## **Step 1: Importing important libraries**

```
In [1]: ▶ import re
            import pandas as pd
            import email
Out[2]: ['__all__',
              '__builtins__',
              ___cached__',
              '__doc__'
              _____,
'__file__',
'__loader__',
              '__name__',
              '__package__',
'__path__',
               __spec__',
              '_encoded_words',
              '_parseaddr',
              '_policybase',
              'base64mime',
              'charset',
              'encoders'
              'errors',
              'feedparser',
              'header',
              'iterators',
              'message',
              'message_from_binary_file',
              'message_from_bytes',
              'message_from_file',
              'message_from_string',
              'parser',
              'quoprimime',
              'utils']
```

# Step 2: create an empty list and then import our corpus.

#### Step 3: Split the corpus so as to have the individual emails

# Step 4: Get senders name and email from the content.

```
In [5]: ▶ | for item in contents:
                emails_dict = {}
                sender = re.search('From:.*', item)
                print(sender.group())
            From: "MR. JAMES NGOLA." <james_ngola2002@maktoob.com>
            From: "Mr. Ben Suleman" <bensul2004nng@spinfinder.com>
            From: "PRINCE OBONG ELEME" <obong_715@epatra.com>
            From: "PRINCE OBONG ELEME" <obong_715@epatra.com>
            From: "Maryam Abacha" <m_abacha03@www.com>
            From: Kuta David <davidkuta@postmark.net>
            From: "Barrister tunde dosumu" <tunde_dosumu@lycos.com>
            From: "William Drallo" <william2244drallo@maktoob.com>
            From: "MR USMAN ABDUL" <abdul_817@rediffmail.com>
            From: "Tunde Dosumu" <barrister_td@lycos.com>
            From: MR TEMI JOHNSON <temijohnson2@rediffmail.com>
            From: "Dr.Sam jordan" <sjordan@diplomats.com>
            From: p_brown2@lawyer.com
            From: mic_k1@post.com
            From: "COL. MICHAEL BUNDU" <mikebunduu1@rediffmail.com>
            From: "MRS MARIAM ABACHA" <elixwilliam@usa.com>
            From: " DR. ANAYO AWKA " <anayoawka@hotmail.com>
            From: " DR. ANAYO AWKA " <anayoawka@hotmail.com>
            From: "Victor Aloma" <victorloma@netscape.net>
            From: "Victor Aloma" <victorloma@netscape.net>
            From: "JAMES NGOLA" <james_ngola2002@maktoob.com>
            AttributeError
                                                     Traceback (most recent call last)
            <ipython-input-5-fb12cc6de30e> in <module>
                  2 emails_dict = {}
                 3 sender = re.search('From:.*', item)
            ---> 4 print(sender.group())
            AttributeError: 'NoneType' object has no attribute 'group'
```

Step 5: find sender's name and email.

```
In [6]: | for item in contents:
                emails_dict = {}
                sender = re.search('From:.*', item)
                #print(sender)
                if sender is not None:
                    s email = re.search('\w\S*@.*\w', sender.group())
                    s_name = re.search('\".*\"', sender.group())
                else:
                    s_email = None
                    s name = None
                if s_email is not None:
                    sender_email = s_email.group()
                else:
                    sender_email = None
                emails_dict['sender_email'] = sender_email
                print(emails_dict)
             {'sender_email': 'temijohnson2@rediffmail.com'}
             {'sender_email': 'sjordan@diplomats.com'}
             {'sender_email': 'p_brown2@lawyer.com'}
             {'sender_email': 'mic_k1@post.com'}
             'sender_email': 'mikebunduu1@rediffmail.com'}
             {'sender_email': 'elixwilliam@usa.com'}
             {'sender_email': 'anayoawka@hotmail.com'}
             {'sender_email': 'anayoawka@hotmail.com'}
             {'sender_email': 'victorloma@netscape.net'}
             { 'sender email': 'victorloma@netscape.net' }
             {'sender_email': 'james_ngola2002@maktoob.com'}
             {'sender_email': None}
             {'sender_email': 'martinchime@usa.com'}
             {'sender_email': 'mboro1555@post.com'}
             {'sender_email': 'martinchime@borad.com'}
             {'sender_email': 'martinchime@borad.com'}
             {'sender_email': 'edema_mb@phantomemail.com'}
             {'sender_email': 'edema_mb@phantomemail.com'}
             {'sender_email': 'adewilliams_ade@lawyer.com'}
            {'sender_email': 'smithkam2@post.com'}
```

```
In [7]: ► for item in contents:
                emails dict = {}
                sender = re.search('From:.*', item)
                #print(sender)
                if sender is not None:
                    s email = re.search('\w\S*@.*\w', sender.group())
                    s_name = re.search('\".*\"', sender.group())
                else:
                    s_email = None
                    s name = None
                if s email is not None:
                    sender_email = s_email.group()
                else:
                    sender_email = None
                emails_dict['sender_email'] = sender_email
                #print(emails dict)
                if s name is not None:
                    sender_name = s_name.group()
                else:
                    sender_name = None
                emails_dict['sender_name'] = sender_name
                print(emails dict)
            {'sender_email': 'james_ngola2002@maktoob.com', 'sender_name': '"MR. JAMES NGOLA."'}
             {'sender_email': 'bensul2004nng@spinfinder.com', 'sender_name': '"Mr. Ben Suleman"'}
            {'sender_email': 'obong_715@epatra.com', 'sender_name': '"PRINCE OBONG ELEME"'}
             {'sender_email': 'obong_715@epatra.com', 'sender_name': '"PRINCE OBONG ELEME"'}
             'sender_email': 'm_abacha03@www.com', 'sender_name': '"Maryam Abacha"'}
             {'sender_email': 'davidkuta@postmark.net', 'sender_name': None}
             {'sender email': 'tunde dosumu@lycos.com', 'sender name': '"Barrister tunde dosumu"'}
             {'sender_email': 'william2244drallo@maktoob.com', 'sender_name': '"William Drallo"'}
             'sender_email': 'abdul_817@rediffmail.com', 'sender_name': '"MR USMAN ABDUL"'}
             {'sender_email': 'barrister_td@lycos.com', 'sender_name': '"Tunde Dosumu"'}
            {'sender_email': 'temijohnson2@rediffmail.com', 'sender_name': None}
             {'sender email': 'sjordan@diplomats.com', 'sender name': '"Dr.Sam jordan"'}
             {'sender_email': 'p_brown2@lawyer.com', 'sender_name': None}
            {'sender_email': 'mic_k1@post.com', 'sender_name': None}
            {'sender_email': 'mikebunduu1@rediffmail.com', 'sender_name': '"COL. MICHAEL BUNDU"'}
             {'sender_email': 'elixwilliam@usa.com', 'sender_name': '"MRS MARIAM ABACHA"'}
             {'sender_email': 'anayoawka@hotmail.com', 'sender_name': '" DR. ANAYO AWKA "'}
            {'sender_email': 'anayoawka@hotmail.com', 'sender_name': '" DR. ANAYO AWKA "'}
            {'sender_email': 'victorloma@netscape.net', 'sender_name': '"Victor Aloma"'}
```

Step 7: Execute similar step to get the emails and names of the recipient. (Classwork)

```
In [8]:
            {'sender_email': 'james_ngola2002@maktoob.com', 'sender_name': '"MR. JAMES NGOLA."', 'recipient_email': 'james_ngola2002@maktoob.com', 'rec
            {'sender_email': 'bensul2004nng@spinfinder.com', 'sender_name': '"Mr. Ben Suleman"', 'recipient_email': 'R@M', 'recipient_name': None}
            {'sender email': 'obong 715@epatra.com', 'sender name': '"PRINCE OBONG ELEME"', 'recipient email': 'obong 715@epatra.com', 'recipient name'
            {'sender_email': 'obong_715@epatra.com', 'sender_name': '"PRINCE OBONG ELEME"', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name'
            {'sender_email': 'm_abacha03@www.com', 'sender_name': '"Maryam Abacha"', 'recipient_email': 'm_abacha03@www.com', 'recipient_name': None}
            {'sender email': 'davidkuta@postmark.net', 'sender name': None, 'recipient email': 'davidkuta@yahoo.com', 'recipient name': None}
            {'sender_email': 'tunde_dosumu@lycos.com', 'sender_name': '"Barrister tunde dosumu"', 'recipient_email': 'tunde_dosumu@lycos.com', 'recipie
             ('sender_email': 'william2244drallo@maktoob.com', 'sender_name': '"William Drallo"', 'recipient_email': 'william2244drallo@maktoob.com', 'r
             {'sender email': 'abdul 817@rediffmail.com', 'sender name': '"MR USMAN ABDUL"', 'recipient email': 'R@M', 'recipient name': None}
             {'sender_email': 'barrister_td@lycos.com', 'sender_name': '"Tunde Dosumu"', 'recipient_email': 'barrister_td@lycos.com', 'recipient_name':
             {'sender email': 'temijohnson2@rediffmail.com', 'sender name': None, 'recipient email': 'R@E', 'recipient name': None}
             {'sender_email': 'sjordan@diplomats.com', 'sender_name': '"Dr.Sam jordan"', 'recipient_email': 'R@M', 'recipient_name': None}
            {'sender_email': 'p_brown2@lawyer.com', 'sender_name': None, 'recipient_email': 'p_brown2@lawyer.com', 'recipient name': None}
             {'sender_email': 'mic_k1@post.com', 'sender_name': None, 'recipient_email': 'R@M', 'recipient_name': None}
             'sender_email': 'mikebunduu1@rediffmail.com', 'sender_name': '"COL. MICHAEL BUNDU"', 'recipient_email': 'R@E', 'recipient_name': None}
             {'sender email': 'elixwilliam@usa.com', 'sender name': '"MRS MARIAM ABACHA"', 'recipient email': 'presidentwifeng@indiatimes.com', 'recipie
             {'sender_email': 'anayoawka@hotmail.com', 'sender_name': '" DR. ANAYO AWKA "', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name':
             {'sender_email': 'anayoawka@hotmail.com', 'sender_name': '" DR. ANAYO AWKA "', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name':
            {'sender_email': 'victorloma@netscape.net', 'sender_name': '"Victor Aloma"', 'recipient_email': 'ntcir-listmem@newns.op.nii.ac.jp', 'recipi
```

Step 8: Getting the date of the email

```
In [10]: ▶ for item in contents:
                 emails_dict = {}
                 sender = re.search('From:.*', item)
                 #print(sender)
                 if sender is not None:
                     s email = re.search('\w\S*@.*\w', sender.group())
                     s_name = re.search('\".*\"', sender.group())
                 else:
                     s_email = None
                     s name = None
                 if s_email is not None:
                     sender_email = s_email.group()
                 else:
                     sender_email = None
                 emails_dict['sender_email'] = sender_email
                 #print(emails_dict)
                 if s name is not None:
                     sender_name = s_name.group()
                 else:
                     sender_name = None
                 emails_dict['sender_name'] = sender_name
                 #print(emails_dict)
                 # Recipient
                 recipient = re.search('To:.*', item)
                 #print(recipient.group())
                 if recipient is not None:
                     r email = re.search('\w\S*@.*\w', recipient.group())
                     r_name = re.search('\".*\"', recipient.group())
                 else:
                     r_email = None
                     r_name = None
                 if r email is not None:
                     recipient_email = r_email.group()
                 else:
                     recipient_email = None
                 emails_dict['recipient_email'] = recipient_email
                 #print(emails_dict)
                 if r_name is not None:
                     recipient_name = r_name.group()
                 else:
                     recipient_name = None
                 emails_dict['recipient_name'] = recipient_name
                 #print(emails_dict)
                 #Getting the date of the emails
                 date_field = re.search("Date:.*", item)
                 #print(date_field.group())
                 if date_field is not None:
                     date = re.search("\d+\s\w+\s\d+", date_field.group())
                 else:
                     date = None
                 if date is not None:
                     date_sent = date.group()
                 else:
                     date_sent = None
                 emails_dict['date_sent'] = date_sent
                 #print(emails dict)
             ( sender_email . etem apacemeonsateane.com , sender_name . etenent Arote , recipiene_email . etem apacameareniing.nee , recipiene_i
```

```
{'sender email': 'g.godwin12345@caramail.com', 'sender name': '"MR GODWIN IGBUNU"', 'recipient email': 'mamu.godwin1@caramail.com', 'recipi
c 2002'}
{'sender email': 'bbell.idr@caramail.com', 'sender name': '"bell.idr bell.idr"', 'recipient email': 'bbell.idr@caramail.com', 'recipient na
{'sender email': 'dav84@mail.com', 'sender name': '"Mr. David Agu"', 'recipient email': 'webmaster@aclweb.org', 'recipient name': None, 'da
{'sender_email': 'kabila72b@37.com', 'sender_name': '"deborah kabila"', 'recipient_email': 'kabila72b@37.com', 'recipient_name': None, 'dat
{'sender_email': 'kss07@go.com', 'sender_name': None, 'recipient_email': 'colibri@let.uu.nl', 'recipient_name': None, 'date_sent': '11 Dec
{'sender_email': 'kss07@go.com', 'sender_name': None, 'recipient_email': 'feldhausen@linguistics.ruhr-uni-bochum.de', 'recipient_name': Nor
{'sender email': 'mohammed bello 1@yahoo.com', 'sender name': '"MOHAMMED BELLO (Ph.D)"', 'recipient email': 'mohammed bello 2@yahoo.com', '
'13 Dec 2002'}
{'sender email': None, 'sender name': None, 'recipient email': 'ga mm1@zwallet.com', 'recipient name': None, 'date sent': '12 Dec 2002'}
{'sender_email': 'cooooofooooooo@rediffmail.com', 'sender_name': '"PETER WILLIAMS"', 'recipient_email': 'R@M', 'recipient_name': None, 'da
{'sender_email': 'olivsav@netscape.net', 'sender_name': '"Oliveira Savimbi"', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name':
{'sender_email': 'tdannosondu@mail.com', 'sender_name': '"Mr. Daniel Osondu"', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name':
{'sender_email': 'williamsdaniel1@phantomemail.com', 'sender_name': None, 'recipient_email': 'williamsdaniel1@phantomemail.com', 'recipient
{'sender_email': None, 'sender_name': None, 'recipient_email': 'R@E', 'recipient_name': None, 'date_sent': None}
 ['sender_email': 'rulagg@excite.com', 'sender_name': '"Ruka"', 'recipient_email': 'rulagg@excite.com', 'recipient_name': None, 'date_sent':
{'sender email': 'idris2003942@aol.com', 'sender name': '"IDRIS ADAMU"', 'recipient email': 'idris adamu@skali.com', 'recipient name': None
```

