**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:

import re

sample\_text = 'Python Exercises, PHP exercises.'

result = re.sub(r'[ ,.]+', ':', sample\_text)

print(result)

**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

import pandas as pd

import re

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

df = pd.DataFrame(data)

df['SUMMARY'] = df['SUMMARY'].apply(lambda x: re.sub(r'[^A-Za-z\s]', '', x))

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.

import re

def find\_long\_words(text):

pattern = re.compile(r'\b\w{4,}\b')

return pattern.findall(text)

sample\_text = "This is a test string with several long words"

print(find\_long\_words(sample\_text))

**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.

import re

def find\_specific\_length\_words(text):

pattern = re.compile(r'\b\w{3,5}\b')

return pattern.findall(text)

sample\_text = "This is a test string with several words of various lengths"

print(find\_specific\_length\_words(sample\_text))

**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

import re

def remove\_parentheses(lst):

pattern = re.compile(r'\s?\(.\*?\)')

return [pattern.sub('', item) for item in lst]

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

print(remove\_parentheses(sample\_text))

**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’]

import re

def remove\_parentheses\_from\_file(filename):

with open(filename, 'r') as file:

content = file.read()

pattern = re.compile(r'\s?\(.\*?\)')

result = pattern.sub('', content)

with open('output.txt', 'w') as file:

file.write(result)

remove\_parentheses\_from\_file('sample.txt')

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

Sample Text: “RegularExpression1IsAn2ImportantTopic3InPython"

Expected Output: RegularExpression 1IsAn 2ImportantTopic 3InPython

import re

def insert\_spaces\_between\_numbers(text):

pattern = re.compile(r'(\d)([A-Za-z])')

return pattern.sub(r'\1 \2', text)

sample\_text = "RegularExpression1IsAn2ImportantTopic3InPython"

print(insert\_spaces\_between\_numbers(sample\_text))

**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

import re

def insert\_spaces\_before\_numbers\_and\_capitals(text):

pattern = re.compile(r'(?<=[0-9A-Za-z])(?=[A-Z0-9])')

return pattern.sub(' ', text)

sample\_text = "RegularExpression1IsAn2ImportantTopic3InPython"

print(insert\_spaces\_before\_numbers\_and\_capitals(sample\_text))

**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>

import pandas as pd

url = 'https://raw.githubusercontent.com/dsrscientist/DSData/master/happiness\_score\_dataset.csv'

df = pd.read\_csv(url)

df['first\_five\_letters'] = df['Country'].str[:6]

print(df.head())

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

import re

def match\_string(text):

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

return bool(pattern.match(text))

sample\_text = "Valid\_String\_123"

print(match\_string(sample\_text))

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

import re

def starts\_with\_number(text):

pattern = re.compile(r'^\d')

return bool(pattern.match(text))

sample\_text = "1ThisStartsWithNumber"

print(starts\_with\_number(sample\_text))

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

import re

def remove\_leading\_zeros(ip\_address):

pattern = re.compile(r'\b0+(\d)')

return pattern.sub(r'\1', ip\_address)

sample\_ip = "192.168.001.001"

print(remove\_leading\_zeros(sample\_ip))

**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.

import re

def extract\_date\_from\_text(filename):

with open(filename, 'r') as file:

content = file.read()

pattern = re.compile(r'\b\w+ \d{1,2}(?:st|nd|rd|th)? \d{4}\b')

return pattern.findall(content)

# Assuming the text is stored in 'sample.txt'

print(extract\_date\_from\_text('sample.txt'))

**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'

import re

def search\_literals(text, words):

pattern = re.compile('|'.join(map(re.escape, words)))

return pattern.findall(text)

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

searched\_words = ['fox', 'dog', 'horse']

print(search\_literals(sample\_text, searched\_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'

import re

def search\_literal\_and\_position(text, word):

pattern = re.compile(re.escape(word))

match = pattern.search(text)

return match.start() if match else -1

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

searched\_word = 'fox'

print(search\_literal\_and\_position(sample\_text, searched\_word))

