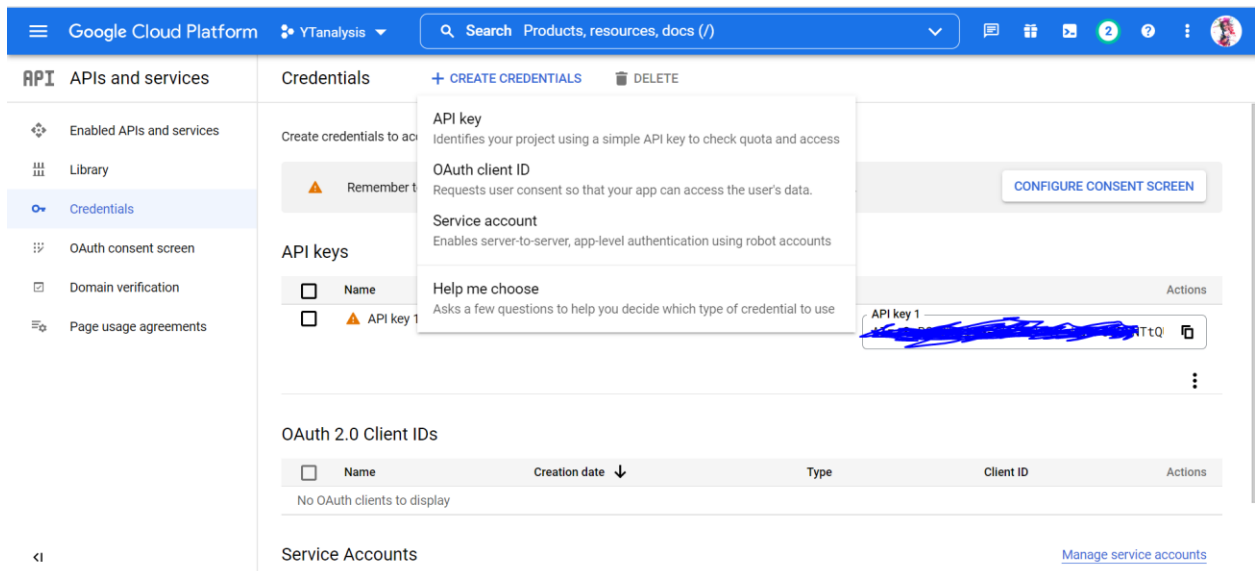# YOUTUBE DATA ANALYSIS

PURPOSE OF THE REPORT:

This report will clarify how to access YouTube data API key, extracting a channel information and detailed information's. Then a basic analysis of data has been performed.

GETTING API ACCESS:

1. Create a google account if you don't have one before. Or skip this step
2. Open the link: https://console.cloud.google.com/getting-started using the google account
3. Click 'Create a Project' (if it's your first project) or 'select a project' on the top navigation bar -> 'Create project' -> add name to the project ->click 'Create'
4. Once the project has been created, search 'Youtube Data API v3' and enable it.

5. After enabling the API, go to credentials tab under the API and click 'create credentials'.
6. Click 'API key' and it will generate an API key (hidden on the picture). Save it for future use.

ENVIRONMENTAL SETUP:

All the codes in this project are developed using Python. Hence any IDE which supports python should be able to execute the codes without any errors. (Author of this documentation used PyCharm)

To run this quickstart, you'll need:

- Python 2.7 or Python 3.5+
- The `pip` package management tool
- The Google APIs Client Library for Python:

```
pip install --upgrade google-api-python-client
```

- pip install pandas
- pip install seaborn
  Packages for Dashboard visualization: (the versions may be subject to change)

```
pip install dash==0.21.1
pip install dash-renderer==0.13.0
pip install dash-html-components==0.11.0
pip install dash-core-components==0.23.0
pip install plotly --upgrade
```

UNDERSTANDING HOW TO USE API KEY TO EXTRACT BASIC INFORMATION ABOUT A CHANNEL:

This part of documentation will be explained in detail to show how the above generated API key can be used to extract information about a YouTube channel. It will be helpful to understand the future codes without much explanation.

Handy guide: https://developers.google.com/youtube/v3/docs/channels/list

This API guide will have all the methods required for scraping the YouTube. This can be our source documentation to go back at anytime to get the appropriate methods for scraping.

Note: API key used in the project has been replaced with 'APIKEY' for privacy reasons. For executing these codes without errors, API key needs to be generated on the user's account and replace the 'APIKEY' string in the codes.

```python
from googleapiclient.discovery import build

api_key='APIKEY'
channel_id='UCnz-ZXXER4jOvuED5trXfEA'
youtube = build(
            'youtube','v3', developerKey=api_key)
```

- Import the required libraries
- Replace the api_key variable with the generated API key and channel_id can also be replaced according to the user needs
- build() method will take three parameters: API service name, API version and API key

```python
def channel_statistics(youtube,channel_id):
    request=youtube.channels().list(part='snippet,contentDetails,statistics',id=channel_id
                                    )
    response=request.execute()
    print(response)
    print('*'*100)
    data=dict(channel_name=response['items'][0]['snippet']['title'],
              subscribers=response['items'][0]['statistics']['subscriberCount'],
              views=response['items'][0]['statistics']['viewCount'],
              videos=response['items'][0]['statistics']['videoCount'])
    return data

print(channel_statistics(youtube,channel_id))
```

The channels() contains the information about the YouTube channel. list() method under the channels resources helps in retrieving all the necessary information about the channel in json format. Hence, the response object can be accessed like a json format and extract the information in a better way (As shown in the figure). The above code block would extract the basic information about the channel in the following way.

```
{'channel_name': 'techTFQ', 'subscribers': '73200', 'views': '3284747', 'videos': '57'}
```

How to get channel_id?

Open any YouTube channel page. The part after channel/*** in the URL is the channel id we are looking for.

youtube.com/channel/UCnz-ZXXER4jOvuED5trXfEA

EXTRACTING VIDEO INFORMATION BASED ON SPECIFIC TOPIC:

More_information.py can be executed by replacing the API key in the code to generate data in csv format. The python script helps in extracting video information based on the provided keyword. Due to restrictions on number of videos can be extracted through API, the user can only pull 50 video information's at a single request. YouTube API has restrictions like quotas (quota value for each method/functionality called) and maximum of 50 video information at a time to be pulled.

```
response = search(youtube, q="ukrainewar", maxResults=50) #use covid19,etc.
items = response.get("items")
data=[]
```

The above code block in the python script can be used to extract contents under specific topics of interest. The parameters: 'q' can be replaced by 'topic' which user wants to extract information and 'maxResults' can be replaced by number of videos need to be pulled.

Below is the example video information retrieved by the API call for different topics like ukrainewar and covid19.

