**Report**

While doing this assignment I faced a lot of issues , firstly because of personal issues such as the fact that I am from a non-IT background .I had very little to no previous knowledge about the subjects that I am leaning in this course. Secondly, I could not find anyone to form a group with so I had to do this individually and figure things out by myself.

On the technical front, firstly an error related to tweepy was showing up many a times , I tried a lot of measures to resolve that like uninstalling and reinstalling it , trying updating it , starting the whole thing from scratch , setting a limit etc .

Secondly, my twitter developer account suspended for some reason which I am not aware of , then I had the credentials provided in the blackboard and it worked out.

Then finally somehow I was able to resolve that , later when finally everything was fine , the code was alright , no errors showed up but yet it was not running ,I tried rewriting the files and some new ones as well ,but still I was unable to figure out the issue because there did’nt seem to be any issue with the code to me . Then later some time I tried again executing the same thing which was not running earlier , this time it did , the tweets showed up and I concluded that maybe due to some technical issue or maybe just the fact that sometimes it takes time to load , was the only reason behind the delay. That’s how I did the twitter part .

1. Installation and setup of zookeeper:

Text

Description automatically generated

2.apache-kafka

Text

Description automatically generated

3.programming files

Graphical user interface, text

Description automatically generated

4.tweets showing up

Text

Description automatically generated

Cleaning and processing data

Firstly error showed up because maybe “date” word is reserved by hive so I can’t use that for creating column name , so I modified it to “tweet\_date” and it worked :

Text

Description automatically generatedGraphical user interface, text

Description automatically generated

For convenience I dropped this table and changed the table name to BigData\_tweets

Graphical user interface, text, application

Description automatically generated with medium confidence

I cleaned the tweets using the following 4 commands:

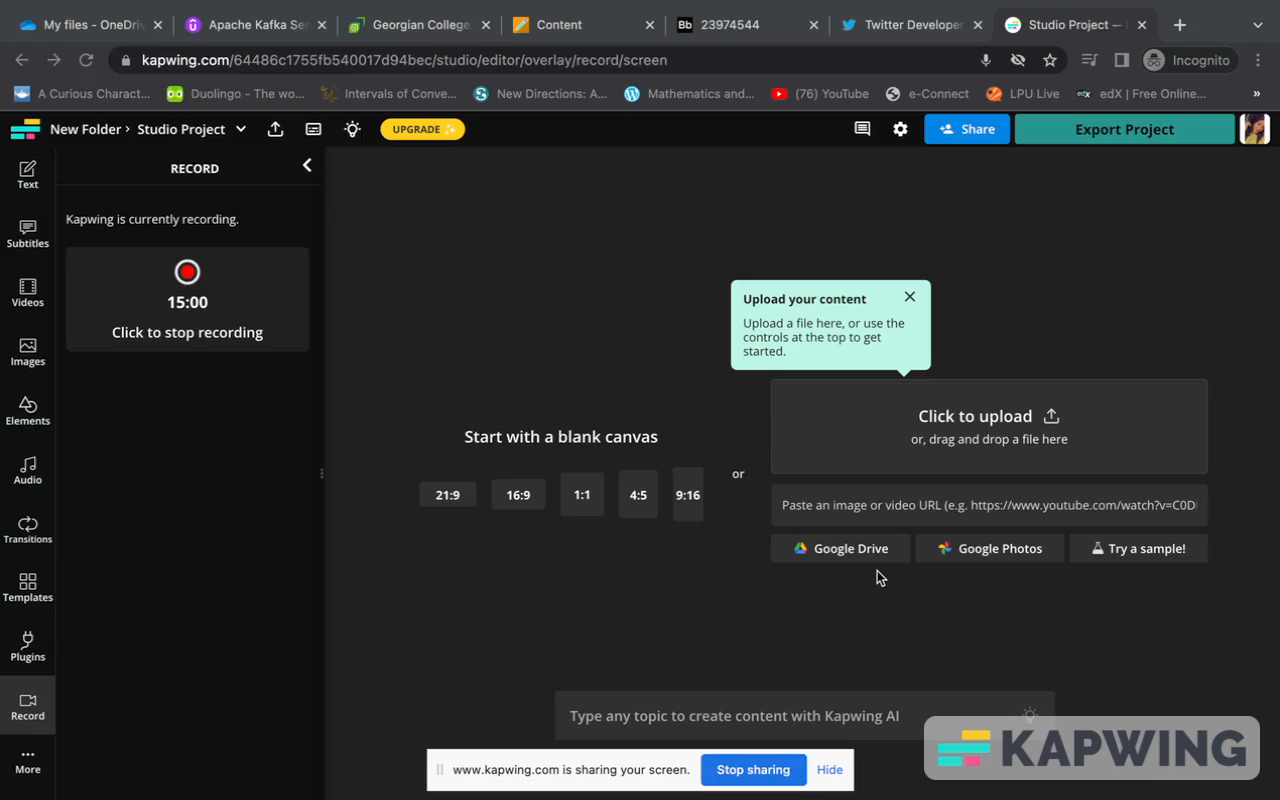
Text

Description automatically generatedText

Description automatically generated

**RECORDINGS:**

(double click to view):



****

Programming Files screenshots:

Consumer.py:

Graphical user interface, text, application

Description automatically generated

Producer-twitter.py:

Graphical user interface, text

Description automatically generated

Write-to-file.shGraphical user interface, text, application

Description automatically generated

Secrets.pyGraphical user interface, text, application

Description automatically generated