Files handling ¶

- File is a collection of records
- Python file handling : create, open, append, Read, Write
  - Create a new file
  - Write the data
  - Read data from a file
  - Append data to a file

Mode of operations

- 'r' - Read mode which is used when the file is only being read
- 'w' - Write mode which is used to edit and write new data to the file(overwrites the existing data)
- 'a' - Append mode ,which is used to add new data to end of the file,new data will autromatically added to existing data
- 'r+' - Special read and write mode, which is used to handle both actions read and write

In Build functions:

- Open : open(path,mode)
- Close : file_name.close()

In [2]:

```
# create a file and writing data to a file

f=open("sample.txt",'w')  # opening file in write mode
f.write("hello world")   # writing data to a file
f.close()
```

In [3]:

```
# write

f=open("sample.txt",'w')
f.write("python programming")
f.close()
```

In [4]:

```
# Read the file data

f=open("sample.txt",'r')
data=f.read()
print(data)
f.close()
```

```
python programming
```

In [5]:

```
# read() -> read all data
# readlines() -> returns all file data lines in list format
# readline() -> return a line from a file
```

In [8]:

```
f=open("sample.txt",'r')
print(f.readlines())
f.readline()
```

```
['python programming\n', 'hello world\n', 'programming languages\n']
```

Out[8]:

```
''
```

In [9]:

```
f=open("sample.txt",'r')
f.readline()
```

Out[9]:

```
'python programming\n'
```

In [10]:

```
# read()

f=open("sample.txt",'r')
f.read()
```

Out[10]:

```
'python programming\nhello world\nprogramming languages\n'
```

In [11]:

```
# append

f=open("sample.txt",'a')
f.write("i am learning data analysis")
f.close()
```

In [12]:

```
f=open("sample.txt",'r')
print(f.read())
f.close()
```

```
python programming
hello world
programming languages
i am learning data analysis
```

In [17]:

```
# read file data line by line  , finding no of lines in a file

f=open("sample.txt",'r')
data=f.readlines()
n=len(data)
print("number of file lines: ",n)
for i in data:
    print(i)
```

```
number of file lines:  4
python programming

hello world

programming languages

i am learning data analysis
```

In [19]:

```
# read file data character by character  , finding total no of characters

f=open("sample.txt",'r')
data=f.read()
l=len(data)
print("no of characters: ",l)
for ch in data:
    print(ch)
```

```
no of characters:  80
p
y
t
h
o
n

p
r
o
g
r
a
m
m
i
n
g


h
e
l
l
o

w
o
r
l
d


p
r
o
g
r
a
m
m
i
n
g

l
a
n
g
u
a
g
e
s


i

a
m
```

```
l
e
a
r
n
i
n
g

d
a
t
a

a
n
a
l
y
s
i
s
```

In [22]:

```
# find no of words inside a file

f=open("sample.txt",'r')
data=f.read()
words=data.split()
nw=len(words)
print(words)
print("no of words: ",nw)
```

```
['python', 'programming', 'hello', 'world', 'programming', 'languages',
'i', 'am', 'learning', 'data', 'analysis']
no of words:  11
```

In [25]:

```
# add 60 students

f1=open("students.txt",'w')
for i in range(1,61):
    f1.write("student %i  \n" %(i))
f1.close()
```

methods in file handling

- tell() -> returns the curent location of the file pointer
- seek() -> we can set the pointer at a specific location
- seek(offset,from_what)
  - offset - no of positions to move forward
  - from_what - defines the pointer reference
  - 0 : set reference pointer at the begining of the file

- 1 : set reference pointer at the current file position
- 2 : set reference pointer at the end of a file

In [28]:

```
f=open("sample.txt",'r')
print(f.tell())
print(f.readline())
print(f.tell())
```

```
0
python programming

19
```

In [29]:

```
f=open("sample.txt",'r')
f.seek(24)
```

Out[29]:

```
24
```

Data Analysis

- Data: data are facts and statistics collected together for reference or analysis
- Data analysis :
  - is a process of obtaining raw data and subsequently converting it into information , useful for decision making
  - is a technique to collect, tranform and organize data to make future predictions
  - all sorts of data analysis, Analyzing the numerical data with numpy , tabular data with pandas, visualization by matplotlib and exploratory data analysis

Data collection:

- Primary Data :
  - refers to first hand data gathered by the research himself/herself
  - real time data
  - Surveys, observations, experiments, questionnaire,personal interviews etc
  - Expensive and takes long time
- Secondary data:
  - data which is collected by someone else earlier
  - past data
  - websites,books,journal , articles, internal records, government publications etc
  - Economical and takes short time

Types of data

- Structed data : row/columns -> tabular format -> excel file,.csv,.tsv etc
- Unstructed data : pictures, videos,text,ppts,pdf,log files, voices

- Semi structed data: xml,json ,html files {key:value}

Data types in Statistics:

- Numerical data: data in the form of numbers
  - 1.Discrete data:

```
                    * is a count that involves only intergers. can't be subd
    ivided into parts.
                    * ex: no of students in a class'
                    * no of workers in  a company etc
```

  - 2.Continous data :

```
                    * Continuos data could be meaningfully divided into fing
    er levels.
                    * It can be a measured on a scale & have a numerical val
    ue
                    * ex: amount of time required to complete a project
                    * height,weight of a children etc..
```

- Categorical data: data made of words
  - 1.Nominal data :

```
                  * nominal data is used for labelling variables
                  * ex: Gender(women,men)'
                  * Hair color
                  * Marital status(Married,single)
```

```
* 2.Ordinal data:
                  * ordinal data is data which is placed into some kind of o
rder by their position on a scale
                  * EX: 1st,2nd,3rd persons in a competetion
                  * economic status: low,medium ,high
```

In [ ]: