Assignment For Regular Expression

Question 1- Write a Python program to replace all occurrences of a space, comma, or dot with a colon.

Sample Text- 'Python Exercises, PHP exercises.'

Expected Output: Python:Exercises::PHP:exercises:

In [95]:
```
import re
input="[ ,.]"
text="Data science, combines math. and statistics, specialized programming, adv
replaced_text= re.sub(input,":", text)
print(replaced_text)
```

```
Data:science::combines:math::and:statistics::specialized:programming::advance
d:analytics
```

Question 2- Create a dataframe using the dictionary below and remove everything (commas (,), !, XXXX, ;, etc.) from the columns except words.

Dictionary- {'SUMMARY' : ['hello, world!', 'XXXXX test', '123four, five:; six...']}

In [1]:
```
import pandas as pd
df=pd.DataFrame({'SUMMARY' : ['hello, world!', 'XXXXX test', '123four, five:; s
df
```

Out[1]:

	SUMMARY
0	hello, world!
1	XXXXX test
2	123four, five:; six...

In [19]:
```
df['SUMMARY']=df['SUMMARY'].str.replace(r'[^a-zA-z\s]+|X{5}',"",regex=True)
```

In [20]:
```
df
```

Out[20]:

	SUMMARY
0	hello world
1	test
2	four five six

Question 3- Create a function in python to find all words that are at least 4 characters long in a string.

The use of the re.compile() method is mandatory.

In [102]:
```
import re
pattern=re.compile(r'\b\w{4,}\b')
text='Data science is the study of data to extract meaningful insights for busi
long_char=pattern.findall(text)
print(long_char)
```

```
['Data', 'science', 'study', 'data', 'extract', 'meaningful', 'insights', 'bu
siness']
```

Question 4- Create a function in python to find all three, four, and five character words in a string.

The use of the re.compile() method is mandatory.

In [28]:
```
import re
pattern=re.compile(r'\b\w{3,5}\b')
text='Data science is the study of data to extract meaningful insights for busi
long_char=pattern.findall(text)
print(long_char)
```

```
['Data', 'the', 'study', 'data', 'for']
```

Question 5- Create a function in Python to remove the parenthesis in a list of strings. The use of the re.compile() method is mandatory.

Sample Text: ["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (Data Science World)", "Data (Scientist)"] Expected Output: example.com hr@fliprobo.com (mailto:hr@fliprobo.com) github.com Hello Data Science World Data Scientist

In [66]:
```
import re
items = ["Datascience (Github.com)","example (.com)", "w3resource", "w3School (
pattern=re.compile(r'[()]')
for item in items:
    print(re.sub(pattern, "", item))
```

```
Datascience Github.com
example .com
w3resource
w3School .com
github .com
stackoverflow .com
```

Question 6- Write a python program to remove the parenthesis area from the text stored in the text file using Regular Expression.

Sample Text: ["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (Data Science World)", "Data (Scientist)"] Expected Output: ["example", "hr@fliprobo", "github", "Hello", "Data"]

In [75]:
```
import re
items = ["Github (.com)", "DataScience", "Fliprobo (.com)", "Naukri (.com)", "H
for item in items:
    print(re.sub(r" ?\([^)]+\)", "", item))
```

```
Github
DataScience
Fliprobo
Naukri
Hello
```

Question 7- Write a regular expression in Python to split a string into uppercase letters.

Sample text: “ImportanceOfRegularExpressionsInPython” Expected Output: [‘Importance’, ‘Of’, ‘Regular’, ‘Expression’, ‘In’, ‘Python’]

In [97]:
```
pattern=r'(?=[A-Z])'
text="ThisIsATextToSplitInUppercase"
split_text=re.split(pattern,text)
print(split_text)
```

```
['', 'This', 'Is', 'A', 'Text', 'To', 'Split', 'In', 'Uppercase']
```

Question 8- Create a function in python to insert spaces between words starting with numbers.

Sample Text: “RegularExpression1IsAn2ImportantTopic3InPython" Expected Output: RegularExpression 1IsAn 2ImportantTopic 3InPython

In [111]:
```
def add_space(string):
   pattern = r'([a-zA-Z])(\d)'
   repl = r'\1 \2'
   modified_string = re.sub(pattern, repl, string)
   return modified_string

text="ThisIsA1Text2ToSplit3InUppercase"
output_string=add_space(text)
print(output_string)
```

```
ThisIsA 1Text 2ToSplit 3InUppercase
```

Question 9- Create a function in python to insert spaces between

words starting with capital letters or with numbers.

Sample Text: “RegularExpression1IsAn2ImportantTopic3InPython" Expected Output:

In [81]:
```
def insert_spaces(text):

  pattern = r'([A-Z][a-z0-9]*)|(\d+)'

  result = re.sub(pattern, r'\1 \2', text)

  result = result.strip()
  return result
```

In [82]:
```
text="ThisIs1Text2ToSplit3InUppercase"
output_string=insert_spaces(text)
print(output_string)
```

```
This Is1 Text2 To Split3 In Uppercase
```

Question 10- Use the github link below to read the data and create a dataframe. After creating the dataframe extract the first 6 letters of each country and store in the dataframe under a new column called first_five_letters.

Github Link-
https://raw.githubusercontent.com/dsrscientist/DSData/master/happiness_score_dataset.csv
(https://raw.githubusercontent.com/dsrscientist/DSData/master/happiness_score_dataset.csv)

In [125]:
```
import pandas as ps
url='https://raw.githubusercontent.com/dsrscientist/DSData/master/happiness_sc
df=pd.read_csv(url)
```

In [126]:
```
df
```

Out[126]:

	Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Fre
0	Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.34951	0.94143	0.(
1	Iceland	Western Europe	2	7.561	0.04884	1.30232	1.40223	0.94784	0.(
2	Denmark	Western Europe	3	7.527	0.03328	1.32548	1.36058	0.87464	0.(
3	Norway	Western Europe	4	7.522	0.03880	1.45900	1.33095	0.88521	0.(
4	Canada	North America	5	7.427	0.03553	1.32629	1.32261	0.90563	0.(
...	...	...	...	...	...	...	...	...	
153	Rwanda	Sub-Saharan Africa	154	3.465	0.03464	0.22208	0.77370	0.42864	0.!
154	Benin	Sub-Saharan Africa	155	3.340	0.03656	0.28665	0.35386	0.31910	0.4
155	Syria	Middle East and Northern Africa	156	3.006	0.05015	0.66320	0.47489	0.72193	0.'
156	Burundi	Sub-Saharan Africa	157	2.905	0.08658	0.01530	0.41587	0.22396	0.
157	Togo	Sub-Saharan Africa	158	2.839	0.06727	0.20868	0.13995	0.28443	0.:

158 rows × 12 columns

In [128]:
```
df['first_six_letters']=df['Country'].str[:6]
```

In [129]:
```
df.head()
```

Out[129]:

	Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freed
0	Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.34951	0.94143	0.665
1	Iceland	Western Europe	2	7.561	0.04884	1.30232	1.40223	0.94784	0.628
2	Denmark	Western Europe	3	7.527	0.03328	1.32548	1.36058	0.87464	0.649
3	Norway	Western Europe	4	7.522	0.03880	1.45900	1.33095	0.88521	0.669
4	Canada	North America	5	7.427	0.03553	1.32629	1.32261	0.90563	0.632

Question 11- Write a Python program to match a string that contains only upper and lowercase letters, numbers, and underscores.

In [138]:
```
import re
def match_string(string):
    pattern=r'^[a-zA-Z0-9_]+$'
    if re.match(pattern,string):
        print("String is matches to Pattern")
    else:
        print("String does not matches to Pattern")


result1=match_string("Hello_DataScience123")

result2=match_string("Hello DataScience!")
```

```
String is matches to Pattern
String does not matches to Pattern
```

Question 12- Write a Python program where a string will start with a specific number.

In [163]:
```
import re
my_string = '123DataScience'
my_number = '1'
m = re.match(my_number, my_string)
if m:
    print('it\'s a match')
else:
    print('no match found')
```

```
it's a match
```

Question 13- Write a Python program to remove leading zeros from an IP address

In [166]:
```
import re
Ip_address='183.234.001.004'
Clean_IP=re.sub('\.0*','.', Ip_address)
print(Clean_IP)
```

```
183.234.1.4
```

Question 14- Write a regular expression in python to match a date string in the form of Month name followed by day number and year stored in a text file.

In [94]:
```
import re
text="2 October 1869 , 30 January 1948 was an Indian lawyer, anti-colonial nati
pattern = r'([\d]{1,2}\s(January|February|March|April|May|June|July|August|Sept
matches=re.findall(pattern,text)
for match in matches:

    print(match)
```

```
('2 October 1869', 'October')
('30 January 1948', 'January')
```

Question 15- Write a Python program to search some literals strings in a string.

Sample text : 'The quick brown fox jumps over the lazy dog.' Searched words : 'fox', 'dog', 'horse'

In [180]:
```
import re
string='The quick brown fox jumps over the lazy dog'
if re.search('cat|dog|fox|horse', string):

    print("The string is match")
else:
    print("The string is not match")
```

```
The string is match
```

Question 16- Write a Python program to search a literals string in a string and also find the location within the original string where the pattern occurs

In [185]:
```
import re
string='a dog and a cat live in the same house, but they do not get along.'
search_string= re.search('\Wdog\W', string)
if search_string:
    print("The string is match, starts on", search_string.start())
else:
    print("The string is not match")
```

```
The string is match, starts on 1
```

Question 17- Write a Python program to find the substrings within a string.

In [1]:
```
import re
text=  'Python exercises, PHP exercises, C# exercises'
sub_text='exercises'
match=re.findall(sub_text,text)
print(match)
```

```
['exercises', 'exercises', 'exercises']
```

Question 18- Write a Python program to find the occurrence and position of the substrings within a string.

In [5]:
```
import re
text='Python is an easy to learn, powerful programming language.'
sub_string='learn'
matches=re.finditer(sub_string,text)
for match in matches:
    print('string \'{}\''.format(sub_string), 'found at position', match.span()
```

```
string 'learn' found at position (21, 26)
```

Question 19- Write a Python program to convert a date of yyyy-mm-dd format to dd-mm-yyyy format.

In [6]:
```
import re
date='2024-05-24'
convert=re.split('-',date)
new_date='-'.join(convert[::-1])
print(new_date)
```

```
24-05-2024
```

Question 20- Create a function in python to find all decimal numbers with a precision of 1 or 2 in a string. The use of the re.compile() method is mandatory.

In [8]:
```
import re
text="01.12 0132.123 2.31875 145.8 3.01 27.25 0.25"
pattern=re.compile(r'\d+\.\d{1,2}')
output=pattern.findall(text)
print(output)
```

```
['01.12', '0132.12', '2.31', '145.8', '3.01', '27.25', '0.25']
```

Question 21- Write a Python program to separate and print the numbers and their position of a given string.

In [59]:
```
import re
string="4 python 45 datascience 1 hello world"
pattern=r'\d+'
matches=re.finditer(pattern,string)
for match in matches:
    print(match.group(0))
    print("found at position :", match.start())
```

```
4
found at position : 0
45
found at position : 9
1
found at position : 24
```

Question 22- Write a regular expression in python program to extract maximum/largest numeric value from a string.

In [14]:
```
import re
string='My marks in each semester are: 947, 896, 926, 524, 734, 950, 642, 980'
regex=re.findall(r'\d+', string)
max_value=max(regex)
print(max_value)
```

```
980
```

Question 23- Create a function in python to insert spaces between words starting with capital letters.

In [17]:
```
text="RegularExpressionIsAnImportantTopicInPython"
pattern=r'(\w)([A-Z])'
insert_spaces=re.sub(pattern, r"\1 \2", text)
print(insert_spaces)
```

```
Regular Expression Is An Important Topic In Python
```

Question 24- Python regex to find sequences of one upper case letter followed by lower case letters

In [19]:
```
import re
text="Python is a general Purpose, interpreted Language"
pattern=r'[A-Z][a-z]+'
matches=re.findall(pattern,text)
print(matches)
```

```
['Python', 'Purpose', 'Language']
```

Question 25- Write a Python program to remove continuous duplicate words from Sentence using Regular Expression.

In [24]:
```
import re
text="python python Language Language"
pattern=r'\b(\w+)(\s+\1\b)+'
output=re.sub(pattern,r'\1', text)
print(output)
```

```
python Language
```

Question 26- Write a python program using RegEx to accept string ending with alphanumeric character.

In [25]:
```
import re
pattern=r'[A-Za-z0-9]+$'
text="pythontutorials123"
if re.match(pattern,text):
    print("The given string is alphanumeric")
else:
    print("Not alphanueric")
```

```
The given string is alphanumeric
```

Question 27-Write a python program using RegEx to extract the hashtags.

In [27]:
```
import re
pattern=r'#\w+'
text="@github.com, #pythonlibraries, #xyz, #datascience, Regex"
extract_hastags=re.findall(pattern,text)
print(extract_hastags)
```

```
['#pythonlibraries', '#xyz', '#datascience']
```

Question 28- Write a python program using RegEx to remove <U+..> like symbols

In [28]:
```
import re
text="@Jags123456 Bharat band on 28??<ed><U+00A0><U+00BD><ed><U+00B8><U+0082>Th
pattern=r'<U\+\w{4}>'
output_string=re.sub(pattern, "", text)
print(output_string)
```

```
@Jags123456 Bharat band on 28??<ed><ed>Those who  are protesting #demonetizat
ion  are all different party leaders
```

Question 29- Write a python program to extract dates from the text stored in the text file.

In [30]:
```
import re
text="Ram was born on 12-09-1995 and he was admitted to school 15-12-1999"
pattern=r'\d{2}-\d{2}-\d{4}'
matches=re.findall(pattern,text)
print(matches)
```

```
['12-09-1995', '15-12-1999']
```

Question 30- Create a function in python to remove all words from a string of length between 2 and 4.

In [44]:
```
import re
text="They rushed out the door, grabbing anything and everything they could thi
pattern=re.compile(r'\b\w{2,4}\b')
output_string=pattern.sub("",text)
print(output_string)
```

```
 rushed   , grabbing anything  everything  could think   might . There
double-check     weren't 3 leaving something important behind.
```

In [ ]:

In [ ]: