**Assignment - 1 of Dialog Systems Course**

**Data link:** https://github.com/mounikamarreddy/Dialogsystems-Course-Assignments/tree/main

**Data description:** Go to the folder “Assignment1” from the above Github link and unzip the subfolder named "subredditlinks.zip", then you will find various files with subreddit names. Each subreddit file consists of few links. Please choose a subreddit file of your preference. One can use all the subreddits links as well. Each link within the file corresponds to a conversation. Utilize the "PRAW" Python library to crawl through these conversations.

The structure of each conversation is as follows:

- Submission Title: The initial post

- Submission Content: Body of the initial post

- Direct replies to the submission

- Nested replies to the submission

**Task description:** Each student is required to complete the following steps to gain an understanding of exploratory data analysis:

1. Utilize the PRAW library to crawl through these conversations from the selected subreddit.

2. Present the statistics for conversations categorized as only text-based content or containing images and videos.

3. For all conversations that you consider:

1. Determine the mean and standard deviation regarding the number of posts.
2. Identify the number of unique users within each conversation.
3. Display the percentage of posts contributed by the top 2 users in each conversation, and calculate the average percentage of posts across all conversations.
4. Calculate the average number of tokens across posts, the average number of posts per submission, and the average number of tokens combining submissions and posts within a conversation.

**Note: Use the given ipython notebook, fill the cells accordingly to get the desired result.**