Summer 2022 Programming Challenge

Estimated Duration: 1.5 - 2 hours
Deadline: Wednesday, June 8, 2022 at 11:59 PM PST

Objective

The goal of this programming exercise is to demonstrate your ability to design a solution to a problem and implement this solution in **Python** using software engineering best practices.

The specific task will be to create a pipeline that collects and analyzes news articles from the web. Given a short list of curated websites, your script should be able to collect the latest news articles (via web scraping) and run them through some basic sentiment analysis.

Requirements

1. **Create** a **GitHub** repo with a **Python 3.8+** environment for this project and start a **requirements.txt** file to capture the external libraries/packages required to run your code. If you use a virtual environment such as conda, specify that in the summary (see step 6). This repository is where you can upload all the files pertaining to your submission.

2. **Collect 10 most recent articles** from https://www.aljazeera.com/where/mozambique/ Include collected articles as a JSON file in your submission repository. The format of the file is up to you, describe this format in your summary (see step 6).

3. **Pre-process** the data. Remove anything that is not part of the article itself, e.g. comments, publishing date, images, etc. Make sure the articles are in English and can be processed by the sentiment analysis library. Use the tqdm package to display progress on the terminal. Use PEP8 Style Guide for your python code.

4. **Compute** the sentiment of each news article. We encourage you to use an off-the-shelf library, but you may create your own if you feel it is appropriate. Document your choice of sentiment analysis approach in your summary (see step 6).

5. **Visualize** the results by plotting the sentiment using the plotly visualization library. Please include a screenshot of this plot in your submission repository.

6. Create a **README.md** file with a **summary of your results** and provide high-level documentation of your code as well as instructions on how to run your code in order to reproduce the results. Include the total operation time of your code.

When you're done, please submit a link to the GitHub repository containing: your solution, a screenshot of the resulting visualization from Requirement 5, and all supporting documentation. Please submit this link at https://forms.gle/S3UpU9niHfsL7rcu9