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

In [2]:
```
f = open("chat.txt",'r',encoding='utf-8')
```

In [3]:
```
data = f.read()
```

In [4]:
```
print(data)
```

```
18/01/21, 15:05 - Akanksha "Team"
19/01/21, 15:05 - You were added
27/08/22, 14:56 - Monika  Hope you candidates are ready to attend *Electric
Vehicle Workshop*

🔜 you will receive meeting link 📎

⚜️ *Topic* - Back EMF & it's significance in Re-Generative Braking System
⏰ *11AM*
📅 *28th August 2022*
27/08/22, 16:28 - Shweta: <Media omitted>

27/08/22, 17:46 - Aknasha: <Media omitted>
27/08/22, 17:46 - Ayushi: hello all
27/08/22, 17:51 - Ashwani : 🙏🙏
27/08/22, 18:07 - Monika : https://forms.gle/dNBpdL5NMsHmhF9k7 (https://for
ms.gle/dNBpdL5NMsHmhF9k7)
27/08/22, 18:07 - Monika : Swati

Register wala form
```

In [5]:
```
pattern = '\d{1,2}/\d{1,2}/\d{2,4},\s\d{1,2}:\d{2}\s-\s'
```

In [6]:
```
messages= re.split(pattern,data)[1:]
```

In [7]:
```
messages
```

Out[7]: ['Akanksha "Team"\n',
 'You were added\n',
 "Monika Hope you candidates are ready to attend *Electric Vehicle Workshop*\n\n🔜 you will receive meeting link 📎\n\n⚜️ *Topic* - Back EMF & it's significance in Re-Generative Braking System\n⏰ *11AM*\n🗓️ *28th August 2022*\n",
 'Shweta: <Media omitted>\n\n',
 'Aknasha: <Media omitted>\n',
 'Ayushi: hello all\n',
 'Ashwani : \U0001fae3😁\n',
 'Monika : https://forms.gle/dNBpdL5NMsHmhF9k7\n', (https://forms.gle/dNBpdL5NMsHmhF9k7\n',)
 'Monika : Swati\n\nRegister wala form\n',
 'Swati : Thankyou \n\n',
 "Aayushi : 🔥 *Founder's Opinion | Student should be upskilled for Electric Vehicle Industry*\n▬▬▬▬▬▬\n🤔💭 Our industry leader\n▬▬▬▬▬▬\n💢 *Click Here to watch* \nhttps://www.youtube.com/watch?v=23fQ9-XUSCU\n▬▬▬▬▬\n💢 *Click Here to watch full Interview Session* \nhttps://www.youtube.com/watch?v=iBIcCGpSpeM&t=574s\n▬▬▬▬▬▬\n💢 To take part in *28 August 2022* Free workshop of and clear your query and doubts, kindly fill this form here 👇👇\nhttps://tinyurl.com/24vee9jt\n\n",
 'Sameera : 😊 Sukriya!!\n',
 'Bhaviya : <Media omitted>\n\n',
 'Monika : Hn sir\n',
 'Shweta : Okay\n',
 'Ayushi : 🔥 *Free Live Workshop Design and Development* \nWatch Live Streaming Now only on YouTube:- https://youtu.be/ZWlyGYWw7Cw\n\nTo (https://youtu.be/ZWlyGYWw7Cw\n\nTo) take part in our next Free workshop of *Electric Vehicle Design and Development* and clear your query and doubts, kindly fill the below form so that we can provide you the useful information. 👇👇\nhttps://forms.gle/Ux6YWQkjMAkhvthR6\n\nFor More details Kindly Join our *Electric Vehicle* Alert *WhatsApp* Group from the below link 👇\nhttps://chat.whatsapp.com/D4VpUnfyafU0dS2Hp3vjuJ\n',
 'Monika : This message was deleted\n',
 'Aayushi : Attendance Sheet | 28th AUGUST 2022 | Electric Vehicle Design & Development Live Free Workshop \nhttps://forms.gle/Ux6YWQkjMAkhvthR6\n\n\nFor More details Kindly Join our Alert *WhatsApp* Group from the below link 👇\nhttps://chat.whatsapp.com/D4VpUnfyafU0dS2Hp3vjuJ\n',
 'Bhavika : Matlab okay?\n\n',
 'Monika : This message was deleted\n',
 "Monika : Thank you for attending today's session. Hope you liked it , your doubts regarding career in EV Industry is cleared. \n▬▬▬▬▬▬\nFor attending upcoming sessions fill the form👇👇\n\nhttps://forms.gle/Ux6YWQkjMAkhvthR6\n",
 "Monika : Thank you for attending today's Workshop. Hope you liked the workshop. \n▬▬▬▬▬▬\nFor attending upcoming sessions, fill the form👇👇\n\nhttps://forms.gle/Ux6YWQkjMAkhvthR6\n\n",
 'Shweta : <Media omitted>\n',
 'Swatika : <Media omitted>\n',
 'Swatika : <Media omitted>\n',
 'Swatika : <Media omitted>\n',
 'Swatika : <Media omitted>\n',
 'Swatika : <Media omitted>\n',
 'Swatika : <Media omitted>\n',
 'Swatika : <Media omitted>\n',
 'Swatika : <Media omitted>\n',
 'Swatika : <Media omitted>\n',

```
 'Swatika : <Media omitted>\n',
 'Swatika : I am joining class\n',
 'Swatika : <Media omitted>\n',
 'Swatika : <Media omitted>\n',
 'Swatika : <Media omitted>\n',
 'Swatika : <Media omitted>\n',
 'Ayushi : <Media omitted>\n',
 'Ayuka : <Media omitted>\n',
 'Ayuka : <Media omitted>\n',
 'Ayuka : <Media omitted>\n',
 'Ayush : <Media omitted>\n',
 'Ayuka : <Media omitted>\n',
 'Ayuka : <Media omitted>\n',
 'Ayuka : <Media omitted>\n',
 'Monika : 😍😍yeeee\n',
 'Ayushi : <Media omitted>\n\n',
 'Ayushi : <Media omitted>\n',
 'Ashish: <Media omitted>\n',
 'Ashish : <Media omitted>\n',
 'Ashish : <Media omitted>\n',
 'Ashish : <Media omitted>\n',
 'Ashish : <Media omitted>\n',
 'Ashish : <Media omitted>\n',
 'Ashish : <Media omitted>\n',
 'Ashish : <Media omitted>\n',
 'Ashish : <Media omitted>\n\n',
 'Bhaviya: Syllabus \n\n',
 'Bhaviya: 👍\n',
 'Ashish : 👌\n',
 'Bhaviya: 👌']
```

In [8]:
```
dates= re.findall(pattern,data)
```

In [9]:
```
dates
```

```
Out[9]: ['18/01/21, 15:05 - ',
         '19/01/21, 15:05 - ',
         '27/08/22, 14:56 - ',
         '27/08/22, 16:28 - ',
         '27/08/22, 17:46 - ',
         '27/08/22, 17:46 - ',
         '27/08/22, 17:51 - ',
         '27/08/22, 18:07 - ',
         '27/08/22, 18:07 - ',
         '27/08/22, 18:09 - ',
         '27/08/22, 20:46 - ',
         '27/08/22, 23:06 - ',
         '28/08/22, 08:40 - ',
         '28/08/22, 08:50 - ',
         '28/08/22, 08:54 - ',
         '28/08/22, 10:31 - ',
         '28/08/22, 10:44 - ',
         '28/08/22, 10:45 - ',
         '28/08/22, 11:20 - ',
         '28/08/22, 12:26 - ',
         '28/08/22, 12:42 - ',
         '28/08/22, 12:44 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 17:23 - ',
         '28/08/22, 19:57 - ',
         '28/08/22, 19:57 - ',
         '28/08/22, 19:57 - ',
         '28/08/22, 19:57 - ',
         '28/08/22, 19:57 - ',
         '28/08/22, 19:57 - ',
         '28/08/22, 19:57 - ',
         '28/08/22, 19:57 - ',
         '28/08/22, 23:43 - ',
         '28/08/22, 23:44 - ',
         '30/08/22, 15:39 - ',
         '30/08/22, 15:41 - ',
         '30/08/22, 15:41 - ',
         '30/08/22, 15:41 - ',
         '30/08/22, 15:41 - ',
         '30/08/22, 15:41 - ',
         '30/08/22, 15:41 - ',
         '30/08/22, 15:42 - ',
         '30/08/22, 15:42 - ',
```

```
 '30/08/22, 15:42 - ',
 '30/08/22, 16:33 - ',
 '30/08/22, 19:04 - ',
 '30/08/22, 19:04 - ',
 '30/08/22, 19:04 - ']
```

In [10]:
```
df = pd.DataFrame({"user_message":messages, "message_date":dates})
```

In [11]:
```
df
```

Out[11]:

	user_message	message_date
0	Akanksha "Team"\n	18/01/21, 15:05 -
1	You were added\n	19/01/21, 15:05 -
2	Monika Hope you candidates are ready to atten...	27/08/22, 14:56 -
3	Shweta: <Media omitted>\n\n	27/08/22, 16:28 -
4	Aknasha: <Media omitted>\n	27/08/22, 17:46 -
...	...	...
57	Ashish : <Media omitted>\n\n	30/08/22, 15:42 -
58	Bhaviya: Syllabus \n\n	30/08/22, 16:33 -
59	Bhaviya: 👍\n	30/08/22, 19:04 -
60	Ashish : 👌\n	30/08/22, 19:04 -
61	Bhaviya: 👌	30/08/22, 19:04 -

62 rows × 2 columns

In [12]:
```
df["message_date"]
```

Out[12]:
```
0     18/01/21, 15:05 -
1     19/01/21, 15:05 -
2     27/08/22, 14:56 -
3     27/08/22, 16:28 -
4     27/08/22, 17:46 -
            ...
57    30/08/22, 15:42 -
58    30/08/22, 16:33 -
59    30/08/22, 19:04 -
60    30/08/22, 19:04 -
61    30/08/22, 19:04 -
Name: message_date, Length: 62, dtype: object
```

In [13]:
```
df["message_date"] =pd.to_datetime(df['message_date'], format = "%d/%m/%y, %H:%
```

In [14]:
```
df['message_date']
```

Out[14]:
```
0    2021-01-18 15:05:00
1    2021-01-19 15:05:00
2    2022-08-27 14:56:00
3    2022-08-27 16:28:00
4    2022-08-27 17:46:00
             ...
57   2022-08-30 15:42:00
58   2022-08-30 16:33:00
59   2022-08-30 19:04:00
60   2022-08-30 19:04:00
61   2022-08-30 19:04:00
Name: message_date, Length: 62, dtype: datetime64[ns]
```

