Assignment:

You are requested to scrap 50 profiles from Instagram fitting the brief given below:

- Target Profiles: Keyword "Fitness Coach" in their Bio
- Geography: USA
- Followers: 50,000+

Objective: The goal of this project is to extract Instagram profiles of fitness coaches/trainers located in the USA and having more than 50,000 followers.

Methodology: To achieve this goal, we will follow the following steps:

1. ***Filtering Profiles:*** Firstly, we will use advanced Google search to filter out the profiles based on the website, location, and the keyword "fitness coach" or "fitness trainer". This will help us to obtain a list of Instagram usernames for fitness coaches and trainers located in the USA.
2. ***Scraping Profiles:*** Next, we will use web scraping techniques to extract the Instagram profiles from the Google search results. We will extract the usernames from the search results and store them in a list.

After webscraping the profiles from Instagram and storing them in a list of dictionaries, the next step is to convert this list into a pandas dataframe for further analysis and processing.

This can be done using the pandas library in Python.

Once the list is converted to a dataframe, we can remove any duplicate profiles using the drop_duplicates() method.

This will ensure that we have a unique list of profiles for our further analysis.

After removing the duplicates, we can extract the profile names from the dataframe and store them in a list format.

This list of profile names can then be used for further processing such as checking their follower count, bio, and location.

3. **Automating the Task**: Once we have the list of Instagram usernames, we will automate the task by,
    - logging into Instagram
    - searching for the username
    - checking the bio for the keywords "fitness coach" or "fitness trainer"
    - verifying if the profile is located in the USA,
    - checking if the profile has more than 50,000 followers.

4. **Saving the Results:** For each profile that meets our criteria, we will extract and save the username, bio, and number of followers in a text file.

Future work: The texts files can be read and emails can be extracted easily for further business process.