Data Collection from Social Networks

**Twitter:** Method1

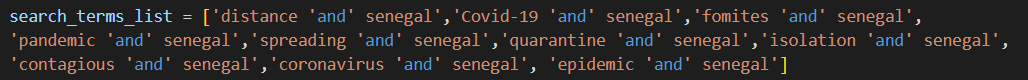
Collecting content from Twitter required obtaining API token access with no restriction to the size of datasets at a time. One requires to have a Twitter account, and then one can apply for developer rights. Having developer rights gives one access to create a web application that allows you to access content from your account for further use. Twitter has limits on data access for each level of developer rights. Basically, all users can obtain 200,000 tweets per month. In order to increase access, one applies for an elevated account that has no limitations to access. We were able to use an account with no monthly restrictions to access public data but had a daily limit of 1500 records. We used the bearer token generated to access user tweets. The steps followed are demonstrated using the algorithm below;

1. State the bearer token



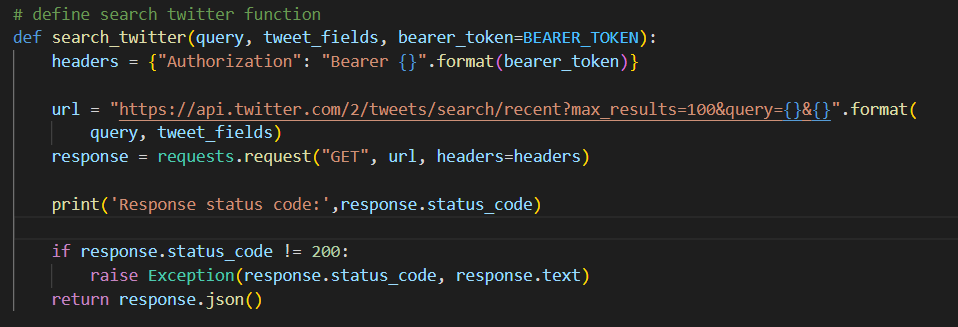
1. Identify the search words

* We used the “AND” logical operator to specify the input string to obtain records about Senegal



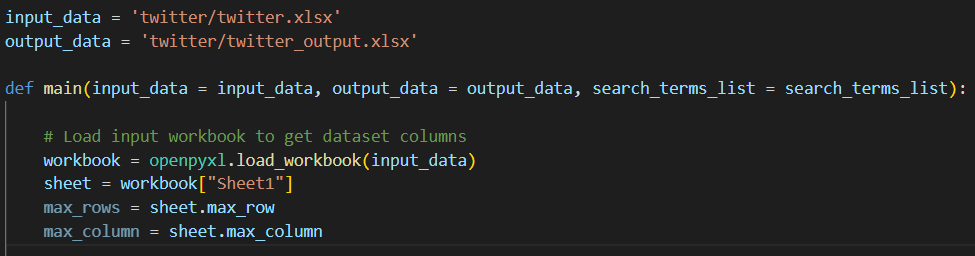
1. Define the procedure to create a query to get tweets using json format

* We set a format of the responses to the json web content



1. Define a procedure to use existing fields of interest to get a dataset with column headings

* We identify an Excel workbook and set column headings. The json data will then be allocated to each of these columns automatically to generate an output file from Excel.
* load input workbook to get dataset columns



* We define a dictionary to hold the column names, state the API response fields we are interested in, and search through all terms in the search terms list for each column.



The dataset file had 1500 records, and it was stored as an Excel file. The columns were automatically computed. We can go ahead to rearrange the columns and remove those not needed for the task. By this method, many records were observed to lack the common keyword of COVID-19. There is need to add more filters to obtain good results.