In [15]:
```
df.shape
```

Out[15]:
```
(62, 2)
```

In [16]:
```
df
```

Out[16]:

	user_message	message_date
0	Akanksha "Team"\n	2021-01-18 15:05:00
1	You were added\n	2021-01-19 15:05:00
2	Monika Hope you candidates are ready to atten...	2022-08-27 14:56:00
3	Shweta: <Media omitted>\n\n	2022-08-27 16:28:00
4	Aknasha: <Media omitted>\n	2022-08-27 17:46:00
...	...	...
57	Ashish : <Media omitted>\n\n	2022-08-30 15:42:00
58	Bhaviya: Syllabus \n\n	2022-08-30 16:33:00
59	Bhaviya: 👍\n	2022-08-30 19:04:00
60	Ashish : 👌\n	2022-08-30 19:04:00
61	Bhaviya: 👌	2022-08-30 19:04:00

62 rows × 2 columns

In [17]:
```
df.rename(columns={'message_date':'date'}, inplace = True)
```

In [18]:
```
df
```

Out[18]:

	user_message	date
0	Akanksha "Team"\n	2021-01-18 15:05:00
1	You were added\n	2021-01-19 15:05:00
2	Monika Hope you candidates are ready to atten...	2022-08-27 14:56:00
3	Shweta: <Media omitted>\n\n	2022-08-27 16:28:00
4	Aknasha: <Media omitted>\n	2022-08-27 17:46:00
...	...	...
57	Ashish : <Media omitted>\n\n	2022-08-30 15:42:00
58	Bhaviya: Syllabus \n\n	2022-08-30 16:33:00
59	Bhaviya: 👍\n	2022-08-30 19:04:00
60	Ashish : 👌\n	2022-08-30 19:04:00
61	Bhaviya: 👌	2022-08-30 19:04:00

62 rows × 2 columns

In [19]:
```
users = []
messages= []

for message in df['user_message']:
    entry = re.split('([\w\W]+?):\s',message)
    if entry[1:]:
        users.append(entry[1])
        messages.append(" ".join(entry[2:]))
    else:
        users.append('group notification')
        messages.append(entry[0])

df['user']= users
df['message']= messages
df.drop(columns=['user_message'], inplace= True)
```

In [20]:
```
df
```

Out[20]:

	date	user	message
0	2021-01-18 15:05:00	group notification	Akanksha "Team"\n
1	2021-01-19 15:05:00	group notification	You were added\n
2	2022-08-27 14:56:00	group notification	Monika Hope you candidates are ready to atten...
3	2022-08-27 16:28:00	Shweta	<Media omitted>\n\n
4	2022-08-27 17:46:00	Aknasha	<Media omitted>\n
...	...	...	...
57	2022-08-30 15:42:00	Ashish	<Media omitted>\n\n
58	2022-08-30 16:33:00	Bhaviya	Syllabus \n\n
59	2022-08-30 19:04:00	Bhaviya	👍\n
60	2022-08-30 19:04:00	Ashish	👌\n
61	2022-08-30 19:04:00	Bhaviya	👌

62 rows × 3 columns

In [21]:
```
df['year']= df['date'].dt.year
```

In [22]:
```
df
```

Out[22]:

	date	user	message	year
0	2021-01-18 15:05:00	group notification	Akanksha "Team"\n	2021
1	2021-01-19 15:05:00	group notification	You were added\n	2021
2	2022-08-27 14:56:00	group notification	Monika Hope you candidates are ready to atten...	2022
3	2022-08-27 16:28:00	Shweta	<Media omitted>\n\n	2022
4	2022-08-27 17:46:00	Aknasha	<Media omitted>\n	2022
...	...	...	...	...
57	2022-08-30 15:42:00	Ashish	<Media omitted>\n\n	2022
58	2022-08-30 16:33:00	Bhaviya	Syllabus \n\n	2022
59	2022-08-30 19:04:00	Bhaviya	👍\n	2022
60	2022-08-30 19:04:00	Ashish	👌\n	2022
61	2022-08-30 19:04:00	Bhaviya	👌	2022

62 rows × 4 columns

In [23]:
```
df['month']= df['date'].dt.month_name()
```

In [24]:
```
df
```

Out[24]:

	date	user	message	year	month
0	2021-01-18 15:05:00	group notification	Akanksha "Team"\n	2021	January
1	2021-01-19 15:05:00	group notification	You were added\n	2021	January
2	2022-08-27 14:56:00	group notification	Monika Hope you candidates are ready to atten...	2022	August
3	2022-08-27 16:28:00	Shweta	<Media omitted>\n\n	2022	August
4	2022-08-27 17:46:00	Aknasha	<Media omitted>\n	2022	August
...	...	...	...	...	...
57	2022-08-30 15:42:00	Ashish	<Media omitted>\n\n	2022	August
58	2022-08-30 16:33:00	Bhaviya	Syllabus \n\n	2022	August
59	2022-08-30 19:04:00	Bhaviya	👍\n	2022	August
60	2022-08-30 19:04:00	Ashish	👌\n	2022	August
61	2022-08-30 19:04:00	Bhaviya	👌	2022	August

62 rows × 5 columns

In [25]:
```
df['day']= df['date'].dt.day
```

In [26]:
```
df
```

Out[26]:

	date	user	message	year	month	day
0	2021-01-18 15:05:00	group notification	Akanksha "Team"\n	2021	January	18
1	2021-01-19 15:05:00	group notification	You were added\n	2021	January	19
2	2022-08-27 14:56:00	group notification	Monika Hope you candidates are ready to atten...	2022	August	27
3	2022-08-27 16:28:00	Shweta	<Media omitted>\n\n	2022	August	27
4	2022-08-27 17:46:00	Aknasha	<Media omitted>\n	2022	August	27
...	...	...	...	...	...	...
57	2022-08-30 15:42:00	Ashish	<Media omitted>\n\n	2022	August	30
58	2022-08-30 16:33:00	Bhaviya	Syllabus \n\n	2022	August	30
59	2022-08-30 19:04:00	Bhaviya	👍\n	2022	August	30
60	2022-08-30 19:04:00	Ashish	👌\n	2022	August	30
61	2022-08-30 19:04:00	Bhaviya	👌	2022	August	30

62 rows × 6 columns

In [27]:
```
df['hour']= df['date'].dt.hour
```

In [28]:
```
df
```

Out[28]:

	date	user	message	year	month	day	hour
0	2021-01-18 15:05:00	group notification	Akanksha "Team"\n	2021	January	18	15
1	2021-01-19 15:05:00	group notification	You were added\n	2021	January	19	15
2	2022-08-27 14:56:00	group notification	Monika Hope you candidates are ready to atten...	2022	August	27	14
3	2022-08-27 16:28:00	Shweta	<Media omitted>\n\n	2022	August	27	16
4	2022-08-27 17:46:00	Aknasha	<Media omitted>\n	2022	August	27	17
...	...	...	...	...	...	...	...
57	2022-08-30 15:42:00	Ashish	<Media omitted>\n\n	2022	August	30	15
58	2022-08-30 16:33:00	Bhaviya	Syllabus \n\n	2022	August	30	16
59	2022-08-30 19:04:00	Bhaviya	👍\n	2022	August	30	19
60	2022-08-30 19:04:00	Ashish	👌\n	2022	August	30	19
61	2022-08-30 19:04:00	Bhaviya	👌	2022	August	30	19

62 rows × 7 columns

In [29]:
```
df['minute']= df['date'].dt.minute
```

In [30]:
```
df
```

Out[30]:

	date	user	message	year	month	day	hour	minute
0	2021-01-18 15:05:00	group notification	Akanksha "Team"\n	2021	January	18	15	5
1	2021-01-19 15:05:00	group notification	You were added\n	2021	January	19	15	5
2	2022-08-27 14:56:00	group notification	Monika Hope you candidates are ready to atten...	2022	August	27	14	56
3	2022-08-27 16:28:00	Shweta	<Media omitted>\n\n	2022	August	27	16	28
4	2022-08-27 17:46:00	Aknasha	<Media omitted>\n	2022	August	27	17	46
...	...	...	...	...	...	...	...	...
57	2022-08-30 15:42:00	Ashish	<Media omitted>\n\n	2022	August	30	15	42
58	2022-08-30 16:33:00	Bhaviya	Syllabus \n\n	2022	August	30	16	33
59	2022-08-30 19:04:00	Bhaviya	👍\n	2022	August	30	19	4
60	2022-08-30 19:04:00	Ashish	👌\n	2022	August	30	19	4
61	2022-08-30 19:04:00	Bhaviya	👌	2022	August	30	19	4

62 rows × 8 columns

In [31]:
```
df.shape[0]
```

Out[31]:
```
62
```

In [32]:
```
df[df['user']=='Bhaviya']
```

Out[32]:

	date	user	message	year	month	day	hour	minute
58	2022-08-30 16:33:00	Bhaviya	Syllabus \n\n	2022	August	30	16	33
59	2022-08-30 19:04:00	Bhaviya	👍\n	2022	August	30	19	4
61	2022-08-30 19:04:00	Bhaviya	👌	2022	August	30	19	4

In [33]:
```
words=[]
for message in df['message']:
    words.extend(message.split())
```

In [34]:
```
len(words)
```

Out[34]:
```
355
```

In [35]:
```
pip install urlextract
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: urlextract in c:\users\lenovo\appdata\roaming
\python\python39\site-packages (1.8.0)
Requirement already satisfied: idna in c:\programdata\anaconda3\lib\site-pack
ages (from urlextract) (3.3)
Requirement already satisfied: filelock in c:\programdata\anaconda3\lib\site-
packages (from urlextract) (3.6.0)
Requirement already satisfied: uritools in c:\users\lenovo\appdata\roaming\py
thon\python39\site-packages (from urlextract) (4.0.1)
Requirement already satisfied: platformdirs in c:\programdata\anaconda3\lib\s
ite-packages (from urlextract) (2.5.2)
Note: you may need to restart the kernel to use updated packages.
```

In [36]:
```
from urlextract import URLExtract

extractor = URLExtract()
urls= extractor.find_urls("Text with url : www.gmail.com")
print(urls)
```

```
['www.gmail.com']
```

In [37]:
```
links = []
for message in df['message']:
    links.extend(extractor.find_urls(message))
```

In [38]:
```
links
```

Out[38]:
```
['https://forms.gle/dNBpdL5NMsHmhF9k7',
 'https://www.youtube.com/watch?v=23fQ9-XUSCU',
 'https://www.youtube.com/watch?v=iBIcCGpSpeM&t=574s',
 'https://tinyurl.com/24vee9jt',
 'https://youtu.be/ZWlyGYWw7Cw',
 'https://forms.gle/Ux6YWQkjMAkhvthR6',
 'https://chat.whatsapp.com/D4VpUnfyafU0dS2Hp3vjuJ',
 'https://forms.gle/Ux6YWQkjMAkhvthR6',
 'https://chat.whatsapp.com/D4VpUnfyafU0dS2Hp3vjuJ',
 'https://forms.gle/Ux6YWQkjMAkhvthR6',
 'https://forms.gle/Ux6YWQkjMAkhvthR6']
```

In [39]:
```
df
```

Out[39]:

	date	user	message	year	month	day	hour	minute
0	2021-01-18 15:05:00	group notification	Akanksha "Team"\n	2021	January	18	15	5
1	2021-01-19 15:05:00	group notification	You were added\n	2021	January	19	15	5
2	2022-08-27 14:56:00	group notification	Monika Hope you candidates are ready to atten...	2022	August	27	14	56
3	2022-08-27 16:28:00	Shweta	<Media omitted>\n\n	2022	August	27	16	28
4	2022-08-27 17:46:00	Aknasha	<Media omitted>\n	2022	August	27	17	46
...	...	...	...	...	...	...	...	...
57	2022-08-30 15:42:00	Ashish	<Media omitted>\n\n	2022	August	30	15	42
58	2022-08-30 16:33:00	Bhaviya	Syllabus \n\n	2022	August	30	16	33
59	2022-08-30 19:04:00	Bhaviya	👍\n	2022	August	30	19	4
60	2022-08-30 19:04:00	Ashish	👌\n	2022	August	30	19	4
61	2022-08-30 19:04:00	Bhaviya	👌	2022	August	30	19	4

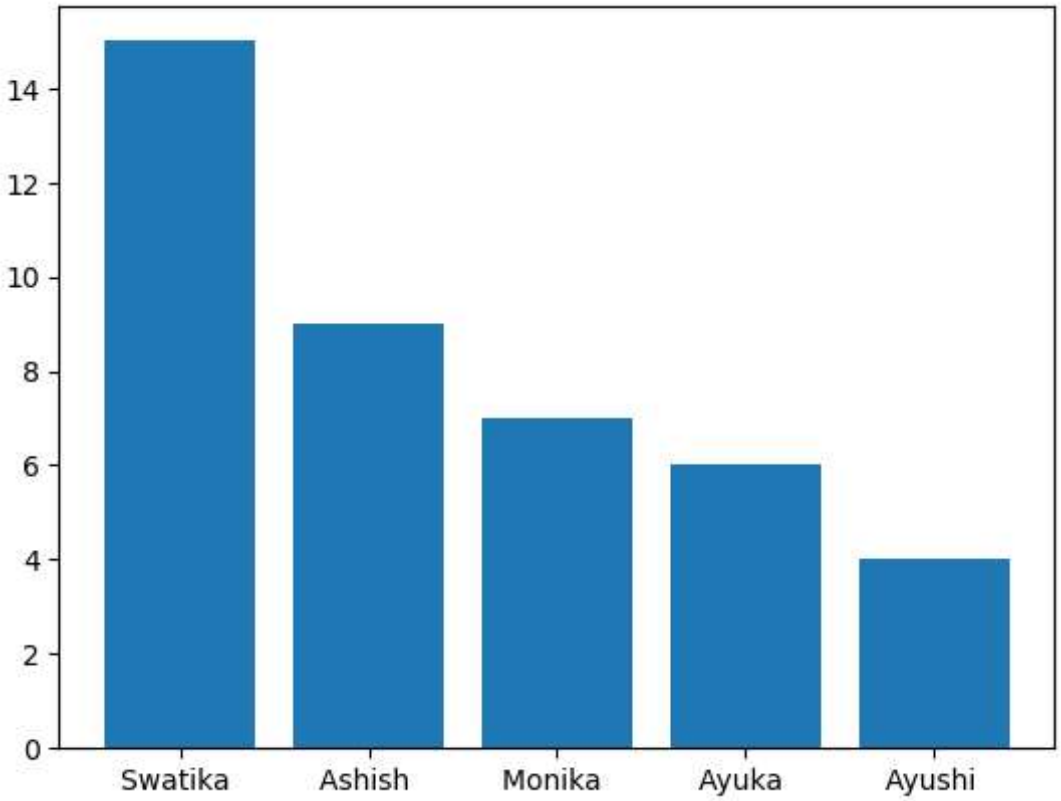
62 rows × 8 columns

In [40]:
```
x = df['user'].value_counts().head()
```

In [41]:
```
import matplotlib.pyplot as plt
```

In [42]:
```
name = x.index
count = x.values
```

In [43]:
```
plt.bar(name,count)
plt.show()
```



In [44]:
```
round((df['user'].value_counts()/df.shape[0])*100,2).reset_index().rename(colum
```

Out[44]:

	name	percent
0	Swatika	24.19
1	Ashish	14.52
2	Monika	11.29
3	Ayuka	9.68
4	Ayushi	6.45
5	group notification	4.84
6	Bhaviya	4.84
7	Aayushi	3.23
8	Shweta	3.23
9	Sameera	1.61
10	Bhaviya	1.61
11	Shweta	1.61
12	Swati	1.61
13	Bhavika	1.61
14	Monika	1.61
15	Ashwani	1.61
16	Ayushi	1.61
17	Ayush	1.61
18	Ashish	1.61
19	Aknasha	1.61

In [60]:
```
temp = df[df['user'] != 'group_notification']
temp= temp[temp['message'] != '<Media omitted>\n']
```

In [61]:
```
f = open('stop_hinglish.txt')
stop_words = f.read()
print(stop_words)
```

```
.
..
...
?
-
--
1
2
3
4
5
6
7
8
9
0
a
aadi
aaj
```

In [64]:
```
words = []

for message in temp['message']:
    for word in message.lower().split():
        if word not in stop_words:
            words.append(word)
```

In [65]:
```
from collections import Counter
pd.DataFrame(Counter(words).most_common(25))
```

Out[65]:

	0	1
0	vehicle	4
1	<media	4
2	omitted>	4
3	▬▬▬▬▬	4
4	workshop	4
5	kindly	4
6	fill	4
7	https://forms.gle/ux6ywqkjmakhvthr6	4
8	attending	4
9	*electric	3
10	link	3
11	august	3
12	form	3
13	|	3
14	💢	3
15	free	3
16	live	3
17	design	3
18	&	2
19	2022*	2
20	🔥	2
21	electric	2
22	industry	2
23	*click	2
24	watch	2

In [66]:
```
!pip install emoji
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: emoji in c:\users\lenovo\appdata\roaming\pytho
n\python39\site-packages (2.2.0)
```

In [67]:
```
import emoji
```

In [68]:
```
emojis=[]

for message in df['message']:
    emojis.extend([c for c in message if c in emoji.EMOJI_DATA])
```

In [70]:
```
pd.DataFrame(Counter(emojis).most_common(len(Counter(emojis))))
```

Out[70]:

	0	1
0	▬	32
1	⬚	13
2	👇	10
3	💢	3
4	🔥	2
5	😍	2
6	👌	2
7	🔜	1
8	🖇	1
9	⚜	1
10	⏰	1
11	🗓	1
12	□	1
13	😊	1
14	😉	1
15	💭	1
16	😊	1
17	👍	1

In [71]:
```
df
```

Out[71]:

	date	user	message	year	month	day	hour	minute
0	2021-01-18 15:05:00	group notification	Akanksha "Team"\n	2021	January	18	15	5
1	2021-01-19 15:05:00	group notification	You were added\n	2021	January	19	15	5
2	2022-08-27 14:56:00	group notification	Monika Hope you candidates are ready to atten...	2022	August	27	14	56
3	2022-08-27 16:28:00	Shweta	<Media omitted>\n\n	2022	August	27	16	28
4	2022-08-27 17:46:00	Aknasha	<Media omitted>\n	2022	August	27	17	46
...	...	...	...	...	...	...	...	...
57	2022-08-30 15:42:00	Ashish	<Media omitted>\n\n	2022	August	30	15	42
58	2022-08-30 16:33:00	Bhaviya	Syllabus \n\n	2022	August	30	16	33
59	2022-08-30 19:04:00	Bhaviya	👍\n	2022	August	30	19	4
60	2022-08-30 19:04:00	Ashish	👌\n	2022	August	30	19	4
61	2022-08-30 19:04:00	Bhaviya	👌	2022	August	30	19	4

62 rows × 8 columns

In [72]:
```
df['month_num'] = df['date'].dt.month
```

In [77]:
```
timeline = df.groupby(['year','month_num','month']).count()['message'].reset_ir
```

In [78]:
```
timeline
```

Out[78]:

	year	month_num	month	message
0	2021	1	January	2
1	2022	8	August	60

In [81]:
```
time = []
for i in range(timeline.shape[0]):
    time.append(timeline['month'][i] + '-' + str(timeline['year'][i]))
```

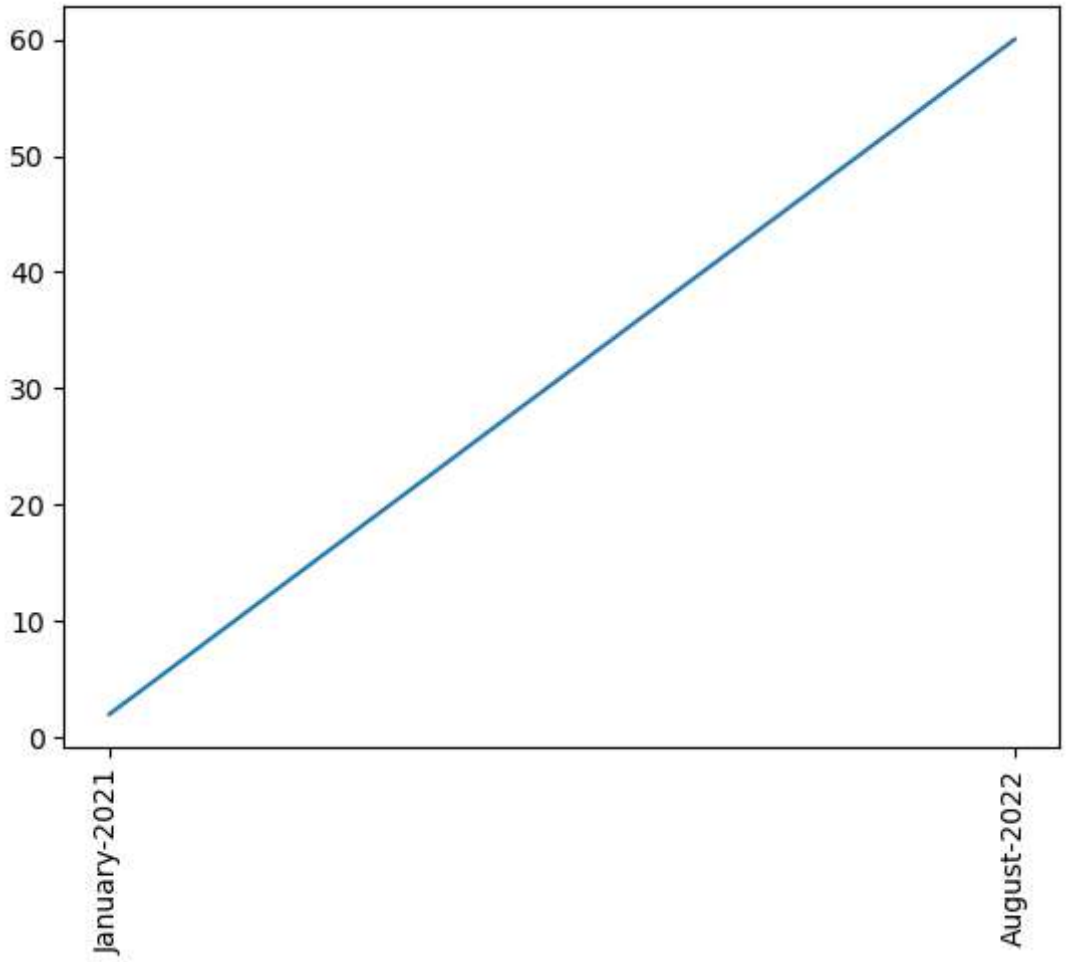
In [82]:
```
timeline['time'] = time
```

