CI328 Introduction to Data Science and Big Data

Final Exam: 3/24/2021 at 12:00 PM (Closed book, Closed Notes, No Computer)

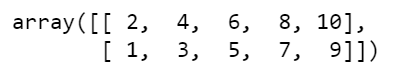
**Final Review**

**Chapter 7**

**Problem 1**

1. **Create a 2-by-5 array containing the even integers from 2 through 10 in the first row by using a list comprehension and the odd integers from 1 through 9 in the second row by using a list comprehension.**

Output





1. **For the two-dimensional array in the Problem 1, display the number of dimensions and shape, type, and size of the array.**

Output

* Number of Dimension





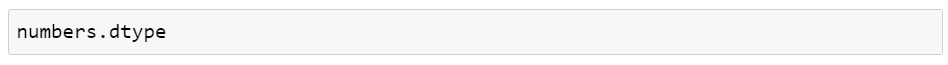
* Shape of Array





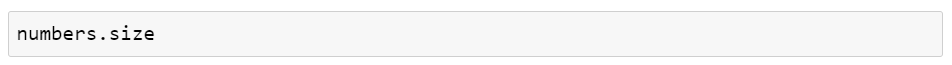
* Element Type of Array





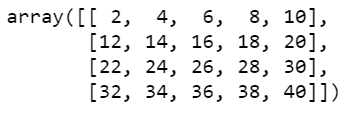
* Array’s Total Number of Elements





1. **Use NumPy function arrange to create an array of 20 even integers from 2 through 40, then reshape the result into a 4-by-5 array.**

Output





**Problem 2**

1. **Use the NumPy’s random-number generation to create an array of five random integers that represent summertime temperatures in the range 60-100, then perform the following tasks:**

**Output**

array([100, 85, 69, 68, 92, 76])



* 1. **Convert the array into Series named temperatures and display it**

**Output**

0 100

1 85

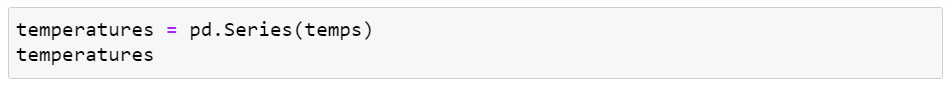
2 69

3 68

4 92

5 76

dtype: int32



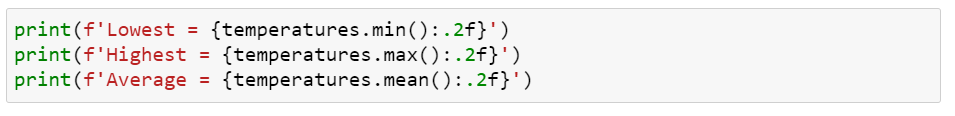
* 1. **Determine the lowest, highest and average temperatures**

**Output**

Lowest = 68.00

Highest = 100.00

Average = 81.67



* 1. **Produce descriptive statistics for the Series.**

**Output**

count 6.000000

mean 81.666667

std 12.909944

min 68.000000

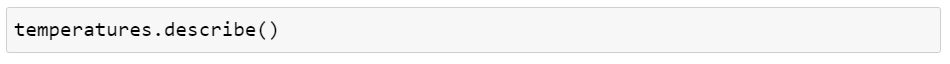
25% 70.750000

50% 80.500000

75% 90.250000

max 100.000000

dtype: float64



1. **Given the following dictionary:**

**Temps = {‘Mon’: [68, 89], ‘Tue’: [71, 93], ‘Mon’: [68, 89], ‘Wed’: [66, 82], ‘Thu’: [75, 97], ‘Fri’: [62, 79]}**

Output

{'Mon': [68, 89],

'Tue': [71, 93],

'Wed': [66, 82],

'Thu': [75, 97],

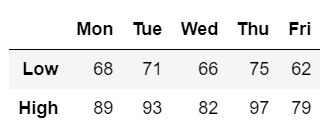
'Fri': [62, 79]}



**Perform the following tasks:**

* 1. **Convert dictionary into DataFrame named temperatures with ‘Low’ and ‘High’ as the indices, then display the DataFrame.**

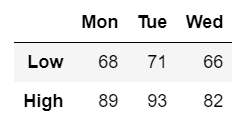
**Output**





* 1. **Use the column names to select only the columns for “Mon through ‘Wed’**

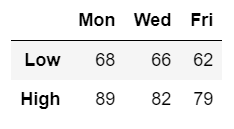
**Output**





* 1. **Use the column names to select only the columns for “Mon ,‘Wed’, and ‘Fri’.**

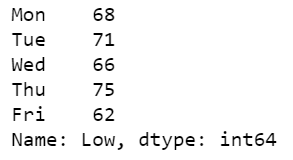
**Output**





* 1. **Use the row index ‘Low’ to select only the low temperatures for each day**

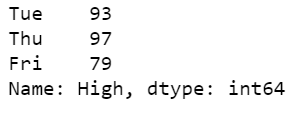
**Output**





* 1. **Select high temperatures for ‘Tue’, ‘Thu’, and ‘Fri’**

**Output**





**Chapter 8**

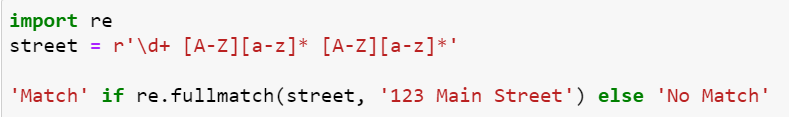
**Problem 1**

1. **Create and test a regular expression that matches a street address consisting of a number with one or more digits followed by two words of one or more character each. Ensure that two words begin with an uppercase letter followed by any number of lowercase letters. The tokens should be separated by one space each, as 123 Main Street.**

