

```

import pandas as pd
import numpy as np
import seaborn as sn
import matplotlib.pyplot as plt
import re
import datetime as dt
%matplotlib inline

f = open('WhatsApp Chat with 2018-2022 IT-B (1).txt','r',encoding='utf-8')

data = f.read()

dummy = data.split('\n')
dummy

"22/08/18, 09:33 - Meet: It's not happening dude..",
"22/08/18, 09:33 - Meet: Give your email id I'll forward u..",
"22/08/18, 09:33 - Meet: That file is not in a PDF format..",
"22/08/18, 09:34 - Jagss: jagadeeshreddy7659@gmail.com",
"22/08/18, 09:34 - +91 99484 61869: harshavardan398@gmail.com",
"22/08/18, 09:35 - Harifriend: kukkadapuharikrishna998@gmail.com",
"22/08/18, 09:36 - Meet: Did u guys received??",
"22/08/18, 09:38 - +91 99484 61869: Unable to access document ani vastundi",
"22/08/18, 09:39 - +91 99484 61869: Send meet ADI open avuthundi",
"22/08/18, 09:45 - Meet: I have forwarded the file plz check your mail's..",
"22/08/18, 09:45 - Harifriend: Meet isn't not coming",
"22/08/18, 09:46 - Meet: Even in mail also?",
"22/08/18, 09:47 - Harifriend: Tq meet",
"22/08/18, 09:47 - Meet: U received it??",
"22/08/18, 10:03 - Preetham: <Media omitted>",
"22/08/18, 11:38 - Chinmayee: <Media omitted>",
"22/08/18, 11:38 - Chinmayee: <Media omitted>",
"22/08/18, 12:15 - You added +91 98498 76810",
"22/08/18, 12:26 - You removed +91 98498 76810",
"22/08/18, 12:56 - Your security code with Jagss changed. Tap to learn more.",
"22/08/18, 16:14 - +91 6309 604 570: Guys for which batch plumbing expt. In workshop lab had completed",
"22/08/18, 16:14 - +91 6309 604 570: Except B1-C0",
"22/08/18, 16:25 - Ali: Why",
"22/08/18, 16:27 - +91 6309 604 570: Bcos we want expt",
"22/08/18, 16:28 - +91 6309 604 570: Not introduction",
"22/08/18, 16:28 - Ali: Kk",
"22/08/18, 22:37 - Your security code with Ke** changed. Tap to learn more.",
"23/08/18, 10:01 - Abhiramitb: <Media omitted>",
"23/08/18, 18:33 - Aditya *New*: https://play.google.com/store/apps/details?id=app.compiler",
"23/08/18, 18:33 - Aditya *New*: guys this is compiler for android",
"23/08/18, 18:33 - +91 6309 604 570: OK mama",
"23/08/18, 18:33 - Aditya *New*: can be done with 21 languages",
"23/08/18, 18:34 - Aditya *New*: c, c++ etc...",
"23/08/18, 18:46 - Ali: I have this app from first on words",
"23/08/18, 18:47 - Ali: Yes this help us to learn and practice c programs",
"23/08/18, 19:09 - Saikiran Nani: For ios bro?????",
"23/08/18, 19:21 - Chinmayee: Hlo guys",
"23/08/18, 19:21 - +91 93980 32560: Hi",
"23/08/18, 19:22 - Chinmayee: Tomorrow A0 to B0 any one of u bring white board marker to write vocabulary words on white board",
"23/08/18, 19:28 - S*2: Pls inform who is getting now",
"23/08/18, 19:28 - Ali: I don't have I can't bring",
"23/08/18, 19:29 - S*2: Ok Irshad",
"23/08/18, 19:29 - S*2: Anybody else who can bring",
"23/08/18, 19:46 - +91 6309 604 570: Not there",
"23/08/18, 20:46 - Aditya *New*: i dont know",
"24/08/18, 04:42 - Your security code with Jagss changed. Tap to learn more.",
"24/08/18, 19:37 - +91 6309 604 570 added +91 88863 07575",
"25/08/18, 18:03 - Irfan: <Media omitted>",
"25/08/18, 19:27 - Your security code with +91 84650 80632 changed. Tap to learn more.",
"25/08/18, 20:43 - Your security code with +91 84650 80632 changed. Tap to learn more.",
"25/08/18, 23:12 - Jagss: <Media omitted>",
"26/08/18, 18:02 - Ali: This message was deleted",
"26/08/18, 18:04 - +91 6309 604 570: Ali this grp",
"26/08/18, 18:04 - +91 6309 604 570: Is only for info",
"26/08/18, 18:04 - +91 6309 604 570: If there is holiday we will inform",
"26/08/18, 18:04 - +91 6309 604 570: Don't keep such type of pics and msgs too",
"26/08/18, 18:04 - +91 6309 604 570: Pls understand",
"26/08/18, 18:10 - Manohar: People,who ever are interested to pay money for drones and quadcopters programme which is held in our college on 7 & 8 of September are requested bring the nominal fee i.e Rs 1000 This is to inform the people who don't know about

pattern = '\d{1,2}/\d{1,2}/\d{2,4},\s\d{1,2}:\d{2}\s-\s'
messages = re.split(pattern,data)[1:]
print(len(messages))

