

# ASSGNMENT –PYTHON –churning email

For Data Analytics Certification

Prepared by Anees Chandrankandy

## STEPS:

### 1. LOAD THE DATA SET

Imported text file to Python from respective file using below code;

```
file= open ("/Users/Admin/Downloads/mbox.txt")  
file.read()
```

**Question 1: Count the number of lines;**

**Answer: 1,32,044**

```
def number_of_lines():  
    path="/Users/Admin/Downloads/mbox.txt"  
    f= open(path)  
    cnt =0  
    for x in f:  
        cnt = cnt+1  
    return(cnt)
```

# ASSGNMENT –PYTHON –churning email

For Data Analytics Certification

Prepared by Anees Chandrankandy

## 2. EXPLORE THE CONTENT OF EMAILS

### Question 1: Count the number of Subject lines

**Answer: 1,797**

```
def number_of_Subject_lines():  
    path="/Users/Admin/Downloads/mbox.txt"  
    f= open(path)  
    cnt =0  
    for x in f:  
        x=x.rstrip()  
        if x.startswith('Subject:'):   
            cnt = cnt+1  
    return(cnt)
```

### Question 2: Find the average spam confidence

**Answer: 0.89**

```
def average_spam_confidence():  
    with open("/Users/Admin/Downloads/mbox.txt") as f:  
        sum=0  
        cnt = 0  
        for c in f:  
            c=c.rstrip()  
            if c.startswith('X-DSPAM-Confidence:'):   
                sp=c.split(':')  
                sum+=float(sp[1])  
                cnt+=1  
        av=float(sum/cnt)  
  
    return av
```

## ASSGNMENT –PYTHON –churning email

For Data Analytics Certification

Prepared by Anees Chandrankandy

### Question 3: Find which day of the week E-mail was sent

**Answer:**

Days	No.Email
SAT	61
FRI	315
THU	392
WED	292
TUE	372
MON	299
SUN	66

```
import re
def find_email_sent_days():
    with open("/Users/Admin/Downloads/mbox.txt") as f:
        content=f.readlines()
        day_dict={}
        dayofweek=''
        for line in content:
            if re.findall('From ',line):
                dayofweek=line.split(' ')[2]
                day_dict[dayofweek]= day_dict.get(dayofweek,0) +1
        return day_dict
print(find_email_sent_days())
```

# ASSGNMENT –PYTHON –churning email

For Data Analytics Certification

Prepared by Anees Chandrankandy

## 3. EXPLORE THE HEADER OF EMAILS

**Question 1: Count no.of messages from each E-mail address**

**Answer :**

```
{'stephen.marquard@uct.ac.za': 58,  
'louis@media.berkeley.edu': 48,  
'zqian@umich.edu': 390,  
'rjlowe@iupui.edu': 180,  
'cwen@iupui.edu': 316,  
'gsilver@umich.edu': 56,  
'wagnermr@iupui.edu': 88,  
'antranig@caret.cam.ac.uk': 36,  
'gopal.ramasammycook@gmail.com': 50,  
'david.horwitz@uct.ac.za': 134,  
'ray@media.berkeley.edu': 64,  
'mmmay@indiana.edu': 322,  
'stuart.freeman@et.gatech.edu': 34,  
'tnguyen@iupui.edu': 12,  
'chmaurer@iupui.edu': 222,  
'aaronz@vt.edu': 220,  
'ian@caret.cam.ac.uk': 192,  
'csev@umich.edu': 38,  
'jimeng@umich.edu': 186,  
'josrodri@iupui.edu': 56,  
'knoop@umich.edu': 10,  
'bkirsch@umich.edu': 54,  
'dlhaines@umich.edu': 168,  
'hu2@iupui.edu': 14,  
'sgithens@caret.cam.ac.uk': 86,  
'arwhyte@umich.edu': 54,  
'gbhatnag@umich.edu': 6,  
'gjthomas@iupui.edu': 88,  
'a.fish@lancaster.ac.uk': 28,  
'ajpoland@iupui.edu': 96,
```

# ASSGNMENT –PYTHON –churning email

For Data Analytics Certification

Prepared by Anees Chandrankandy

```
'lance@indiana.edu': 16,  
'ssmail@indiana.edu': 10,  
'jlrenfro@ucdavis.edu': 2,  
'nuno@ufp.pt': 56,  
'zach.thomas@txstate.edu': 34,  
'ktsao@stanford.edu': 24,  
'ostermmg@whitman.edu': 34,  
'john.ellis@rsmart.com': 16,  
'jleasia@umich.edu': 4,  
'ggolden@umich.edu': 16,  
'thoppaymallika@fhda.edu': 2,  
'kimsooil@bu.edu': 28,  
'bahollad@indiana.edu': 8,  
'jzaremba@unicon.net': 18,  
'mbreuker@loi.nl': 18,  
'colin.clark@utoronto.ca': 2}
```

## **Code**

```
def count_message_from_email():  
    with open("/Users/Admin/Downloads/mbox.txt") as f:  
        s=' '  
        d={}  
        for line in f:  
            if line.startswith('From') and '@' in line:  
                #print(line)  
                sp=line.split()  
                email=sp[1]  
                d[email]=d.get(email, 0)+1  
        return d  
count_message_from_email()
```

## ASSGNMENT –PYTHON –churning email

For Data Analytics Certification

Prepared by Anees Chandrankandy

### Question 2: Count no.of messages from each domain

#### Answer :

```
{'uct.ac.za': 96,  
'media.berkeley.edu': 56,  
'umich.edu': 491,  
'iupui.edu': 536,  
'caret.cam.ac.uk': 157,  
'gmail.com': 25,  
'indiana.edu': 178,  
'et.gatech.edu': 17,  
'vt.edu': 110,  
'lancaster.ac.uk': 14,  
'ucdavis.edu': 1,  
'ufp.pt': 28,  
'txstate.edu': 17,  
'stanford.edu': 12,  
'whitman.edu': 17,  
'rsmart.com': 8,  
'fhda.edu': 1,  
'bu.edu': 14,  
'unicon.net': 9,  
'loi.nl': 9,  
'utoronto.ca': 1}
```

Code;

```
def count_message_from_domain():  
    lineslist=[]  
    domaindict={}  
    with open("/Users/Admin/Downloads/mbox.txt") as f:  
        for line in f:  
            lineslist = line.split()  
            if line.startswith('From:'):   
                email=lineslist[1]  
                domain = email.split('@')[1]  
                if domain not in domaindict:  
                    domaindict[domain] = 1  
                else:  
                    domaindict[domain] += 1  
    return domaindict
```