Using Selenium to Scrape Amazon Product Reviews

1. Prerequisite:
   1. Register one Amazon account
   2. Install Chrome Browser
   3. Run your Jupyter Notebook from Chrome Browser
2. Download the following notebooks and copy to your Jupypter Notebook folder:
   1. “01 Install for python selenium package and chrome driver.ipynb”
   2. “02 Amazon reviews scrape.ipynb”
3. Follow the instructions in the “01 Install for python selenium package and chrome driver.ipynb”
4. Scrape Amazon reviews using the “02 Amazon reviews scrape.ipynb”
   1. Before running the codes
      1. Create a folder to store all the scraped reviews
      2. In “3. define the weblist you want to scrape”
         * Change the value of “product\_folder” to the above folder path
         * Change the item value in the list “web\_dict” to the product name and also product reviews link you want to scrape (as like in the previous Amazon Scraping code using the Beautiful Soup)
         * Change the item value in the list “web\_pages” to the product name (the same as above value) and “number of reviews”/10 (this is equivalent to the total review pages)
   2. Run step 1, 2, 3 and 4
      1. When you run the step 4 part 2 for the first time, a new web page will be launched by the code in a new window
      2. You just need to login to Amazon using your account email and password. Leave this page as it is, don’t close it.
      3. After login, rerun the step 4
      4. You will see the page in “i" keep refreshing, this means that the code is simulating the browsing and scrape the reviews from the page.
      5. Just wait until the page no longer refreshing.
      6. Check the “product\_folder”, a new csv file will be created with all the scraped reviews for the product.
   3. Continue to scrape the reviews for another product by changing the values in “web\_dict” and “web\_pages” and rerun the step 3 and 4.