```

```
dates = re.findall(pattern,data)
print(len(dates))
```

```
9203
9203
```

```
string = '16/08/18, 20:09 - '
string = string.split(',')
date,time = string[0],string[1]
time = time.split('-')
time = time[0].strip()
print(date+" and "+time)
```

```
16/08/18 and 20:09
```

```
def gettimeanddate(string):
    string = string.split(',')
    date,time = string[0],string[1]
    time = time.split('-')
    time = time[0].strip()

    return date+" "+time
```

```
df = pd.DataFrame({'user_messages':messages,
                  'message_date':dates})
```

```
df['message_date'] = df['message_date'].apply(lambda text:gettimeanddate(text))
df.rename(columns={'message_date':'date'},inplace=True)
```

```
df.head()
```

```

      user_messages      date
0  Messages and calls are end-to-end encrypted. N... 16/08/18 20:09
1  Chinmayee created group "AGI OFFICIAL@IT 2018-... 16/08/18 20:09
2                Chinmayee added you\n              16/08/18 20:09
3  Chinmayee: Hello this is official grup\n          16/08/18 20:09
4                Jagss: Chusinam leaaa\n            16/08/18 20:10
```

```
df['user_messages']
```

```

0  Messages and calls are end-to-end encrypted. N...
1  Chinmayee created group "AGI OFFICIAL@IT 2018-...
2                Chinmayee added you\n
3  Chinmayee: Hello this is official grup\n
4                Jagss: Chusinam leaaa\n
...
9198 +91 79950 77079: 12B1 - registration Completed...
9199 Asha Ma'am: 1275 and 1288 do respond please ....
9200                Manoj: 75 registered mam\n
9201                +91 6304 605 703 left\n
9202 Nagalakshmi Ma'am 💰 $: Dear students,\nGood mo...
Name: user_messages, Length: 9203, dtype: object
```

```
users = []
messages = []
```

```
for message in df['user_messages']:
```

```
    entry = re.split('([\w\W]+?):\s', message)
    if entry[1:]:
        users.append(entry[1])
        messages.append(entry[2])
```

```
    else:
        users.append('Group Notification')
        messages.append(entry[0])
```

```
df['User'] = users
df['message'] = messages

def getstring(text):
    return text.split('\n')[0]

df['message'] = df['message'].apply(lambda text:getstring(text))

df = df.drop(['user_messages'],axis=1)
df = df[['message','date','User']]

df = df.rename(columns={'message':'Message',
                        'date':'Date'})
```

```
df.head()
```