1. Ukraine war video information

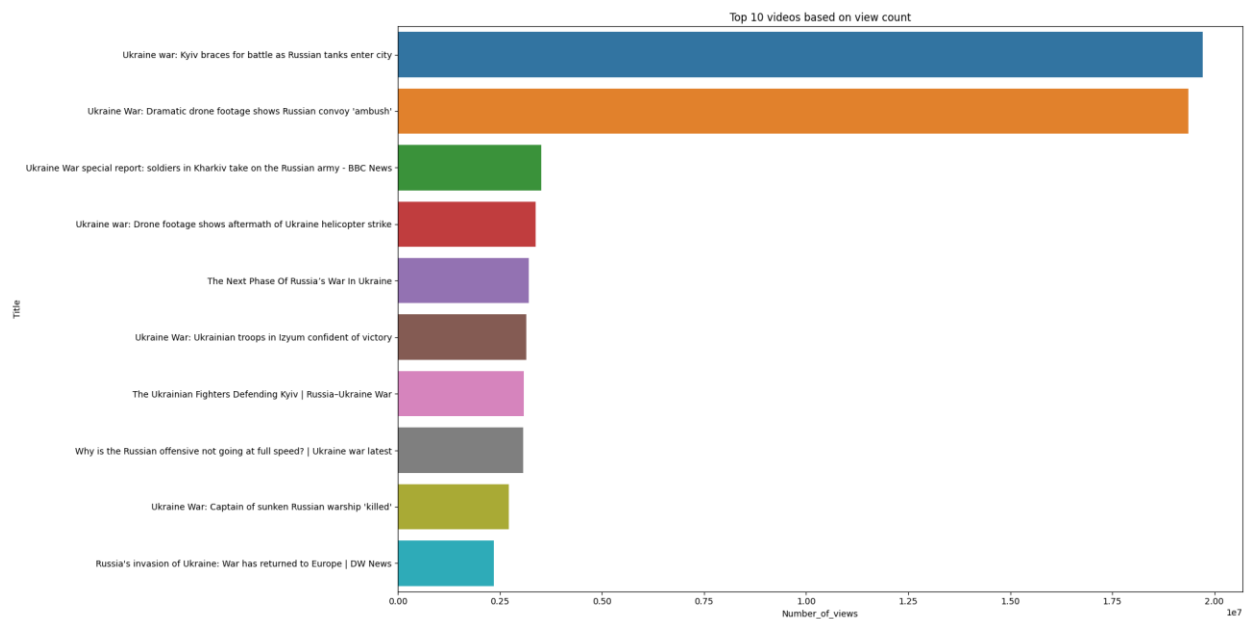|	A	B	C	D	E	F	G	H	I
		Channel_1	Title	Descriptio	Publish_d	Number_c	Number_c	Number_c	Video_url
0		BBC News	Investigati	Newsnigh	########	1520435	17624	2799	https://www.youtube.com/watch?v=sTQ5ZGHV9Zs
1		The New Y	Under Att	Our	########	2326484	25143	4850	https://www.youtube.com/watch?v=ODU_QasrkcQ
2		6abc Phila	US warnin	Ukraine	########	306912	1517	1451	https://www.youtube.com/watch?v=7dJ3rMG_AZk
3		Sky News	Ukraine W	Military	########	587163	5393	1368	https://www.youtube.com/watch?v=sWrf0WxsHcw
4		Sky News	Ukraine W	Retired	5/8/2022	937744	9890	2604	https://www.youtube.com/watch?v=TP4RFfg_EKU
5		The New Y	The Ukrair	As Russia	########	3087409	33539	5523	https://www.youtube.com/watch?v=gCldVMfFim8
6		PBS NewsI	As Ukraine	Ukraine	########	862526	9567	0	https://www.youtube.com/watch?v=PRzrf-WI1gA
7		CRUX	Russia-Ukr	A	5/8/2022	1009141	10379	6255	https://www.youtube.com/watch?v=qgWYBfrhTls
8		VICE News	The Next F	Note:	5/6/2022	3202288	69664	13485	https://www.youtube.com/watch?v=IvNc9yDox8U
9		DW News	Ukraine w	US	########	999653	8583	2626	https://www.youtube.com/watch?v=gDhldXoVvlU
10		Sky News	Ukraine W	President	5/9/2022	806229	9713	2048	https://www.youtube.com/watch?v=MXa0JVgpVtE
11		DW News	Russia's in	Ukrainian	########	2350504	15503	6127	https://www.youtube.com/watch?v=YtqAQIOfpWY
12		Sky News	Ukraine W	Countries	########	771987	10984	2481	https://www.youtube.com/watch?v=jNBPdFqb6fY
13		BBC News	Ukraine W	Ukraineâ	########	3508227	39283	6891	https://www.youtube.com/watch?v=bfvFKUmoGrM
14		Sky News	Ukraine W	Drone	########	19370193	79960	6692	https://www.youtube.com/watch?v=dcYOjbyttvM
15		Channel 4	Russia Ukr	Ukrainian	5/7/2022	465821	3876	1676	https://www.youtube.com/watch?v=sFRs860L6ZI
16		ABC News	Finland to	While	########	169288	1660	1429	https://www.youtube.com/watch?v=i9RbE8rh-bw
17		Sky News	Ukraine W	Sky	########	930642	8469	1148	https://www.youtube.com/watch?v=E5Bibr9nR7Q
18		CBS Evenir	First Russi	The first	########	97806	1180	531	https://www.youtube.com/watch?v=MgdBG6s3ooY
19		Sky News	Ukraine W	Voluntee	4/9/2022	1731318	17563	2724	https://www.youtube.com/watch?v=vXkaoO_dXUQ
20		DW News	Why is the	The	3/8/2022	3067889	23839	5420	https://www.youtube.com/watch?v=paKweV2pdJs

2. Covid related information

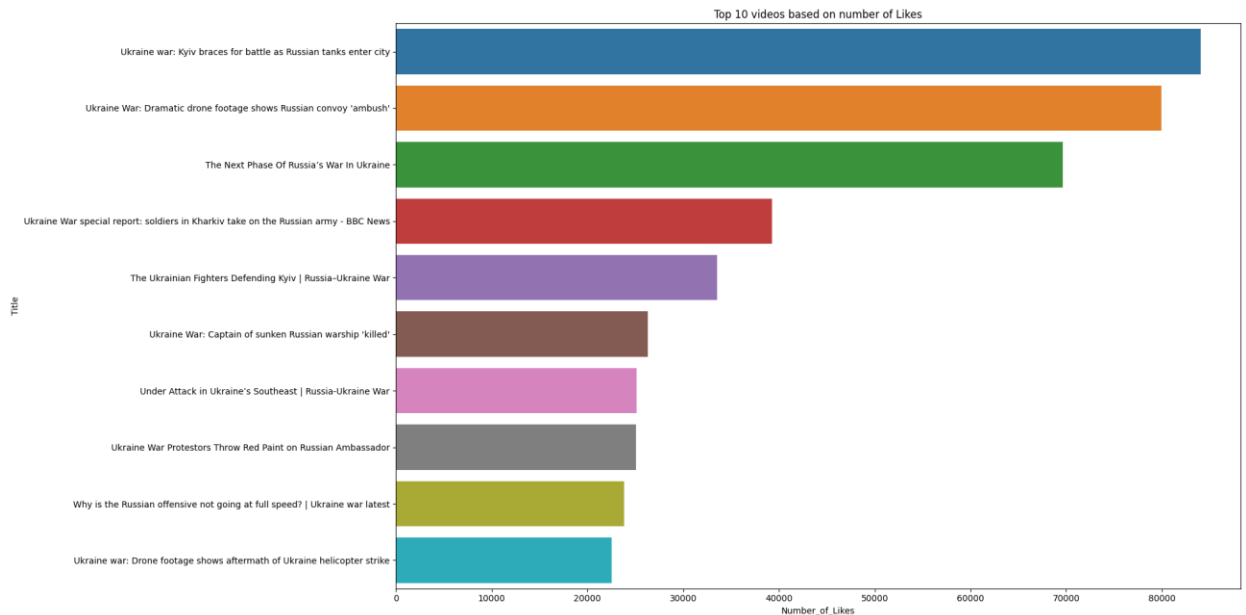|	A	B	C	D	E	F	G	H	I	J	K	L	M	N
		Channel_T	Title	Descriptio	Publish_d	Number_c	Number_c	Number_c	Video_url					
0		DW News	North Kor	North	########	111147	1106	434	https://www.youtube.com/watch?v=k1E6fGyg3LE					
1		ABC News	Breakthro	Breakthr	########	127943	1247	894	https://www.youtube.com/watch?v=wYdB1xydpBU					
2		ABC News	Growing p	Many	########	61915	560	696	https://www.youtube.com/watch?v=5ftewn-T51o					
3		Nucleus M	COVID-19	For	########	3.39E+08	1174363	50936	https://www.youtube.com/watch?v=5DGwOJXSxqg					
4		TODAY	Contagiou	COVID-19	5/7/2022	55411	569	317	https://www.youtube.com/watch?v=DkbXt4Akc8A					
5		ä¸–å¤©é›»	ã€ä¸–å¤©©(	#COVID1	########	9373	288	1	https://www.youtube.com/watch?v=ADwdk-MIlyU					
6		DW News	The hidde	How	########	233126	2668	791	https://www.youtube.com/watch?v=rJfwN-XhSiw					
7		ABC News	COVID-19	There are	5/8/2022	39079	442	416	https://www.youtube.com/watch?v=TnnbvoNh124					
8		World Hea	Coronavir	What do	########	1706168	16154	1190	https://www.youtube.com/watch?v=i0ZabxXmH4Y					
9		Polimer N	Today Hea	Today	########	30152	349	17	https://www.youtube.com/watch?v=MvLartjbAFE					
10		ABC News	COVID-19	ABC	5/5/2022	20351	192	253	https://www.youtube.com/watch?v=ZxxGDhNLF8w					
11		MedCram	Update 16	Roger	########	107916	3853	698	https://www.youtube.com/watch?v=6sdvw_0R240					
12		detikcom	Epidemiol	Tonton	########	3372	11	20	https://www.youtube.com/watch?v=fpeGmQYxz78					
13		TODAY	COVID-19	A	########	19452	684	99	https://www.youtube.com/watch?v=V467lwb8DqY					
14		ABC Actio	Health off	Top health	########	2500	37	73	https://www.youtube.com/watch?v=Xd5rXp8CwXX					

ANALYSIS OF DATA:

This is a basic analysis of knowing what the top 10 videos are based on the views, likes and comments. Like other codes, this script can also be used to for future analysis by just replacing the dataset associated with the scraping.
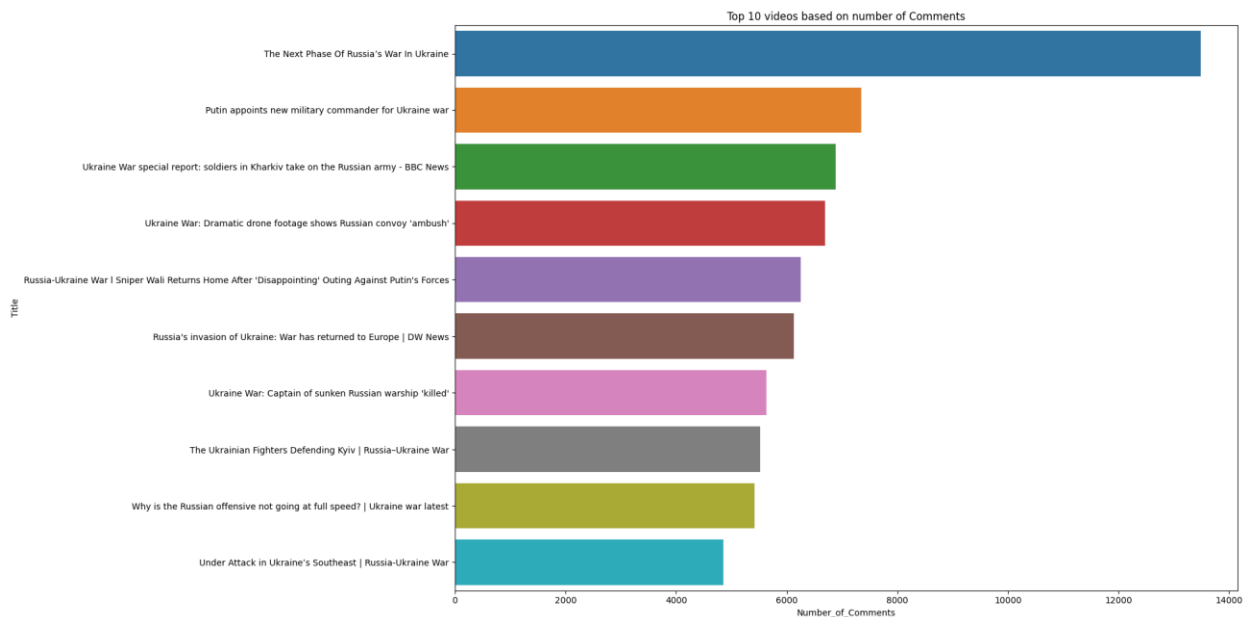
1. Based on View count

2. Based on likes

Top 10 videos based on number of Likes

3. Based on comments

Top 10 videos based on number of Comments

DASHBOARD OVERVIEW:

Designed application is a simple dashboard which helps the user to analyze their data. The navigations in this app are user friendly and the components in the dashboard are in preliminary stage. The idea behind creating this was to explore how an interactive app can be created for future projects.

- Get the path of the dataset and replace it at the following part of Dashboard.py file.

```
#Load dataset - replace the data file path according to the scraped data
data=pd.read_csv('data.csv')
```

- Execute the python script like other python scripts in this project and click on the url generated in the console to launch the app.

```
 * Running on all addresses.
   WARNING: This is a development server. Do not use it in a production deployment.
 * Running on http://192.168.1.129:8033/ (Press CTRL+C to quit)
```

<u>Navigating through the app:</u>

First tab: Basic information about the dataset. Has options to view column names from dataset, first few rows of dataset and channels & number of videos extracted for each channel based on the keyword used. The user can visualize these fundamentals of the dataset by clicking on the appropriate radio buttons.

Dashboard for Analysis

Basic Information	Top videos

Click one option to understand the basic information about the data!

⦿Column Names ○Display data ○Channels and their video count

```
Unnamed: 0
Channel_Title
Title
Description
Publish_date
Number_of_views
Number_of_Likes
Number_of_Comments
Video_url
```

Second tab: This tab has an option to get a revised version of the dataset. User can select based on what column the dataset needs to be sorted (descending) and get it downloaded. For example: Sorting based on 'Number of views' will download the sorted data based on top views.

Dashboard for Analysis

Basic Information	Top videos

Select an option to sort the dataset accordingly!

Number_of_Comments	× ▾

Click to download the sorted dataset!

Download CSV

FUTURE SCOPE:

- Other way of extracting the information is using authorized client ID. Creating credentials using '**Create an OAuth 2.0 client ID**' to receive requests from authorized users**.** For example, you need user authorization to retrieve information about the currently authenticated user's YouTube channel.
- Using more parameters to pull requests. For example, extracting comments for the video. The comment extraction methods were attempted but due to time crunch and recent deprecations in API methods it was not completely implemented. Hence, it was not discussed in the report
- Using NLP techniques to scrape more information from video descriptions and perform sentiment analysis on the data to understand the trends in topics discussed.
- Adding more layouts and analysis to the dashboard.

REFERENCES:

- https://developers.google.com/youtube/v3/quickstart/python
- https://stackoverflow.com/questions/63948396/issue-when-trying-to-register-an-app-for-consent-screen-in-google-oauth
- https://www.thepythoncode.com/article/using-youtube-api-in-python
- https://developers.google.com/youtube/v3/docs/channels
- https://www.youtube.com/watch?v=SwSbnmqk3zY