**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:

A**ns:** import re

y = r"[ ,.]"

text ='Python Exercises, PHP exercises.'

pattern = re.sub(y,":",text)

print (pattern)

**Output:** 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

ANS: import pandas as pd

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

df = pd.DataFrame(data)

df['SUMMARY'] = df['SUMMARY'].str.replace('[^a-zA-Z\s]', '', regex=True)

print(df)

OUTPUT: SUMMARY

0 hello world

1 XXXXX 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.

Ans: import re

text = 'devi eyes colour is blue,thats why she look like a doll'

pattern = re.findall(r"\b\w{4}\b", text)

print (pattern)

output: ['devi', 'eyes', 'blue', 'look', 'like', 'doll']

**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.

Ans: text = 'devi eyes colour is blue,thats why she look like a doll'

pattern = re.findall(r"\b\w{3,5}\b", text)

print (pattern)

output: ['devi', 'eyes', 'blue', 'thats', 'why', 'she', 'look', 'like', 'doll']

**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

github.com

Hello Data Science World

Data Scientist

ANS: import re

items = ["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (Data Science World)", "Data (Scientist)"]

for item in items:

print(re.sub(r"\(|\)", "", item))

OUTPUT: example .com

hr@fliprobo .com

github .com

Hello Data Science World

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.

ANS: import re

items = ["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (Data Science World)", "Data (Scientist)"]

for item in items:

print(re.sub(r" \([^)]+\)", "", item))

OUTPUT: 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’]

ANS: text = "ImportanceOfRegularExpressionsInPython"

pattern = re.findall('[A-Z][^A-Z]\*', text)

print (pattern)

OUTPUT: ['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

ANS: text = "RegularExpression1IsAn2ImportantTopic3InPython"

matches = re.sub(r'(\d)([a-zA-Z 0-9])', r'\1 \2',text)

print (matches)

OUTPUT: RegularExpression1 IsAn2 ImportantTopic3 InPython

**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

ANS: text = "RegularExpression1IsAn2ImportantTopic3InPython"

matches = re.sub(r"(\w)([A-Z0-9])", r"\1 \2",text)

print (matches)

OUTPUT: Regular Expression 1Is An 2Important Topic 3In 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>

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

ANS: pattern = r"[a-zA-Z0-9\_]\*$"

text = "India is the best\_123"

x = re.search ("India is the best\_123",text)

print (x)

OUTPUT: <re.Match object; span=(0, 21), match='India is the best\_123'>

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

ANS: pattern ='\d+'

text = '50 apples & 60 mango in 1 basket'

matches = re.findall(pattern,text)

print (matches)

OUTPUT: ['50', '60', '1']

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

ANS: import re

ip = "218.08.096.196"

pattern = re.sub('\.[0]\*', '.', ip)

print(pattern)

OUTPUT: 218.8.96.196

**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.

**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'

ANS: patterns = [ 'fox', 'dog', 'horse' ]

text = 'The quick brown fox jumps over the lazy dog.'

for pattern in patterns:

print('Searching for "%s" in "%s" ->' % (pattern, text),)

if re.search(pattern, text):

print('Matched!')

else:

print('Not Matched!')

OUTPUT: Searching for "fox" in "The quick brown fox jumps over the lazy dog." ->

Matched!

Searching for "dog" in "The quick brown fox jumps over the lazy dog." ->

Matched!

Searching for "horse" in "The quick brown fox jumps over the lazy dog." ->

Not Matched!

**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'

ANS: pattern = 'fox'

text = 'The quick brown fox jumps over the lazy dog.'

match = re.search(pattern, text)

s = match.start()

e = match.end()

print('Found "%s" in "%s" from %d to %d ' % \

(match.re.pattern, match.string, s, e))

OUTPUT: Found "fox" in "The quick brown fox jumps over the lazy dog." from 16 to 19

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

**Sample text :** 'Python exercises, PHP exercises, C# exercises'

**Pattern :** 'exercises'.

ANS: text = 'Python exercises, PHP exercises, C# exercises'

pattern = 'exercises'

match = re.findall(pattern, text)

print (match)

OUTPUT: ['exercises', 'exercises', 'exercises']

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

ANS: ext = 'Python exercises, PHP exercises, C# exercises'

pattern = 'exercises'

for match in re.finditer(pattern, text):

s = match.start()

e = match.end()

print('Found "%s" at %d:%d' % (text[s:e], s, e))

OUTPUT: Found "exercises" at 7:16

Found "exercises" at 22:31

Found "exercises" at 36:45

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

ANS: dt1 = "2026-01-02"

pattern = re.sub(r'(\d{4})-(\d{1,2})-(\d{1,2})', '\\3-\\2-\\1', dt1)

print (pattern)

OUTPUT: 02-01-2026

**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']

ANS: pattern = r'\b\d+\.\d{1,2}\b'

input\_string = "01.12 0132.123 2.31875 145.8 3.01 27.25 0.25"

matches = re.findall(pattern, input\_string )

print(matches)

OUTPUT: ['01.12', '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.

ANS: import re

text = 'I have 2 cars & 1 bike'

pattern = re.finditer(r"\d+", text)

for m in pattern:

print(m.group(0))

print("Index position:", m.start())

OUTPUT: 2

Index position: 7

1

Index position: 16

**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

**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

ANS: text = 'RegularExpressionIsAnImportantTopicInPython'

pattern = re.sub(r"(\w)([A-Z])", r"\1 \2", text)

print (pattern)

OUTPUT: Regular Expression Is An Important Topic In Python

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

ANS: text = 'Devi Eyes colour is Blue,thats why she look like a Doll'

pattern = re.findall('[A-Z]+[a-z]+$',text)

print (pattern)

OUTPUT: ['Doll']

**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

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

**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 <ed><U+00A0><U+00BD><ed><U+00B1><U+0089> "acquired funds" No wo"""

**Expected Output:** ['#Doltiwal', '#xyzabc', '#Demonetization']

ANS: text = """RT @kapil\_kausik: #Doltiwal I mean #xyzabc is "hurt" by #Demonetization as the

same has rendered USELESS <ed><U+00A0><U+00BD><ed><U+00B1><U+0089> "acquired funds" No wo"""

pattern = re.compile(r'#\w+')

hashtags = pattern.findall(text)

print(hashtags)

OUTPUT: ['#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??<ed><U+00A0><U+00BD><ed><U+00B8><U+0082>Those who are protesting #demonetization are all different party leaders"

**Expected Output:** @Jags123456 Bharat band on 28??<ed><ed>Those who are protesting #demonetization 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.

**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.

ANS: text = """The following example creates an ArrayList with a capacity of 50 elements.

4 elements"""

shortword = re.compile(r'\W\*\b\w{1,3}\b')

print(shortword.sub('', text))

OUTPUT: following example creates ArrayList with capacity elements elements