

```
In [2]: import re
import pandas as pd
```

```
In [3]: f = open('WhatsApp Chat with N.N Office IHH Team.txt','r',encoding='utf-8')
```

```
In [4]: data=f.read()
```



```
In [5]: print(data)
```

18/01/21, 15:05 - Aknasha Di Ihh created group "N.N Office IHH Team"  
18/01/21, 15:05 - You were added  
27/08/22, 14:17 - Monika Didi Alt No: This message was deleted  
27/08/22, 14:25 - Monika Didi Alt No: This message was deleted  
27/08/22, 14:25 - Monika Didi Alt No: Attendance Sheet | 28th AUGUST 2022 |  
Electric Vehicle Design & Development Live Free Workshop | IHPET: Industri  
es Helping Hands

<https://forms.gle/Ux6YWQkjMAkhvthR6> (<https://forms.gle/Ux6YWQkjMAkhvthR6>)

27/08/22, 14:55 - Monika Didi Ihh: This message was deleted  
27/08/22, 14:55 - Monika Didi Ihh: !! \*JOIN NOW\* !!

Your \*Free Live Master Class of Electric Vehicle Component Design and Devel  
opment\* is Schedule on

 \*28th August 2022\*  
 \*11:00 AM\*

~~~~~  
\*! ! ! ! \*

```
In [6]: pattern = '\d{1,2}/\d{1,2}/\d{2,4},\s\d{1,2}:\d{2}\s-\s'
```

```
In [7]: messages = re.split(pattern, data)[1:]
messages
```

```
Out[7]: ['Aknasha Di Ihh created group "N.N Office IHH Team"\n',
'You were added\n',
'Monika Didi Alt No: This message was deleted\n',
'Monika Didi Alt No: This message was deleted\n',
'Monika Didi Alt No: Attendance Sheet | 28th AUGUST 2022 | Electric Vehicle Design & Development Live Free Workshop | IHHPET: Industries Helping Hands \n\n https://forms.gle/Ux6YWQkjMAKhvthR6\n', (https://forms.gle/Ux6YWQkjMAKhvthR6\n',)
'Monika Didi Ihh: This message was deleted\n',
'Monika Didi Ihh: !! *JOIN NOW* !!\n\nYour *Free Live Master Class of Electric Vehicle Component Design and Development* is Schedule on\n\n 📅 *28th August 2022* \n 🕒 *11:00 AM* \n~~~~~\n *Link* 🔗\nhhttps://meet.google.com/nen-yy-mz-ept \n\n*Meeting Code*-\n nen-yy-mz-ept\n~~~~~\n~~~~\nKindly Join session on time.\n🚫 *Note* - We have limited seats for this *Live Master Class* If you will unable to join, kindly watch our *LIVE SESSION* on *YouTube LIVE STREAMING*\n\n(You will receive *Link🔗 of YouTube 📺 Streaming* in this WhatsApp group) \n\n *Regards*\n *IHHPET: Industries Helping Hands*\n',
'Aayush Bhaiya Alt No: ek msg aur bnaiyega jo 9:30 am ko bhejengey\n',
...]
```

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

```
27/08/22, 14:55 - ',
'27/08/22, 14:55 - ',
'27/08/22, 14:55 - ',
'27/08/22, 14:56 - ',
'27/08/22, 16:28 - ',
'27/08/22, 16:31 - ',
'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, 20:37 - ',
'27/08/22, 20:42 - ',
'27/08/22, 20:46 - ',
'27/08/22, 21:45 - ',
'27/08/22, 21:50 - ',
'27/08/22, 21:50 - ',
'27/08/22, 21:50 - ',
'27/08/22, 21:51 - ',
...]
```

```
In [9]: df = pd.DataFrame({'user_message': messages, 'message_date': dates})
```

```
In [10]: df
```

```
Out[10]:
```

|      | user_message                                      | message_date      |
|------|---------------------------------------------------|-------------------|
| 0    | Aknasha Di Ihh created group "N.N Office IHH T... | 18/01/21, 15:05 - |
| 1    | You were added\n                                  | 18/01/21, 15:05 - |
| 2    | Monika Didi Alt No: This message was deleted\n    | 27/08/22, 14:17 - |
| 3    | Monika Didi Alt No: This message was deleted\n    | 27/08/22, 14:25 - |
| 4    | Monika Didi Alt No: Attendance Sheet   28th AU... | 27/08/22, 14:25 - |
| ...  | ...                                               | ...               |
| 7709 | Bhai: <Media omitted>\n                           | 06/05/23, 14:55 - |
| 7710 | Ayush Bhaiya: Ye nai h\n                          | 06/05/23, 14:56 - |
| 7711 | Ayush Bhaiya: Dhundha mai\n                       | 06/05/23, 14:56 - |
| 7712 | Bhai: Linkedin me connect ho kar next free LIV... | 06/05/23, 14:57 - |
| 7713 | Aayush Bhaiya Alt No: ok\n                        | 06/05/23, 14:57 - |

7714 rows × 2 columns

```
In [11]: df['message_date']
```

```
Out[11]:
```

|      |                   |
|------|-------------------|
| 0    | 18/01/21, 15:05 - |
| 1    | 18/01/21, 15:05 - |
| 2    | 27/08/22, 14:17 - |
| 3    | 27/08/22, 14:25 - |
| 4    | 27/08/22, 14:25 - |
| ...  | ...               |
| 7709 | 06/05/23, 14:55 - |
| 7710 | 06/05/23, 14:56 - |
| 7711 | 06/05/23, 14:56 - |
| 7712 | 06/05/23, 14:57 - |
| 7713 | 06/05/23, 14:57 - |

Name: message\_date, Length: 7714, dtype: object

```
In [12]: df['message_date'] = pd.to_datetime(df['message_date'],format = "%d/%m/%y, %H:%M")
```

In [13]: df

Out[13]:

|      | user_message                                      | message_date        |
|------|---------------------------------------------------|---------------------|
| 0    | Aknasha Di Ihh created group "N.N Office IHH T... | 2021-01-18 15:05:00 |
| 1    | You were added\n                                  | 2021-01-18 15:05:00 |
| 2    | Monika Didi Alt No: This message was deleted\n    | 2022-08-27 14:17:00 |
| 3    | Monika Didi Alt No: This message was deleted\n    | 2022-08-27 14:25:00 |
| 4    | Monika Didi Alt No: Attendance Sheet   28th AU... | 2022-08-27 14:25:00 |
| ...  | ...                                               | ...                 |
| 7709 | Bhai: <Media omitted>\n                           | 2023-05-06 14:55:00 |
| 7710 | Ayush Bhaiya: Ye nai h\n                          | 2023-05-06 14:56:00 |
| 7711 | Ayush Bhaiya: Dhundha mail\n                      | 2023-05-06 14:56:00 |
| 7712 | Bhai: Linkedin me connect ho kar next free LIV... | 2023-05-06 14:57:00 |
| 7713 | Aayush Bhaiya Alt No: ok\n                        | 2023-05-06 14:57:00 |

7714 rows × 2 columns

In [14]: df.shape

Out[14]: (7714, 2)

In [15]: df.rename(columns={'message\_date': 'date'}, inplace=True)

In [16]: df

Out[16]:

|      | user_message                                      | date                |
|------|---------------------------------------------------|---------------------|
| 0    | Aknasha Di Ihh created group "N.N Office IHH T... | 2021-01-18 15:05:00 |
| 1    | You were added\n                                  | 2021-01-18 15:05:00 |
| 2    | Monika Didi Alt No: This message was deleted\n    | 2022-08-27 14:17:00 |
| 3    | Monika Didi Alt No: This message was deleted\n    | 2022-08-27 14:25:00 |
| 4    | Monika Didi Alt No: Attendance Sheet   28th AU... | 2022-08-27 14:25:00 |
| ...  | ...                                               | ...                 |
| 7709 | Bhai: <Media omitted>\n                           | 2023-05-06 14:55:00 |
| 7710 | Ayush Bhaiya: Ye nai h\n                          | 2023-05-06 14:56:00 |
| 7711 | Ayush Bhaiya: Dhundha mail\n                      | 2023-05-06 14:56:00 |
| 7712 | Bhai: Linkedin me connect ho kar next free LIV... | 2023-05-06 14:57:00 |
| 7713 | Aayush Bhaiya Alt No: ok\n                        | 2023-05-06 14:57:00 |

7714 rows × 2 columns

```
In [17]: users = []
messages = []
for message in df['user_message']:
    entry = re.split('([\w\W]+?):\s', message)
    if entry[1:]: # user name
        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 [18]: df.head()
```

```
Out[18]:
```

|   | date                | user               | message                                           |
|---|---------------------|--------------------|---------------------------------------------------|
| 0 | 2021-01-18 15:05:00 | group_notification | Aknasha Di Ihh created group "N.N Office IHH T... |
| 1 | 2021-01-18 15:05:00 | group_notification | You were added\n                                  |
| 2 | 2022-08-27 14:17:00 | Monika Didi Alt No | This message was deleted\n                        |
| 3 | 2022-08-27 14:25:00 | Monika Didi Alt No | This message was deleted\n                        |
| 4 | 2022-08-27 14:25:00 | Monika Didi Alt No | Attendance Sheet   28th AUGUST 2022   Electri...  |

