# COURSE ASSIGNMENT

## SOFTWARE ENGINEERING DEPARTMENT


## INTRODUCTION TO DATASCIENCE

## ASSIGNMENT # 01

**SUBMITTED BY**

**MUSHARRAF RAZA KHAN | 52024**

BS-SE 8th

**INSTRUCTOR**

**MA'AM SANIYA ASHRAF**

Lecturer, Software Engineering Department,

Faculty of Information & Communication Technology

BUITEMS Quetta,

**ASSIGNMENT # 02** **INTRODUCTION TO DATA SCIENCE**

"Session Spring-2023"

# "IDS-ASSIGNMENT

# BY-MUSHARRAF RAZA KHAN | 52024"

## CODE

**The Youtube Data API**

```
api_key = "AIzaSyDp0v6n0UFdgroIYskMui5u6-K01GYm7n4"
# Replace this dummy api key with your own.

from apiclient.discovery import build
youtube = build('youtube', 'v3', developerKey=api_key)

import pandas as pd

ID = "i_LwzRVP7bg" # Replace this YouTube video ID with your own.

box = [['Name', 'Comment', 'Time', 'Likes', 'Reply Count']]


def scrape_comments_with_replies():
   data = youtube.commentThreads().list(part='snippet', videoId=ID, maxResults='100',
textFormat="plainText").execute()

   for i in data["items"]:

      name = i["snippet"]['topLevelComment']["snippet"]["authorDisplayName"]
      comment = i["snippet"]['topLevelComment']["snippet"]["textDisplay"]
      published_at = i["snippet"]['topLevelComment']["snippet"]['publishedAt']
      likes = i["snippet"]['topLevelComment']["snippet"]['likeCount']
      replies = i["snippet"]['totalReplyCount']

      box.append([name, comment, published_at, likes, replies])

      totalReplyCount = i["snippet"]['totalReplyCount']

      if totalReplyCount > 0:

         parent = i["snippet"]['topLevelComment']["id"]

         data2 = youtube.comments().list(part='snippet', maxResults='100', parentId=parent,
                                         textFormat="plainText").execute()

         for i in data2["items"]:
```

```
            name = i["snippet"]["authorDisplayName"]
            comment = i["snippet"]["textDisplay"]
            published_at = i["snippet"]['publishedAt']
            likes = i["snippet"]['likeCount']
            replies = ""

            box.append([name, comment, published_at, likes, replies])

    while ("nextPageToken" in data):

        data = youtube.commentThreads().list(part='snippet', videoId=ID,
pageToken=data["nextPageToken"],
                                             maxResults='100', textFormat="plainText").execute()

        for i in data["items"]:
            name = i["snippet"]['topLevelComment']["snippet"]["authorDisplayName"]
            comment = i["snippet"]['topLevelComment']["snippet"]["textDisplay"]
            published_at = i["snippet"]['topLevelComment']["snippet"]['publishedAt']
            likes = i["snippet"]['topLevelComment']["snippet"]['likeCount']
            replies = i["snippet"]['totalReplyCount']

            box.append([name, comment, published_at, likes, replies])

            totalReplyCount = i["snippet"]['totalReplyCount']

            if totalReplyCount > 0:

                parent = i["snippet"]['topLevelComment']["id"]

                data2 = youtube.comments().list(part='snippet', maxResults='100', parentId=parent,
                                                textFormat="plainText").execute()

                for i in data2["items"]:
                    name = i["snippet"]["authorDisplayName"]
                    comment = i["snippet"]["textDisplay"]
                    published_at = i["snippet"]['publishedAt']
                    likes = i["snippet"]['likeCount']
                    replies = ''

                    box.append([name, comment, published_at, likes, replies])

    df = pd.DataFrame({'Name': [i[0] for i in box], 'Comment': [i[1] for i in box], 'Time': [i[2]
for i in box],
                       'Likes': [i[3] for i in box], 'Reply Count': [i[4] for i in box]})

    df.to_csv('youtube-comments1.csv', index=False, header=False)
```

```
    return "Successful! Check the CSV file that you have just created."

scrape_comments_with_replies()
```

## Explanation of The Code:

The given code is for extracting comments and replies from a YouTube video using the YouTube Data API. Let's go through the code step by step;

The code starts by importing the necessary libraries. The apiclient.discovery module is imported from the apiclient library, and the build function is used to create a YouTube API client object. The pandas library is also imported as pd.