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

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

**Pattern :** 'exercises'.

import re

def find\_substrings(text, pattern):

regex = re.compile(re.escape(pattern))

return regex.findall(text)

sample\_text = 'Python exercises PHP exercises C# exercises'

pattern = 'exercises'

print(find\_substrings(sample\_text, pattern))

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

import re

def find\_occurrences\_with\_positions(text, pattern):

matches = [(match.start(), match.group()) for match in re.finditer(re.escape(pattern), text)]

return matches

sample\_text = 'Python exercises PHP exercises C# exercises'

pattern = 'exercises'

print(find\_occurrences\_with\_positions(sample\_text, pattern))

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

import re

def convert\_date\_format(date\_str):

pattern = re.compile(r'(\d{4})-(\d{2})-(\d{2})')

return pattern.sub(r'\3-\2-\1', date\_str)

sample\_date = '2024-08-22'

print(convert\_date\_format(sample\_date))

**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']

import re

def find\_decimal\_numbers(text):

pattern = re.compile(r'\b\d+\.\d{1,2}\b')

return pattern.findall(text)

sample\_text = "01.12 0132.123 2.31875 145.8 3.01 27.25 0.25"

print(find\_decimal\_numbers(sample\_text))

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

import re

def separate\_numbers\_and\_positions(text):

matches = [(match.start(), match.group()) for match in re.finditer(r'\d+', text)]

return matches

sample\_text = 'There are 3 apples, 4 oranges, and 5 bananas.'

print(separate\_numbers\_and\_positions(sample\_text))

**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

import re

def extract\_max\_numeric\_value(text):

numbers = map(int, re.findall(r'\d+', text))

return max(numbers, default=None)

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

print(extract\_max\_numeric\_value(sample\_text))

**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

import re

def insert\_spaces\_before\_capitals(text):

pattern = re.compile(r'(?<=[a-z])(?=[A-Z])')

return pattern.sub(' ', text)

sample\_text = "RegularExpressionIsAnImportantTopicInPython"

print(insert\_spaces\_before\_capitals(sample\_text))

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

import re

def find\_uppercase\_followed\_by\_lowercase(text):

pattern = re.compile(r'[A-Z][a-z]+')

return pattern.findall(text)

sample\_text = "This is an Example Of Mixed Case Letters"

print(find\_uppercase\_followed\_by\_lowercase(sample\_text))

**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

import re

def remove\_duplicate\_words(text):

pattern = re.compile(r'\b(\w+)\s+\1\b', re.IGNORECASE)

return pattern.sub(r'\1', text)

sample\_text = "Hello hello world world"

print(remove\_duplicate\_words(sample\_text))

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

import re

def ends\_with\_alphanumeric(text):

pattern = re.compile(r'[a-zA-Z0-9]$')

return bool(pattern.search(text))

sample\_text = "This ends with a character1"

print(ends\_with\_alphanumeric(sample\_text))

**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']

import re

def extract\_hashtags(text):

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

return pattern.findall(text)

sample\_text = """RT @kapil\_kausik: #Doltiwal I mean #xyzabc is "hurt" by #Demonetization..."""

print(extract\_hashtags(sample\_text))

**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

import re

def remove\_u\_symbols(text):

pattern = re.compile(r'<U\+\w+>')

return pattern.sub('', text)

sample\_text = """@Jags123456 Bharat band on 28??<ed><U+00A0><U+00BD><ed>..."""

print(remove\_u\_symbols(sample\_text))

**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.

import re

def extract\_dates\_from\_file(filename):

with open(filename, 'r') as file:

content = file.read()

pattern = re.compile(r'\b\d{2}-\d{2}-\d{4}\b')

return pattern.findall(content)

# Assuming the text is stored in 'sample.txt'

print(extract\_dates\_from\_file('sample.txt'))

**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.

import re

def remove\_words\_of\_length(text):

pattern = re.compile(r'\b\w{2,4}\b')

return pattern.sub('', text)

sample\_text = "The following example creates an ArrayList with a capacity of 50 elements..."

print(remove\_words\_of\_length(sample\_text))