```
In [19]: df['year'] = df['date'].dt.year
```

In [20]: df

Out[20]:

|      | date                   | user                 | message                                           | year |
|------|------------------------|----------------------|---------------------------------------------------|------|
| 0    | 2021-01-18<br>15:05:00 | group_notification   | Aknasha Di Ihh created group "N.N Office IHH T... | 2021 |
| 1    | 2021-01-18<br>15:05:00 | group_notification   | You were added\n                                  | 2021 |
| 2    | 2022-08-27<br>14:17:00 | Monika Didi Alt No   | This message was deleted\n                        | 2022 |
| 3    | 2022-08-27<br>14:25:00 | Monika Didi Alt No   | This message was deleted\n                        | 2022 |
| 4    | 2022-08-27<br>14:25:00 | Monika Didi Alt No   | Attendance Sheet   28th AUGUST 2022   Electri...  | 2022 |
| ...  | ...                    | ...                  | ...                                               | ...  |
| 7709 | 2023-05-06<br>14:55:00 | Bhai                 | <Media omitted>\n                                 | 2023 |
| 7710 | 2023-05-06<br>14:56:00 | Ayush Bhaiya         | Ye nai h\n                                        | 2023 |
| 7711 | 2023-05-06<br>14:56:00 | Ayush Bhaiya         | Dhundha mai\n                                     | 2023 |
| 7712 | 2023-05-06<br>14:57:00 | Bhai                 | Linkedin me connect ho kar next free LIVE work... | 2023 |
| 7713 | 2023-05-06<br>14:57:00 | Aayush Bhaiya Alt No | ok\n                                              | 2023 |

7714 rows × 4 columns

In [21]: df['month'] = df['date'].dt.month\_name()

In [22]: df.head()

Out[22]:

|   | date                   | user               | message                                           | year | month   |
|---|------------------------|--------------------|---------------------------------------------------|------|---------|
| 0 | 2021-01-18<br>15:05:00 | group_notification | Aknasha Di Ihh created group "N.N Office IHH T... | 2021 | January |
| 1 | 2021-01-18<br>15:05:00 | group_notification | You were added\n                                  | 2021 | January |
| 2 | 2022-08-27<br>14:17:00 | Monika Didi Alt No | This message was deleted\n                        | 2022 | August  |
| 3 | 2022-08-27<br>14:25:00 | Monika Didi Alt No | This message was deleted\n                        | 2022 | August  |
| 4 | 2022-08-27<br>14:25:00 | Monika Didi Alt No | Attendance Sheet   28th AUGUST 2022   Electri...  | 2022 | August  |

In [23]: df['day'] = df['date'].dt.day

```
In [24]: df.head()
```

```
Out[24]:
```

|   | date                   | user                  | message                                              | year | month   | day |
|---|------------------------|-----------------------|------------------------------------------------------|------|---------|-----|
| 0 | 2021-01-18<br>15:05:00 | group_notification    | Aknasha Di Ihh created group "N.N Office<br>IHH T... | 2021 | January | 18  |
| 1 | 2021-01-18<br>15:05:00 | group_notification    | You were added\n                                     | 2021 | January | 18  |
| 2 | 2022-08-27<br>14:17:00 | Monika Didi Alt<br>No | This message was deleted\n                           | 2022 | August  | 27  |
| 3 | 2022-08-27<br>14:25:00 | Monika Didi Alt<br>No | This message was deleted\n                           | 2022 | August  | 27  |
| 4 | 2022-08-27<br>14:25:00 | Monika Didi Alt<br>No | Attendance Sheet   28th AUGUST 2022  <br>Electri...  | 2022 | August  | 27  |

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

```
In [26]: df.head()
```

```
Out[26]:
```

|   | date                   | user                  | message                                              | year | month   | day | hour |
|---|------------------------|-----------------------|------------------------------------------------------|------|---------|-----|------|
| 0 | 2021-01-18<br>15:05:00 | group_notification    | Aknasha Di Ihh created group "N.N<br>Office IHH T... | 2021 | January | 18  | 15   |
| 1 | 2021-01-18<br>15:05:00 | group_notification    | You were added\n                                     | 2021 | January | 18  | 15   |
| 2 | 2022-08-27<br>14:17:00 | Monika Didi Alt<br>No | This message was deleted\n                           | 2022 | August  | 27  | 14   |
| 3 | 2022-08-27<br>14:25:00 | Monika Didi Alt<br>No | This message was deleted\n                           | 2022 | August  | 27  | 14   |
| 4 | 2022-08-27<br>14:25:00 | Monika Didi Alt<br>No | Attendance Sheet   28th AUGUST<br>2022   Electri...  | 2022 | August  | 27  | 14   |

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

```
In [28]: df.head()
```

```
Out[28]:
```

|   | date                   | user                  | message                                              | year | month   | day | hour | minute |
|---|------------------------|-----------------------|------------------------------------------------------|------|---------|-----|------|--------|
| 0 | 2021-01-18<br>15:05:00 | group_notification    | Aknasha Di Ihh created<br>group "N.N Office IHH T... | 2021 | January | 18  | 15   | 5      |
| 1 | 2021-01-18<br>15:05:00 | group_notification    | You were added\n                                     | 2021 | January | 18  | 15   | 5      |
| 2 | 2022-08-27<br>14:17:00 | Monika Didi Alt<br>No | This message was deleted\n                           | 2022 | August  | 27  | 14   | 17     |
| 3 | 2022-08-27<br>14:25:00 | Monika Didi Alt<br>No | This message was deleted\n                           | 2022 | August  | 27  | 14   | 25     |
| 4 | 2022-08-27<br>14:25:00 | Monika Didi Alt<br>No | Attendance Sheet   28th<br>AUGUST 2022   Electri...  | 2022 | August  | 27  | 14   | 25     |

```
In [29]: df[df['user'] == 'Swati Dhote'].shape
```

```
Out[29]: (533, 8)
```

```
In [30]: words = []  
for message in df['message']:  
    words.extend(message.split())
```

