Problem Statement

Design a web scraper to read articles off theverge.com using Python

The script should be able to perform the following:

- Read the headline, get the link of the article, the author, and the date of each of the articles found on "theverge.com"
- Store these in a CSV file titled `ddmmyyy_verge.csv`, with the following header `id, URL, headline, author, date`.
- Create an SQLite database to store the same data, and make sure that the id is the primary key
- Run this script on a cloud service (preferably AWS)
- Save the articles (and de-duplicate them) daily on the server in a SQL Database.

Test Csv File

id,url,headline,author,date
0,https://www.theverge.com/2022/4/3/23008668/tesla-shanghai-factory-closed-lockdown-covid-china,Tesla's Shanghai factory stays closed as COVID restrictions remain in place,Emma Roth,2022/4/3
1,https://www.theverge.com/2022/4/2/22999741/fortnite-chapter-3-season-2-building-returns-zero-build-mode,Fortnite brings back building,Andrew Webster,2022/4/3
....
....
37,https://www.theverge.com/2022/3/31/23004599/activision-blizzard-overwatch-anniversary-event,Overwatch sixth anniversary event offers 'remixes' of popular skins,Ash Parrish,2022/3/31

Coding Guidelines

- Please write modular code using best OOP practices
- Share your code on GitHub publicly, to make sure we can review it.
- Add test cases to catch nefarious bugs in the code

P.S. If you prefer something other than python, please feel free to use that coding language. Lastly, we are not affiliated with The Verge and are using this website just to test your coding skills; we will not be using the output of this test in any form whatsoever.