Indeed.com Web Scraping

Extract: Available jobs in Virginia were scraped from indeed.com from a series of cities identified by the Bureau of Labor and Statistics site of relevant cities. These cities were looped through with the requests library and BeautifulSoup library for 50 pages, totaling 500 record per city. Only records with both a job title and location were used.

Transform: Both the job title and location where stripped of return characters from the respective parts of the div locations. The job title was cleaned with two functions. The first function cleaned emojis and other odd special characters. The second function cleaned numbers and other special characters. The location was split at the ‘,’ character into two pieces and the first piece was then reassigned to the location variable.

Load: Each complete entry with a job title and a location were pulled into a list, identifying the job. That job is then pulled into a list, creating a list of lists. This was then converted into a pandas dataframe with columns identified and matching the Bureau of Labor and Statistics sheets.