**Test the following addresses**

* **123 Main Street : Match**
* **816 S Michigan : Match**
* **816 S. Michigan : No Match**
* **Main Street : No Match**

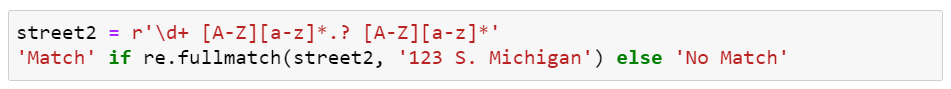
**Answer**



**2. Modify regular expression for problem 1 to match following addresses**

* 1. **123 Main Street : Match**
  2. **816 S Michigan : Match**
  3. **816 S. Michigan : Match**
  4. **Main Street : No Match**

**Answer**



**3. Use a regular expression and split function to split the following string at one or more adjacent $ characters.**

* **123$Main$$Street**

**Output**

['123', 'Main', 'Street']

**Answer**



**Problem**

1. **Use regular expressions and the findall function to count the number of digits, non-digit characters, whitespace characters and words in a string**

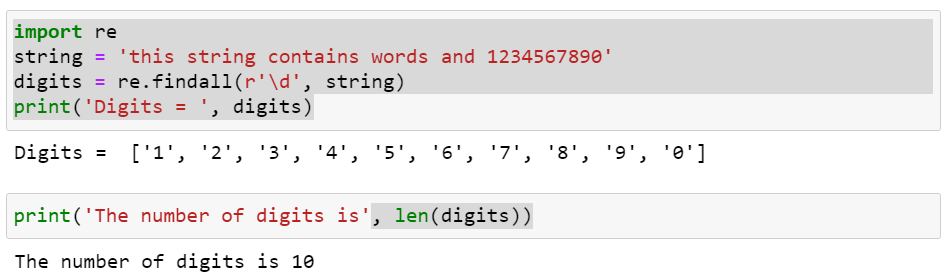
**string = 'this string contains words and 1234567890'**

**Output**

Digits = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '0']

The number of digits is 10

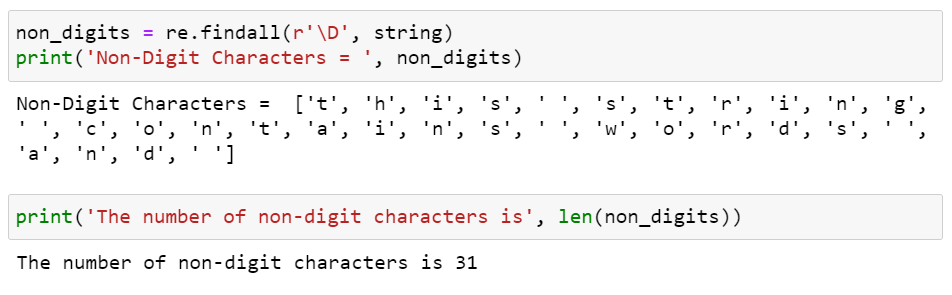
**Answer**



Non-Digit Characters = ['t', 'h', 'i', 's', ' ', 's', 't', 'r', 'i', 'n', 'g', ' ', 'c', 'o', 'n', 't', 'a', 'i', 'n', 's', ' ', 'w', 'o', 'r', 'd', 's', ' ', 'a', 'n', 'd', ' ']

The number of non-digit characters is 31

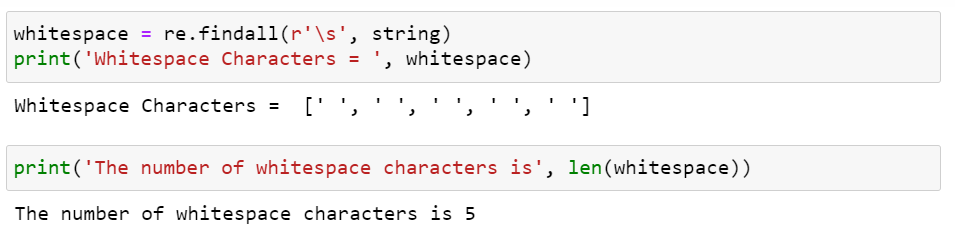
**Answer**



Whitespace Characters = [' ', ' ', ' ', ' ', ' ']

The number of whitespace characters is 5

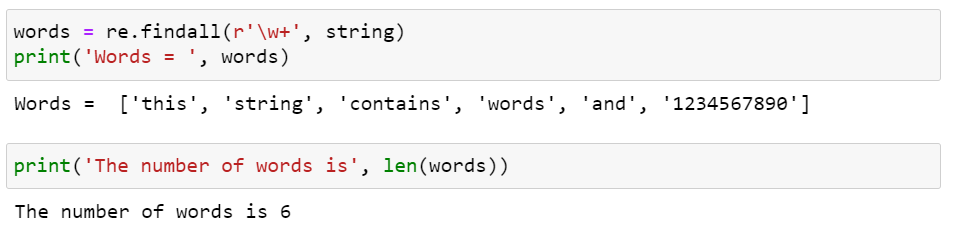
**Answer**



Words = ['this', 'string', 'contains', 'words', 'and', '1234567890']

The number of words is 6

**Answer**



1. **Use a regular expression to search through a string and to locate all valid URLs. For this exercise, assume that a valid URL has the form**

[**http://www.domain\_name.extension**](http://www.domain_name.extension)**, where extension must be two or more characters.**

**Search the following text:**

**text = """Our web address is http://www.deitel.com and**

**Pearson's web address is http://www.pearson.com,**

**but http://www.fake.c is not a valid URL"""**

**Output**

['http://www.deitel.com', 'http://www.pearson.com']

**Answer**



**Programs**

* **Lecture for 3-8-2021 (9.5 Serialization with JSON)**
* **Lecture for 3-17-2021**