functions

June 14, 2021

1 Functions

1.1 basic functions

[3]:
```
def printme( str ):
   "This prints a passed string into this function"
   print (str)
   return

# Now you can call printme function
printme("This is my first function call")
printme("This is my second function call")
```

```
This is my first function call
This is my second function call
```

[4]:
```
def changeme( mylist ):
   "This changes a passed list into this function"
   print ("Values inside the function before change: ", mylist)

   mylist[2]=50
   print ("Values inside the function after change: ", mylist)
   return

# Now you can call changeme function
mylist = [10,20,30]
changeme( mylist )
print ("Values outside the function: ", mylist)
```

```
Values inside the function before change:  [10, 20, 30]
Values inside the function after change:  [10, 20, 50]
Values outside the function:  [10, 20, 50]
```

1.2 Regular Expressions

[6]:
```
fhand = open('Emails.txt')
for line in fhand:
    line = line.rstrip()
    if not '@uct.ac.za' in line :
```

```
        continue
        print (line)
```

```
---------------------------------------------------------------------------
FileNotFoundError                         Traceback (most recent call last)
<ipython-input-6-0a938bb3c97e> in <module>
----> 1 fhand = open('Emails.txt')
      2 for line in fhand:
      3     line = line.rstrip()
      4     if not '@uct.ac.za' in line :
      5         continue

FileNotFoundError: [Errno 2] No such file or directory: 'Emails.txt'
```

[ ]:
```
print("\nSearching Through a File\n")
fhand = open('Emails.txt')
for line in fhand:
    line = line.rstrip()
    if line.startswith('From:') :
        print (line)
```

[ ]:
```
# ext starting with a capital X followed by any character repeated zero or more␣
↪times and ending with a colon (:).
```

[ ]:
```
import re
print ("\nRegular Expressions\n'^X.*:' \n") hand =
open('Data.txt')
for line in hand:
    line = line.rstrip()
    y = re.findall('^X.*:',line)
    print (y)
```

[8]:
```
#Extracting Numerical Values and Specific Characters
import re
print ("\n Matching and Extracting Data \n")
x = 'My 2 favorite numbers are 19 and 42'
y = re.findall('[0-9]+',x)
print (y)
```

```
 Matching and Extracting Data

['2', '19', '42']
```

[10]:
```
#Python finds a string starting with F and containing any number of characters␣
↪up to a colon and then stops when it reaches the end of the line.
import re
```

```
print ("\nGreedy Matching \n")
x = 'From: Using the : character'
y = re.findall('^F.+:', x)
print (y)
```

```
Greedy Matching

['From: Using the :']
```

[11]:
```
# In the second example, re.findall('^F.+?:', x) asks Python to retrieve␣
 ↪characters starting with an F and ending with the first occurrence of a␣
 ↪delimiter, which is a colon regardless of whether it reached the end of the␣
 ↪line or not.
print ("\nNon-Greedy Matching \n")
x = 'From: Using the : character'
y = re.findall('^F.+?:', x)
print (y)
```

```
Non-Greedy Matching

['From:']
```

1.3 The Use of Methods vs. Regular Expressions

[13]:
```
import re
print ("\nFine-Tuning String Extraction \n")
mystr="From ossama.embarak@hct.ac.ae Sat Jun 5 08:14:16 2018"
Extract = re.findall('\S+@\S+',mystr)
print (Extract)
E_xtracted = re.findall('^From.*? (\S+@\S+)',mystr)
print (E_xtracted)
print (E_xtracted[0])
```

```
Fine-Tuning String Extraction

['ossama.embarak@hct.ac.ae']
['ossama.embarak@hct.ac.ae']
ossama.embarak@hct.ac.ae
```

[14]:
```
mystr="From ossama.embarak@hct.ac.ae Sat Jun 5 08:14:16 2018"
atpos = mystr.find('@')
sppos = mystr.find(' ',atpos) # find white space starting from atpos
host = mystr[atpos+1 : sppos]
print (host)
```

```
usernamepos = mystr.find(' ')
username = mystr[usernamepos+1 : atpos]
print (username)
```

```
hct.ac.ae
ossama.embarak
```

[15]:
```
print ("\n The Regex Version\n")
import re
mystr="From ossama.embarak@hct.ac.ae Sat Jun 5 08:14:16 2018"
Extract = re.findall('@([^ ]*)',mystr)
print (Extract)
Extract = re.findall('^From .*@([^ ]*)',mystr)
print (Extract)
```

```
 The Regex Version

['hct.ac.ae']
['hct.ac.ae']
```

[16]:
```
print ("\nScape character \n")
mystr = 'We just received $10.00 for cookies and $20.23 for juice'
Extract = re.findall('\$[0-9.]+',mystr)
print (Extract)
```

```
Scape character

['$10.00', '$20.23']
```

1.4 Defining and Calling Functions for Data Cleaning

[17]:
```
def CleanData_Sales(cell):
    if (cell=="n.a." or cell=="-1" or cell=="not avilable"):
        return 0
    return cell

def CleanData_REGION(cell):
    if (cell=="n.a." or cell=="-1" or cell=="not avilable"):
        return 'AbuDhabi'
    return cell
```

[ ]:
```
sales = pd.read_csv("Sales.csv", nrows=7, converters={
                            "SALES_BY_REGION": CleanData_REGION,
                            "JANUARY": CleanData_Sales,
                            "FEBRUARY": CleanData_Sales,
```

```
                            "APRIL": CleanData_Sales,
                            })
sales.head(20)
```