Author: Ellis
Task: Job Scraping Indeed
Script: `PhaseTwo/Scraping/new_england_indeed_scraper.ipynb`
Data: `PhaseTwo/Data/new_england_indeed_jobs.csv`

Methodology:

- Used package JobSpy from SpeedyApply on GitHub
  - Takes advantage of official Indeed API to avoid rate limiting
  - Job attributes seen here
  - Limited to 1000 jobs per search as mentioned here
- Created a function "get_jobs" which takes an input of the sitename and the states, then loops through each state and appends their results together into one DataFrame which is then returned
- Called function for Indeed and all New England states and got 6K results (1K per state)
- Dropped jobs that were posted before the last 30 days resulting in 5.4K jobs (-600)

Notes:

- The output from JobSpy does not include "expiration date" like our schema suggests, however we will still use these results and keep the attribute. Our rationale is the user should be able to tell if a job is scraped via tags like "External" and jobs posted on the website by recruiters are given the option to provide such information.