Python Regular Expression Tutorial Exericse

```
In [1]:  import re
```

1. Extract all twitter handles from following text. Twitter handle is the text that appears after https://twitter.com/ and is a single word. Also it contains only alpha numeric characters i.e. A-Z a-z , o to 9 and underscore _

```
In [2]:  text = '''
         Follow our leader Elon musk on twitter here: https://twitter.com/elonmusk, more
         on Tesla's products can be found at https://www.tesla.com/. Also here are leadin
         for tesla related news,
         https://twitter.com/teslarati
         https://twitter.com/dummy_tesla
         https://twitter.com/dummy_2_tesla
         '''
```

```
In [3]:  pattern =r'[a-z]*://[a-z]*.com/[a-zA-Z0-9_]+'
         re.findall(pattern,text)
```

```
Out[3]:  ['https://twitter.com/elonmusk',
          'https://twitter.com/teslarati',
          'https://twitter.com/dummy_tesla',
          'https://twitter.com/dummy_2_tesla']
```

OR

```
In [4]:  #Now I want only names
         pattern = r'https://twitter\.com/([a-zA-Z0-9_]+)'
         re.findall(pattern,text)
```

```
Out[4]:  ['elonmusk', 'teslarati', 'dummy_tesla', 'dummy_2_tesla']
```

OR

```
In [5]:  pattern =r'[a-z]*://[a-z]*.com/([a-zA-Z0-9_]+)'
         re.findall(pattern,text)
```

```
Out[5]:  ['elonmusk', 'teslarati', 'dummy_tesla', 'dummy_2_tesla']
```

```
In [ ]:
```

2. Extract Concentration Risk Types. It will be a text that appears after "Concentration Risk:", In below example, your regex should extract these two strings (1) Credit Risk (2) Supply Rish

```
In [6]:  text = '''
         Concentration of Risk: Credit Risk
         Financial instruments that potentially subject us to a concentration of credit r
         restricted cash, accounts receivable, convertible note hedges, and interest rate
         or on deposit at high credit quality financial institutions in the U.S. These de
         and December 31, 2020, no entity represented 10% or more of our total accounts r
         hedges and interest rate swaps is mitigated by transacting with several highly-r
         Concentration of Risk: Supply Risk
         We are dependent on our suppliers, including single source suppliers, and the in
         products in a timely manner at prices, quality levels and volumes acceptable to
```

```
suppliers, could have a material adverse effect on our business, prospects, fina
'''
```

In [7]:
```python
pattern ='Concentration of Risk:([^\n]+)'
re.findall(pattern,text)
```

Out[7]: `[' Credit Risk', ' Supply Risk']`

In [ ]:

3. Companies in europe reports their financial numbers of semi annual basis and you can have a document like this. To exatract quarterly and semin annual period you can use a regex as shown below Hint: you need to use (?:) here to match everything enclosed

In [8]:
```python
text = '''
Tesla's gross cost of operating lease vehicles in FY2021 Q1 was $4.85 billion.
BMW's gross cost of operating vehicles in FY2021 S1 was $8 billion.
'''
```

In [9]:
```python
pattern ='FY(\d* (?:Q[1-4]|S[1-2]))'
re.findall(pattern,text)
```

```
<>:1: SyntaxWarning: invalid escape sequence '\d'
<>:1: SyntaxWarning: invalid escape sequence '\d'
C:\Users\sriha\AppData\Local\Temp\ipykernel_25832\2632939244.py:1: SyntaxWarning:
invalid escape sequence '\d'
  pattern ='FY(\d* (?:Q[1-4]|S[1-2]))'
```

Out[9]: `['2021 Q1', '2021 S1']`

In [ ]:

In [ ]:

In [ ]: