In [ ]:

```
whenever you want to read any file

type of file

- .txt

- .csv     (comma seperted value)

- .xlsx

- .json   (dictionay format)

- .xml    (IOT data)

- .parquet  (encoded files)

- .delta    (encoded files)

- .pdf

- .jpg

- .png
```

In [ ]:

```
- where is your file located:  <location>

- what is your file name

- what is the type of file (extention)
```

In [ ]:

```
C:\Users\omkar\OneDrive\Documents\Data science\Naresh IT\Datafiles

mbox-short

.txt
```

In [ ]:

```
file_location="C:\Users\omkar\OneDrive\Documents\Data science\Naresh IT\Data
file_name="mbox-short"
extention=".txt"
```

In [ ]:

```
file_location+file_name+extention
```

In [2]:

```
"C:\Users\omkar\OneDrive\Documents\Data science\Naresh IT\Datafiles"
```

```
  Cell In[2], line 1
    "C:\Users\omkar\OneDrive\Documents\Data science\Naresh IT\Datafiles"
                                                                        ^

SyntaxError: (unicode error) 'unicodeescape' codec can't decode bytes in p
osition 2-3: truncated \UXXXXXXXX escape
```

```
In [ ]: - unicode error : single slash

        - in order to read the file , we need provide '\\'
```

```
In [ ]: "C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Datafiles\\
```

```
In [ ]: "C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Datafiles\\
```

```
In [3]: file_path="C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\D
        file_path
```

```
Out[3]: 'C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Datafiles
        \\mbox-short.txt'
```

```
In [4]: open(file_path,encoding='utf-8')

        # mode: read: r/write:w
        # emcoding: cp1252/utf-8
```

```
Out[4]: <_io.TextIOWrapper name='C:\\Users\\omkar\\OneDrive\\Documents\\Data scien
        ce\\Naresh IT\\Datafiles\\mbox-short.txt' mode='r' encoding='cp1252'>
```

- whenever eny file has spl charcters it will not read
- that time you need to provide encoding value
- different emcoded files has different encoding method
- 'utf-8'/'cp-1252'

```
In [14]: file_path="C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\D
         file=open(file_path,encoding='utf-8-sig')
         print(file.read())

         #\n: new line

         #\t: tab
```

```
From stephen.marquard@uct.ac.za Sat Jan  5 09:14:16 2008
Return-Path: <postmaster@collab.sakaiproject.org>
Received: from murder (mail.umich.edu [141.211.14.90])
         by frankenstein.mail.umich.edu (Cyrus v2.3.8) with LMTPA;
         Sat, 05 Jan 2008 09:14:16 -0500
X-Sieve: CMU Sieve 2.3
Received: from murder ([unix socket])
         by mail.umich.edu (Cyrus v2.2.12) with LMTPA;
         Sat, 05 Jan 2008 09:14:16 -0500
Received: from holes.mr.itd.umich.edu (holes.mr.itd.umich.edu [141.211.1
4.79])
        by flawless.mail.umich.edu () with ESMTP id m05EEFR1013674;
        Sat, 5 Jan 2008 09:14:15 -0500
Received: FROM paploo.uhi.ac.uk (app1.prod.collab.uhi.ac.uk [194.35.219.
184])
        BY holes.mr.itd.umich.edu ID 477F90B0.2DB2F.12494 ;
         5 Jan 2008 09:14:10 -0500
Received: from paploo.uhi.ac.uk (localhost [127.0.0.1])
        by paploo.uhi.ac.uk (Postfix) with ESMTP id 5F919BC2F2;
        Sat,  5 Jan 2008 14:10:05 +0000 (GMT)
```

In [ ]:

```

```