BDP - HW 4
Deadline - 12 April, 2019

Total Marks - 140

Instructions
1. Preferably use **Python3** for this assignment.
2. You'll have to collect your **own dataset** for this assignment, so please start early.
2. The final submission must consist of all your source code, data collected and a **PDF report** containing all your findings.
3. In case of any doubts/issues, please post on the Moodle forum.
4. Do not copy. This could result in an **'F'** grade in the course.

Part 1 (60M)

Use Twitter's official API only. No other method is allowed for this part.

1. Collect 3,200 most recent tweets from the BJP's official Twitter handle
    a. Which tweet has the most likes, retrieve it's IDs and post a screenshot of the tweet (5)
    b. Which has the most retweets, retrieve it's IDs and post a screenshot of the tweet (5)
    c. List the top 5 most popular tweets. Devise your own algorithm to calculate popularity based on likes, retweets or any other metric. (10)
    d. List down the top 5 most used hashtags by the BJP's official handle. (10)

2. Collect 3,200 most recent tweets from the Indian National Congress' official Twitter handle.
    a. How many tweets consist of images? Plot a pie chart representing the share of images in the tweets as compared to videos and text-only tweets. (10)
    b. Using the same algorithm devised earlier to calculate popularity, plot a word cloud of the top 1000 most popular tweets. Which words occur most frequently and why? (10)
    c. List down the top 5 most used hashtags by the INC's official handle. (10)

Part 2 (80M)

Using any tool of your choice, 'scrape' data from Instagram for this part of the assignment.

1. Collect 500 most recent posts from the Indian National Congress' official Instagram handle.
    a. Which post has the most likes, and which has the most comments? Retrieve it's IDs and post a screenshot of the post in the final report. (10)
    b. Show the top 5 most popular posts on the report. Devise your own algorithm to calculate popularity based on likes, comments, views or any other metric. (10)
    c. List down the top 5 most used hashtags by the INC's official handle. (10)

2. Collect 1,000 most recent posts from the BJP's official Instagram handle.
    a. How many posts consist of videos? Plot a pie chart representing the share of images in the tweets as compared to videos. (10)
    b. Using the same algorithm devised earlier to calculate popularity, plot a word cloud of the top 200 most popular posts' descriptions. Which words occur most frequently and why? (10)
    c. List down the top 5 most used hashtags by the BJP's official handle. (10)

3. Collect 3,000 posts from Instagram's 'explore' feed.
    a. Which are the top 5 most popular posts? Post screenshots of the posts in your report (10)
    b. Plot a pie chart comparing number of posts with just a single image/video in it, and number of posts with multiple images/videos in it. (10)