	Message	Date	User
0	Messages and calls are end-to-end encrypted. N...	16/08/18 20:09	Group Notification
1	Chinmayee created group "AGI OFFICIAL@IT 2018-22"	16/08/18 20:09	Group Notification
2	Chinmayee added you	16/08/18 20:09	Group Notification
3	Hello this is official grup	16/08/18 20:09	Chinmayee
4	Chusinam leaaa	16/08/18 20:10	Jagss

16/08/18 20:09

YYYY-MM-DD

```
df.shape
```




(9203, 3)

```
df['Only date'] = pd.to_datetime(df['Date']).dt.date
df['Year'] = pd.to_datetime(df['Date']).dt.year
df['Month_num'] = pd.to_datetime(df['Date']).dt.month
df['Month'] = pd.to_datetime(df['Date']).dt.month_name()
df['Day'] = pd.to_datetime(df['Date']).dt.day
df['Day_name'] = pd.to_datetime(df['Date']).dt.day_name()


df['Hour'] = pd.to_datetime(df['Date']).dt.hour
df['Minute'] = pd.to_datetime(df['Date']).dt.minute
```

```
df.head()
```



	Message	Date	User	Only date	Year	Month_num	Month	Day	Day_name	Hour	Minute
0	Messages and calls are end-to-end encrypted. N...	16/08/18 20:09	Group Notification	2018-08-16	2018	8	August	16	Thursday	20	9
1	Chinmayee created group "AGI OFFICIAL@IT 2018-22"	16/08/18 20:09	Group Notification	2018-08-16	2018	8	August	16	Thursday	20	9
2	Chinmayee added you	16/08/18 20:09	Group Notification	2018-08-16	2018	8	August	16	Thursday	20	9

```
df[df['User'] == 'Jagss']
```



	Message	Date	User	Only	date	Year	Month_num	Month	Day	Day_name	Hour	Minute
4	Chusinam leaaa	16/08/18 20:10	Jagss	2018-08-16	2018		8	August	16	Thursday	20	10
12	Adhea kadha cheppindhiii	16/08/18 20:11	Jagss	2018-08-16	2018		8	August	16	Thursday	20	11
60	Not more	17/08/18 10:02	Jagss	2018-08-17	2018		8	August	17	Friday	10	2
118	Room no	17/08/18 23:01	Jagss	2018-08-17	2018		8	August	17	Friday	23	1
153	We are in B bloCK	18/08/18 09:55	Jagss	2018-08-18	2018		8	August	18	Saturday	9	55
...
8859	1297 done sir	13/09/21 17:04	Jagss	2021-09-13	2021		9	September	13	Monday	17	4
8873	18h61a1297	13/09/21 18:34	Jagss	2021-09-13	2021		9	September	13	Monday	18	34
9014	1292 registered mam	16/09/21 13:35	Jagss	2021-09-16	2021		9	September	16	Thursday	13	35
9065	18h61a1297	16/09/21 22:30	Jagss	2021-09-16	2021		9	September	16	Thursday	22	30
9150	1297_Completed mam 19-0921	20/09/21 11:53	Jagss	2021-09-20	2021		9	September	20	Monday	11	53


162 rows × 11 columns

```
words = []
for message in df['Message']:
    words.extend(message.split())
```

```
print(len(words))
```

 44880

words



```
['Messages',
 'and',
 'calls',
 'are',
 'end-to-end',
 'encrypted.',
 'No',
 'one',
 'outside',
 'of',
 'this',
 'chat,',
 'not',
 'even',
 'WhatsApp,',
 'can',
 'read',
 'or',
 'listen',
 'to',
 'them.',
 'Tap',
 'to',
 'learn',
 'more.',
 'Chinmayee',
 'created',
 'group',
 '"AGI',
 'OFFICIAL@IT',
 '2018-22"',
 'Chinmayee',
 'added',
 'you',
 'Hello',
 'this',
 'is',
 'official',
 'grup',
 'Chusinam',
 'leaaa',
 'Admin',
 'cheyyi',
 'First',
 'Kkkk',
 'No',
 'msg',
```

```
'anything',
'except',
'studies',
'Hello',
'sirs',
'kuda',
'unarru',
'Chusi',
'matladu',
"You're",
'now'

df[(df['Message'] == '<Media omitted>') & (df['User'] == 'Jagss')].shape

(27, 11)

df['Message'].unique()

array(['Messages and calls are end-to-end encrypted. No one outside of this chat, not even WhatsApp, can read or listen to them. Tap to learn more.',
      'Chinmayee created group "AGI OFFICIAL@IT 2018-22"',
      'Chinmayee added you', ..., '75 registered mam',
      '+91 6304 605 703 left', 'Dear students,'], dtype=object)
```

df[df['Message']=='<Media omitted>']

	Message	Date	User	Only date	Year	Month_num	Month	Day	Day_name	Hour	Minute
21	<Media omitted>	16/08/18 21:40	Meet	2018-08-16	2018	8	August	16	Thursday	21	40
111	<Media omitted>	17/08/18 17:05	+91 99083 50196	2018-08-17	2018	8	August	17	Friday	17	5
134	<Media omitted>	18/08/18 07:50	+91 6309 604 570	2018-08-18	2018	8	August	18	Saturday	7	50
160	<Media omitted>	18/08/18 10:00	Jagss	2018-08-18	2018	8	August	18	Saturday	10	0
198	<Media omitted>	18/08/18 20:17	Harifriend	2018-08-18	2018	8	August	18	Saturday	20	17
...
9020	<Media omitted>	16/09/21 15:27	Ravi Raju Sir	2021-09-16	2021	9	September	16	Thursday	15	27
9021	<Media omitted>	16/09/21 15:28	Ravi Raju Sir	2021-09-16	2021	9	September	16	Thursday	15	28
9097	<Media omitted>	18/09/21 10:56	Prudhvi Raj	2021-09-18	2021	9	September	18	Saturday	10	56
9099	<Media omitted>	19/09/21 10:07	Ravi Raju Sir	2021-09-19	2021	9	September	19	Sunday	10	7
9176	<Media omitted>	22/09/21 13:14	Ananad Sir	2021-09-22	2021	9	September	22	Wednesday	13	14

1393 rows × 11 columns

Start coding or [generate](#) with AI.

```
df['User'].value_counts().head()

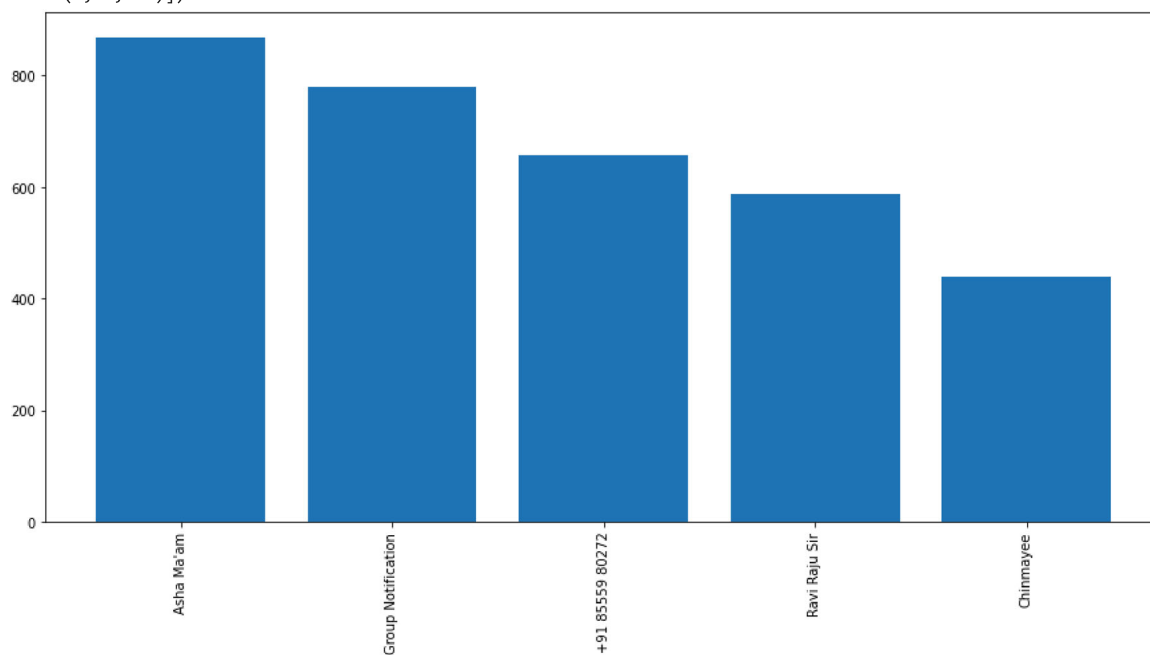
Asha Ma'am      869
Group Notification  778
+91 85559 80272  657
Ravi Raju Sir    588
Chinmayee        438
Name: User, dtype: int64

counts = df['User'].value_counts().head()
plt.figure(figsize=(15,7))
plt.bar(counts.index,counts.values)
plt.xticks(rotation = 'vertical')
```

```

([0, 1, 2, 3, 4],
 [Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, '')]