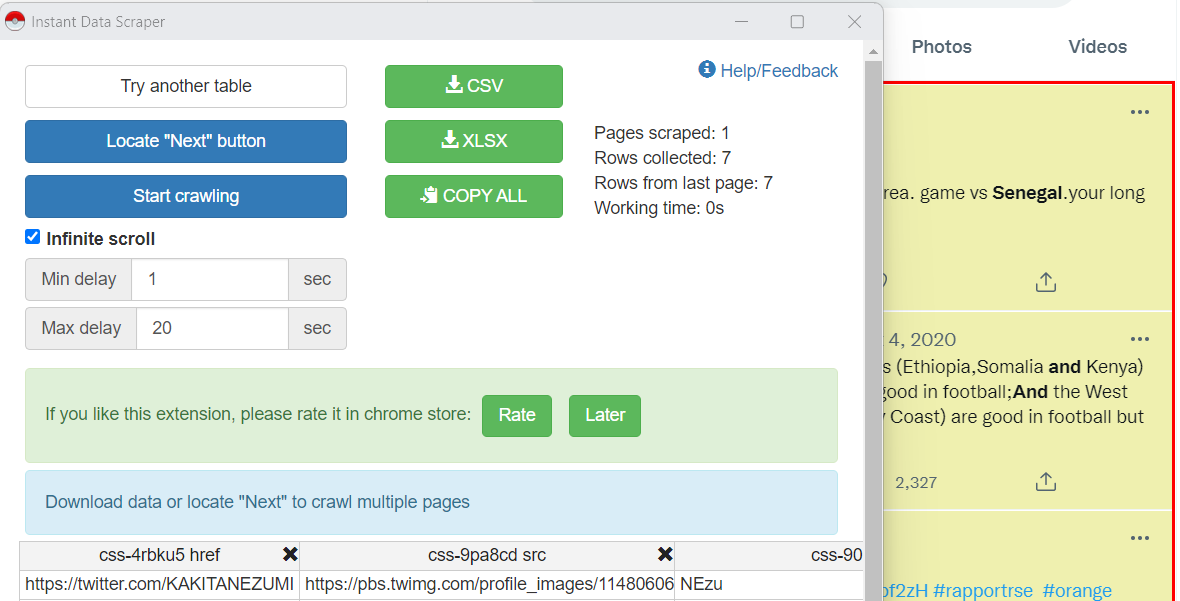
Method 2

In order to increase the data obtained, we also used one web crawler extension named Instant Data Scrapper, among others like Data Scrapper, Web Scrapper, and more. And so we used the web crawler to collect more data to obtain datasets with higher numbers of records. The steps taken include;

1. Log into your Twitter account
2. Set one search string to enter into the web page at time for all the ten search words



1. We use the web crawler extension called Instant Data Scrapper to extract data. You will need to highlight the sector which contains the comments to extract.



We start crawling the data from the highlighted sector. A file is then downloaded as either csv or xlsx file. The process was repeated for each search string and files for all ten search words are aggregated into one. The total number of records obtained from the merged file was 2066.

To increase the volume of records, the process of data collection can be repeated for a number of days since new comments are obtained after each second of a day. For our study, we maintained both files and aggregated the columns into one.

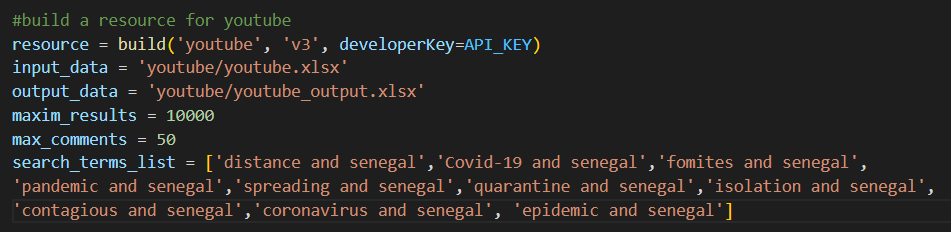
**YouTube:** Method1

Collecting data from the YouTube tool required obtaining a YouTube API key. This requires one to have an email with Google. One can then search for [developers](https://developers.google.com/youtube/v3) from the Google platform. After opening a project, one will have access to the API key. A number of steps that follow are included below:

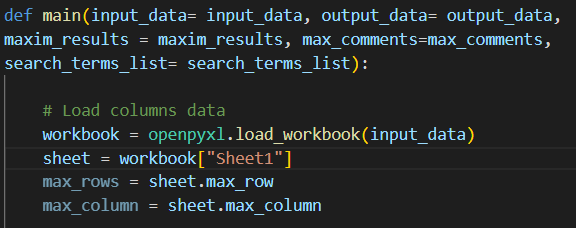
1. Copy your API key and store it in a txt file



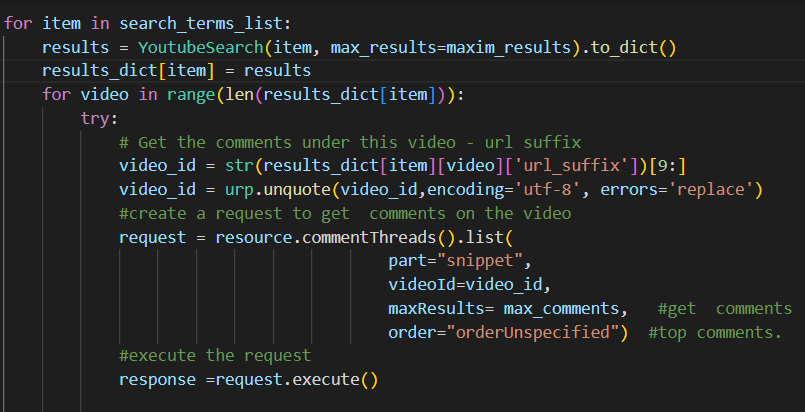
1. We define the parameters needed to request the data using the API key. This includes highlighting the input data file that contains column titles, an empty output file, setting the maximum number of 10000 comments and 50 to be the maximum number of replies



