Experiment 4

Aim : To Perform Preprocessing of collected data and store it

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 save data

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
    "CSHZKyawb77ixDdsGog4iWA",  # Lex Fridman
]
```

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)
collected_data
```

```
[{'Channel_name': 'CodeWithHarry',
  'Subscribers': '3760000',
  'Views': '486418246',
  'Total_videos': '1990',
  'playlist_id': 'UUeVMnSShP_Iviwkknt83cww'},
 {'Channel_name': 'NDTV India',
  'Subscribers': '15000000',
  'Views': '6300439714',
  'Total_videos': '78706',
  'playlist_id': 'UU9CYT9gSNLevX5ey2_6CK0Q'},
 {'Channel_name': 'The Kiffness',
  'Subscribers': '1560000',
  'Views': '313347027',
  'Total_videos': '195',
  'playlist_id': 'UUFy846QdKs3LbLgBpSqPcdg'},
 {'Channel_name': 'Tanmay Bhat',
  'Subscribers': '4410000',
  'Views': '1243326796',
  'Total_videos': '900',
  'playlist_id': 'UU0rE2qq81of4fojo-KhO5rg'}]
```

Structuring the collected data

```
structured_form = pd.DataFrame(collected_data)
structured_form
```

```
    Channel_name Subscribers       Views Total_videos  \
0  CodeWithHarry     3760000   486418246         1990
1     NDTV India    15000000  6300439714        78706
2   The Kiffness     1560000   313347027          195
3    Tanmay Bhat     4410000  1243326796          900

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

Storing the collected data

```
structured_form.to_csv("collected_yt_channels_data.csv")
```