

# Your Whatsapp chats can tell a lot about you

Visualising WhatsApp chats using Python

Lately, I was looking for some small and exciting visualisation projects to explore the data visualisation field. Then I came across a feature in WhatsApp about exporting your chats in a text file, which is quite handy and easy to work with. WhatsApp claims that nearly 65 billion WhatsApp messages sent per day, or 29 million per minute in 2020.

I started using WhatsApp frequently after getting into college in 2016, so I thought of collecting and visualising my last four years of chats. I obtained around 50 text files from my WhatsApp, including personal conversations with friends and family members, and some group chats.

For people who aren't interested in the code, you can enjoy the images. For others, I have also uploaded the entire code in my **GitHub** repository. I used **Google Colaboratory** for this project, in case you are using this or some other platform then change the path of the files accordingly.

## Loading the Messages

The messages in the text file are of the format - {Date}, {Time} — {Author}: {Message}

09/12/17, 10:20 pm - Sheet: Bhai sun na...

The plain text files will have to be converted in a meaningful manner to store in a Pandas data frame. I saved all my conversations in a data folder so I could list, load, and merge them into one dataframe. Now our dataframe is ready and looks something like this.

	index	date	am-pm	name	msg	msg_len	date1
0	0	2019-09-26 21:20:00	pm	sachin	Helo	4	2019-09-26
1	1	2019-09-26 21:20:00	pm	Diwakar Awaaz 1	I m 🤖	5	2019-09-26

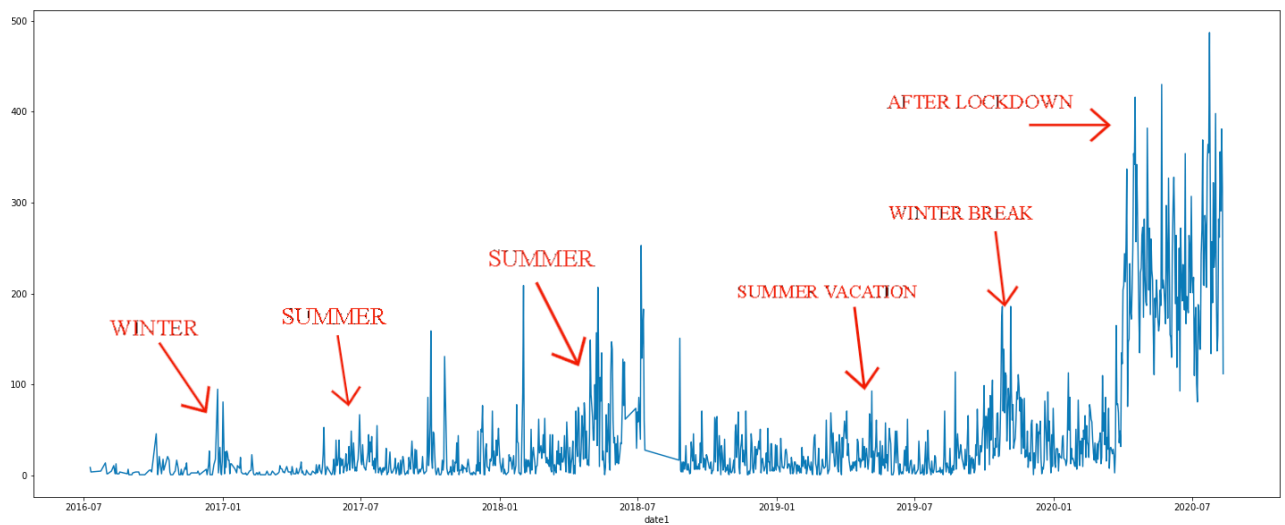
## Some Statistics

How many messages have I sent in the last four years? How many different people have I talked to within the past four years?

In my case, I sent over 58k messages, talked to over 350 different people. I also checked my AM-PM messages frequency PM-43820, AM-14185.

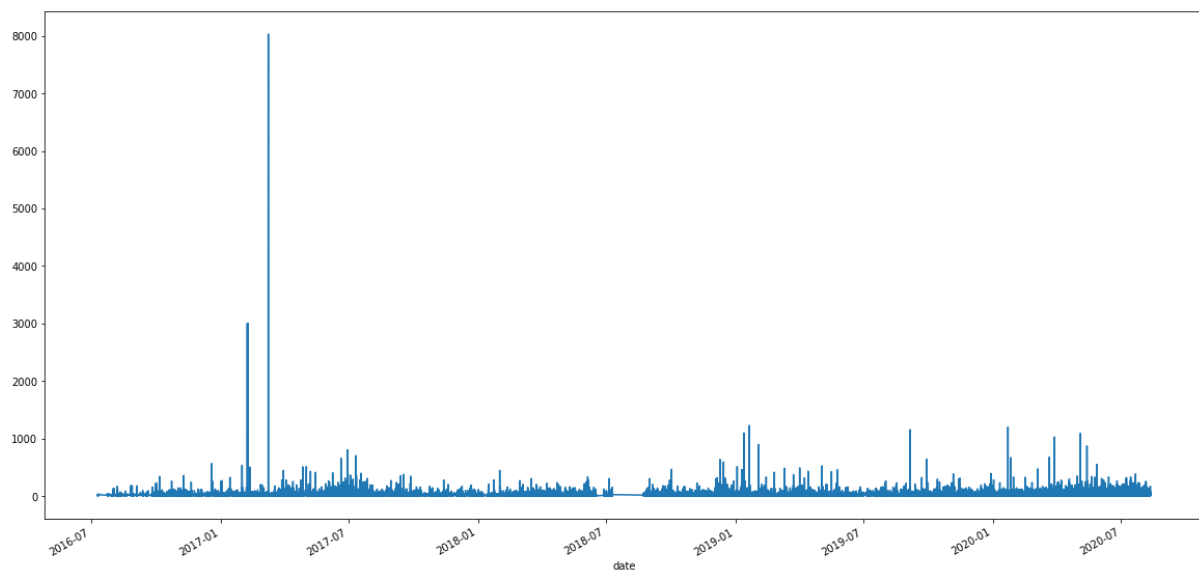
## Data Exploration

This is the most exciting part of this article - Data Exploration. Let's dig out all the fascinating stories that these data are trying to tell us.



This plot is pretty impressive as it can very quickly identify when I am on vacation at my home or in college. The effect of coronavirus on my texting pattern can be determined very quickly (I guess everyone is going through the same situation). Apart from this, some peaks in the plots over a few months (May to July and December) can be justified by Summer vacations and Winter breaks of my college.

I also find it funny to see the length of my messages over the years.

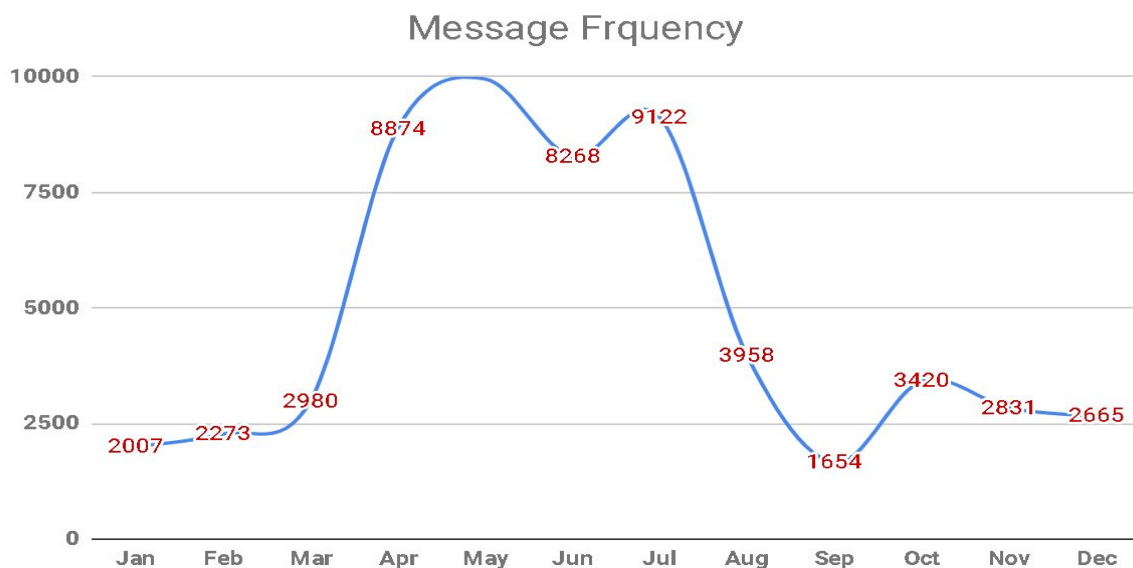


So I tried to find that outlier, and I got this -

```
array(['Google drive movies
https://drive.google.com/drive/u/0/mobile/folders/0B6FjKMQKynZILtlwZHI4ajUwcFU Programming
language collection on google drive -
https://drive.google.com/drive/folders/0ByWO0aO1eI_MN1BEd3VNRUZENkU Books -
```

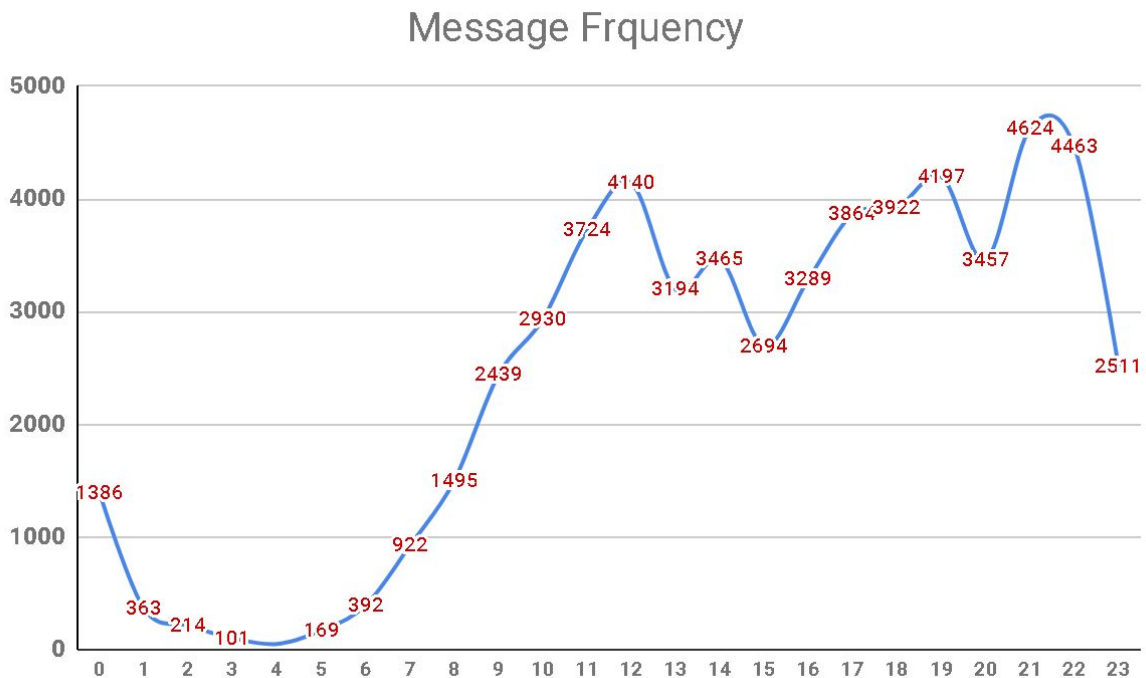
[https://drive.google.com/drive/folders/0B0hgUX3me1\\_RNi1KTXBzXzdXSzA](https://drive.google.com/drive/folders/0B0hgUX3me1_RNi1KTXBzXzdXSzA) Books for reading -  
<https://drive.google.com/drive/folders/0B09qtt10aqV1SGxRVXBWYmNIS2M> Books (novels) -  
<https://drive.google.com/drive/folders/0B1v9Iy1jH3FXdlNDeUNHNEVsZIE> Books -  
<https://drive.google.com/drive/folders/0B1Ef5shqGHDNRGEwd3BYb0N3Um8C> [Udemy] the complete digital marketing course -  
<https://drive.google.com/drive/folders/0B1qoi1IIEKwaM2tSMFBmOGUyNzg> Udemy course google drive link -  
<https://drive.google.com/drive/folders/0B1HQDi7EkA9XNIR3STF5SVJIVUk> Programming books google drive -  
<https://drive.google.com/drive/folders/0B2VhgL11XGQoamJCNGprcUJ2Zms> Books for reading...]

After this, I plotted my Message frequency over the months.



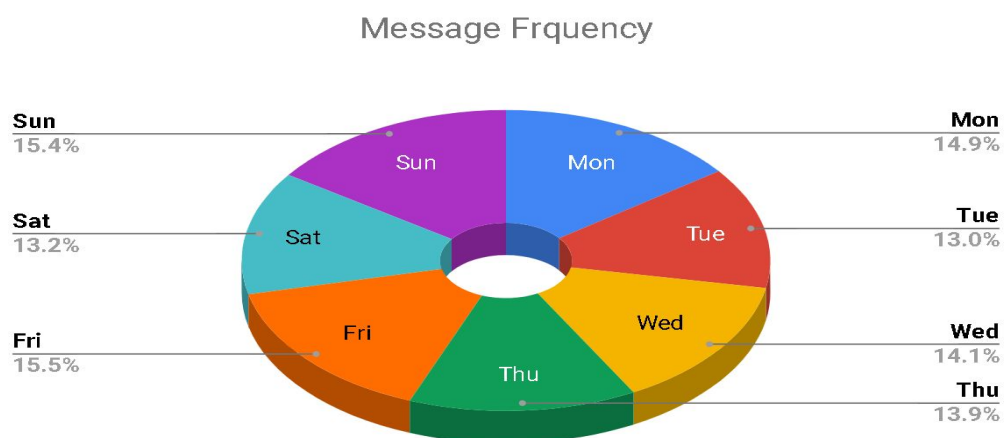
I was amazed to see this distribution. I guess this is also suggesting that I chat a lot in summers 😄.

Next plot is fascinating and tells about your sleeping time.



It suggests that I am not a Night-Owl and prefers to sleep after 11 or midnight. Also, the number of messages increases from morning to noon. It is contrary to the fact that during these hours one should work or study more 🤪.

Next, I also plotted the frequency over the days in a week. Well, it looks like it doesn't matter if it's Sunday or Monday, I chat almost in the same amount every day.

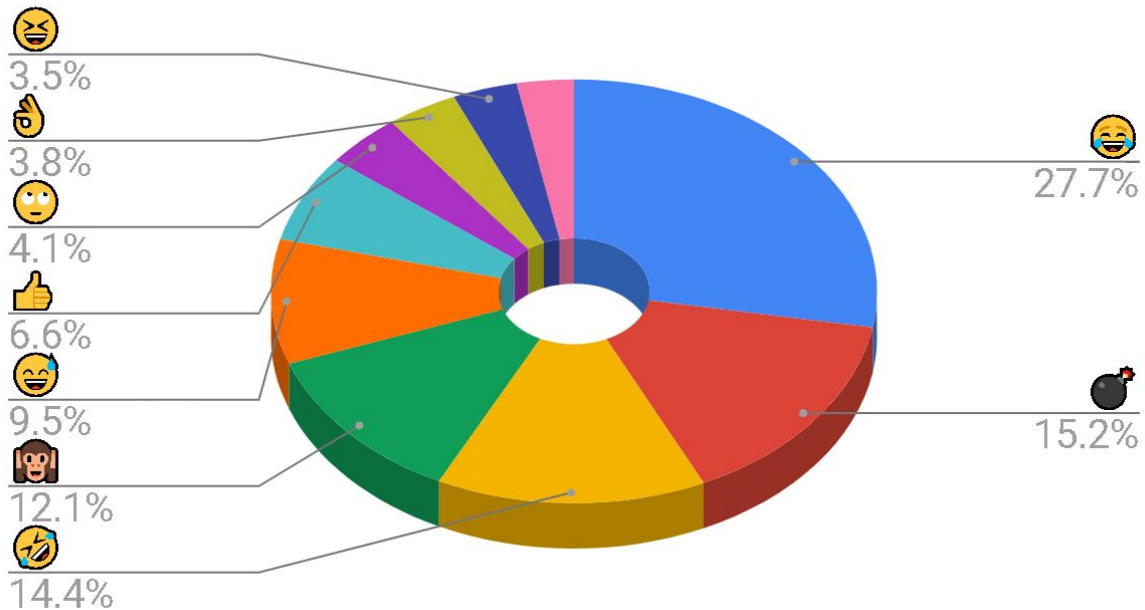


Now comes the most interesting plot of this article.

## Let's Plot Emojis

The complete code for emoji is in my GitHub. I used Emoji library by Python to plot this. You can install this library using this line - `pip install emoji --upgrade`

Emoji Count



Well, let's first talk about the second most frequent emoji used by me 💣. I generally use it while quoting some dialogues from the movie "Gangs of Wasseypur" or sometimes for no reason. But I never realised that I have become obsessed with this emoji so much 💣. Also, I was glad to see that evidently, my life is full of laughter, surprises and bombs!

## Conclusion

It looks like this analysis while answering some questions has opened up a lot of new problems which can be further solved. Stay tuned for the next part of this article where I will try to use the concepts of NLP like Topic Modelling.

Did you find any of these insights useful? Or do you have suggestions about some valuable insights that I missed? Feel free to add your comments below.