**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

def replace\_colon(text):

result=re.sub(r"[ ,.]",":",text)

return result

ts='Python Exercises, PHP exercises.'

replace\_colon(ts)

**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)

print(df)

def mod\_string(text):

new\_tab=re.sub("[^a-z\s]","",text)

return new\_tab

df["SUMMARY"]=df["SUMMARY"].apply(mod\_string)

print(df)

**Question 3-** Create a function inpython to find all words that are at least 4 characters long in a string.The use of the re.compile() method is mandatory.

def word\_func(x):

pattern=re.compile("\w{4,}")

match=pattern.findall(x)

return match

ts="Today class on 28-1-24 ensemble model which is further divided into 2 type homogenous and hetrogenous."

final=word\_func(ts)

print(final)

**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.

ef mtt\_word(input\_string):

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

result=pattern.findall(input\_string)

return result

target\_string="Today class on 28-1-24 ensemble model which is further divided into 2 type homogenous and hetrogenous."

mtt\_word(target\_string)

**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

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

def neat\_text(word):

result\_text=[]

for url in word:

pattern=re.compile("[^\w\s@.]")

subs=re.sub(pattern,"",url)

result\_text.append(subs)

return result\_text

neat\_text(data)

**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.

import re

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

modified\_text=[]

for line in sample\_text:

pattern=re.compile("\([^()]\*\)")

sub=re.sub(pattern,"",line)

modified\_text.append(sub)

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

for line in modified\_text:

file.write(line + "\n")

**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’]

sample\_text= "ImportanceOfRegularExpressionsInPython"

import re

def split\_uppcase(text):

pattern=(r"(?<=[a-z])(?=[A-Z])")

split=re.split(pattern,text)

return split

split\_uppcase(sample\_text)

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

Sample Text: “RegularExpression1IsAn2ImportantTopic3InPython"

Expected Output: RegularExpression 1IsAn 2ImportantTopic 3InPython

sample\_text ="RegularExpression1IsAn2ImportantTopic3InPython"

import re

def word\_numb\_space(line):

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

sub=re.sub(pattern,r" ", line)

return sub

word\_numb\_space(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 create\_space(word):

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

result=re.sub(pattern,r" \1", word)

return result

in\_put="RegularExpression1IsAn2ImportantTopic3InPython"

create\_space(in\_put)

**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

import re

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

ds=pd.read\_csv("happiness.csv")

ds

df=pd.DataFrame(ds)

df['four\_five\_six'] = df['Country'].str.extract(r'(\b\w{6}\b)')

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

import re

string= "The\_Ram\_mandir18@ is located @ \_Ayodhya, the date of pran prastisha is 22\_."

def match\_character(text):

pattern=r"\w+"

match=re.match(pattern,string)

return match

match\_character(string)

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

import re

string ="The 1vaccine derive for 1family is always 1to the 1free 1dose "

def numb\_string(text):

pattern="1[\w]+"

result=re.findall(pattern,string)

return result

numb\_string(string)

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

import re

def rem\_lead\_zero(ip):

modified\_ip=re.sub(r"\b0\*(\d+)",r"\1",ip)

return modified\_ip

input\_ip="216.08.094.196"

rem\_lead\_zero(input\_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.

sample="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’."

result=re.findall("A[\w\s]+.[\d+\b]",sample)

print(result)

main\_result=str(result)

with open("date.txt","w") as file:

file.write(main\_result)

**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'

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

import re

comp\_pat=re.compile("fox|dog|horse")

for match in comp\_pat.finditer(text):

print(match)

print(match.group())

**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

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

compile\_pattern= re.compile("fox")

search= re.search(compile\_pattern,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.

ts="The island of Superman"

pattern=re.compile("Super(hero)\*man")

search=pattern.search(ts)

print(search)

print(search.group())

ts="The island of Superheroman"

pattern=re.compile("Super(hero)\*man")

search=pattern.search(ts)

print(search)

print(search.group())

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

import re

def change\_date\_format(dt):

mod\_date=re.sub("(\d{4})-(\d{1,2})-(\d{1,2})",r"\3-\2-\1",dt)

return mod\_date

dt1="2024-02-03"

print("Year month day format",dt1)

print("The Req format",change\_date\_format(dt1))

**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']

def decimal\_value(line):

comp\_pattern=re.compile(r"\b\d+\.\d{1,2}\b")

result=comp\_pattern.findall(line)

return result

text="01.12 0132.123 2.31875 145.8 3.01 27.25 0.25"

print("The Expected output:",decimal\_value(text))

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

import re

def sep\_print\_position(line):

result=re.finditer(r"\d+",line)

for match in result:

print(match)

print(match.group())

text="Yashavi jaiswal has scored 200 in 1 match of his test debut."

sep\_print\_position(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

def max\_numeric\_value(val):

result=re.findall(r"\d+",val)

return max(result)

string="947, 896, 926, 524, 734, 950, 642"

print("Maximun number:",max\_numeric\_value(string))

**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

def insert\_space(text):

pattern=r"([A-Z][a-z]\*)"

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

return result

string= "RegularExpressionIsAnImportantTopicInPython"

insert\_space(string)

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

import re

def case\_letter(val):

pattern="([A-Z][a-z]+)"

search=re.findall(pattern,val)

print(search)

input\_text="Python Regex to Find Sequence of one Upper CASE Letter Followed BY Lower Case LETters"

case\_letter(input\_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

def rem\_dup\_words(text):

pattern=r"\b(\w+)(?:\W+\1\b)+"

sub=re.sub(pattern,r"\1",text)

print(sub)

string="Hello hello world world"

rem\_dup\_words(string)

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

def alpha\_numeric(text):

regex=r"\w$"

if re.search(regex,text):

print("Accept")

else:

print("Decline")

string1="Anil@123"

string2="Anil#"

string3="Anil\_"

alpha\_numeric(string1)

alpha\_numeric(string2)

alpha\_numeric(string3)

**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']

def extract\_hashtag(text):

regex=r"\#[\w]+"

find=re.findall(regex,text)

print(find)

string="""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"""

extract\_hashtag(string)

**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

def clean\_text(val):

pattern=r"<U\+[\w]+>”

sub=re.sub(pattern, "",val)

print(sub)

sample\_text="@Jags123456 Bharat band on28??<ed><U+00A0><U+00BD><ed><U+00B8><U+0082>Those who are protesting #demonetization are all different party leaders"

clean\_text(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.

sample\_text="Ron was born on 12-09-1992 and he was admitted to school 15-12-1999."

with open("sample\_text.txt","w") as file:

file.write(sample\_text)

def extract\_dates(file\_path):

with open(file\_path,"r") as file:

text=file.read()

regex=re.compile(r"(\d{2})-(\d{2})-(\d{4})")

dates= regex.findall(text)

return dates

ed=extract\_dates("sample\_text.txt")

print("Extracted Dates:",ed)

**Question 30-** Create a function inpython 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 trim\_words(line):

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

result=pattern.sub("", line)

print(result)

input\_string="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."

trim\_words(input\_string)