**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:

**Answer :-**

**x='Python Exercises, PHP exercises.'**

**c=x.replace(" ",":")**

**d=c.replace(".",":")**

**e=d.replace(",",":")**

**print(e)**

**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

**Answer:-**

**import pandas as pd**

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

**df=pd.DataFrame(dictnary)**

**print(df)**

**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.

**Answer:-**

**Import regex as re**

**a="Virat Kohli is a prominent Indian cricketer, widely regarded as one of the best batsmen in the world. Born on November 5, 1988, in Delhi, India, Kohli made his international debut for the Indian cricket team in 2008. Since then, he has established himself as a key player in all formats of the game."**

**b=r"\w{4}"**

**c=re.compile(b)**

**d=c.findall(a)**

**print(d)**

**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.

**Answer :-**

**Import regex as re**

**a="Virat Kohli is a prominent Indian cricketer, widely regarded as one of the best batsmen in the world. Born on November 5, 1988, in Delhi, India, Kohli made his international debut for the Indian cricket team in 2008. Since then, he has established himself as a key player in all formats of the game."**

**b=r"\w{3,5}"**

**c=re.compile(b)**

**d=c.findall(a)**

**print(d)**

**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

**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.

**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’]

**Answer:-**

**import regex as re**

**sample\_text=”ImportanceOfRegularExpressionsInPython”**

**output=re.findal(“[A-Z][^A-Z]+”,sample\_text)**

**print(output)**

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

Sample Text: “RegularExpression1IsAn2ImportantTopic3InPython"

Expected Output: 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

**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.

**Answer:-**

**Import regex as re**

**string= ”The\_quick\_brown\_fox\_jumps\_over\_the\_lazy\_dog\_1234567890."**

**def text\_match(string):**

**pattern='^[a-zA-Z0-9\_]'**

**if re.search(pattern,string):**

**return 'matched'**

**else:**

**return 'not matched'**

**print(text\_match(string))**

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

**Answer:-**

**String=' 1947 August 15th that India was declared independent from British colonialism, and the reins of control were handed over to the leaders of the Country.’**

**Result=re.match(r”\d+”,string)**

**Print(result)**

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

**Answer :-**

**import regex as re**

**ip\_address="1270.00.00.010"**

**def remove\_zeros\_from\_ip(ip\_address):**

**return re.sub(r'(^|\.)0+(?=[^.])',r'\1',ip\_address)**

**print(remove\_zeros\_from\_ip(ip\_address))**

**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'

**Answer:-**

**import regex as re**

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

**pattern=”fox|dog|horse”**

**words=re.findall(pattern,string)**

**print(words)**

**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'

**Answer:-**

**import regex as re**

**sample\_text=”The quick brown fox jumps over the lazy dog.”**

**for word in sample\_text:**

**search=re.search(“fox”,sample\_text)**

**print(search)**

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

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

**Pattern :** 'exercises'.

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

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

**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']

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

**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

**Answer:-**

**import regex as re**

**sample\_input='My marks in each semester are: 947, 896, 926, 524, 734, 950, 642'**

**number=number = re.findall('\d+', sample\_input)**

**number = map(int, number)**

**print("Max\_value:",max(number))**

**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

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

**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

**Answer:-**

**Import regex as re,**

**Sample\_text = "Hello Hello hello hello world world"**

**Regular\_expressions= r'\b(\w+)(?:\W+\1\b)+'**

**Expected\_output= re.sub(Regular\_expressions,r'\1',Sample\_text)**

**print(Expected\_output)**

**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']

**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.