**AML-3204 Social Media Analytics**

**Assignment-1 (20%)**

**Analyzing YouTube Data**

In this assignment, you will write a Python script to download user interaction data from YouTube for each “youtubeId” provided in the “vdoLinks.csv” file. For each of the YouTube video id, you will extract the following information:

1. 100 comments
2. Description of the video
3. View Count
4. Like Count
5. Dislike Count
6. Comment Count
7. Duration of the video
8. Favorite Count

While extracting the data, if any of the “youtubeId” does not work, please ignore that ID and proceed with the following “youtubeId”.

After the data extraction is done, then please find out the following:

1. A list of the top-10 videos based on the total views
2. A list of the bottom-10 videos based on the total views
3. The most liked video
4. The least liked video
5. The video with the highest duration
6. Apply sentiment analysis on the downloaded comments for each of the videos. For sentiment analysis, you can use VADER or any other tool of your choice. Your program should list the sentiments score for each of the movies

Data from YouTube can be extracted in many ways. You can use any of the available approaches. However, I have found the following tutorial is handy:

<https://www.geeksforgeeks.org/how-to-extract-youtube-comments-using-youtube-api-python/>

**Submission:**

To submit, please do the following:

1. Write a short report on less than 12 pages for the assignment. The report should have the followings:
   1. Data collection procedure
   2. Data cleansing steps
   3. Using a bar diagram, show the top-10 videos based on total views
   4. Using a bar diagram, show the bottom-10 videos based on total views
   5. The title of the video that has the most likes
   6. The title of the video that has the least likes
   7. The title of the video that has the highest duration
   8. Using a bar diagram, show the top-10 videos that have the highest positive sentiment scores, which you calculated from the comments
   9. Make sure to include the proper explanations, analysis of results etc in your report as practiced in other courses
2. Please upload the report in MS word file and the Python file (ipynb)