

GROUP ASSIGNMENT

Due: Jan 31, 11:59pm

(1) Do the following IN YOUR BROWSER, no programming required. Please write down your answers (concise written answers please) (20pts).

Use your browsers development tools. Open the network tab and analyze the network for the following:

- a) go to <https://www.ebay.com> and search for "iphone 6s"
- b) what type of search request is eBay using, GET or POST? (2 pts)
- c) which URL variable represents the search term? (2 pts)
- d) can you come up with a shorter URL that produces the same search result page? (4 pts)
- e) click on the next search result page and observe how the URL changes. what variable in the URL identifies the page number? (2 pts)
- f) what is the feature common to each item in the search results page? i.e., what item do i need to select to obtain each item among the search results? (4 pts)
- g) identify the "SPONSORED" search results. what do they have in common? how does it look in the HTML source code? why is there no easy way to identify them? can perhaps regex help? (6 pts)

NOTE: eBay does a bunch of A/B testing. if you don't see "SPONSORED" search results, you may be in their control group. in this case, please change the browser + IP address you are using or ask a group member to work on this together.

(2) Let's program! (40pts)

- a) Use the URL identified above and write code that loads eBay's search result page for "iphone 6s". save the result to file. (Please give it a meaningful filename. E.g., "ebay_iphone6s_01.htm".) (10pts)
- b) Take your code in (a) and write a loop that will download the first 10 pages of search results. save each of these pages. IMPORTANT: Each page request needs to be followed by at least a 10 second pause! remember, you want your program to mimic your behavior as a human and help you make good purchasing decisions. (10pts)

c) Write a separate piece of code that loops through the pages you downloaded in (b) and opens and parses them into a Python or Java xxxsoup-object. Next find the sponsored items on each search result page and print their URL to the screen. (20pts)