In [83]:
```
timeline
```

Out[83]:

	year	month_num	month	message	time
0	2021	1	January	2	January-2021
1	2022	8	August	60	August-2022

In [86]:
```
plt.plot(timeline['time'], timeline['message'])
plt.xticks(rotation = 'vertical')
plt.show()
```



In [87]:
```
df['only_date'] = df['date'].dt.date
```

In [91]:
```
daily_timeline = df.groupby('only_date').count()['message'].reset_index()
```

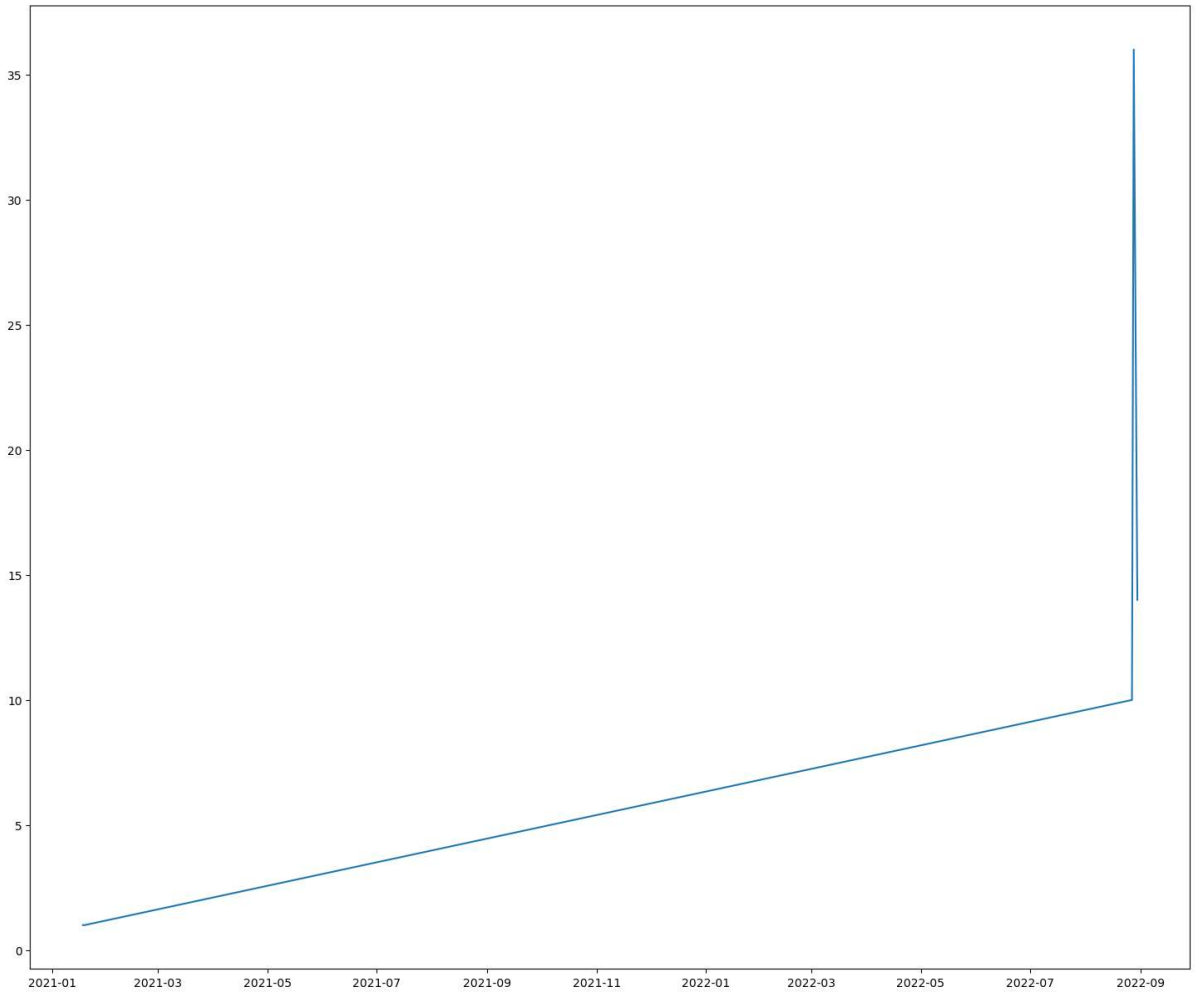
In [92]:
```
daily_timeline
```

Out[92]:

	only_date	message
0	2021-01-18	1
1	2021-01-19	1
2	2022-08-27	10
3	2022-08-28	36
4	2022-08-30	14

In [93]:
```
plt.figure(figsize= (18,15))
plt.plot(daily_timeline['only_date'],daily_timeline['message'])
```

Out[93]:
```
[<matplotlib.lines.Line2D at 0x1ccbed11130>]
```



In [94]:
```
df.head()
```

Out[94]:

	date	user	message	year	month	day	hour	minute	month_num	only_date
0	2021-01-18 15:05:00	group notification	Akanksha "Team"\n	2021	January	18	15	5	1	2021-01-18
1	2021-01-19 15:05:00	group notification	You were added\n	2021	January	19	15	5	1	2021-01-19
2	2022-08-27 14:56:00	group notification	Monika Hope you candidates are ready to atten...	2022	August	27	14	56	8	2022-08-27
3	2022-08-27 16:28:00	Shweta	<Media omitted>\n\n	2022	August	27	16	28	8	2022-08-27
4	2022-08-27 17:46:00	Aknasha	<Media omitted>\n	2022	August	27	17	46	8	2022-08-27

In [95]:
```
df['day_name'] =df['date'].dt.day_name()
```

In [96]:
```
df.head()
```

Out[96]:

	date	user	message	year	month	day	hour	minute	month_num	only_date	d
0	2021-01-18 15:05:00	group notification	Akanksha "Team"\n	2021	January	18	15	5	1	2021-01-18	
1	2021-01-19 15:05:00	group notification	You were added\n	2021	January	19	15	5	1	2021-01-19	
2	2022-08-27 14:56:00	group notification	Monika Hope you candidates are ready to atten...	2022	August	27	14	56	8	2022-08-27	
3	2022-08-27 16:28:00	Shweta	<Media omitted>\n\n	2022	August	27	16	28	8	2022-08-27	
4	2022-08-27 17:46:00	Aknasha	<Media omitted>\n	2022	August	27	17	46	8	2022-08-27	

In [98]:
```
df['day_name'].value_counts()
```

Out[98]:
```
Sunday       36
Tuesday      15
Saturday     10
Monday        1
Name: day_name, dtype: int64
```

In [99]:
```
df['month'].value_counts()
```

Out[99]:
```
August     60
January     2
Name: month, dtype: int64
```

In [100]:
```
df
```

Out[100]:

	date	user	message	year	month	day	hour	minute	month_num	only_date
0	2021-01-18 15:05:00	group notification	Akanksha "Team"\n	2021	January	18	15	5	1	2021-01-18
1	2021-01-19 15:05:00	group notification	You were added\n	2021	January	19	15	5	1	2021-01-19
2	2022-08-27 14:56:00	group notification	Monika Hope you candidates are ready to atten...	2022	August	27	14	56	8	2022-08-27
3	2022-08-27 16:28:00	Shweta	<Media omitted>\n\n	2022	August	27	16	28	8	2022-08-27
4	2022-08-27 17:46:00	Aknasha	<Media omitted>\n	2022	August	27	17	46	8	2022-08-27
...	...	...	...	...	...	...	...	...	...	...
57	2022-08-30 15:42:00	Ashish	<Media omitted>\n\n	2022	August	30	15	42	8	2022-08-30
58	2022-08-30 16:33:00	Bhaviya	Syllabus \n\n	2022	August	30	16	33	8	2022-08-30
59	2022-08-30 19:04:00	Bhaviya	👍\n	2022	August	30	19	4	8	2022-08-30
60	2022-08-30 19:04:00	Ashish	👌\n	2022	August	30	19	4	8	2022-08-30
61	2022-08-30 19:04:00	Bhaviya	👌	2022	August	30	19	4	8	2022-08-30

62 rows × 11 columns

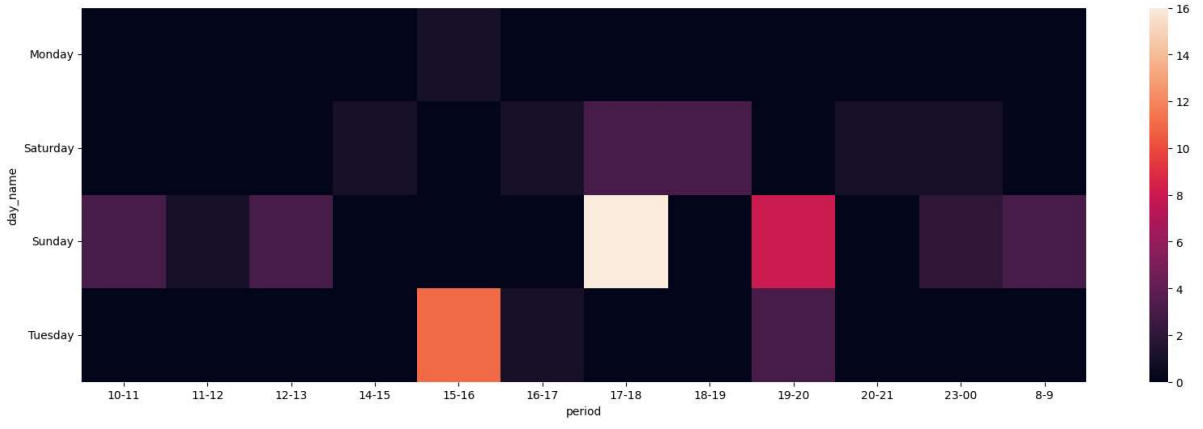
In [103]:
```
period=[]
for hour in df[['day_name','hour']]['hour']:
    if hour ==23:
        period.append(str(hour) + "-" +str("00"))
    elif hour ==0:
        period.append(str("00") + '-'+str(hour+1))
    else:
        period.append(str(hour) + '-'+str(hour+1))
```

In [104]:
```
df['period'] = period
```

In [106]:
```
df['period']
```

Out[106]:
```
0     15-16
1     15-16
2     14-15
3     16-17
4     17-18
      ...
57    15-16
58    16-17
59    19-20
60    19-20
61    19-20
Name: period, Length: 62, dtype: object
```

In [107]:
```
import seaborn as sns
plt.figure(figsize = (20,6))
sns.heatmap(df.pivot_table(index = 'day_name', columns='period',values="message
plt.yticks(rotation= 'horizontal')
plt.show()
```



In [ ]: