

# Regular Expression Project

```
In [1]: import pandas as pd
import numpy as np
import re
import matplotlib as matlab
```

Question 1- Write a Python program to replace all occurrences of a space, comma, or dot with a colon.

```
In [2]: text="Python Exercises, PHP exercises."
print(re.sub(r"[ ,.]", ":", text))
```

Python:Exercises::PHP:exercises:

In [ ]: Question 2- Create a data frame using the dictionary below **and** remove everything (commas (,), !, XXXX, ;, etc

```
In [7]: import pandas as pd
text={'SUMMARY' : ['hello, world!', 'XXXXX test', '123four, five;; six...']}
df=pd.dataframe(data)
df['SUMMARY']=df['SUMMARY'].str.replace(r'^a-zA-Z\s', '', regex=True)
print(df)
```

-----  
**AttributeError**

Traceback (most recent call last)

Cell In[7], line 3

```
1 import pandas as pd
2 text={'SUMMARY' : ['hello, world!', 'XXXXX test', '123four, five;; six...']}
----> 3 df=pd.dataframe(data)
4 df['SUMMARY']=df['SUMMARY'].str.replace(r'^a-zA-Z\s', '', regex=True)
5 print(df)
```

**AttributeError:** module 'pandas' has no attribute 'dataframe'

In [ ]: Question 3- Create a function in python to find all words that are at least 4 characters long in a string. The

```
In [11]: import re
text="Im Fareed Ahmad im from Saharanpur City is located on Western Uttar Pradesh"
print (re.findall(r"\b\w{4,}\b",text))

['Fareed', 'Ahmad', 'from', 'Saharanpur', 'City', 'located', 'Western', 'Uttar', 'Pradesh']
```

In [ ]: Question 4- Create a function in python to find all three, four, and five character words in a string. The use

```
In [16]: import re
text="Im Fareed Ahmad Im from Saharanpur city is located on Western Uttar Pradesh"
print(re.findall(r"\b\w{3,5}\b",text))

['Ahmad', 'from', 'city', 'Uttar']
```

In [ ]: Question 5- Create a function in Python to remove the parenthesis in a list of strings. The use of the re.com

```
In [32]: import re
items=["chaudharyfareed(.com)","info@angels(.com)","github(.com)","Hello(.Data Science)"]
for item in items :
    print(re.sub(r"\([^\)]+\)", "", item))

chaudharyfareed
info@angels
github
Hello
```

In [ ]: Question 6- Write a python program to remove the parenthesis area from the text stored in the text file using

```
In [33]: import re
items=["chaudharyfareed(.com)","info@angels(.com)","github(.com)","Hello(.Data Science)"]
for item in items :
    print(re.sub(r"\([^)]*\)", "", item))
```

```
chaudharyfareed
info@angels
github
Hello
```

In [ ]: Question 7- Write a regular expression in Python to split a string into uppercase letters.

```
In [34]: import re
text= "ImportanceOfRegularExpressionInPython"
print(re.findall('[A-Z][^A-Z]*',text))
```

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

In [ ]: Question 8- Create a function in python to insert spaces between words starting with numbers.

```
In [46]: target_string=" RegularExpression1IsAn2ImportantTopic3InPython "
print(target_string)
res_str=re.sub("r\s+", "",target_string)
print(res_str)
```

```
RegularExpression1IsAn2ImportantTopic3InPython
RegularExpression1IsAn2ImportantTopic3InPython
```

In [ ]: Question 9- Create a function in python to insert spaces between words starting with capital letters or with

```
In [50]: import re
string=" RegularExpression1IsAn2ImportantTopic3InPython "
words=re.findall('[A-Z][a-z]*',string)
print(' '.join((words)))
```

```
Regular Expression Is An Important Topic In Python
```

In [ ]: Question 10- Use the github link below to read the data and create a dataframe. After creating the dataframe

```
In [13]: import pandas as pd
url='https://raw.githubusercontent.com/dsrscientist/DSDData/master/happiness_score_dataset.csv'
df=pd.read_csv(url,index_col=0)
#df=pd.read_csv(url)
print(df.head(5))
```

