Experiment 5

Aim : Analyze and visualize the social media data collected

Course Outcome : Learner will be able to collect, preprocess , analyze and visualize social media data from multiple platforms

```
from googleapiclient.discovery import build # for using api to send
requests
import pandas as pd                         # To structure data
import seaborn as sns

KEY = "AIzaSyBT3GkJ8WUvdum1PGd3aFAXHc3eXgs19Sg"
youtube = build('youtube','v3',developerKey=KEY)

channel_ids = [
    "UCFy846QdKs3LbLgBpSqPcdg", # The Kiffness
    "UCeVMnSShP_Iviwkknt83cww", # Code with harry
    "UC0rE2qq81of4fojo-KhO5rg", # Tanmay bhatt
    "UC9CYT9gSNLevX5ey2_6CK0Q", # NDTV
    ]
```

Requesting API to give channel statistics

```
def getChannelStats(yoututbe,channel_ids):
  all_data = []
  request =
youtube.channels().list(part='snippet,contentDetails,statistics',id=',
'.join(channel_ids))
  response = request.execute()
  # PreProcessing
  for i in range(len(response['items'])):
    data = dict(
        Channel_name = response['items'][i]['snippet']['title'],
        Subscribers = response['items'][i]['statistics']
['subscriberCount'],
        Views = response['items'][i]['statistics']['viewCount'],
        Total_videos = response['items'][i]['statistics']
['videoCount'],
        playlist_id = response['items'][i]['contentDetails']
['relatedPlaylists']['uploads'],
        )
    all_data.append(data)

  return all_data
```

collected data is semi structured

```
collected_data = getChannelStats(youtube,channel_ids)
structured_form = pd.DataFrame(collected_data)
structured_form
```

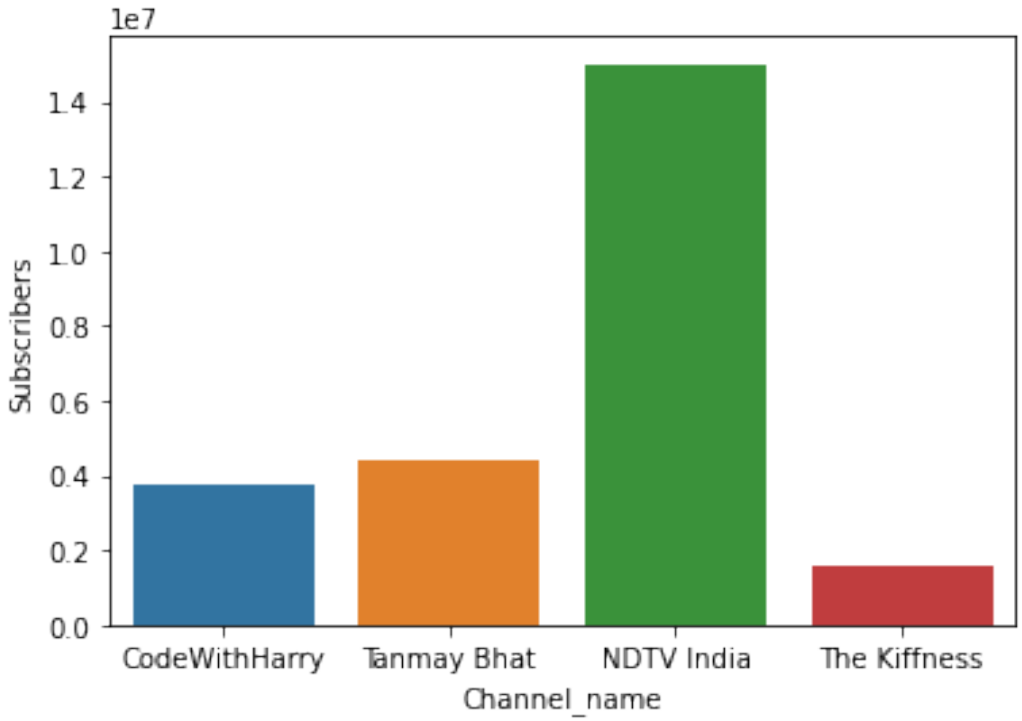
```
    Channel_name Subscribers       Views Total_videos  \
0  CodeWithHarry     3760000   487054143         1990
1    Tanmay Bhat     4410000  1243975502          900
2     NDTV India    15000000  6300439714        78711
3   The Kiffness     1560000   313347027          195

                playlist_id
0  UUeVMnSShP_Iviwkknt83cww
1  UU0rE2qq81of4fojo-KhO5rg
2  UU9CYT9gSNLevX5ey2_6CK0Q
3  UUFy846QdKs3LbLgBpSqPcdg
```

```
structured_form['Subscribers'] =
pd.to_numeric(structured_form['Subscribers'])
structured_form['Views'] = pd.to_numeric(structured_form['Views'])
structured_form['Total_videos'] =
pd.to_numeric(structured_form['Total_videos'])
structured_form.dtypes
```

```
Channel_name    object
Subscribers      int64
Views            int64
Total_videos     int64
playlist_id     object
dtype: object
```

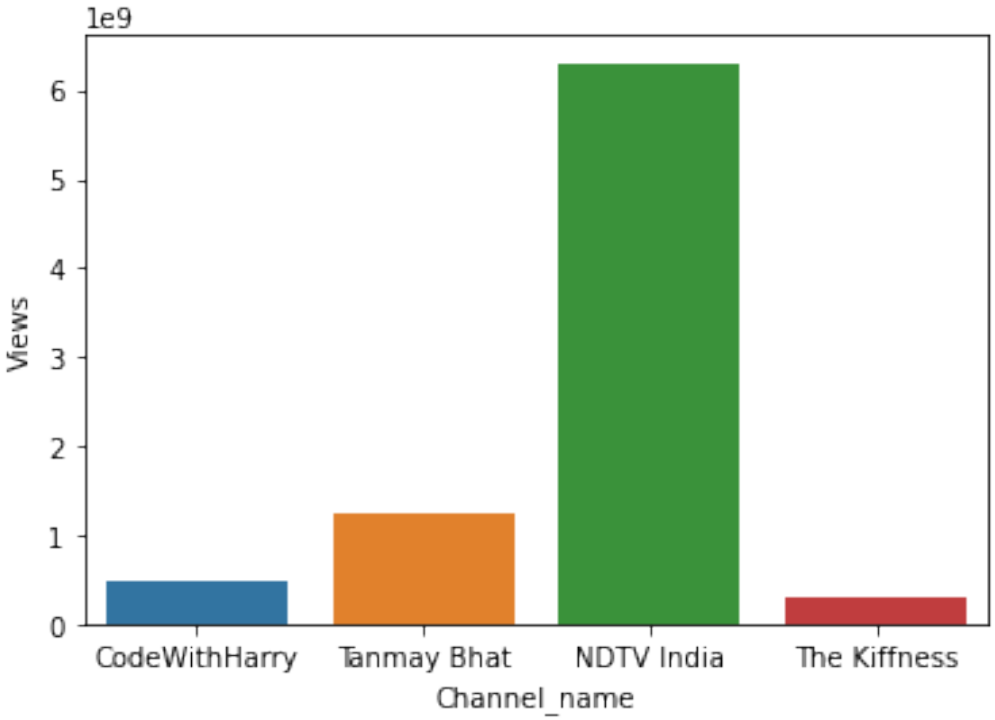
Subscribers Count

```
ax = sns.barplot(x='Channel_name', y='Subscribers',
data=structured_form)
```



View count

```
ax = sns.barplot(x='Channel_name', y='Views', data=structured_form)
```



Video Count

```
ax = sns.barplot(x='Channel_name', y='Total_videos',
data=structured_form)
```