1. An API key is defined as a string variable named api_key. You would need to replace this dummy API key with your own valid API key obtained from the Google Cloud Console.
2. The youtube object is created using the build function from the YouTube API client. The first argument specifies the API service name ('youtube') and the API version ('v3'). The developerKey parameter is set to the api_key variable.
3. A two-dimensional list named box is initialized with the header row containing column names: 'Name', 'Comment', 'Time', 'Likes', and 'Reply Count'.
4. The scrape_comments_with_replies() function is defined. This function is responsible for extracting comments and replies from the YouTube video.
5. Inside the function, the youtube.commentThreads().list() method is called to retrieve the comments associated with the specified video ID (ID). The part parameter is set to 'snippet' to include relevant data in the response. The maxResults parameter is set to '100' to limit the number of comments per page, and the textFormat parameter is set to 'plainText' to retrieve the comment text without any formatting.
6. The retrieved comments are iterated using a for loop. For each comment, the name of the commenter, the comment text, the publication timestamp, the number of likes, and the total reply count are extracted and stored in separate variables.
7. The extracted data is appended to the box list as a new row.

8. If the total reply count for the comment is greater than 0, an additional API request is made to retrieve the replies to that comment. The youtube.comments().list() method is called with the part parameter set to 'snippet', and the parentId parameter is used to specify the comment ID for which replies are being fetched.
9. The retrieved replies are iterated, and the same data (name, comment, publication timestamp, likes) is extracted and appended to the box list, leaving the 'replies' field as an empty string.
10. The above steps are repeated as long as there are more pages of comments. The while loop checks if the "nextPageToken" is present in the response data and retrieves the next page of comments accordingly.
11. After all the comments and replies have been retrieved, the box list is converted to a pandas DataFrame named df, where each column corresponds to the data extracted from the comments and replies.
12. The DataFrame is then saved to a CSV file named 'youtube-comments1.csv' using the to_csv() function of pandas.
13. Finally, the function returns a success message indicating the completion of the data extraction process.
14. The scrape_comments_with_replies() function is called at the end to initiate the data extraction process.

**NOTE:** Make sure you have the necessary API key and the required libraries installed before running the code.

**ASSIGNMENT # 02** **INTRODUCTION TO DATA SCIENCE**

youtube-comments1.csv

1 to 100 of 492 entries Filter

| Name | Comment | Time | Likes | Reply Count |
|---|---|---|---|---|
| Sangeetha Studio | You are literally the best, I've been looking for a tutorial for three days and yours works | 2022-09-26T23:50:54Z | 4 | 0 |
| 0xWxE | I just stated my journey today | 2022-09-26T16:24:00Z | 0 | 1 |
| 1.234.567 PokemGaming | Honestly, she is beautiful. The explanation is also perfect. Thank you | 2023-03-01T06:24:13Z | 0 | 0 |
| 4th Reich | I cant see ur eyes | 2023-05-22T18:32:47Z | 0 | 0 |
| 4th Reich | Baby girl | 2023-05-22T18:32:25Z | 0 | 0 |
| 4th Reich | I lobe u | 2023-05-22T18:32:05Z | 0 | 0 |
| 585ghz | 😳 thanks for all 💕💕 | 2022-09-27T01:01:02Z | 0 | 0 |
| ? | thanks =) | 2023-04-04T06:50:19Z | 0 | 0 |
| A Andres Riera | I cannot focus on the course you 're so pretty while explain, my bad lol. Great content and clear explanation. | 2022-10-12T16:46:16Z | 0 | 0 |
| A K RAGHAVENDRAN | thanks | 2022-10-10T09:04:26Z | 0 | 0 |
| ABU BAKAR HASNATH | 28:31 35:28 train, validation, test dataset 1:15:33 | 2022-12-29T15:50:22Z | 0 | 0 |
| AHMAD MPONDA | She's excellent.I wish she do deep learning tutorials | 2022-10-21T03:35:39Z | 1 | 0 |
| AMTV | Nice! | 2022-09-26T16:04:38Z | 3 | 0 |

**NOTE:** Link To Github Repository

[GitHub - Musharraf-Raza-Khan/Introduction-To-Data-Science-IDS-: Welcome To My (IDS) Repository & Happy Learning !]

**“IDS-ASSIGNMENT**

**BY-MUSHARRAF RAZA KHAN | 52024”**