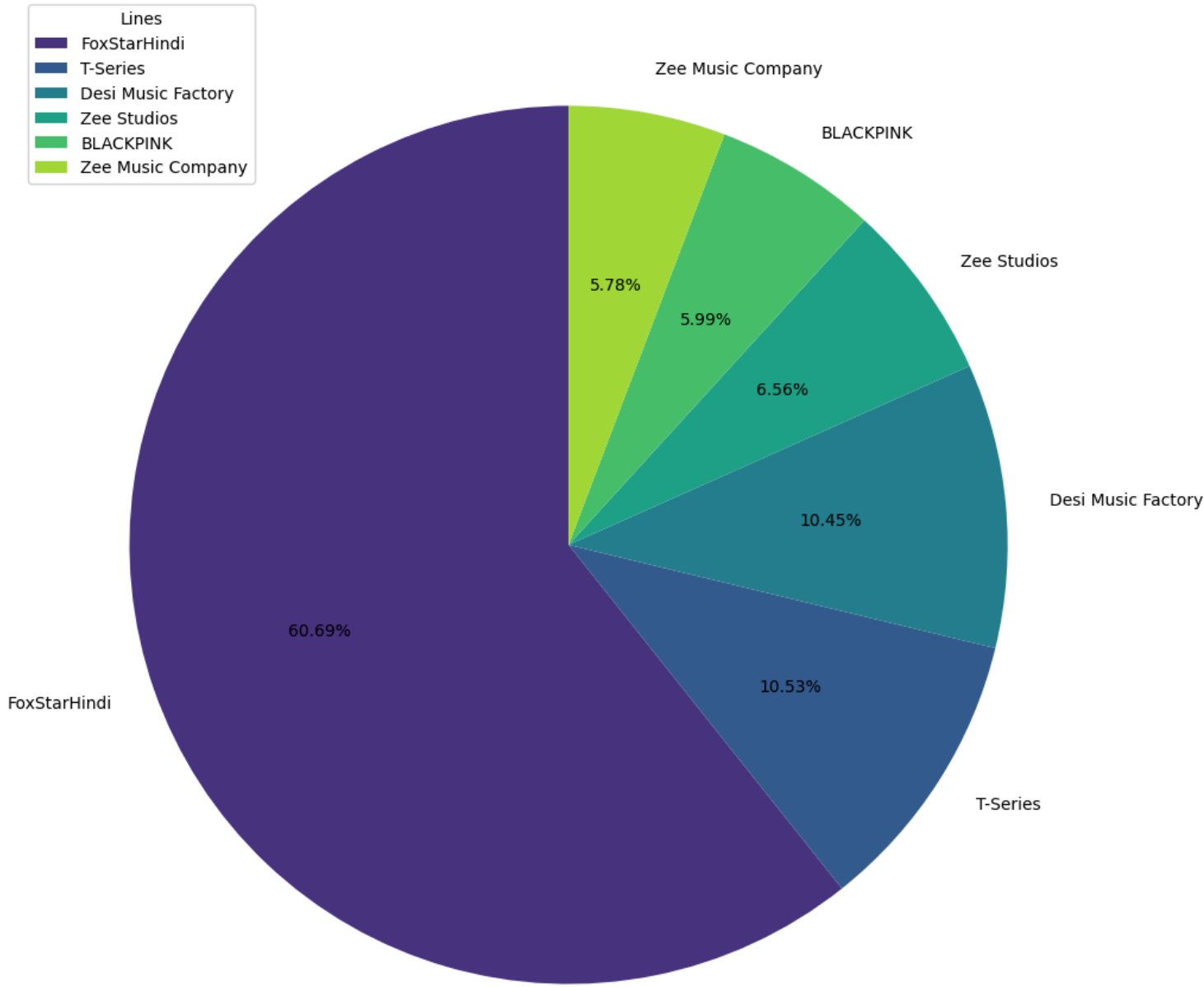
# Generate colors using seaborn palette
colors = sns.color_palette("viridis", len(top_6_channels_likes_in_billions))

# Plot the pie chart with formatted labels showing values in billions
plt.figure(figsize=(12, 12))
plt.pie(top_6_channels_likes_in_billions, labels=top_6_channels_likes_in_billions.index,
        autopct='%1.2f%%', colors=colors, startangle=90)

# Customize the plot
plt.title("Top 6 Most Disliked YouTube Channels")
plt.legend(title="Lines",loc='upper left')
plt.show()
```



Top 6 Most Disliked YouTube Channels



Question_10:top 5 channel view.

```
gb=gb.sort_values("view_count",ascending=False)
gb.head()
```



	view_count
channelTitle	
T-Series	26997737316
MrBeast	15748817791
Sourav Joshi Vlogs	11761288222
HAR PAL GEO	11403761002
BLACKPINK	8788111105

Top 5 comment youtube channel.

```
gb1 = df.groupby("title").agg({"view_count":"sum"})
gb1 = gb1.sort_values(by = "view_count", ascending = False)
gb1 = gb1.head()
gb1
```



		view_count
title		
BTS (방탄소년단) 'Dynamite' Official MV		1887077945