```
In [31]: df[df['message'] == '<Media Omitted>/n'].shape
```

```
Out[31]: (0, 8)
```

```
In [32]: 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: filelock in c:\programdata\anaconda3\lib\site-packages (from urlextract) (3.6.0)  
Requirement already satisfied: platformdirs in c:\programdata\anaconda3\lib\site-packages (from urlextract) (2.5.2)  
Requirement already satisfied:uritools in c:\users\lenovo\appdata\roaming\python\python39\site-packages (from urlextract) (4.0.1)  
Requirement already satisfied: idna in c:\programdata\anaconda3\lib\site-packages (from urlextract) (3.3)  
Note: you may need to restart the kernel to use updated packages.

```
In [33]: from urlextract import URLExtract  
  
extractor = URLExtract()  
urls = extractor.find_urls("Text with URLs. Let's have URL janlipovsky.cz as an example")  
print(urls)
```

```
['janlipovsky.cz', 'example.www.gmail.com']
```



```
In [34]: links = []
for message in df['message']:
    print(extractor.find_urls(message))
```

```
[]
[]
[]
[]
['https://forms.gle/Ux6YWQkjMAkhvthR6']
[]
['https://meet.google.com/nen-yymz-ept']
[]
[]
[]
[]
[]
[]
[]
['https://forms.gle/dNBpdL5NMsHmhF9k7']
[]
[]
[]
[]
```

```
In [35]: links = []
for message in df['message']:
    links.extend(extractor.find_urls(message))
```

```
In [36]: links
```

```
Out[36]: ['https://forms.gle/Ux6YWQkjMAkhvthR6',
'https://meet.google.com/nen-yymz-ept',
'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://forms.gle/Ux6YWQkjMAkhvthR6',
'https://youtu.be/ZWlyGYWw7Cw',
'https://forms.gle/Ux6YWQkjMAkhvthR6',
'https://chat.whatsapp.com/D4VpUnfyafU0dS2Hp3vjuJ',
'https://forms.gle/Ux6YWQkjMAkhvthR6',
'https://forms.gle/Ux6YWQkjMAkhvthR6',
'https://youtu.be/xRwaI7BFb1Y',
'https://youtu.be/9DeZjxeODxE',
'https://forms.gle/FPjUDqHjW6SshXSh8',
'https://forms.gle/6CedKGFbV1qzFVp8']
```

```
In [37]: len(links)
```

```
Out[37]: 1115
```

```
In [38]: x = df['user'].value_counts().head()
```

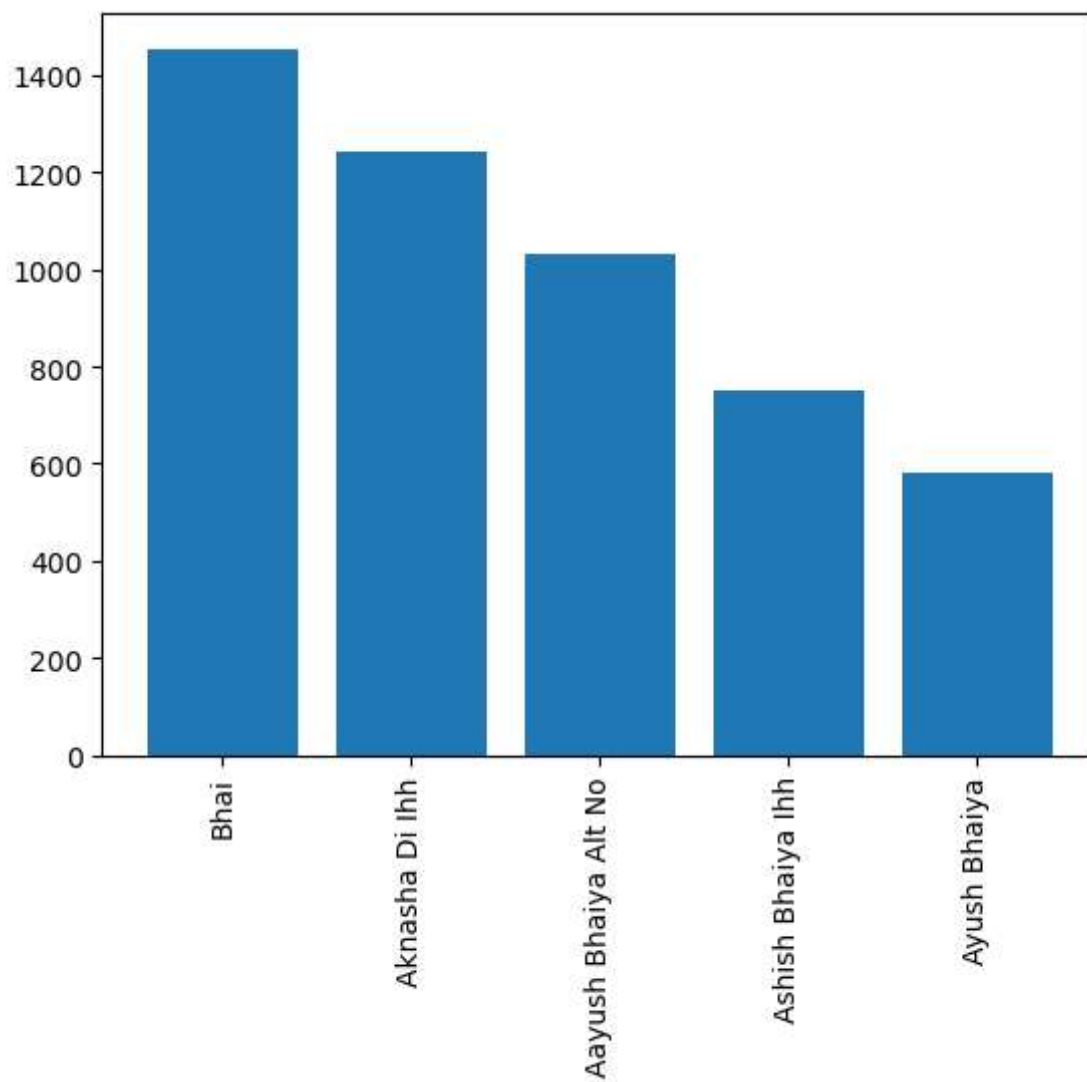
```
In [39]: x
```

```
Out[39]: Bhai                1455  
Aknasha Di Ihh            1244  
Aayush Bhaiya Alt No      1033  
Ashish Bhaiya Ihh         749  
Ayush Bhaiya              581  
Name: user, dtype: int64
```

```
In [40]: import matplotlib.pyplot as plt
```

```
In [41]: name = x.index  
count = x.values
```

```
In [42]: plt.bar(name,count)
plt.xticks(rotation = 'vertical')
plt.show()
```



```
In [43]: round((df['user'].value_counts()/df.shape[0])*100,2).reset_index().rename(column
```

