Author: Jered Kalombo
Task: Job Scraping - Live and Work in Maine
Script: *PhaseTwo/Scraping/liveandworkinmaineScraper.ipynb*
Data: *PhaseTwo/Data/liveandworkinmaine_jobs.csv*

Methodology:

- Used the requests and BeautifulSoup packages to scrape job postings directly from the Live and Work in Maine careers site.
- Collected job titles and URLs from the main search page, then visited each job detail page to extract additional attributes.
- Extracted key fields including posted date, street address, city, state, zip code, country, and industry.
- The scraper navigates through HTML elements such as *<div class="title">* for job links and *<span id=" lblOutAddress">* for address information, parsing *
* tags to separate lines.
- Each job entry is appended to a structured dataset, then exported as a CSV for further cleaning and analysis.
- Since the site lacks a public API, the scraper is rate-limited by inserting short pauses between requests to avoid overloading the server.

Notes:

- The *"posted date"* and "*industry*" fields are only present on individual job pages, not the main listings, so the scraper must fetch each page separately.
- Address formats vary slightly between listings, requiring flexible parsing logic for city, state, and Zip extraction.
- If the site changes its HTML structure, minor code updates may be needed.
- The scraper also avoids duplicate listings and skips incomplete listings.

Results:

- Most listings included complete address information and an identifiable industry tag.
- Successfully scraped roughly 100+ current job listings from the *Live and Work in Maine* portal.
- Filters out duplicates and outdated jobs.