

Homework #5. Exploratory Data Analysis

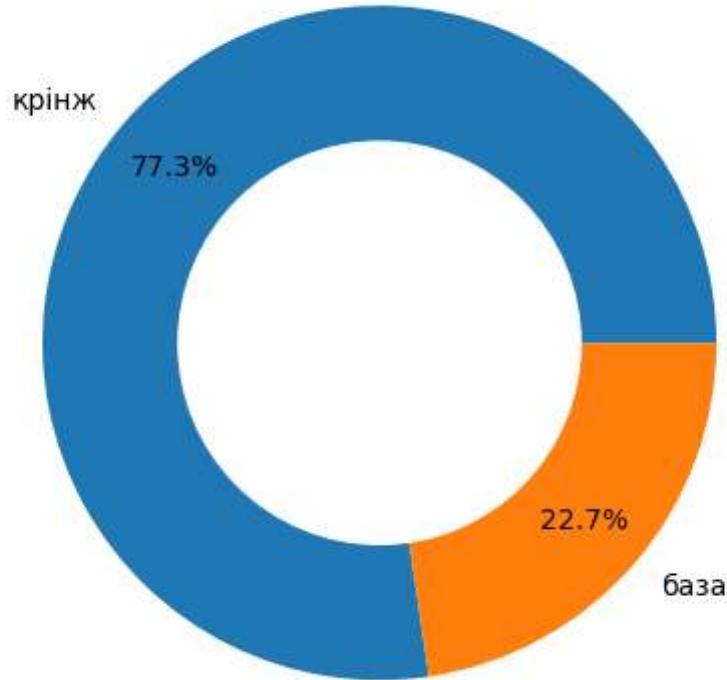
Author: Danylo Ustymenko

Total time spent on h/w (in minutes): 1500

My messages analysis

1) My based/cringe usage

Based/cringe chart

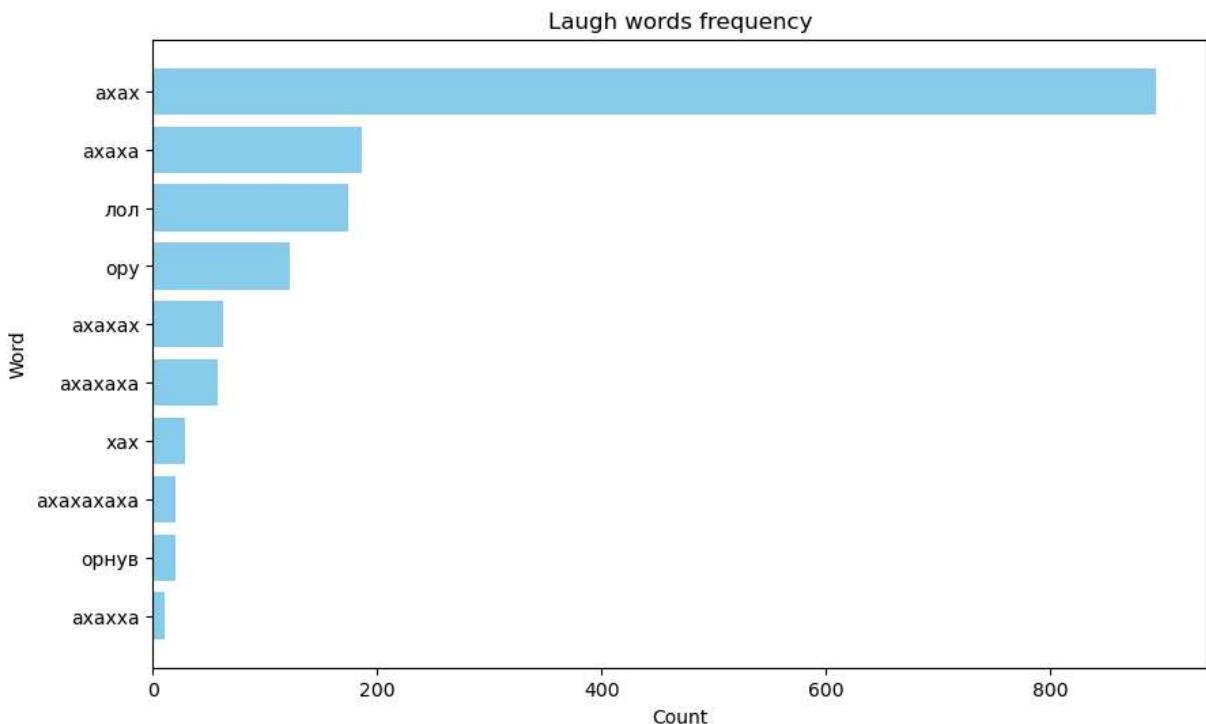


Seems like I have more cringe in my life (or maybe because I use "based" gifs instead)

2) Top 10 ways to laugh

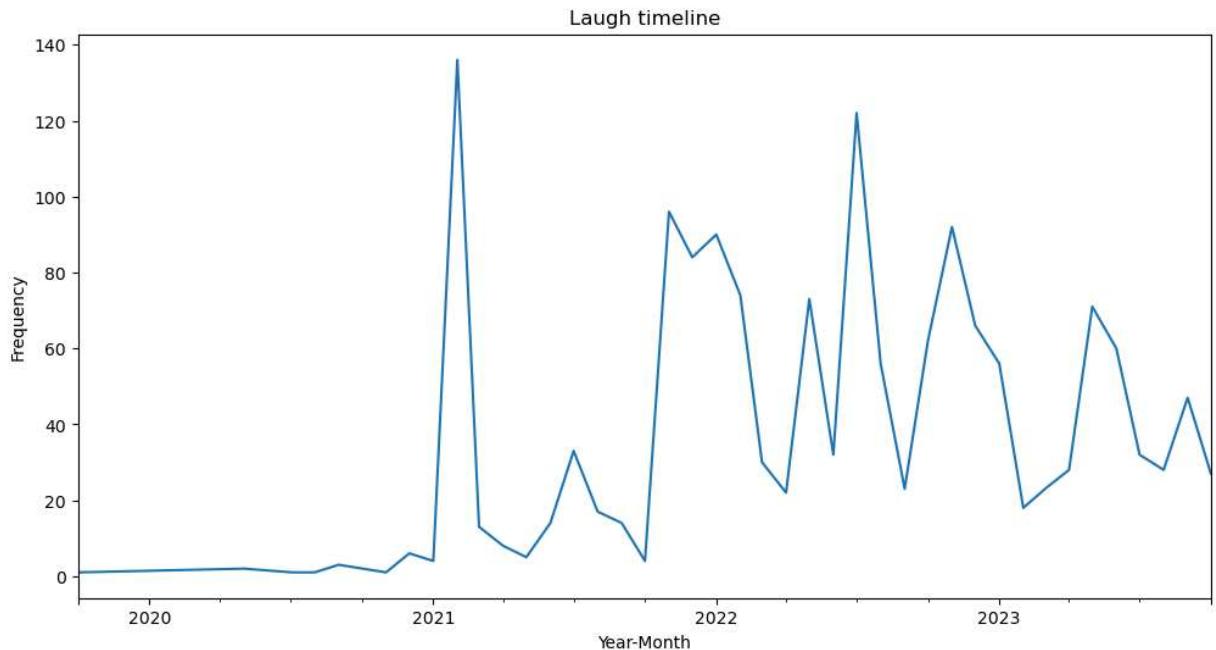
Out[5]:

	Word	Count
550	ахах	2141
3508	лол	1757
1421	ору	1738
4922	ахахах	1373
562	ахаха	1098
584	ахахаха	788
1920	хай	688
55	ахахахах	674
17028	ахахахх	417
7680	оп	389
4098	ахахха	326
378	ахахахаха	293
13983	ахахахахах	278
16556	хахахаха	258
8061	хахаха	252
12237	пхд	246
5709	ахаххаха	243
20669	хату	207
4520	хах	205
5971	ахахахха	205



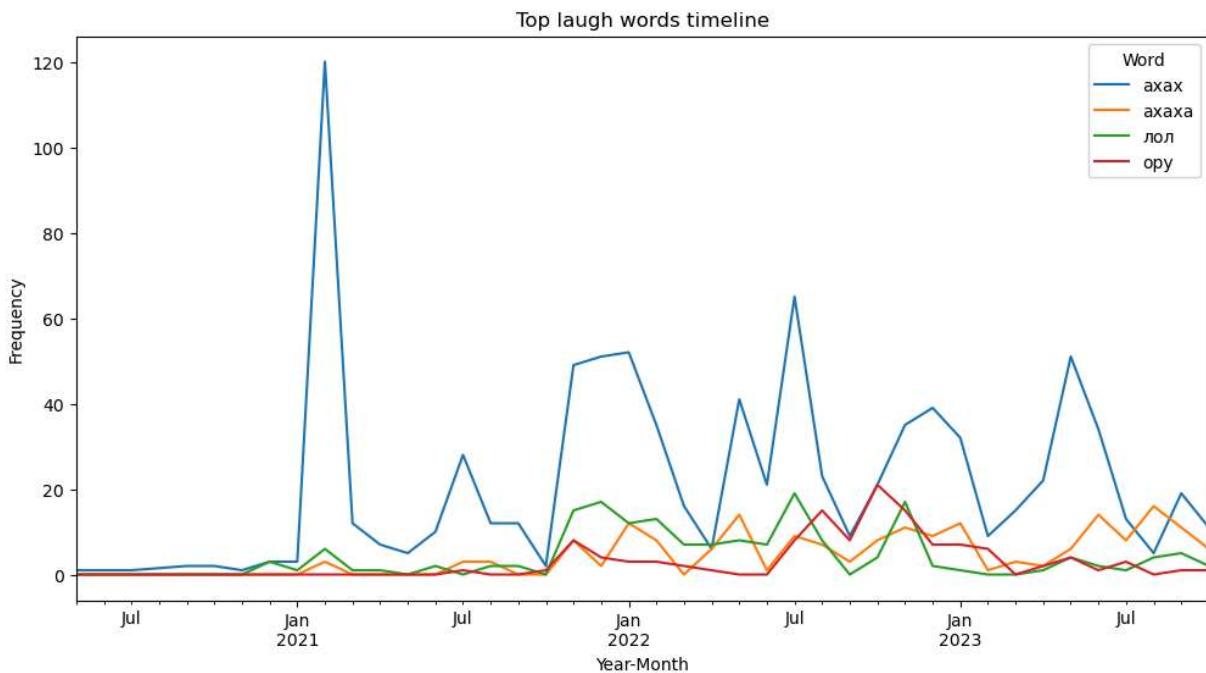
As shown the most popular laugh is "axax", so if I react this way it's usually something not very funny

3) Laugh frequency over the last 4 years



The 2020 year was sad based on tg data, in 2021 I rapidly laugh more, and now I'm laughing on average 50 times a month

4) Laugh behaviour changes over time



So "axax" was always very popular, "axaxa" is growing popularity. "opy" was popular mostly in 2022 and is rarely used now.

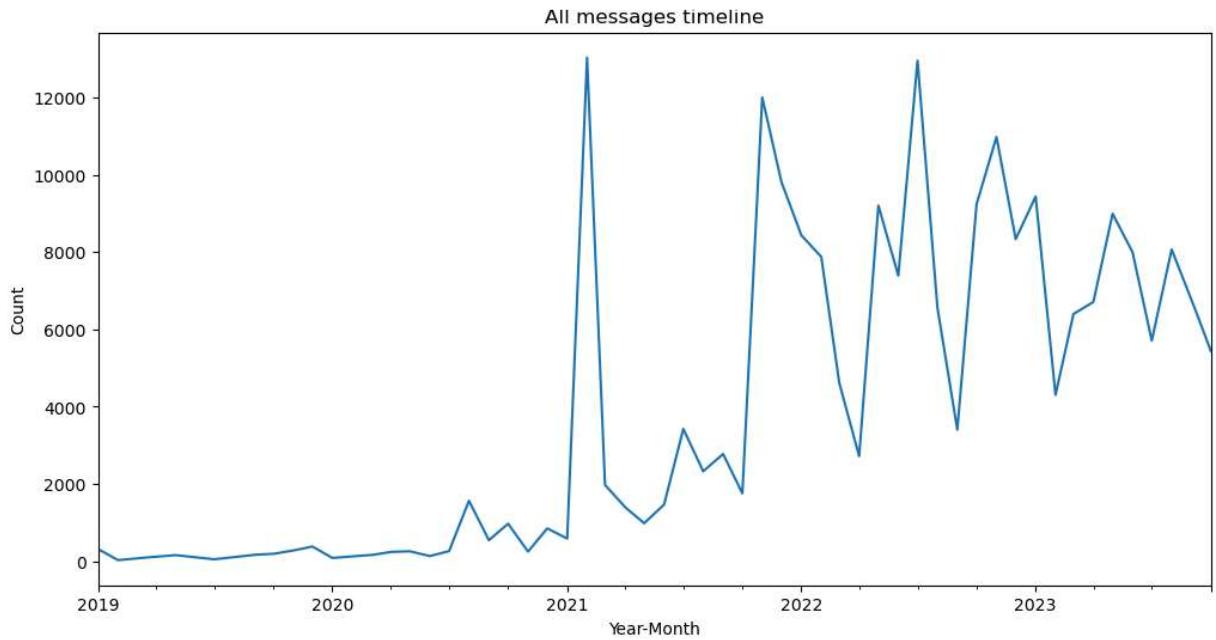
5) My word cloud



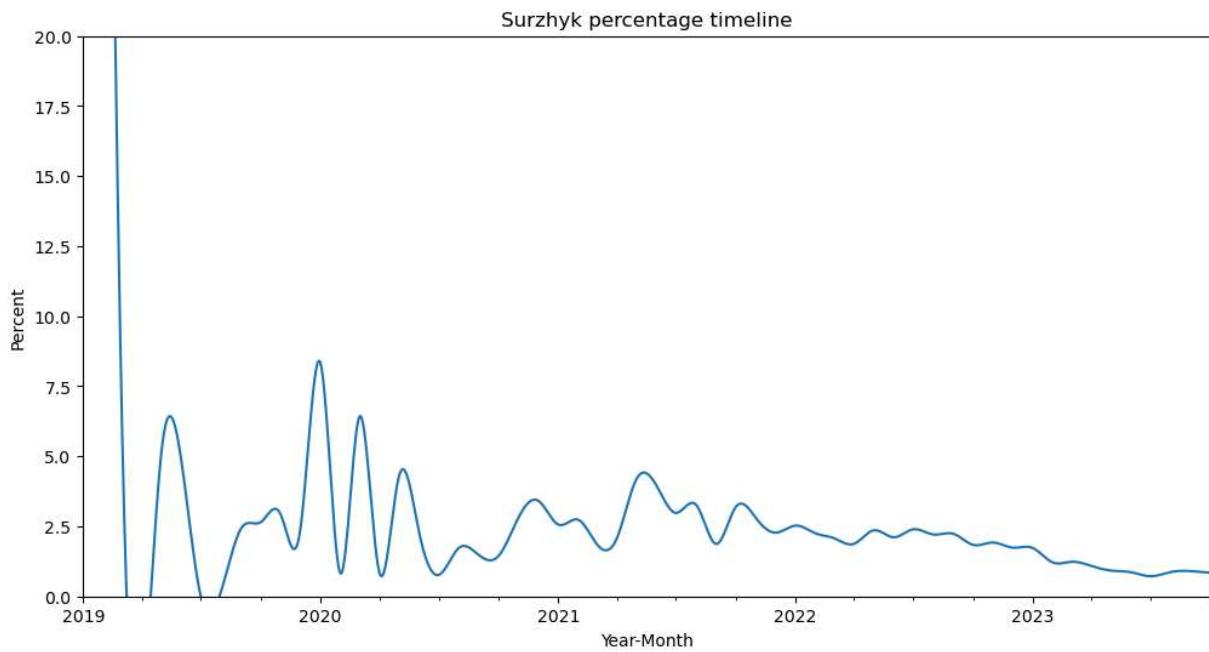
Here I can see some of my parasite words like "типу", "треш", "взагалі", "нуу", "крінж". Another interesting

fact is that I moved to Toronto, so there is a word for it.

6) Number of all my messages over time

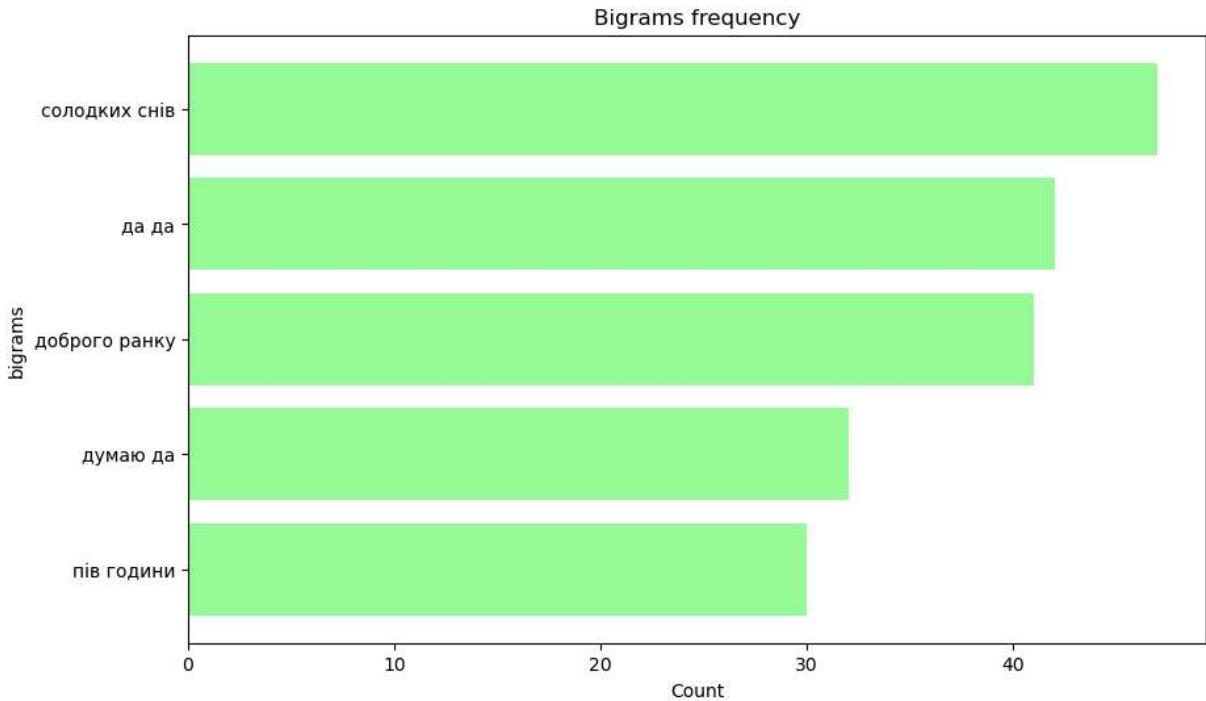


7) Usage of surzhyk timeline



It seems that there is a positive tendency of reducing surzhyk

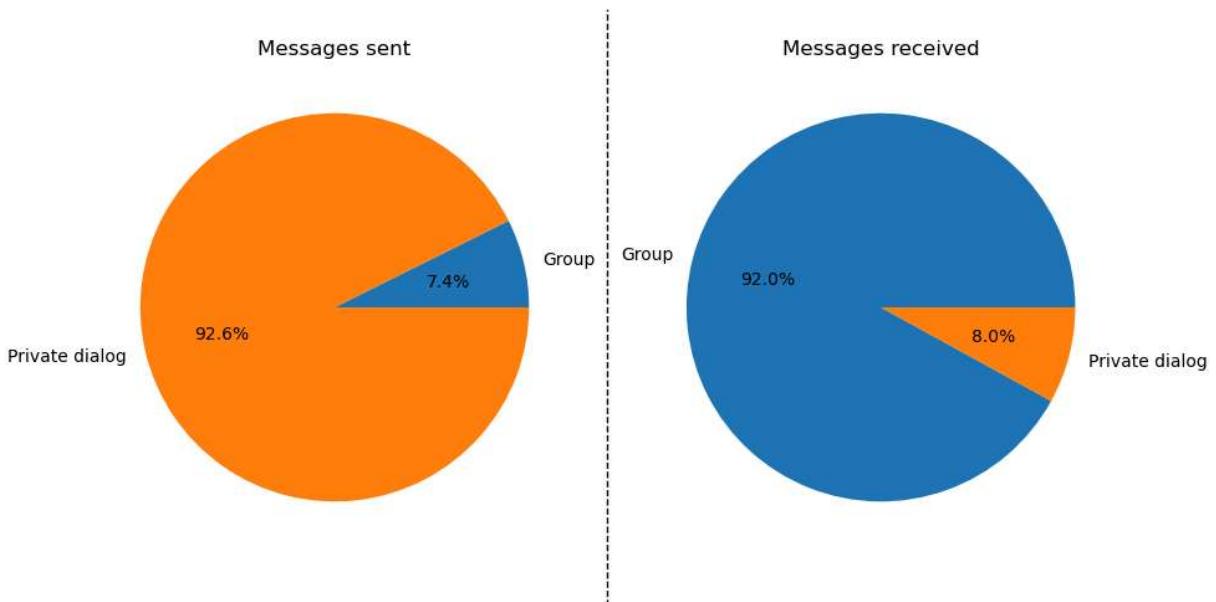
8) My most popular bigrams



"солодких снів" and "доброго ранку" are the most popular phrases

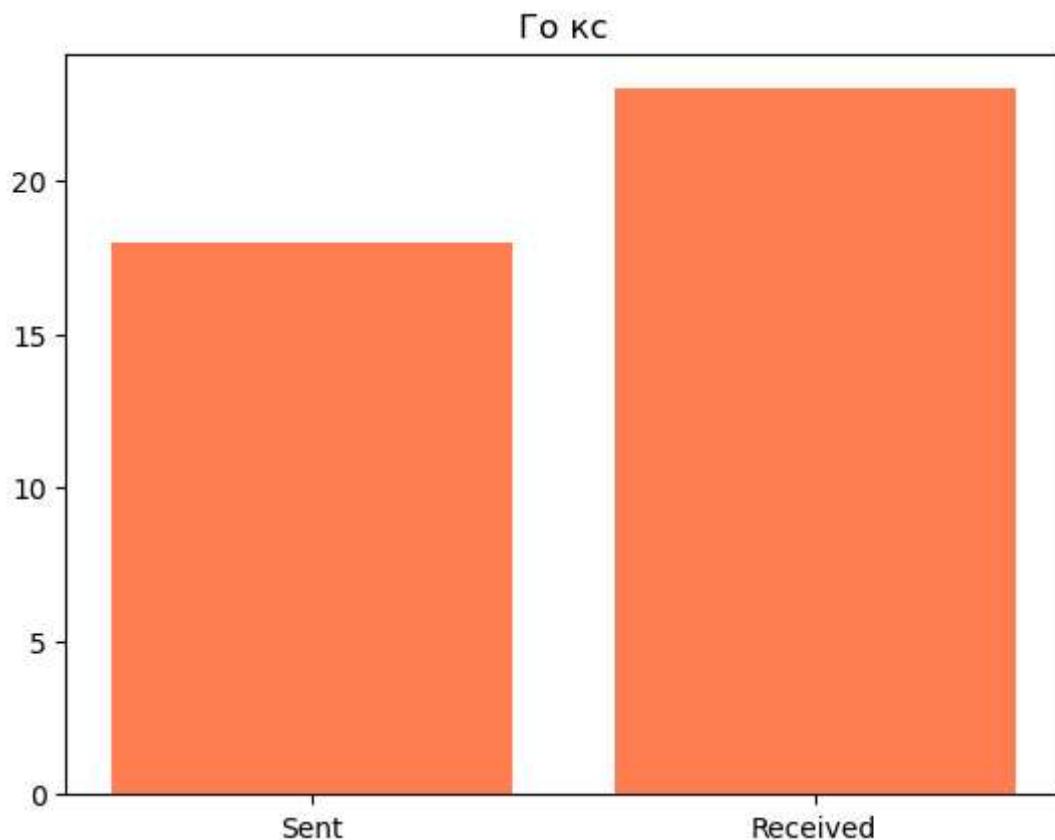
My behaviour in chats

9) Comparison of sent/recieved messages in private dialogs and groups



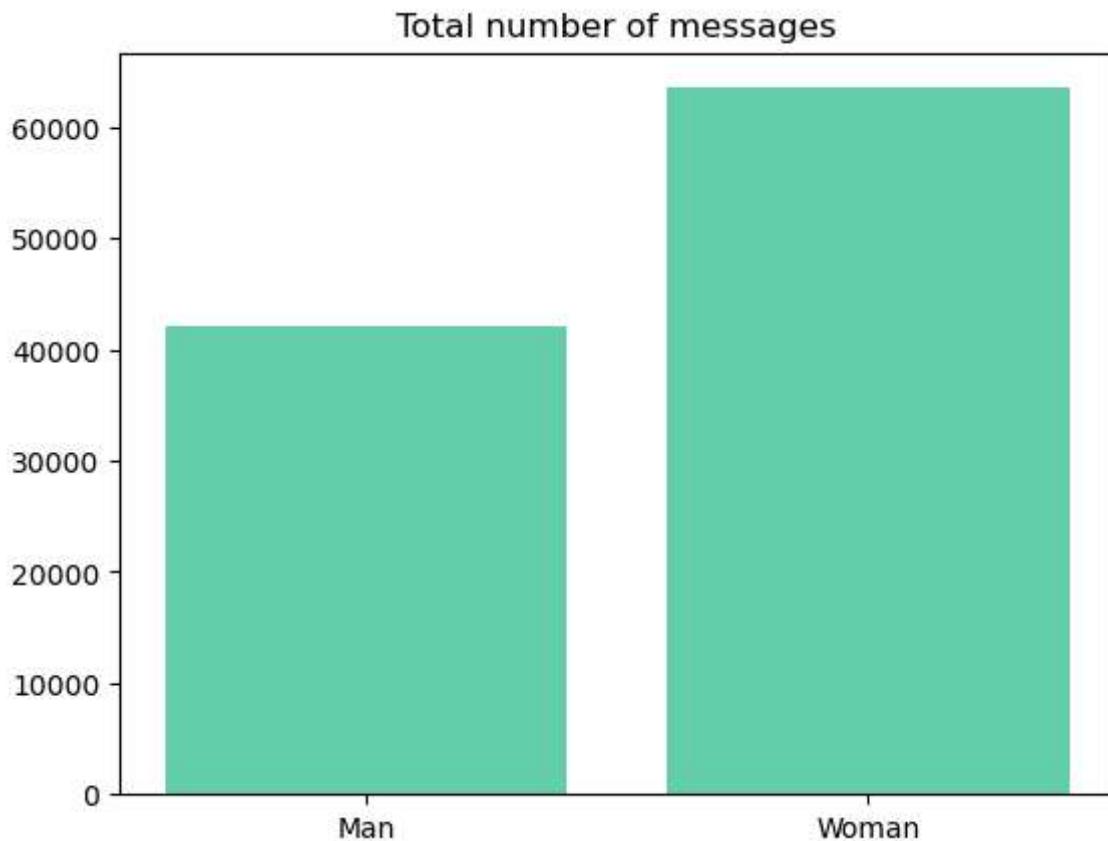
I like to chat in private dialogs more, but I'm a part of many groups

10) Sent/received invites to play cs go



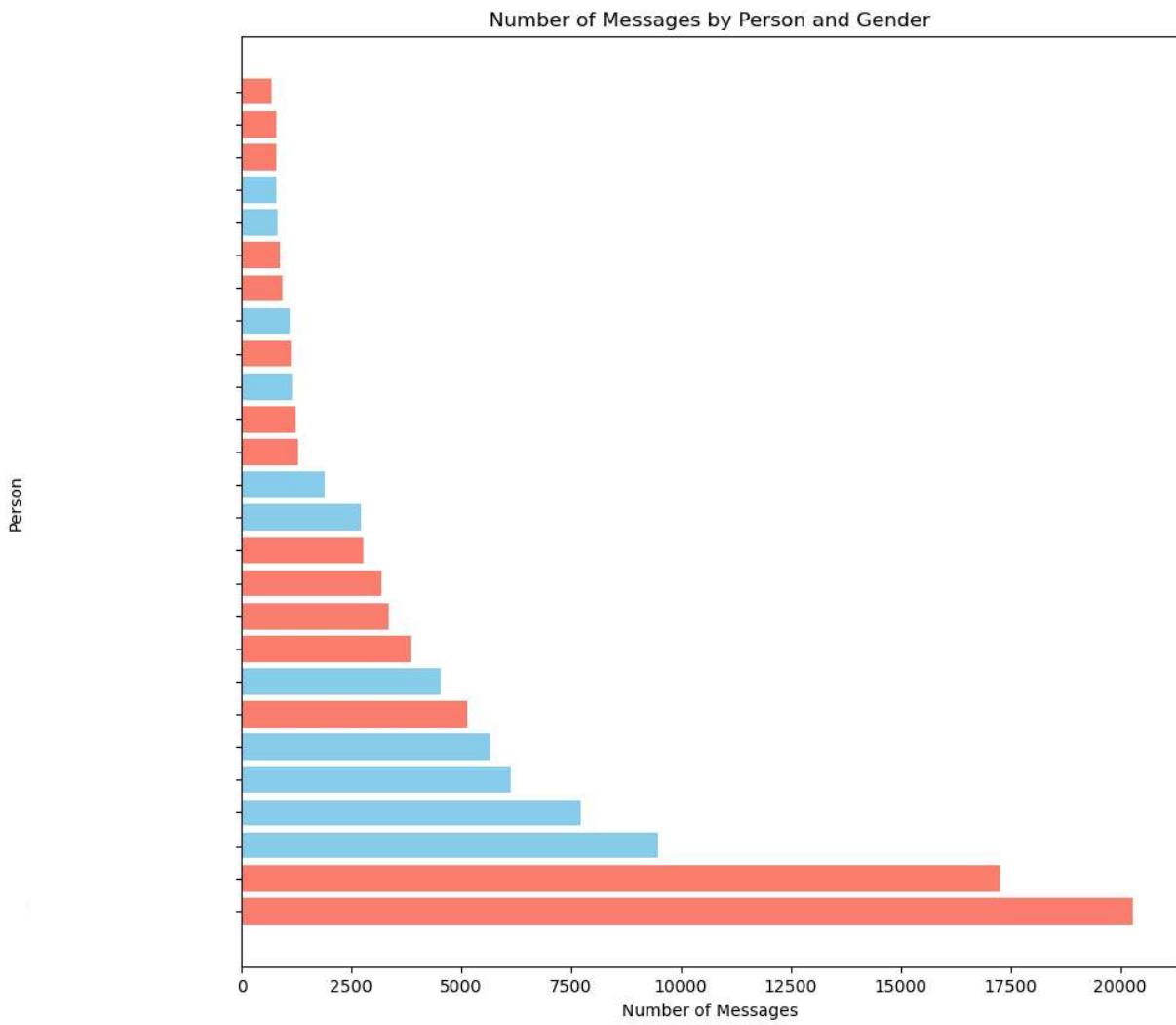
Looks like I was a valuable player, and people wanted to play with me

11) Total messages in dialogs by gender



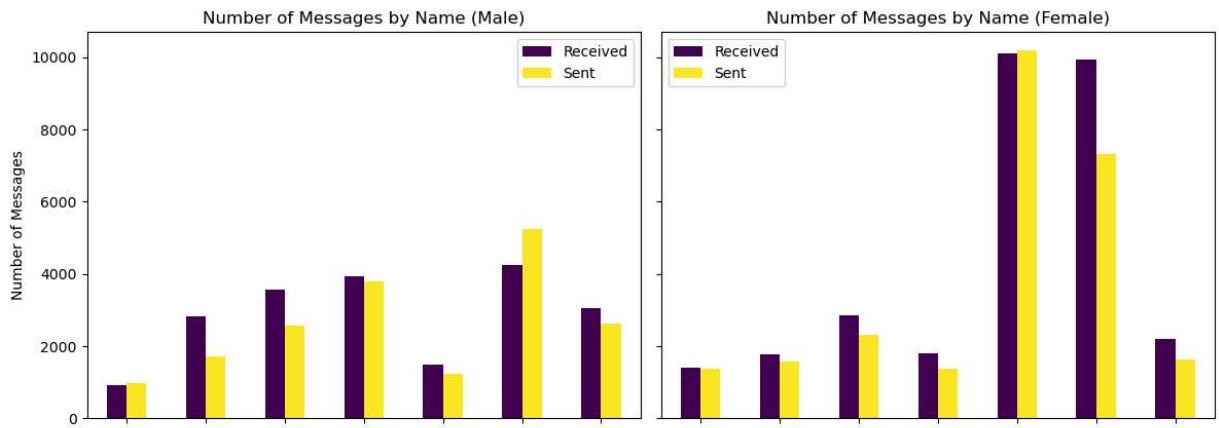
I chatted with girls in Telegram more. I think the reason is that I often chat with my boys in Discord

12) Top dialogs number of messages with gender data



Top two are my ex-girlfriends, but other top 4 friends that I chat with are boys

13) Comparison of send/received with my girl and boys dialogs



It looks pretty balanced. In most cases I have more received messages than sent for both genders.

14) Wordclouds for my top dialogs



Word Cloud for



Word Cloud for



Word Cloud for



Word Cloud for

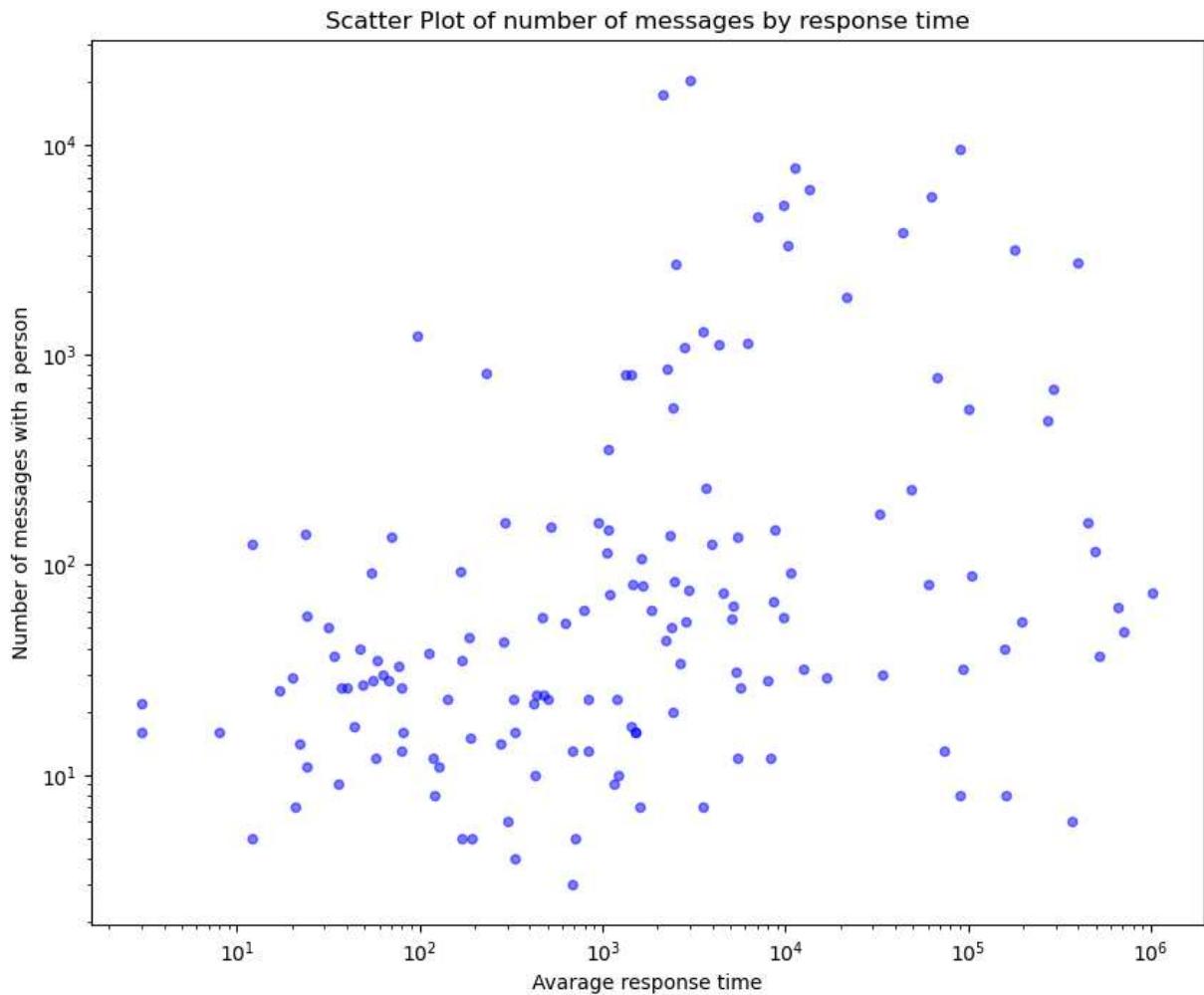


Word Cloud for



There are some words that correspond to common interests like: "бібліо", "аніме", "навчання", "гурто", "клуб"

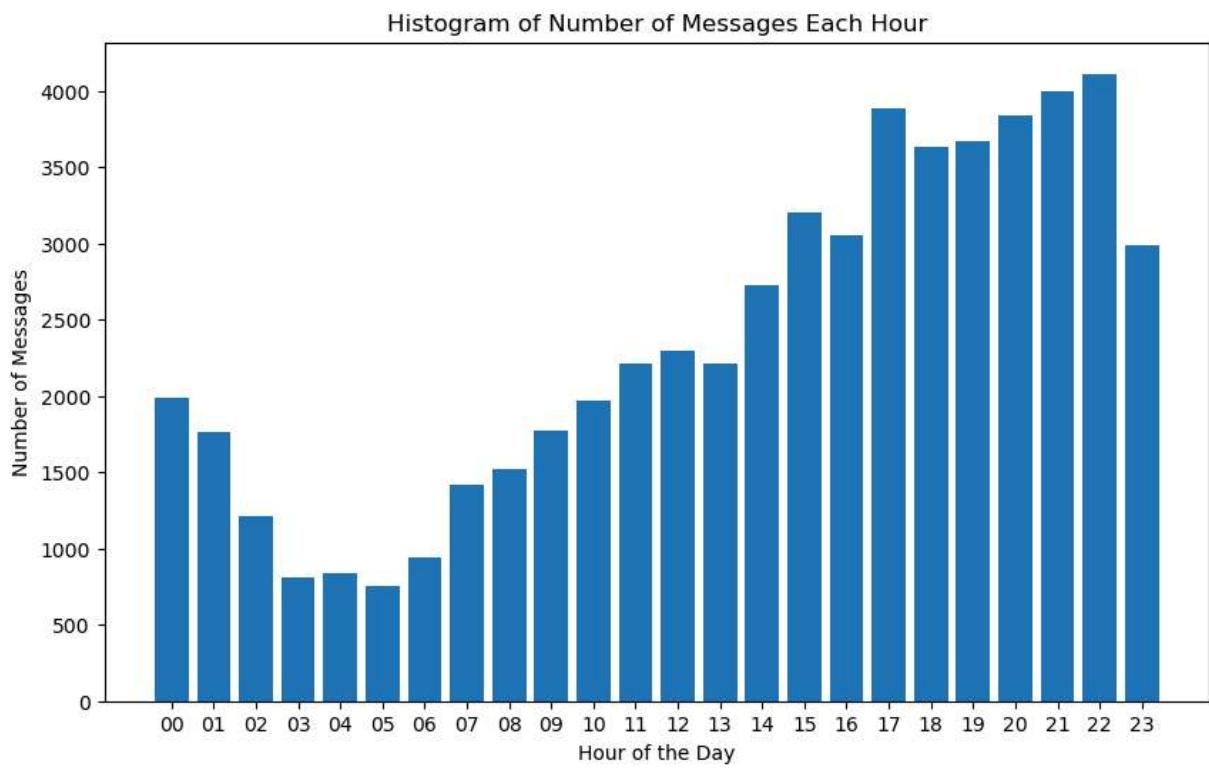
15) My response time after greeting correlation with number of messages for the person



I expected to see that I respond to my close friends faster, but it looks like it's pretty random. I can even see the small tendency to answer to random people a bit faster.

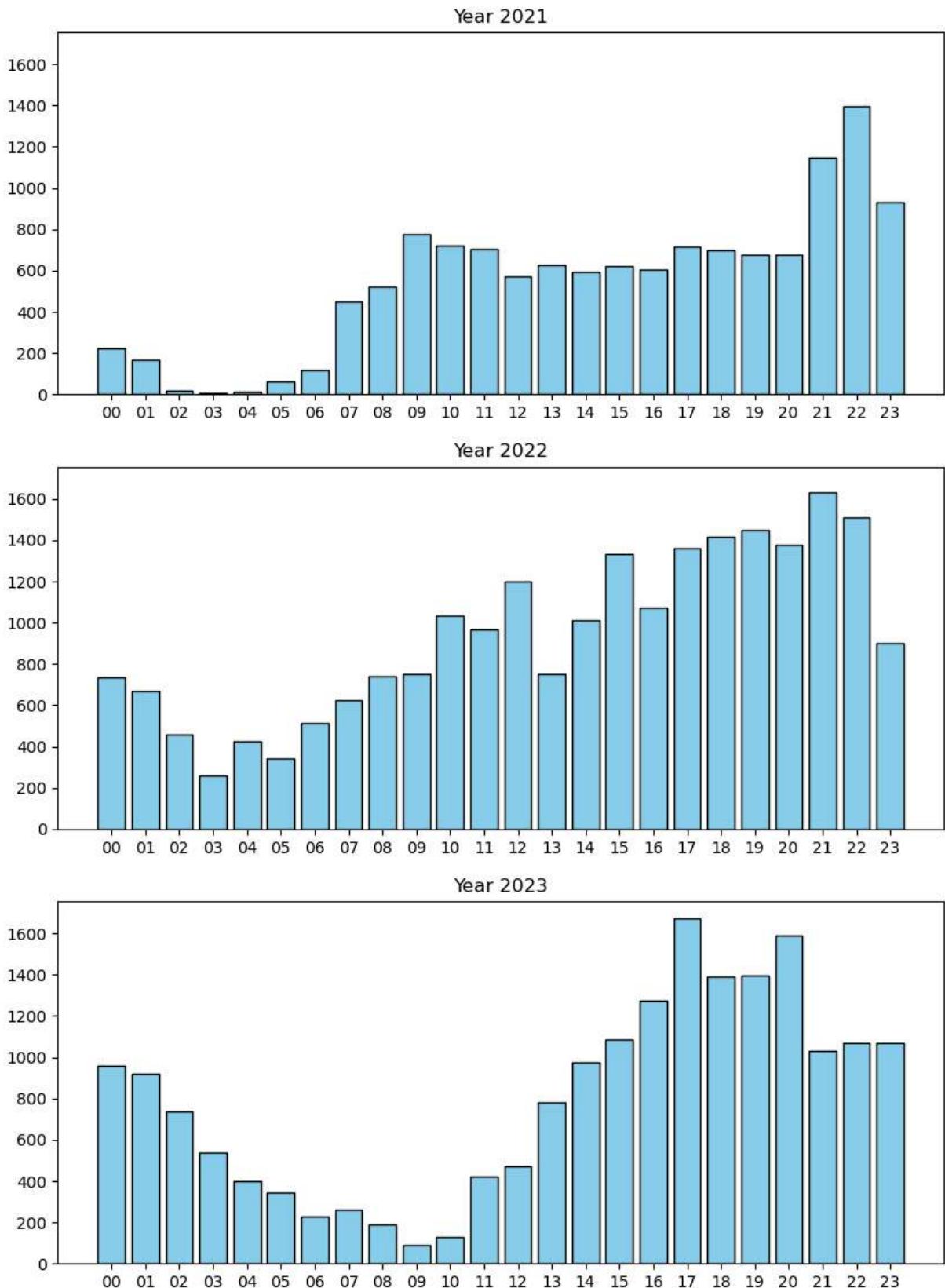
My activity during the day

16) Activity during the day by hours



There are some most active hours. It looks like I chat even during the night

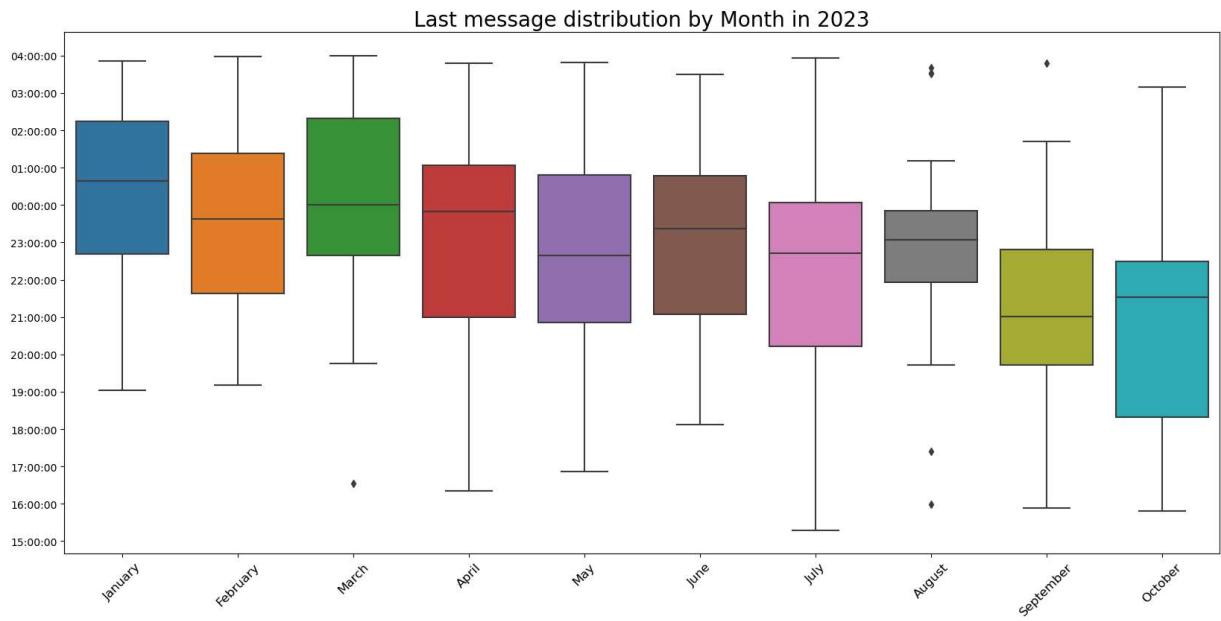
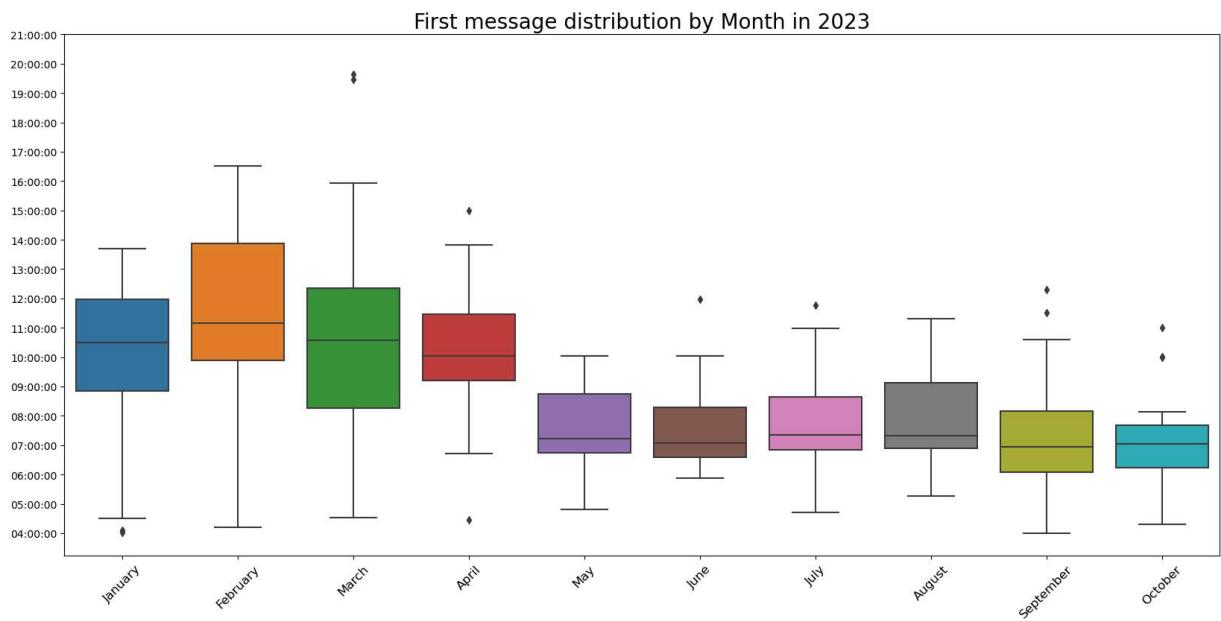
17) Activity during the day each year



Now the activity makes more sense. There are most active hours before the night, and inactive hours in the night. There is a shift of active hours in 2023, which

corresponds to changing the timezone, since I moved to Canada in 2022

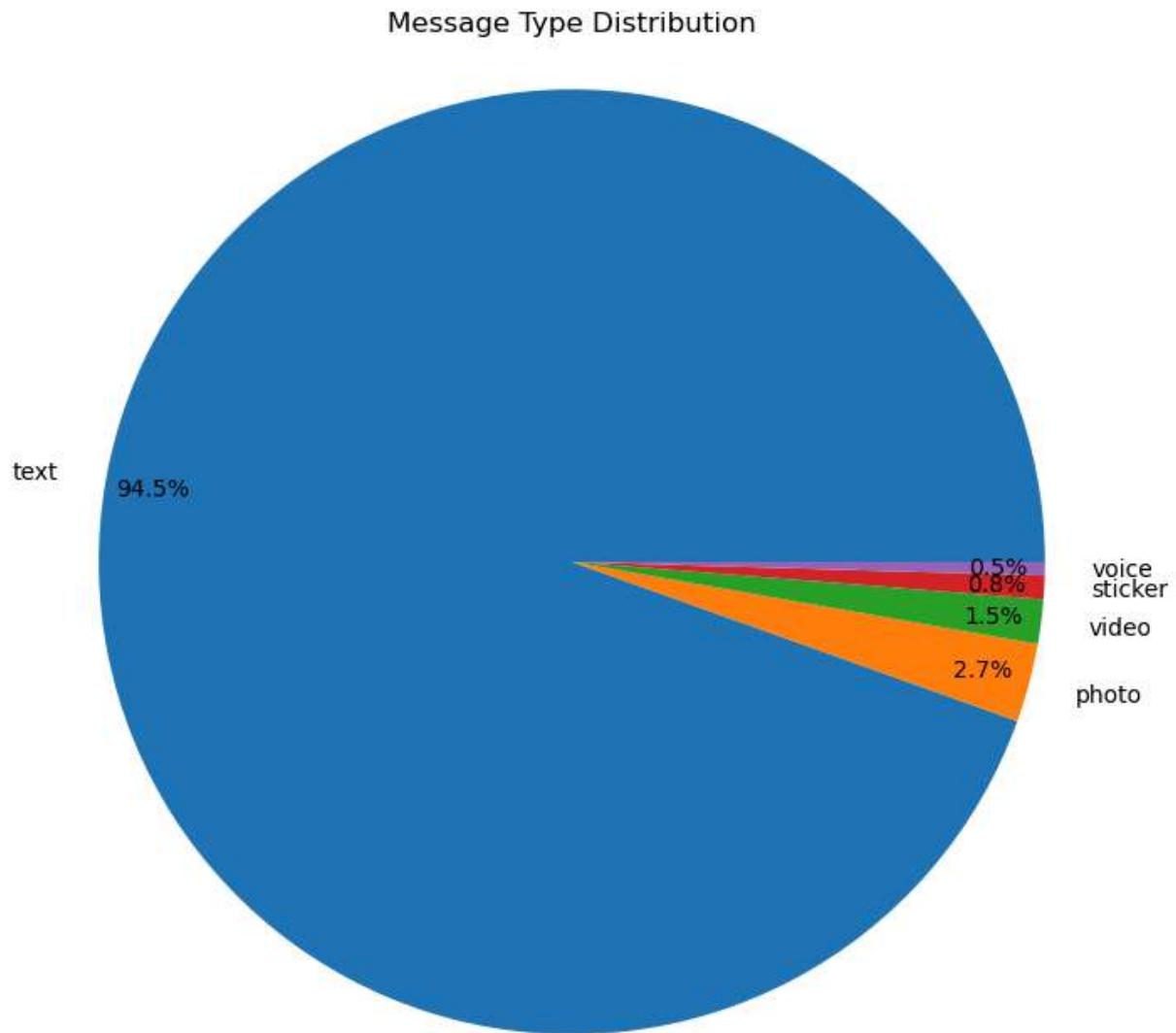
18) First/last message of the day in 2023



The distribution of first message is narrow when compared to last message. Since I go to the university or work at fixed time. In winter I gone to bed later, but now my sleep hours are more consistent.

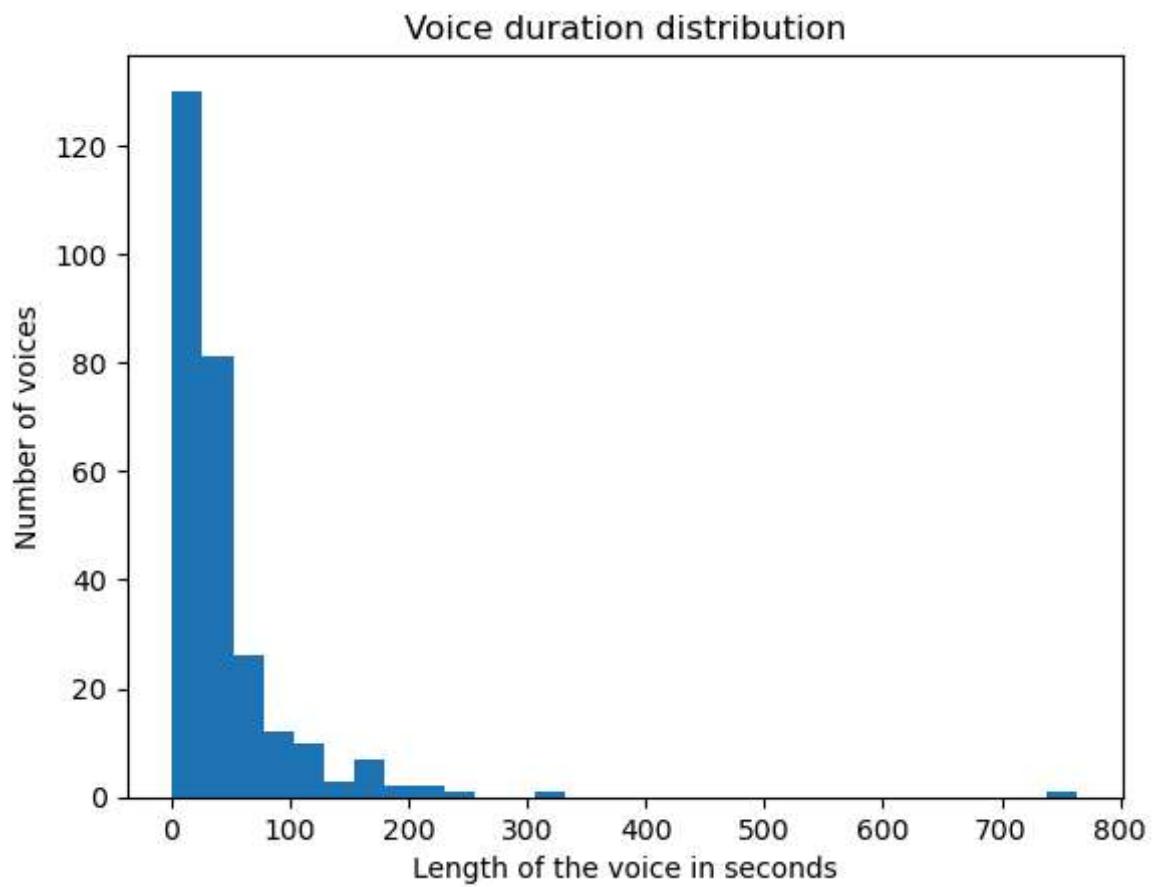
My voices

19) All the messages distribution by type



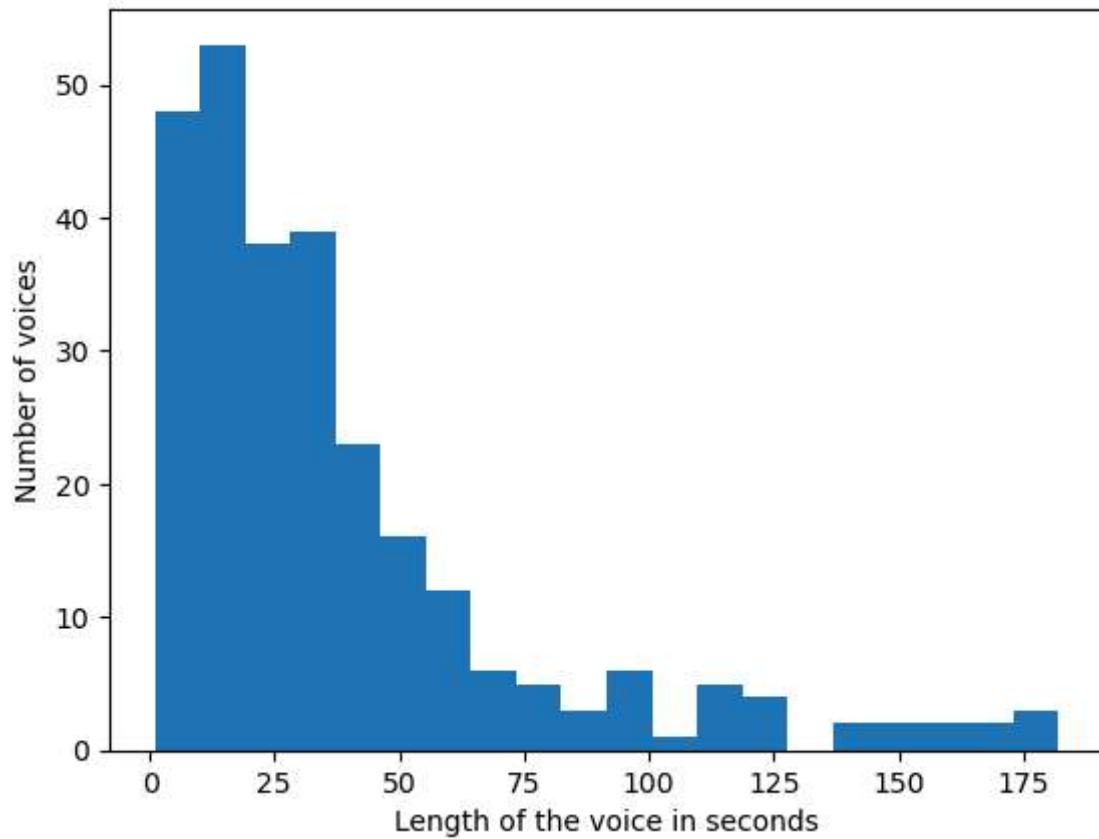
Text is obviously the biggest part of all messages. Voice is only 0.5% of all. I don't like sending and receiving them

20) The distribution of the length of voice messages



With outliers excluded

Voice duration distribution



We see exponential distribution. I send short voices more often than long ones