Out[43]:

|    | name                            | percent |
|----|---------------------------------|---------|
| 0  | Bhai                            | 18.86   |
| 1  | Aknasha Di Ihh                  | 16.13   |
| 2  | Aayush Bhaiya Alt No            | 13.39   |
| 3  | Ashish Bhaiya Ihh               | 9.71    |
| 4  | Ayush Bhaiya                    | 7.53    |
| 5  | Monika Didi Ihh                 | 7.10    |
| 6  | Swati Dhote                     | 6.91    |
| 7  | Bhawesh Bhaiya Ihh              | 6.49    |
| 8  | Monika Didi Alt No              | 5.59    |
| 9  | Sameer Interview                | 2.54    |
| 10 | Suryansh Ihh                    | 1.37    |
| 11 | Menika Ihh WhatsApp             | 0.73    |
| 12 | St Rajkaran                     | 0.70    |
| 13 | group_notification              | 0.61    |
| 14 | Jyotirmay Ihh                   | 0.39    |
| 15 | Priyanka (Ihh Altranate Number  | 0.38    |
| 16 | Rukhsar IHH                     | 0.32    |
| 17 | Akshay Ihh                      | 0.27    |
| 18 | Ruchita Ihh                     | 0.25    |
| 19 | Manish Ihh                      | 0.18    |
| 20 | Baljeet Ihhpet                  | 0.10    |
| 21 | +91 95162 42658                 | 0.10    |
| 22 | +91 99931 64904                 | 0.09    |
| 23 | +91 79993 96547                 | 0.08    |
| 24 | Sonali Industries Helping Hands | 0.06    |
| 25 | Pankaj Interview                | 0.04    |
| 26 | Vandana Naidu                   | 0.04    |
| 27 | +91 73892 19668                 | 0.01    |
| 28 | Bhavesh Bhaiya Alt No           | 0.01    |

```
In [44]: temp = df[df['user'] != 'group_notification']
temp=temp[temp['message'] != '<Media omitted>\n']
```

```
In [45]: f = open('stop_hinglish.txt','r')
stop_words = f.read()
print(stop_words)
```

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

```
In [46]: words = []

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

```
In [47]: from collections import Counter
pd.DataFrame(Counter(words).most_common(25))
```

```
Out[47]:
```

|    | 0           | 1   |
|----|-------------|-----|
| 0  |             | 508 |
| 1  | message     | 346 |
| 2  | deleted     | 327 |
| 3  | link        | 284 |
| 4  | helping     | 267 |
| 5  | =====       | 264 |
| 6  | class       | 237 |
| 7  | video       | 234 |
| 8  | vehicle     | 231 |
| 9  | call        | 231 |
| 10 | pm          | 227 |
| 11 | electric    | 219 |
| 12 | live        | 204 |
| 13 | &           | 180 |
| 14 | workshop    | 175 |
| 15 | *free       | 173 |
| 16 | !!          | 167 |
| 17 | join        | 163 |
| 18 | kindly      | 158 |
| 19 | dena        | 157 |
| 20 | office      | 153 |
| 21 | *industries | 153 |
| 22 | hands*      | 142 |
| 23 | =====       | 141 |
| 24 | gya         | 138 |

```
In [48]: !pip install emoji
```

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

```
In [49]: import emoji
```

```
In [50]: emojis = []
for message in df['message']:
    emojis.extend([c for c in message if c in emoji.EMOJI_DATA ])
```

```
In [51]: pd.DataFrame(Counter(emojis).most_common(len(Counter(emojis))))
```

```
Out[51]:
```

|     | 0   | 1    |
|-----|-----|------|
| 0   | ==  | 2564 |
| 1   | 📄   | 961  |
| 2   | 👉   | 374  |
| 3   | ~   | 309  |
| 4   | 🤖   | 185  |
| ... | ... | ...  |
| 208 | 🎧   | 1    |
| 209 | 👁   | 1    |
| 210 | 📺   | 1    |
| 211 | 🛡   | 1    |
| 212 | 🙏   | 1    |

213 rows × 2 columns

```
In [52]: df['month_num'] =df['date'].dt.month
```

```
In [53]: timeline = df.groupby(['year','month_num','month']).count()['message'].reset_index
```

```
In [54]: timeline
```

```
Out[54]:
```

|    | year | month_num | month     | message |
|----|------|-----------|-----------|---------|
| 0  | 2021 | 1         | January   | 2       |
| 1  | 2022 | 8         | August    | 467     |
| 2  | 2022 | 9         | September | 733     |
| 3  | 2022 | 10        | October   | 762     |
| 4  | 2022 | 11        | November  | 466     |
| 5  | 2022 | 12        | December  | 1295    |
| 6  | 2023 | 1         | January   | 1508    |
| 7  | 2023 | 2         | February  | 705     |
| 8  | 2023 | 3         | March     | 847     |
| 9  | 2023 | 4         | April     | 798     |
| 10 | 2023 | 5         | May       | 131     |

```
In [55]: time = []
         for i in range(timeline.shape[0]):
             time.append(timeline['month'][i] + "-" + str(timeline['year'][i]))
```

```
In [56]: timeline['time']=time
```

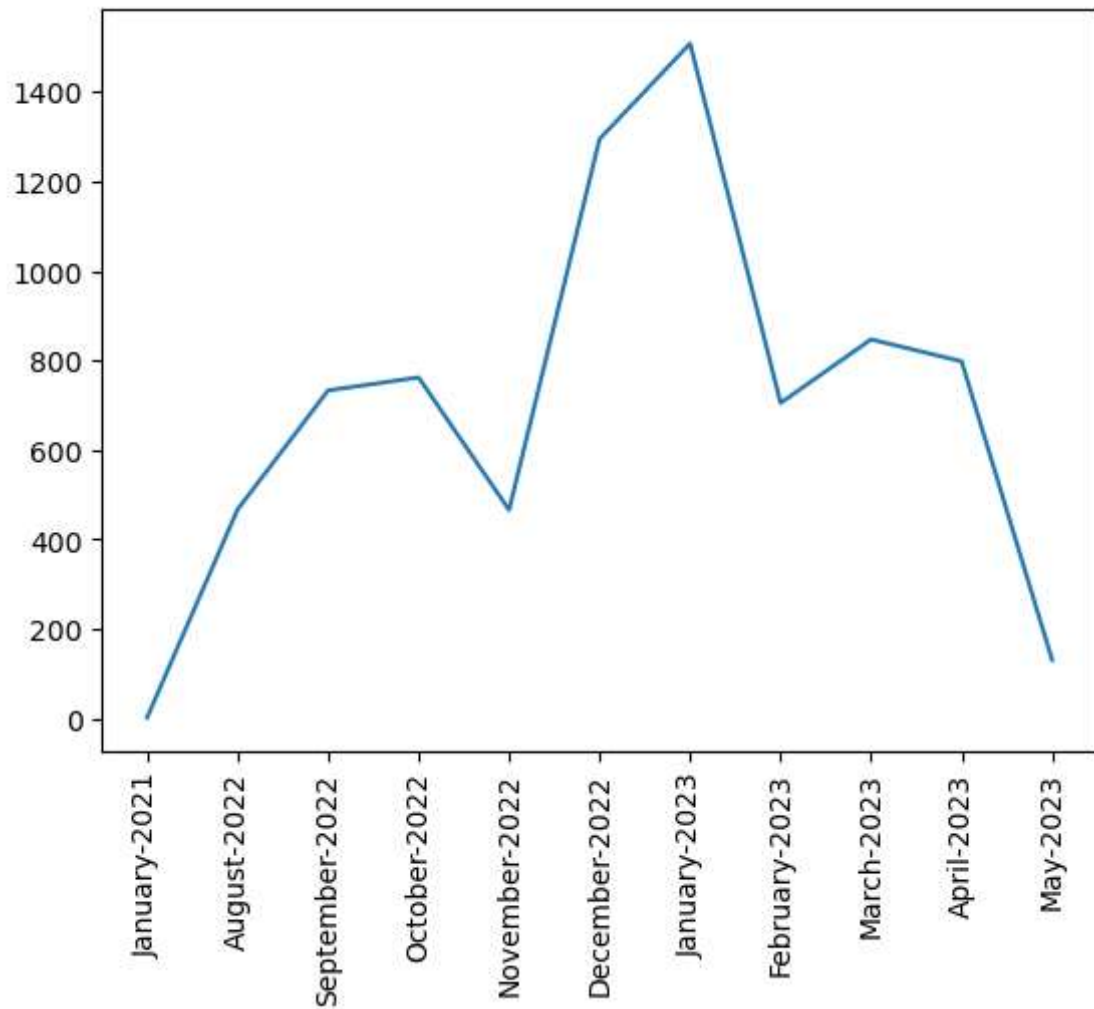
```
In [57]: timeline
```

