Assignment: Web Scraping and Data Extraction

Objective:

The goal of this assignment is to develop a Python script that scrapes job data from the provided website and organizes it into a structured format.

Website to Scrape:

https://realpython.github.io/fake-jobs/

Data to Extract:

Your script should extract the following information for each job listing:

1. **Job Title:** The title of the job.
2. **Company Name:** The name of the company offering the job.
3. **Location (City):** The city where the job is located.
4. **Location (State):** The state where the job is located.
5. **Date Posted:**
   - The original date string (e.g., "2021-04-08") should be converted into a Python `datetime` object.
   - From the `datetime` object, extract the following:
     - **Day of Week, Day and Month (e.g., "Thursday, 8 April")**
     - **Year (e.g. 2021)**

Data Output:

The extracted data should be organized into a structured format, such as a Pandas DataFrame or a list of dictionaries.

Specific Requirements:

- Use Python for web scraping.
- You are free to use libraries such as `requests` and `Beautiful Soup`.
- The date posted must be converted to datetime object.

- The Date posted must be further parsed to create two additional columns. One containing the day of week, day and month, and the second column containing the year.
- Ensure your code is well-documented and easy to understand.
- Provide clear comments explaining the logic behind your code.
- The final result must be clearly outputed, either as a dataframe, or a list of dictionaries.

Example Output (Illustrative):

If using pandas, your dataframe should contain the following columns:

Job Title	Company Name	Location (City)	Location (State)	Date Posted (Day, Month, Day of Week)	Date Posted (Year)
Software Engineer	Lux, Narnia	Atlanta	GA	Wednesday, 7 April	2021
Data Analyst	Initech	New York	NY	Thursday, 8 April	2021

Submission:

- Create a GitHub repository for your project.
- Push your Python script to the repository.
- Include a `README.md` file in your repository that provides:
  - A brief description of the project.
  - Instructions on how to run the script.
  - Submit the link to your GitHub repository.

Submit the outputted data, either as a CSV or a printed dataframe alongside the link to your repository.

Evaluation Criteria – Rubric:

- Accuracy of data extraction - 30.

- Correct date conversion and formatting - 10.
- Code clarity and documentation - 20.
- Efficiency of the script - 30.
- Adherence to the output format requirements - 10.