```



```

wordmap = {}
for message in df['Message']:
    dummy = message.split()
    for word in dummy:
        if word in wordmap:
            wordmap[word] += 1
        else:
            wordmap[word] = 1

```

```
wordmap = dict(sorted(wordmap.items(), key = lambda x:x[1], reverse = True))
```

```
wordmap
```



```
students, : 04,
'mail': 61,
'complete': 61,
'2': 60,
'there': 60,
'lab': 60,
'left': 59,
'submit': 59,
'take': 58,
'test': 58,
'come': 57,
'attend': 57,
'get': 56,
'tomorrow': 56,
'79950': 56,
'.': 56,
'77079': 55,
'classes': 55,
'You': 53,
'joined': 53,
'Only': 52,
'respond': 50,
'plz': 49,
'time': 49,
'received': 49,
'completed': 49,
'85559': 49,

from collections import Counter
words = []
for message in df['Message']:
    words.extend(message.split())

wordmap = dict(Counter(words).most_common(20))
wordmap = dict(sorted(wordmap.items(),key=lambda x:x[1],reverse=True))
newdf = pd.DataFrame(wordmap.items())
newdf
```



	0	1
0	to	1512
1	<Media	1393
2	omitted>	1393
3	the	786
4	with	715
5	for	606
6	learn	602
7	code	602
8	Tap	597
9	Your	595
10	more.	594
11	security	592
12	changed.	592
13	This	508
14	join	506
15	message	501
16	deleted	488
17	are	481
18	was	480
19	and	458

```
a,an,the,you,ok,...

temp = df[(df['User']!='Group Notification') | (df['User']!='<Media omitted>')]
temp
```



	Message	Date	User	Only date	Year	Month_num	Month	Day	Day_name	Hour	Minute
0	Messages and calls are end-to-end encrypted. N...	16/08/18 20:09	Group Notification	2018-08-16	2018	8	August	16	Thursday	20	9
1	Chinmayee created group "AGI OFFICIAL@IT 2018-22"	16/08/18 20:09	Group Notification	2018-08-16	2018	8	August	16	Thursday	20	9
2	Chinmayee added you	16/08/18 20:09	Group Notification	2018-08-16	2018	8	August	16	Thursday	20	9
3	Hello this is official grup	16/08/18 20:09	Chinmayee	2018-08-16	2018	8	August	16	Thursday	20	9
4	Chusinam leaaa	16/08/18 20:10	Jagss	2018-08-16	2018	8	August	16	Thursday	20	10
...
9198	12B1 - registration Completed sir	23/09/21 16:18	+91 79950 77079	2021-09-23	2021	9	September	23	Thursday	16	18
9199	1275 and 1288 do respond please ...done with r...	23/09/21 21:53	Asha Ma'am	2021-09-23	2021	9	September	23	Thursday	21	53

```
def gettopwords(selecteduser,df):

    file = open('stop_hinglish.txt','r')
    stopwords = file.read()
    stopwords = stopwords.split('\n')

    if selecteduser!='Overall':
        df = df[df['User']==selecteduser]

    temp = df[(df['User']!='Group Notification') |
              (df['User']!='<Media omitted>')]

    words = []

    for message in df['Message']:
        for word in message.lower().split():
            if word not in stopwords:
                words.append(word)

    mostcommon = pd.DataFrame(Counter(words).most_common(20))
    return mostcommon

mostcommon = gettopwords('Overall',df)

mostcommon
```




	0	1
0	<media	1393
1	omitted>	1393
2	join	689
3	learn	603
4	code	602
5	tap	597
6	more.	594
7	security	592
8	changed.	592
9	message	502
10	deleted	488
11	+91	376
12	immediately	339
13	class	312
14	requested	266
15	dear	211
16	students	209
17	mam	198
18	u	178
19	ur	133

```
import emoji
```

```
def emojihelp(selecteduser,df):
```


```
    if selecteduser!='Overall':
        df = df[df['User']==selecteduser]
```

```
    emojis = []
    for message in df['Message']:
        emojis.extend([c for c in message if c in emoji.UNICODE_EMOJI['en']])
```

```
    emojiidf = pd.DataFrame(Counter(emojis).most_common(len(Counter(emojis))))
```

```
    return emojiidf
```


```
dfmoji = emojihelp('Overall',df)
dfmoji
```



	0	1
0	👉	183
1	📄	88
2	😂	41
3	👍	39
4	👉	38
...
69	🌟	1
70	📌	1
71	👉	1
72	👉	1
73	👉	1

74 rows × 2 columns


```
dfmoji.columns = ['Emoji','Count']
dfmoji
```



	Emoji	Count
0	👉	183
1	📄	88
2	😂	41
3	👍	39
4	👉	38
...
69	🌟	1
70	📌	1
71	👉	1
72	👉	1
73	👉	1

74 rows × 2 columns

```
df.head()
```



	Message	Date	User	Only date	Year	Month_num	Month	Day	Day_name	Hour	Minute
0	Messages and calls are end-to-end encrypted. N...	16/08/18 20:09	Group Notification	2018-08-16	2018	8	August	16	Thursday	20	9
1	Chinmayee created group "AGI OFFICIAL@IT 2018-22"	16/08/18 20:09	Group Notification	2018-08-16	2018	8	August	16	Thursday	20	9
2	Chinmayee added you	16/08/18 20:09	Group Notification	2018-08-16	2018	8	August	16	Thursday	20	9

```
def monthtimeline(selecteduser,df):

    temp = df.groupby(['Year','Month_num','Month']).count()['Message'].reset_index().sort_values(['Message'])

    time = []
    for i in range(temp.shape[0]):
        time.append(temp['Month'][i]+"-"+str(temp['Year'][i]))

    temp['Time'] = time


    return temp
```



	Year	Month_num	Month	Message
0	2018	1	January	47
1	2018	2	February	35
2	2018	3	March	78
3	2018	4	April	66
4	2018	5	May	29
5	2018	6	June	25
6	2018	7	July	47
7	2018	8	August	520
8	2018	9	September	264
9	2018	10	October	208
10	2018	11	November	247
11	2018	12	December	99
12	2019	1	January	222
13	2019	2	February	207
14	2019	3	March	201
15	2019	4	April	152
16	2019	5	May	123
17	2019	6	June	115
18	2019	7	July	211
19	2019	8	August	200
20	2019	9	September	150
21	2019	10	October	184
22	2019	11	November	82
23	2019	12	December	134
24	2020	1	January	142
25	2020	2	February	56
26	2020	3	March	101
27	2020	4	April	149
28	2020	5	May	154
29	2020	6	June	89
30	2020	7	July	171
31	2020	8	August	354
32	2020	9	September	186
33	2020	10	October	148
34	2020	11	November	155
35	2020	12	December	223
36	2021	1	January	204
37	2021	2	February	84
38	2021	3	March	268
39	2021	4	April	615
40	2021	5	May	444
41	2021	6	June	315
42	2021	7	July	272
43	2021	8	August	721
44	2021	9	September	547
45	2021	10	October	70