```
Out[57]:
```

|    | year | month_num | month     | message | time           |
|----|------|-----------|-----------|---------|----------------|
| 0  | 2021 | 1         | January   | 2       | January-2021   |
| 1  | 2022 | 8         | August    | 467     | August-2022    |
| 2  | 2022 | 9         | September | 733     | September-2022 |
| 3  | 2022 | 10        | October   | 762     | October-2022   |
| 4  | 2022 | 11        | November  | 466     | November-2022  |
| 5  | 2022 | 12        | December  | 1295    | December-2022  |
| 6  | 2023 | 1         | January   | 1508    | January-2023   |
| 7  | 2023 | 2         | February  | 705     | February-2023  |
| 8  | 2023 | 3         | March     | 847     | March-2023     |
| 9  | 2023 | 4         | April     | 798     | April-2023     |
| 10 | 2023 | 5         | May       | 131     | May-2023       |



```
In [58]: plt.plot(timeline['time'], timeline['message'])  
plt.xticks(rotation="vertical")  
plt.show()
```

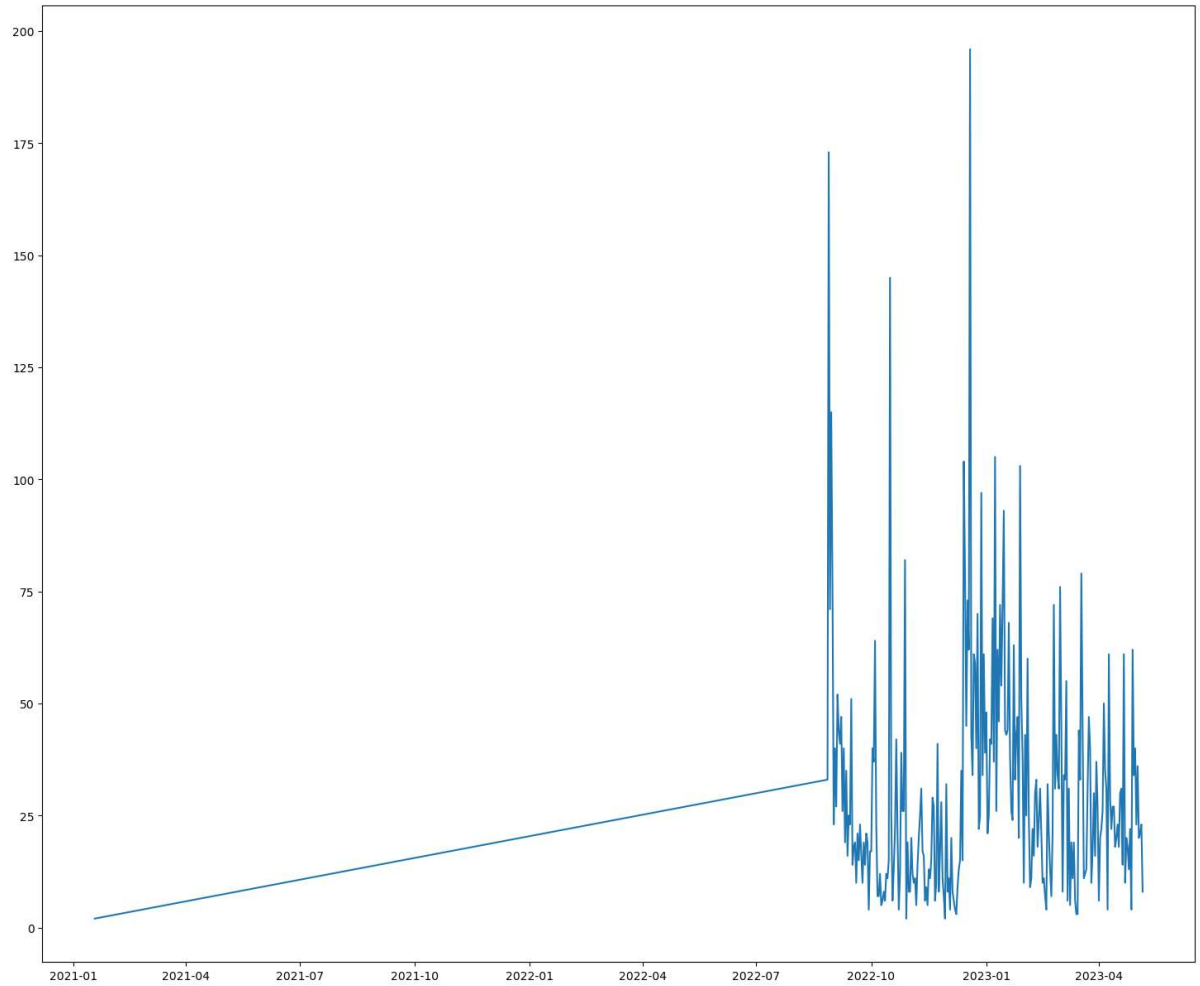


```
In [59]: df['only_date'] = df['date'].dt.date
```

```
In [66]: daily_timeline = df.groupby('only_date').count()['message'].reset_index()
```

```
In [69]: plt.figure(figsize=(18,15))  
plt.plot(daily_timeline['only_date'],daily_timeline['message'])
```

```
Out[69]: [<matplotlib.lines.Line2D at 0x24ae87cd310>]
```



```
In [70]: df.head()
```

```
Out[70]:
```

|   | date                | user               | message                                           | year | month   | day | hour | minute | month_num | only_da |
|---|---------------------|--------------------|---------------------------------------------------|------|---------|-----|------|--------|-----------|---------|
| 0 | 2021-01-18 15:05:00 | group_notification | Aknasha Di Ihh created group "N.N Office IHH T... | 2021 | January | 18  | 15   | 5      | 1         | 2021-0  |
| 1 | 2021-01-18 15:05:00 | group_notification | You were added\n                                  | 2021 | January | 18  | 15   | 5      | 1         | 2021-0  |
| 2 | 2022-08-27 14:17:00 | Monika Didi Alt No | This message was deleted\n                        | 2022 | August  | 27  | 14   | 17     | 8         | 2022-0  |
| 3 | 2022-08-27 14:25:00 | Monika Didi Alt No | This message was deleted\n                        | 2022 | August  | 27  | 14   | 25     | 8         | 2022-0  |
| 4 | 2022-08-27 14:25:00 | Monika Didi Alt No | Attendance Sheet   28th AUGUST 2022   Electri...  | 2022 | August  | 27  | 14   | 25     | 8         | 2022-0  |

```
In [72]: df['day_name']=df['date'].dt.day_name()
```

```
In [73]: df['day_name'].value_counts()
```

```
Out[73]: Sunday      1428
Friday      1179
Wednesday   1157
Monday      1030
Thursday    1019
Saturday     989
Tuesday     912
Name: day_name, dtype: int64
```

```
In [74]: df['month'].value_counts()
```

```
Out[74]: January      1510
December    1295
March        847
April        798
October      762
September    733
February     705
August       467
November     466
May          131
Name: month, dtype: int64
```

```
In [75]: period = []
for hour in df[['day_name', 'hour']][ 'hour']:
    if hour == 23:
        period.append(str(hour) + "-" +str('00'))

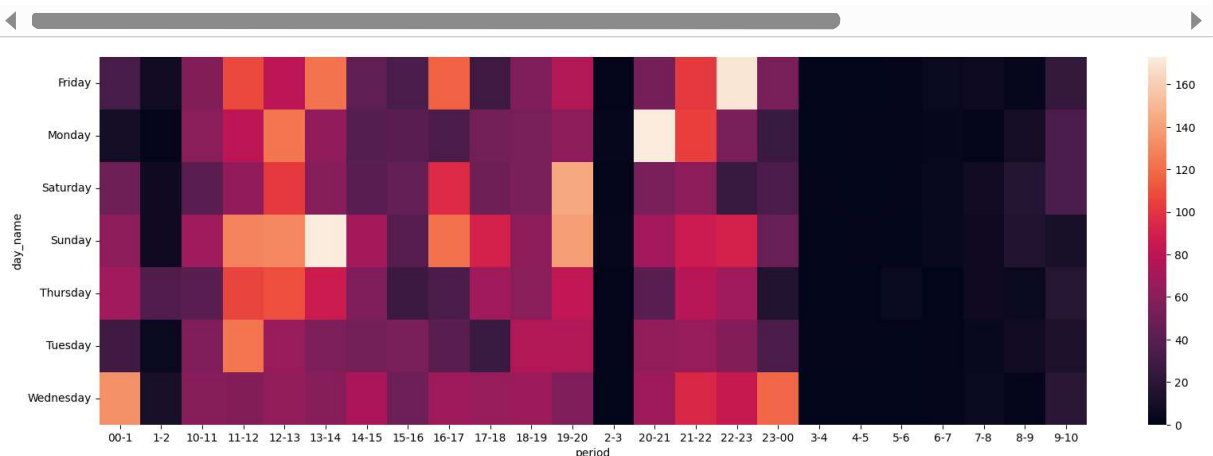
    elif hour ==0:
        period.append(str("00")+ "-" +str(hour+1))
    else:
        period.append(str(hour)+ "-" +str(hour+1))
```

```
In [76]: df['period']=period
```

```
In [79]: df['period']
```

```
Out[79]: 0      15-16
1      15-16
2      14-15
3      14-15
4      14-15
...
7709   14-15
7710   14-15
7711   14-15
7712   14-15
7713   14-15
Name: period, Length: 7714, dtype: object
```

```
In [80]: import seaborn as sns
plt.figure(figsize=(20,6))
sns.heatmap(df.pivot_table(index = 'day_name', columns = 'period', values="message_count"))
plt.xticks(rotation='horizontal')
plt.show()
```



```
In [ ]:
```

