Regular Expressions

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 [5]:
```
import re
import pandas as pd
import regex as re
```

In [7]:
```
text='Python Exercises, PHP exercises.'
result= re.sub(r"[ ,.]",r":",text)
print(result)
```

```
Python:Exercises::PHP:exercises:
```

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...']} Expected output- 0 hello world 1 test 2 four five six

In [3]:
```
df=pd.DataFrame({'SUMMARY' : ['hello, world!', 'XXXXX test', '123four, five
```

In [4]:
```
df
```

Out[4]:

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

In [5]:
```
df['SUMMARY']=df['SUMMARY'].str.replace('[^a-z\s]','',regex=True)
```

In [6]:
```
print(df)
```

```
         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 [8]:
```
target_string='Let the melody of birdsong fill your soul.'
pattern=re.compile(r"\b\w{4,}\b",flags=0)
result=pattern.findall(target_string)
print(result)
```

```
['melody', 'birdsong', 'fill', 'your', 'soul']
```

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 [9]:
```
text='Let the melody of birdsong fill your soul.'
pattern=re.compile(r"\b\w{3,5}\b",flags=0)
result=pattern.findall(text)
print(result)
```

```
['Let', 'the', 'fill', 'your', 'soul']
```

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 [1]:
```
import re

def remove_parenthesis(text):
    pattern=re.compile(r'[()]')
    x=[pattern.sub('', text) for text in text]
    return x

text=["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (Data
result=remove_parenthesis(text)
print(result)
```

```
['example .com', 'hr@fliprobo .com', 'github .com', 'Hello Data Science Wo
rld', 'Data Scientist']
```

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"] Note- Store given sample text in the text file and then to remove the parenthesis area from the text.

In [40]:
```
text=["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (Data
for text in text:
    result=re.sub(r'\(.*?\)'," ",text)
    print(result)
```

```
example
hr@fliprobo
github
Hello
Data
```

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 [6]:
```
text="ImportanceOfRegularExpressionsInPython"
pattern=r'(?=[A-Z])'
x=re.split(pattern,text)
print(x)
```

```
['', 'Importance', 'Of', 'Regular', 'Expressions', 'In', 'Python']
```

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 [7]:
```
def insert_space(text):
    pattern=re.compile(r'(?=[0-9a-z])(?=[0-9A-Z])')
    x=re.sub(pattern," ",text)
    return x
text='RegularExpression1IsAn2ImportantTopic3InPython'
result=insert_space(text)
print(result)
```

```
RegularExpression 1IsAn 2ImportantTopic 3InPython
```

Question 9- Create a function in python to insert spaces between words starting with capital letters or with numbers. Sample Text: “RegularExpression1IsAn2ImportantTopic3InPython" Expected Output: RegularExpression 1 IsAn 2 ImportantTopic 3 InPython

In [8]:
```
def insert_space(text):
    pattern=re.compile(r'(?<=[0-9a-z])(?=[0-9A-Z])')
    x=re.sub(pattern," ",text)
    return x
text='RegularExpression1IsAn2ImportantTopic3InPython'
result=insert_space(text)
print(result)
```

```
Regular Expression 1 Is An 2 Important Topic 3 In Python
```

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 [9]:
```
Github_link="https://raw.githubusercontent.com/dsrscientist/DSData/master/h

df=pd.read_csv(Github_link)

df['first_five_letters']=df['Country'].str[:5]

print(df.head())
```

```
       Country          Region  Happiness Rank  Happiness Score  \
0  Switzerland  Western Europe               1            7.587
1      Iceland  Western Europe               2            7.561
2      Denmark  Western Europe               3            7.527
3       Norway  Western Europe               4            7.522
4       Canada   North America               5            7.427

   Standard Error  Economy (GDP per Capita)   Family  \
0         0.03411                   1.39651  1.34951
1         0.04884                   1.30232  1.40223
2         0.03328                   1.32548  1.36058
3         0.03880                   1.45900  1.33095
4         0.03553                   1.32629  1.32261

   Health (Life Expectancy)  Freedom  Trust (Government Corruption)  \
0                   0.94143  0.66557                        0.41978
1                   0.94784  0.62877                        0.14145
2                   0.87464  0.64938                        0.48357
3                   0.88521  0.66973                        0.36503
4                   0.90563  0.63297                        0.32957

   Generosity  Dystopia Residual first_five_letters
0     0.29678            2.51738              Switz
1     0.43630            2.70201              Icela
2     0.34139            2.49204              Denma
3     0.34699            2.46531              Norwa
4     0.45811            2.45176              Canad
```

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

In [11]:
```
string=input('Enter a string')
pattern='[a-zA-Z0-9_]+'
result=re.search(pattern,string)
if result:
    print("Match found")
else:
    print("Match not found")
```

```
Enter a string"RegularExpression1IsAn2ImportantTopic3InPython"
Match found
```

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

In [15]:
```
string=input('Enter a string')
number=input('Enter a number')
result=re.match(number,string)
if result:
    print('String starts with specific number')
else:
    print('String does not match with specific number')
```

```
Enter a string22 roses are pink
Enter a number22
String starts with specific number
```

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

In [16]:
```
ip="102.67.520.001"
result=re.sub('\.[0]*','.',ip)
print(result)
```

```
102.67.520.1
```

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. Sample text : ' On August 15th 1947 that India was declared independent from British colonialism, and the reins of control were handed over to the leaders of the Country'. Expected Output- August 15th 1947 Note- Store given sample text in the text file and then extract the date string asked format.

In [17]:
```
text="On August 15th 1947 that India was declared independent from British
pattern=r'\b(January|February|March|April|May|June|July|August|September|Oc
result=re.findall(pattern,text)
print(result)
```

```
[('August', '15', 'th', '1947')]
```

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 [18]:
```
text='The quick brown fox jumps over the lazy dog.'
x= re.search('fox|dog|horse',text)
if x:
    print('Its a match')
else:
    print('No match found')
```

```
Its a 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 Sample text : 'The quick brown fox jumps over the lazy dog.' Searched words : 'fox'

In [21]:
```
text='The quick brown fox jumps over the lazy dog.'
x=re.search('\Wfox\W',text)
if x:
    print('Its a match,starts on:',x.start())
else:
    print('No match found')
```

```
Its a match,starts on: 15
```

Question 17- Write a Python program to find the substrings within a string. Sample text : 'Python exercises, PHP exercises, C# exercises' Pattern : 'exercises'.

In [22]:
```
string='Python exercises,PHP exercises,C# exercises'
substring='exercises'
x=re.findall(substring,string)
if x:
    print('Its a match')
else:
    print('No match found')
```

```
Its a match
```

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

In [23]:
```
string='Python exercises,PHP exercises,C# exercises'
substring='exercises'
x=re.finditer(substring,string)
for match in x:
    print('string\{}\''.format(substring),'Found at position',match.span())
```

```
string\exercises' Found at position (7, 16)
string\exercises' Found at position (21, 30)
string\exercises' Found at position (34, 43)
```

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

In [25]:
```
from datetime import datetime

def convert_date_format(date_string):
    x=datetime.strptime(date_string, '%Y-%m-%d')
    y=x.strftime('%d-%m-%Y')
    return y

date='2024-05-19'
result=convert_date_format(date)
print(result)
```

```
19-05-2024
```

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. Sample Text: "01.12 0132.123 2.31875 145.8 3.01 27.25 0.25" Expected Output: ['01.12', '145.8', '3.01', '27.25',

'0.25']

In [26]:
```
def find_decimal_number(string):
    pattern=re.compile(r'\d+\.\d{1,2}')
    decimal_numbers=re.findall(pattern,string)
    return decimal_numbers
sample_text="01.12 0132.123 2.31875 145.8 3.01 27.25 0.25"
result=find_decimal_number(sample_text)
print(result)
```

```
['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 [27]:
```
text="There are 50 flowers in the garden"
for x in re.finditer("\d+",text):
    print(x.group(0))
    print("Index position: ",x.start(),x.end())
```

```
50
Index position:  10 12
```

Question 22- Write a regular expression in python program to extract maximum/largest numeric value from a string. Sample Text: 'My marks in each semester are: 947, 896, 926, 524, 734, 950, 642' Expected Output: 950

In [28]:
```
text="My marks in each semester are: 947, 896, 926, 524, 734, 950, 642"
marks=re.findall(r'\d+',text)
max_marks=max(marks)
print(max_marks)
```

```
950
```

Question 23- Create a function in python to insert spaces between words starting with capital letters. Sample Text: “RegularExpressionIsAnImportantTopicInPython" Expected Output: Regular Expression Is An Important Topic In Python

In [30]:
```
text="RegularExpressionIsAnImportantTopicInPython"
pattern=r'(?=[A-Z][a-z])'
result=re.split(pattern,text)
print(result)
```

```
['', 'Regular', 'Expression', 'Is', 'An', 'Important', 'Topic', 'In', 'Pyt
hon']
```

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

In [32]:
```
def text_match(text):
    pattern='[A-Z]+[a-z]+$'
if re.search(pattern,text):
    print('Match found')
else:
    print('Match not found')

print(text_match("Rashmi"))
```

```
Match found
None
```

Question 25- Write a Python program to remove continuous duplicate words from Sentence using Regular Expression. Sample Text: "Hello hello world world" Expected Output: Hello hello world

In [33]:
```
text="Hello hello world world"
regex=r'\b(\w+)(?:\W+\1\b)+'
x=re.sub(regex,r'\1',text)
print(x)
```

```
Hello hello world
```

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

In [34]:
```
def check_string(string):
    pattern='[a-zA-Z0-9]$'
    if(re.search(pattern,string)):
        print("String ends with an alphanumeric character")
    else:
        print("String does not ends with an alphanumeric character")

print(check_string("Rashmi@123"))
print(check_string("Hit_Ay"))
print(check_string("ABC_"))
```

```
String ends with an alphanumeric character
None
String ends with an alphanumeric character
None
String does not ends with an alphanumeric character
None
```

Question 27-Write a python program using RegEx to extract the hashtags. Sample Text: """RT @kapil_kausik: #Doltiwal I mean #xyzabc is "hurt" by #Demonetization as the same has rendered USELESS <U+00A0><U+00BD><U+00B1><U+0089> "acquired funds" No wo""" Expected Output: ['#Doltiwal', '#xyzabc', '#Demonetization']

In [36]:
```
def extract_hashtags(text):
    hashtags=re.findall(r'#\w+',text)
    return hashtags

text="""RT @kapil_kausik: #Doltiwal I mean #xyzabc is "hurt" by #Demonetiza
result=extract_hashtags(text)
print(result)
```

```
['#Doltiwal', '#xyzabc', '#Demonetization']
```

Question 28- Write a python program using RegEx to remove <U+..> like symbols Check the below sample text, there are strange symbols something of the sort <U+..> all over the place. You need to come up with a general Regex expression that will cover all such symbols. Sample Text: "@Jags123456 Bharat band on 28??<U+00A0><U+00BD> <U+00B8><U+0082>Those who are protesting #demonetization are all different party leaders" Expected Output: @Jags123456 Bharat band on 28??Those who are protesting #demonetization are all different party leaders

In [37]:
```
text="@Jags123456 Bharat band on 28??<ed><U+00A0><U+00BD><ed><U+00B8><U+008
pattern=r'<U\+\w{4}>'
result=re.sub(pattern,"",text)
print(result)
```

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

Question 29- Write a python program to extract dates from the text stored in the text file. Sample Text: Ron was born on 12-09-1992 and he was admitted to school 15-12-1999. Note- Store this sample text in the file and then extract dates.

In [38]:
```
text="Ron was born on 12-09-1992 and he was admitted to school 15-12-1999."
pattern=r'\d{2}-\d{2}-\d{4}'
dates=re.findall(pattern,text)
print(dates)
```

```
['12-09-1992', '15-12-1999']
```

Question 30- Create a function in python to remove all words from a string of length between 2 and 4. The use of the re.compile() method is mandatory. Sample Text: "The following example creates an ArrayList with a capacity of 50 elements. 4 elements are then added to the ArrayList and the ArrayList is trimmed accordingly." Expected Output: following example creates ArrayList a capacity elements. 4 elements added ArrayList ArrayList trimmed accordingly.

In [39]:
```
def remove_words(text):
    pattern=re.compile(r'\b\w{2,4}\b')
    modified_text=re.sub(pattern,'',text)
    print(modified_text)

text="The following example creates an ArrayList with a capacity of 50 elem
result=remove_words(text)
print(result)
```

```
 following example creates  ArrayList  a capacity   elements. 4 elements
added   ArrayList   ArrayList  trimmed accordingly.
None
```

In [ ]:
```

```