46	2021	11	November	49
47	2021	12	December	40

```
df.head()
```



	Message	Date	User	Only date	Year	Month_num	Month	Day	Day_name	Hour	Minute
0	Messages and calls are end-to-end encrypted. N...	16/08/18 20:09	Group Notification	2018-08-16	2018	8	August	16	Thursday	20	9
1	Chinmayee created group "AGI OFFICIAL@IT 2018-22"	16/08/18 20:09	Group Notification	2018-08-16	2018	8	August	16	Thursday	20	9
2	Chinmayee added you	16/08/18 20:09	Group Notification	2018-08-16	2018	8	August	16	Thursday	20	9

1,2,2,3,4,4,4,5,6,7,3,4

1| 2 2 | 3 3| 4 4 4 4 | 5 | 6 | 7

{4 : 4, 3:2 , 2:2,}

```
temp = df.groupby(['Year','Month_num','Month']).count()['Message'].reset_index()
temp
```



	Year	Month_num	Month	Message
0	2018	1	January	47
1	2018	2	February	35
2	2018	3	March	78
3	2018	4	April	66
4	2018	5	May	29
5	2018	6	June	25
6	2018	7	July	47
7	2018	8	August	520
8	2018	9	September	264
9	2018	10	October	208
10	2018	11	November	247
11	2018	12	December	99
12	2019	1	January	222
13	2019	2	February	207
14	2019	3	March	201
15	2019	4	April	152
16	2019	5	May	123
17	2019	6	June	115
18	2019	7	July	211
19	2019	8	August	200
20	2019	9	September	150
21	2019	10	October	184
22	2019	11	November	82
23	2019	12	December	134
24	2020	1	January	142
25	2020	2	February	56
26	2020	3	March	101
27	2020	4	April	149
28	2020	5	May	154
29	2020	6	June	89
30	2020	7	July	171
31	2020	8	August	354
32	2020	9	September	186
33	2020	10	October	148
34	2020	11	November	155
35	2020	12	December	223
36	2021	1	January	204
37	2021	2	February	84
38	2021	3	March	268
39	2021	4	April	615
40	2021	5	May	444
41	2021	6	June	315
42	2021	7	July	272
43	2021	8	August	721
44	2021	9	September	547
45	2021	10	October	70

46	2021	11	November	49
47	2021	12	December	40

```
temp = df.groupby(['Year', 'Month_num', 'Month']).count()['Message'].reset_index()
temp
```



	Year	Month_num	Month	Message
0	2018	1	January	47
1	2018	2	February	35
2	2018	3	March	78
3	2018	4	April	66
4	2018	5	May	29
5	2018	6	June	25
6	2018	7	July	47
7	2018	8	August	520
8	2018	9	September	264
9	2018	10	October	208
10	2018	11	November	247
11	2018	12	December	99
12	2019	1	January	222
13	2019	2	February	207
14	2019	3	March	201
15	2019	4	April	152
16	2019	5	May	123
17	2019	6	June	115
18	2019	7	July	211
19	2019	8	August	200
20	2019	9	September	150
21	2019	10	October	184
22	2019	11	November	82
23	2019	12	December	134
24	2020	1	January	142
25	2020	2	February	56
26	2020	3	March	101
27	2020	4	April	149
28	2020	5	May	154
29	2020	6	June	89
30	2020	7	July	171
31	2020	8	August	354
32	2020	9	September	186
33	2020	10	October	148
34	2020	11	November	155
35	2020	12	December	223
36	2021	1	January	204
37	2021	2	February	84
38	2021	3	March	268
39	2021	4	April	615
40	2021	5	May	444
41	2021	6	June	315
42	2021	7	July	272
43	2021	8	August	721
44	2021	9	September	547
45	2021	10	October	70