**Step 9: Getting email subject** 

```
In [11]: ▶ for item in contents:
                 emails_dict = {}
                 sender = re.search('From:.*', item)
                 #print(sender)
                 if sender is not None:
                     s email = re.search('\w\S*@.*\w', sender.group())
                     s_name = re.search('\".*\"', sender.group())
                 else:
                     s_email = None
                     s name = None
                 if s_email is not None:
                     sender_email = s_email.group()
                 else:
                     sender_email = None
                 emails_dict['sender_email'] = sender_email
                 #print(emails_dict)
                 if s name is not None:
                     sender_name = s_name.group()
                 else:
                     sender_name = None
                 emails_dict['sender_name'] = sender_name
                 #print(emails_dict)
                 # Recipient
                 recipient = re.search('To:.*', item)
                 #print(recipient.group())
                 if recipient is not None:
                     r_email = re.search('\w\S*@.*\w', recipient.group())
                     r_name = re.search('\".*\"', recipient.group())
                 else:
                     r_email = None
                     r_name = None
                 if r email is not None:
                     recipient_email = r_email.group()
                     recipient_email = None
                 emails_dict['recipient_email'] = recipient_email
                 #print(emails_dict)
                 if r_name is not None:
                     recipient_name = r_name.group()
                 else:
                     recipient_name = None
                 emails_dict['recipient_name'] = recipient_name
                 #print(emails_dict)
                 #Getting the date of the emails
                 date_field = re.search("Date:.*", item)
                 #print(date_field.group())
                 if date_field is not None:
                     date = re.search("\d+\s\w+\s\d+", date_field.group())
                 else:
                     date = None
                 if date is not None:
                     date_sent = date.group()
                 else:
                     date_sent = None
                 emails_dict['date_sent'] = date_sent
                 #print(emails_dict)
```

```
#Getting the subject
    subject field = re.search('Subject:.*', item)
    if subject field is not None:
        subject = re.sub('Subject:', '', subject_field.group())
    else:
        subject = None
    emails dict['subject'] = subject
    print(emails_dict)
{'sender email': 'mikebunduu1@rediffmail.com', 'sender name': '"COL. MICHAEL BUNDU"', 'recipient email': 'R@E', 'recipient name': None, 'da
 *****SPAM***** IMPORTANT'}
{'sender_email': 'elixwilliam@usa.com', 'sender_name': '"MRS MARIAM ABACHA"', 'recipient_email': 'presidentwifeng@indiatimes.com', 'recipie
v 2002', 'subject': ' TRUST TRANSACTION'}
{'sender email': 'anayoawka@hotmail.com', 'sender name': '" DR. ANAYO AWKA "', 'recipient email': 'webmaster@aclweb.org', 'recipient name':
'subject': ' REQUEST FOR YOUR UNRESERVED ASSISTANCE'}
{'sender_email': 'anayoawka@hotmail.com', 'sender_name': '" DR. ANAYO AWKA "', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name':
'subject': ' REQUEST FOR YOUR UNRESERVED ASSISTANCE'}
{'sender_email': 'victorloma@netscape.net', 'sender_name': '"Victor Aloma"', 'recipient_email': 'ntcir-listmem@newns.op.nii.ac.jp', 'recipi
ov 2002', 'subject': ' Urgent Assistance'}
{'sender_email': 'victorloma@netscape.net', 'sender_name': '"Victor Aloma"', 'recipient_email': 'ntcir-outgoing@mail.op.nii.ac.jp', 'recipi
ov 2002', 'subject': ' Urgent Assistance'}
{'sender_email': 'james_ngola2002@maktoob.com', 'sender_name': '"JAMES NGOLA"', 'recipient_email': 'james_ngola2002@maktoob.com', 'recipier
2002', 'subject': ' URGENT BUSINESS PROPOSAL'}
{'sender_email': None, 'sender_name': None, 'recipient_email': 'oshea@UM', 'recipient_name': None, 'date_sent': None, 'subject': ' WANTED F
{'sender email': 'martinchime@usa.com', 'sender_name': '"MARTIN CHIME"', 'recipient_email': 'martinchime@usa.com', 'recipient_name': None,
t': ' URGENT'}
{'sender_email': 'mboro1555@post.com', 'sender_name': '"Mr George Mboro"', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name': Nor
ject': ' Offer'}
S'sandan amail' 'mantinchima@hanad cam' 'sandan nama' '"MADTTN CHTME"' 'nasiniant amail' 'mantinchima@hanad cam' 'nasiniant nama' N
```

Step 10: Getting the body of the email.

4/22/2021

```
In [12]: ▶ for item in contents:
                 emails_dict = {}
                 sender = re.search('From:.*', item)
                 #print(sender)
                 if sender is not None:
                     s email = re.search('\w\S*@.*\w', sender.group())
                     s_name = re.search('\".*\"', sender.group())
                 else:
                     s_email = None
                     s name = None
                 if s_email is not None:
                     sender_email = s_email.group()
                 else:
                     sender_email = None
                 emails_dict['sender_email'] = sender_email
                 #print(emails_dict)
                 if s name is not None:
                     sender_name = s_name.group()
                 else:
                     sender_name = None
                 emails_dict['sender_name'] = sender_name
                 #print(emails_dict)
                 # Recipient
                 recipient = re.search('To:.*', item)
                 #print(recipient.group())
                 if recipient is not None:
                     r_email = re.search('\w\S*@.*\w', recipient.group())
                     r_name = re.search('\".*\"', recipient.group())
                 else:
                     r_email = None
                     r_name = None
                 if r email is not None:
                     recipient_email = r_email.group()
                 else:
                     recipient_email = None
                 emails_dict['recipient_email'] = recipient_email
                 #print(emails_dict)
                 if r_name is not None:
                     recipient_name = r_name.group()
                 else:
                     recipient_name = None
                 emails_dict['recipient_name'] = recipient_name
                 #print(emails_dict)
                 #Getting the date of the emails
                 date_field = re.search("Date:.*", item)
                 #print(date_field.group())
                 if date_field is not None:
                     date = re.search("\d+\s\w+\s\d+", date_field.group())
                 else:
                     date = None
                 if date is not None:
                     date_sent = date.group()
                 else:
                     date_sent = None
                 emails_dict['date_sent'] = date_sent
                 #print(emails_dict)
```

