

DSC 80 Discussion 6 Worksheet

Name: _____

1 FA23 Final Exam Problem 7

Alan set up a web page for his DSC 80 note with the following HTML:

```
<html><body>
<div id = "hero">DSC 80 NOTES</div>
  <div class="notes">
    <div class="notes">
      <p>Lecture 1: 5/5 stars!</p>
    </div>
    <div class="lecture_notes">
      <p>Lecture 2: 6/5 stars!!</p>
    </div>
  </div>
  <div class="lecture">
    <p>Lecture 3: 10/5 stars!!!!</p>
  </div>
</body></html>
```

Assume that the web page is parsed into a BeautifulSoup object named `soup`. Fill in each of the expressions below to evaluate to the desired string. Pay careful attention to the indexes after each call to `find_all`!

Desired string: "Lecture 1: 5/5 stars!"

`soup.find_all(_____)[0].text`

Desired string: "Lecture 2: 6/5 stars!!"

`soup.find_all(_____)[3].text`

Desired string: "Lecture 3: 10/5 stars!!!!"

`soup.find_all(_____)[1].text`

2 Spring 2023 Final Exam Problem 1

Consider the following line. Choose the regex pattern that, when filled in the blank, will return the desired matches.

`re.findall(r'_____', 'my cat is hungry, concatenate!, catastrophe! What a cat!')`

Output: ['my', 'a']

[] `\b([a-z]*)cat\b`
[] `\b[a-z]*\scat\b`
[] `([a-z]*)\scat\b`
[] `\b([a-z]*\scat)\b`

Output: ['concatenate']

[] `\b.*cat.*\b`
[] `[a-z]*cat[a-z]*`
[] `[a-z]+cat[a-z]+`
[] `\b[a-z]*cat[a-z]*\b`

Output: ['cat', 'concatenate', 'catastrophe', 'cat']

[] `.*cat.*`
[] `\b.*cat.*\b`
[] `\b[a-z]*cat[a-z]*\b`
[] `\b[a-z]+cat[a-z]+\b`

3 WI23 Final Exam Problem 4

Nicole decides to read research papers on data science. While reading these papers, she notices that there are many citations interspersed that refer to other research papers, and she'd like to read the cited papers as well. An excerpt from one such paper is stored in the string `s` below.

```
s = '''In DSC 10 [3], you learned about babypandas, a strict subset
of pandas [15][4]. It was designed [5] to provide programming
beginners [3][91] just enough syntax to be able to perform
meaningful tabular data analysis [8] without getting lost in
100s of details.'''
```

For each expression below, select the list it evaluates to, or select "None of the above."

```
list1 = ['10', '100']
list2 = ['3', '15', '4', '5', '3', '91', '8']
list3 = ['10', '3', '15', '4', '5', '3', '91', '8', '100']
list4 = ['[3]', '[15]', '[4]', '[5]', '[3]', '[91]', '[8]']
list5 = ['1', '0', '3', '1', '5', '4', '5', '3',
        '9', '1', '8', '1', '0', '0']
```

<code>re.findall(r'\d+', s)</code>	<code>re.findall(r'[\d+]', s)</code>	<code>re.findall(r'\[(\d+)\]', s)</code>
<input type="checkbox"/> list1	<input type="checkbox"/> list1	<input type="checkbox"/> list1
<input type="checkbox"/> list2	<input type="checkbox"/> list2	<input type="checkbox"/> list2
<input type="checkbox"/> list3	<input type="checkbox"/> list3	<input type="checkbox"/> list3
<input type="checkbox"/> list4	<input type="checkbox"/> list4	<input type="checkbox"/> list4
<input type="checkbox"/> list5	<input type="checkbox"/> list5	<input type="checkbox"/> list5

4 WI23 Final Exam Problem 5

Suppose `soup` is a BeautifulSoup object instantiated using the following HTML document.

```
<college>Your score is ready!</college>
<sat verbal="ready" math="ready">Your percentiles are as follows:
  <scorelist listtype="percentiles">
    <scorerow kind="verbal" subkind="per">Verbal: <scorenum>84</scorenum></
    scorerow>
    <scorerow kind="math" subkind="per">Math: <scorenum>99</scorenum></scorerow>
  </scorelist>
  And your actual scores are as follows:
  <scorelist listtype="scores">
    <scorerow kind="verbal"> Verbal: <scorenum>680</scorenum> </scorerow>
    <scorerow kind="math"> Math: <scorenum>800</scorenum> </scorerow>
  </scorelist></sat>
```

Which expression evaluates to "verbal"?

- `soup.find("scorerow").get("kind")`
- `soup.find("sat").get("ready")`
- `soup.find("scorerow").text.split(":")[0].lower()`
- `[s.get("kind") for s in soup.find_all("scorerow")][-2]`
- `soup.find("scorelist", attrs="listtype":"scores").get("kind")`
- None of the above