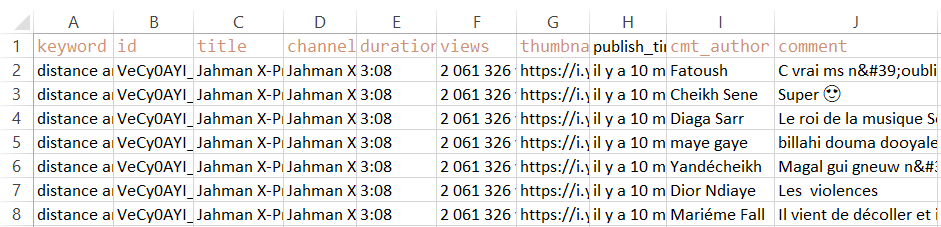
1. We set the request to collect data in the json format. The content is then converted to excel format using openpyxl library. The dictionary of the



1. We then create a result dictionary and perform another request for the top most common comments about each comment in the result.



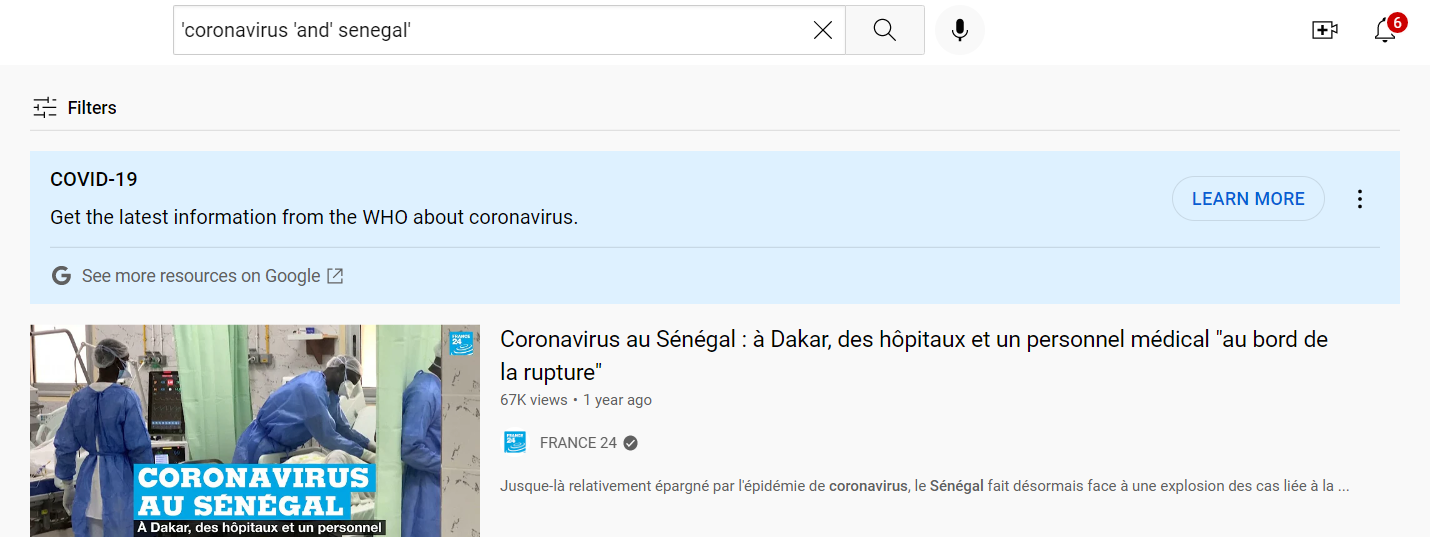
We then automatically add all content to their respective column titles. The dataset had 21035 records. By this method, many records lacked the common keyword of Covid-19. This could be solved by increasing the filters to eliminate irrelevant records.



**Method 2:**

In order to increase the data obtained, we also used one web crawler extension named Instant Data Scrapper, among others like Data Scrapper, Web Scrapper, and more. And so we used the web crawler to collect more data to obtain datasets with higher numbers of records. The steps taken include;

1. Using Google search, type YouTube videos and locate the official website
2. Set one search string to enter into the web page at a time for all the ten search words



1. Using the web crawler extension called Instant Data Scrapper to extract data. You will need to highlight the sector which contains the comments to extract.
2. We start crawling the data from the highlighted sector. A file is then downloaded as either csv or xlsx file. The process is repeated for the search string and files for all ten search words are aggregated into one. The total number of records obtained from the merged file was 3661. The records were fewer than those of the prior method, but most of the records had the keyword of Covid-19. For our study, we maintained both files and aggregated the columns into one.

**Facebook:** Method1