

## Step 1: Importing important libraries

```
In [1]: ➤ import re
import pandas as pd
import email
```

```
In [2]: ➤ dir(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.

```
In [3]: ➤ emails = []
fh = open('fradulent_emails.txt').read()
```

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

```
In [4]: ➤ contents = re.split('From r', fh)
contents.pop(0)
```

```
Out[4]: ''
```

## 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', 'recipient_name': None}
{'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': None}
{'sender_email': 'obong_715@epatra.com', 'sender_name': '"PRINCE OBONG ELEME"', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name': None}
{'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', 'recipient_name': None}
{'sender_email': 'william2244drallo@maktoob.com', 'sender_name': '"William Drallo"', 'recipient_email': 'william2244drallo@maktoob.com', 'recipient_name': None}
{'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': None}
{'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', 'recipient_name': None}
{'sender_email': 'anayoawka@hotmail.com', 'sender_name': '" DR. ANAYO AWKA "', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name': None}
{'sender_email': 'anayoawka@hotmail.com', 'sender_name': '" DR. ANAYO AWKA "', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name': None}
{'sender_email': 'victorloma@netscape.net', 'sender_name': '"Victor Aloma"', 'recipient_email': 'ntcir-listmem@news.op.nii.ac.jp', 'recipient_name': None}
```

## 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': 'clem.apate@consulcare.com', 'sender_name': 'CLEMENT APATE', 'recipient_email': 'clem.apate@earthlink.net', 'recipient_
2' }

```

```
{'sender_email': 'g.godwin12345@caramail.com', 'sender_name': '"MR GODWIN IGBUNU"', 'recipient_email': 'mamu.godwin1@caramail.com', 'recipient_name': 'mamu.godwin1', 'date_sent': '12 Dec 2002'}  
{'sender_email': 'bbell.idr@caramail.com', 'sender_name': '"bell.idr bell.idr"', 'recipient_email': 'bbell.idr@caramail.com', 'recipient_name': 'bbell.idr', 'date_sent': '12 Dec 2002'}  
{'sender_email': 'dav84@mail.com', 'sender_name': '"Mr. David Agu"', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name': None, 'date_sent': '12 Dec 2002'}  
{'sender_email': 'kabila72b@37.com', 'sender_name': '"deborah kabila"', 'recipient_email': 'kabila72b@37.com', 'recipient_name': None, 'date_sent': '12 Dec 2002'}  
{'sender_email': 'kss07@go.com', 'sender_name': None, 'recipient_email': 'colibri@let.uu.nl', 'recipient_name': None, 'date_sent': '11 Dec 2002'}  
{'sender_email': 'kss07@go.com', 'sender_name': None, 'recipient_email': 'feldhausen@linguistics.ruhr-uni-bochum.de', 'recipient_name': None, 'date_sent': '11 Dec 2002'}  
{'sender_email': 'mohammed_bello_1@yahoo.com', 'sender_name': '"MOHAMMED BELLO (Ph.D)"', 'recipient_email': 'mohammed_bello_2@yahoo.com', 'recipient_name': 'mohammed_bello_2', 'date_sent': '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': 'cooooofooooooooo@rediffmail.com', 'sender_name': '"PETER WILLIAMS"', 'recipient_email': 'R@M', 'recipient_name': None, 'date_sent': '12 Dec 2002'}  
{'sender_email': 'olivsav@netscape.net', 'sender_name': '"Oliveira Savimbi"', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name': None, 'date_sent': '12 Dec 2002'}  
{'sender_email': 'tdannosondu@mail.com', 'sender_name': '"Mr. Daniel Osondu"', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name': None, 'date_sent': '12 Dec 2002'}  
{'sender_email': 'williamsdaniel1@phantomemail.com', 'sender_name': None, 'recipient_email': 'williamsdaniel1@phantomemail.com', 'recipient_name': None, 'date_sent': '12 Dec 2002'}  
{'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': '12 Dec 2002'}  
{'sender_email': 'idris2003942@aol.com', 'sender_name': '"IDRIS ADAMU"', 'recipient_email': 'idris adamu@skali.com', 'recipient_name': None, 'date_sent': '12 Dec 2002'}
```

## 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()
    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)

```

```

{'sender_email': 'mikebunduu1@rediffmail.com', 'sender_name': '"COL. MICHAEL BUNDU"', 'recipient_email': 'R@E', 'recipient_name': None, 'date_sent': None, 'subject': '*****SPAM***** IMPORTANT'}
{'sender_email': 'elixwilliam@usa.com', 'sender_name': '"MRS MARIAM ABACHA"', 'recipient_email': 'presidentwifeng@indiatimes.com', 'recipient_name': None, 'date_sent': None, 'subject': 'v 2002', 'subject': ' TRUST TRANSACTION'}
{'sender_email': 'anayoawka@hotmail.com', 'sender_name': '" DR. ANAYO AWKA "', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name': None, 'date_sent': None, 'subject': ' REQUEST FOR YOUR UNRESERVED ASSISTANCE'}
{'sender_email': 'anayoawka@hotmail.com', 'sender_name': '" DR. ANAYO AWKA "', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name': None, 'date_sent': None, 'subject': ' REQUEST FOR YOUR UNRESERVED ASSISTANCE'}
{'sender_email': 'victorloma@netscape.net', 'sender_name': '"Victor Aloma"', 'recipient_email': 'ntcir-listmem@newns.op.nii.ac.jp', 'recipient_name': None, 'date_sent': None, 'subject': ' Urgent Assistance'}
{'sender_email': 'victorloma@netscape.net', 'sender_name': '"Victor Aloma"', 'recipient_email': 'ntcir-outgoing@mail.op.nii.ac.jp', 'recipient_name': None, 'date_sent': None, 'subject': ' Urgent Assistance'}
{'sender_email': 'james_ngola2002@maktoob.com', 'sender_name': '"JAMES NGOLA"', 'recipient_email': 'james_ngola2002@maktoob.com', 'recipient_name': None, 'date_sent': None, 'subject': ' URGENT BUSINESS PROPOSAL'}
{'sender_email': None, 'sender_name': None, 'recipient_email': 'oshea@UM', 'recipient_name': None, 'date_sent': None, 'subject': ' WANTED FOR INFORMATION'}
{'sender_email': 'martinchime@usa.com', 'sender_name': '"MARTIN CHIME"', 'recipient_email': 'martinchime@usa.com', 'recipient_name': None, 'date_sent': None, 'subject': ' URGENT'}
{'sender_email': 'mboro1555@post.com', 'sender_name': '"Mr George Mboro"', 'recipient_email': 'webmaster@aclweb.org', 'recipient_name': None, 'date_sent': None, 'subject': ' Offer'}
{'sender_email': 'martinchime@honorad.com', 'sender_name': '"MARTIN CHIME"', 'recipient_email': 'martinchime@honorad.com', 'recipient_name': None, 'date_sent': None, 'subject': ' Offer'}

```

## Step 10: Getting the body of the email.

```

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)

```



```

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)
```

```
In [18]:  emails_df = pd.DataFrame(emails)
```

```
In [19]:  emails_df
```

Out[19]:

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

3977 rows × 7 columns

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]:

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

3977 rows × 7 columns

In [ ]: ▶