{'sender email': 'bensul2004nng@spinfinder.com', 'sender name': '"Mr. Ben Suleman"', 'recipient email': 'R@M', 'recipient name': None, 'dat ' URGENT ASSISTANCE /RELATIONSHIP (P)', 'email\_body': 'Dear Friend,\n\nI am Mr. Ben Suleman a custom officer and work as Assistant controll ment Of the Federal Ministry of Internal Affairs stationed at the Murtala Mohammed International Airport, Ikeja, Lagos-Nigeria.\n\nAfter th f state of Nigeria General Sanni Abacha on June 8th 1998 his aides and immediate members of his family were arrested while trying to escape Saudi Arabia with 6 trunk boxes Marked "Diplomatic Baggage". Acting on a tip-off as they attempted to board the Air Craft, my officials carr r craft and discovered that the 6 trunk boxes contained foreign currencies amounting to US\$197,570,000.00(One Hundred and Ninety-Seven Mil United States Dollars).\n\nI declared only (5) five boxes to the government and withheld one (1) in my custody containing the sum of (US\$30). States Dollars Only, which has been disguised to prevent their being discovered during transportation process. Due to several media reports the money him and his co-government officials stole from our government treasury amounting\nto US\$55 Billion Dollars (ref:ngrguardiannews.c on times of July 1998 reported that General Abacha has over US\$3.Billion dollars in one account overseas. We decided to conceal this one (1 quite on the issue. The box was thus deposited with a security company here in Nigeria and tagged as "Precious Stones and Jewellry" in other covered. Now that all is calm, we (myself and two of my colleagues in the operations team) are now ready to move this box out of the country which is the safest means. \n\nHowever as government officials the Civil Service Code of Conduct does not allow us by law to operate any fc ment and the amount of money that can be found in our account\ncannot be more than our salary on the average, thus our handicapp and our ne ect and keep safely in your account this money.\n\nTherefore we want you to assist us in moving this money out of Nigeria. We shall definit he assistance. We can do this by instructing the Security Company here in Nigeria to\nmove the consignment to their affiliate branch office means and the consignment will be termed as Precious Stones and Jewelleries" which you bought during your visit to Nigeria and is being tra or safe keeping. Then we can arrange to meet at the destination country to take the delivery of the consignment. You will thereafter open a those and gradually instruct nomittance to your Country. InInThis business is 1889 nick from for you so places theat this matter with utmost

# Step 11: Append the dictionaries to a list

```
In [16]: ▶ for item in contents:
                 emails_dict = {}
                 sender = re.search('From:.*', item)
                 #print(sender)
                 if sender is not None:
                     s email = re.search('\w\S*@.*\w', sender.group())
                     s_name = re.search('\".*\"', sender.group())
                 else:
                     s_email = None
                     s name = None
                 if s_email is not None:
                     sender_email = s_email.group()
                 else:
                     sender_email = None
                 emails_dict['sender_email'] = sender_email
                 #print(emails_dict)
                 if s name is not None:
                     sender_name = s_name.group()
                 else:
                     sender_name = None
                 emails_dict['sender_name'] = sender_name
                 #print(emails_dict)
                 # Recipient
                 recipient = re.search('To:.*', item)
                 #print(recipient.group())
                 if recipient is not None:
                     r_email = re.search('\w\S*@.*\w', recipient.group())
                     r_name = re.search('\".*\"', recipient.group())
                 else:
                     r_email = None
                     r_name = None
                 if r email is not None:
                     recipient_email = r_email.group()
                 else:
                     recipient_email = None
                 emails_dict['recipient_email'] = recipient_email
                 #print(emails_dict)
                 if r_name is not None:
                     recipient_name = r_name.group()
                 else:
                     recipient_name = None
                 emails_dict['recipient_name'] = recipient_name
                 #print(emails_dict)
                 #Getting the date of the emails
                 date_field = re.search("Date:.*", item)
                 #print(date_field.group())
                 if date_field is not None:
                     date = re.search("\d+\s\w+\s\d+", date_field.group())
                 else:
                     date = None
                 if date is not None:
                     date_sent = date.group()
                 else:
                     date_sent = None
                 emails_dict['date_sent'] = date_sent
                 #print(emails_dict)
```

```
#Getting the subject
                  subject_field = re.search('Subject:.*', item)
                  if subject field is not None:
                       subject = re.sub('Subject:', '', subject_field.group())
                  else:
                       subject = None
                  emails_dict['subject'] = subject
                  #print(emails_dict)
                  #Getting email body
                  full_email = email.message_from_string(item)
                  #print(full_email.get_payload())
                  body = full_email.get_payload()
                  emails_dict['email_body'] = body
                  #print(emails dict)
                  emails.append(emails dict)
              emails df = pd.DataFrame(emails)
In [18]:
In [19]:
           emails_df
   Out[19]:
                                                                                                                                                           subject
                                   sender email
                                                        sender_name
                                                                                 recipient_email recipient_name
                                                                                                              date_sent
                                                                                                                 31 Oct
                                                                                                                                  URGENT BUSINESS ASSISTANCE AND
                  0 james_ngola2002@maktoob.com
                                                 "MR. JAMES NGOLA." james_ngola2002@maktoob.com
                                                                                                        None
                                                                                                                  2002
                                                                                                                                                     PARTNERSHIP
                                                                                                                 31 Oct
                                                                                         R@M
                                                                                                                               URGENT ASSISTANCE / RELATIONSHIP (P)
                     bensul2004nng@spinfinder.com
                                                    "Mr. Ben Suleman"
                                                                                                        None
                                                                                                                  2002
                                                     "PRINCE OBONG
                                                                                                                 31 Oct
                                                                                                                                                 GOOD DAY TO YOU
                           obong_715@epatra.com
                                                                           obong_715@epatra.com
                                                                                                        None
                                                            ELEME"
                                                                                                                  2002
                                                     "PRINCE OBONG
                                                                                                                 31 Oct
                                                                                                                                                 GOOD DAY TO YOU
                 3
                           obong_715@epatra.com
                                                                           webmaster@aclweb.org
                                                                                                        None
                                                            ELEME"
                                                                                                                  2002
                                                                                                                  1 Nov
                  4
                           m_abacha03@www.com
                                                     "Maryam Abacha"
                                                                           m_abacha03@www.com
                                                                                                                                               I Need Your Assistance.
                                                                                                        None
                                                                                                                   2002
                                                                                                                                                    =?iso-8859-1?Q?
                                                                                                                  None
               3972 michealagu0255@zipmail.com.br
                                                              None
                                                                                          None
                                                                                                        None
                                                                                                                                CONTACT=20GLOBAL=20MAX=20SHIPI...
                                                                                                                 17 Sep
               3973
                                                                                                                                                 TREAT AS URGENT.
                           ali_sherif252@hotmail.fr
                                                                         ali_sherif105@yahoo.co.uk
                                                              None
                                                                                                        None
                                                                                                                  2007
```

drusmanibrahim.tg@homs.cc

motherdorisk9@yahoo.com.hk

None

None

None

None

18 Sep

19 Sep

2007 19 Sep

2007

2007

None

None

None

From Dr Usman Ibrahim / Mr Wahid Yoffe property.

My Beloved In Christ.

**ATTENTION** 

3977 rows × 7 columns

3974

3975

3976

drusmanibrahimtg08@hotmail.fr

motherdorisk61@hotmail.com

None

## The whole script

```
In [20]: ▶ import re
             import pandas as pd
             import email
             emails = []
             fh = open('fradulent_emails.txt').read()
             contents = re.split('From r', fh)
             contents.pop(0)
             for item in contents:
                 emails_dict = {}
                 sender = re.search('From:.*', item)
                 #print(sender)
                 if sender is not None:
                     s_email = re.search('\w\S*@.*\w', sender.group())
                     s_name = re.search('\".*\"', sender.group())
                 else:
                     s_email = None
                     s name = None
                 if s_email is not None:
                     sender_email = s_email.group()
                 else:
                     sender_email = None
                 emails_dict['sender_email'] = sender_email
                 #print(emails_dict)
                 if s_name is not None:
                     sender_name = s_name.group()
                 else:
                     sender_name = None
                 emails_dict['sender_name'] = sender_name
                 #print(emails_dict)
                 # Recipient
                 recipient = re.search('To:.*', item)
                 #print(recipient.group())
                 if recipient is not None:
                     r_email = re.search('\w\S*@.*\w', recipient.group())
                     r_name = re.search('\".*\"', recipient.group())
                 else:
                     r_email = None
                     r_name = None
                 if r email is not None:
                     recipient_email = r_email.group()
                 else:
                     recipient_email = None
                 emails_dict['recipient_email'] = recipient_email
                 #print(emails_dict)
                 if r name is not None:
                     recipient_name = r_name.group()
                 else:
                     recipient_name = None
                 emails_dict['recipient_name'] = recipient_name
                 #print(emails dict)
                 #Getting the date of the emails
                 date_field = re.search("Date:.*", item)
                 #print(date_field.group())
                 if date_field is not None:
                     date = re.search("\d+\s\w+\s\d+", date_field.group())
                 else:
                     date = None
```

```
if date is not None:
        date_sent = date.group()
    else:
        date_sent = None
    emails_dict['date_sent'] = date_sent
    #print(emails_dict)
    #Getting the subject
    subject_field = re.search('Subject:.*', item)
    if subject_field is not None:
        subject = re.sub('Subject:', '', subject_field.group())
    else:
        subject = None
   emails_dict['subject'] = subject
    #print(emails_dict)
    #Getting email body
    full_email = email.message_from_string(item)
    #print(full_email.get_payload())
    body = full_email.get_payload()
    emails_dict['email_body'] = body
    #print(emails_dict)
    emails.append(emails_dict)
emails_df = pd.DataFrame(emails)
emails_df
```

#### Out[20]:

	subject	date_sent	recipient_name	recipient_email	sender_name	sender_email	
	URGENT BUSINESS ASSISTANCE AND PARTNERSHIP	31 Oct 2002	None	james_ngola2002@maktoob.com	"MR. JAMES NGOLA."	james_ngola2002@maktoob.com	0
	URGENT ASSISTANCE /RELATIONSHIP (P)	31 Oct 2002	None	R@M	"Mr. Ben Suleman"	bensul2004nng@spinfinder.com	1
FI	GOOD DAY TO YOU	31 Oct 2002	None	obong_715@epatra.com	"PRINCE OBONG ELEME"	obong_715@epatra.com	2
FI	GOOD DAY TO YOU	31 Oct 2002	None	webmaster@aclweb.org	"PRINCE OBONG ELEME"	obong_715@epatra.com	3
	I Need Your Assistance.	1 Nov 2002	None	m_abacha03@www.com	"Maryam Abacha"	m_abacha03@www.com	4
	=?iso-8859-1?Q? CONTACT=20GLOBAL=20MAX=20SHIPI	None	None	None	None	michealagu0255@zipmail.com.br	3972
	TREAT AS URGENT.	17 Sep 2007	None	ali_sherif105@yahoo.co.uk	None	ali_sherif252@hotmail.fr	3973
	From Dr Usman Ibrahim / Mr Wahid Yoffe property.	18 Sep 2007	None	drusmanibrahim.tg@homs.cc	None	drusmanibrahimtg08@hotmail.fr	3974
	My Beloved In Christ.	19 Sep 2007	None	motherdorisk9@yahoo.com.hk	None	motherdorisk61@hotmail.com	3975
	ATTENTION	19 Sep 2007	None	None	None	None	3976

3977 rows × 7 columns

In [ ]: ▶