	Region	Happiness Rank	Happiness Score	Standard Error	\
Country					
Switzerland	Western Europe	1	7.587	0.03411	
Iceland	Western Europe	2	7.561	0.04884	
Denmark	Western Europe	3	7.527	0.03328	
Norway	Western Europe	4	7.522	0.03880	
Canada	North America	5	7.427	0.03553	

	Economy (GDP per Capita)	Family	Health (Life Expectancy)	\
Country				
Switzerland	1.39651	1.34951	0.94143	
Iceland	1.30232	1.40223	0.94784	
Denmark	1.32548	1.36058	0.87464	
Norway	1.45900	1.33095	0.88521	
Canada	1.32629	1.32261	0.90563	

	Freedom	Trust (Government Corruption)	Generosity	\
Country				
Switzerland	0.66557	0.41978	0.29678	
Iceland	0.62877	0.14145	0.43630	
Denmark	0.64938	0.48357	0.34139	
Norway	0.66973	0.36503	0.34699	
Canada	0.63297	0.32957	0.45811	

	Dystopia Residual
Country	
Switzerland	2.51738
Iceland	2.70201
Denmark	2.49204
Norway	2.46531
Canada	2.45176

```
In [14]: df
print(df.head(5))
```

	Region	Happiness Rank	Happiness Score	Standard Error	\
Country					
Switzerland	Western Europe	1	7.587	0.03411	
Iceland	Western Europe	2	7.561	0.04884	
Denmark	Western Europe	3	7.527	0.03328	
Norway	Western Europe	4	7.522	0.03880	
Canada	North America	5	7.427	0.03553	

	Economy (GDP per Capita)	Family	Health (Life Expectancy)	\
Country				
Switzerland	1.39651	1.34951	0.94143	
Iceland	1.30232	1.40223	0.94784	
Denmark	1.32548	1.36058	0.87464	
Norway	1.45900	1.33095	0.88521	
Canada	1.32629	1.32261	0.90563	

	Freedom	Trust (Government Corruption)	Generosity	\
Country				
Switzerland	0.66557	0.41978	0.29678	
Iceland	0.62877	0.14145	0.43630	
Denmark	0.64938	0.48357	0.34139	
Norway	0.66973	0.36503	0.34699	
Canada	0.63297	0.32957	0.45811	

	Dystopia Residual
Country	
Switzerland	2.51738
Iceland	2.70201
Denmark	2.49204
Norway	2.46531
Canada	2.45176

```
In [ ]: Question 11- Write a Python program to match a string that contains only upper and lowercase letters, numbers
```

```
In [26]: def text_match(text):  
         patterns='^[a-zA-Z0-9_]*$'  
         re.match(patterns,text)is not None  
input_text="Im fareed ahmad from im from saharanpur"  
print(text_match(input_text))
```

None

In [ ]: Question 12- Write a Python program where a string will start with a specific number.

```
In [46]: import re  
def match_num(string):  
    pattern=re.compile(r"^5")  
    if pattern.match(string):  
        return True  
    else:  
        return False  
  
print(match_num("5-5465465"))  
print(match_num("6-5646585"))
```

True  
False

In [ ]: Question 13- Write a Python program to remove leading zeros from an IP address

```
In [49]: def remove_zeros_from_ip(ip_adr):  
         return '.'.join(p.lstrip('0') or '0' for p in ip_adr.split('.'))  
ip = '10.000.002.30'  
print(remove_zeros_from_ip(ip))
```

10.0.2.30

In [ ]: Question 14- Write a regular expression in python to match a date string in the form of Month name followed by



```
In [63]: import re
target_string = "On August 15th 1947 that India was declared independent from British colonialism, and the re
pattern = r"([A-Z][a-z]+) \d{1,2}\d{4}\b"
matches = re.findall(pattern, target_string)
print(matches)
```



[]

In [ ]: Question 15- Write a Python program to search some literals strings in a string.

```
In [64]: text='The quick brown fox jumps over the lazy dog.'
```

```
In [66]: import re
match=re.findall(r"fox|dog|horse",text)
print(match)
```

['fox', 'dog']

In [ ]: Question 16- Write a Python program to search a literals string in a string and also find the location within



```
In [67]: text1='The quick brown fox jumps over the lazy dog.'
```

```
In [71]: import re
match = re.search(r"fox",text1)
print(match)
```

<re.Match object; span=(16, 19), match='fox'>

In [ ]: Question 17- Write a Python program to find the substrings within a string.

In [70]: text2='Python exercises, PHP exercises, C# exercises'

```
In [75]: pattern='exercises'
match = re.findall(pattern,text2)
print(pattern)
```

exercises

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

In [78]: text3="Im fareed ahmad Im from saharanpur city "

```
In [83]: pattern = 'ahmad'
for match in re.finditer(pattern,text3):
    s=match.start()
    e=match.end()
print(pattern)
```

ahmad

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

In [89]: date='2000-08-12'

```
In [94]: import datetime
input='2000/08/12'
format='%Y/%m/%d'
datetime=datetime.datetime.strptime(input,format)
print(datetime.date())
```

2000-08-12



In [ ]: Question 20- Create a function in python to find all decimal numbers with a precision of 1 or 2 in a string.

In [96]: decimals='01.12 0132.123 2.31875 145.8 3.01 27.25 0.25'

```
In [103]: import re
def find_decimal_numbers(decimals):
    pattern= re.compile(r'\d+\.\d{1,2}')
    decimal_numbers= re.findall(pattern,decimals)
    return decimal_numbers
decimals ="01.12 0132.123 2.31875 145.8 3.01 27.25 0.25"
matching_numbers=find_decimal_numbers(decimals)
print(matching_numbers)

['01.12', '0132.12', '2.31', '145.8', '3.01', '27.25', '0.25']
```

```
In [106]: def is_decimal(decimals):
import re
dnumre=re.compile("^[0-9]+(\.[0-9]{1,2})?$")
result=dnumre.search(decimals)
return bool(result)
print(decimals)
```

01.12 0132.123 2.31875 145.8 3.01 27.25 0.25

In [ ]: Question 21- Write a Python program to separate and print the numbers and their position of a given string.

```
In [3]: import re
text = "Thirtyone 31, Fortytwo 42, Fiftyfive 55 "
result=re.search("\D+",text)
print(text,result)
```

Thirtyone 31, Fortytwo 42, Fiftyfive 55 <re.Match object; span=(0, 10), match='Thirtyone '>

In [ ]: Question 22- Write a regular expression in python program to extract maximum/largest numeric value from a string.

```
In [7]: import re
Sample_Text='My marks in each semester are: 947, 896, 926, 524, 734, 950, 642'
number =re.findall('\d+',Sample_Text)
print("Max_number:",max(number))

Max_number: 950
```

In [ ]: Question 23- Create a function in python to insert spaces between words starting with capital letters.

```
In [8]: import re
string="RegularExpressionIsAnImportantTopicInPython"
words=re.findall('[A-Z][a-z]*',string)
print(' '.join((words)))

Regular Expression Is An Important Topic In Python
```

In [ ]: Question 24- Python regex to find sequences of one upper case letter followed by lower case letters.

```
In [9]: text="Im Fareed Ahmad im from Saharanpur City is Located in Western Uttarpradesh"
```

```
In [12]: import re
pattern = r'[A-Z][a-z]+'
matches= re.findall(pattern,text)
print(matches)

['Im', 'Fareed', 'Ahmad', 'Saharanpur', 'City', 'Located', 'Western', 'Uttarpradesh']
```

In [ ]: Question 25- Write a Python program to remove continuous duplicate words from Sentence using Regular Expression

```
In [13]: text1="Hello hello world world"
```

```
In [27]: import re
def Remove_Duplicates(text):
    Pattern = r"\b(\w+)(?:\W\1\b)+"
    return re.sub(Pattern, r"\1",text,flags=re.IGNORECASE)
print(Remove_Duplicates(text1))
```

Hello world

In [ ]:

In [ ]:

In [ ]: