Bohdan Hurskyi

course: Computational Social Science
teacher: Andrew Kurochkin
20.05.2022

Plan:

- Intro
- How to get data?
- Data analysis
- What to do next?
- Conclusion

How to get data?

We've got two major csv files:
dialogs_users_all.csv
dialogs_data_all.csv

STATS

number of msgs: **1 592 043**
number of users: **7 867**
combined size: **331 mb**
time spent to download: **7.5 hours**

3

dialogs_users_all.csv contents

My daily activity by hours

OBSERVATION

I have peak activity around **9pm**. It's time when I usually end up all my work/study activity and am able to check some messages.

Obviously, the lowest activity is between **1-4am**. We can see the strong rise around **5-6am**, usually I schedule my day in my "saved messages".
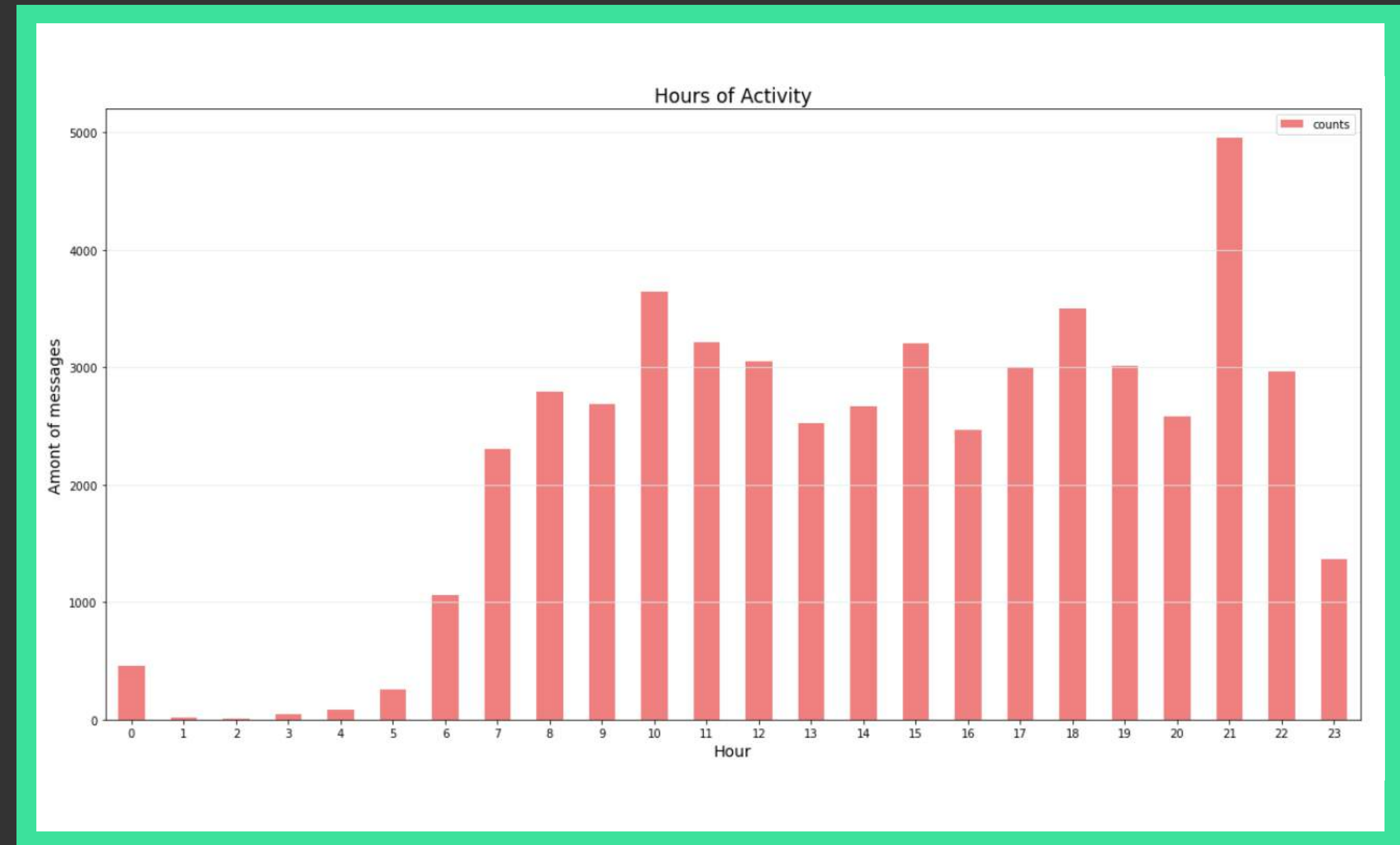
My favorite sticker?

OBSERVATION

This plot is pretty simple.
Just my favorite stickers.

My most used stickers

4
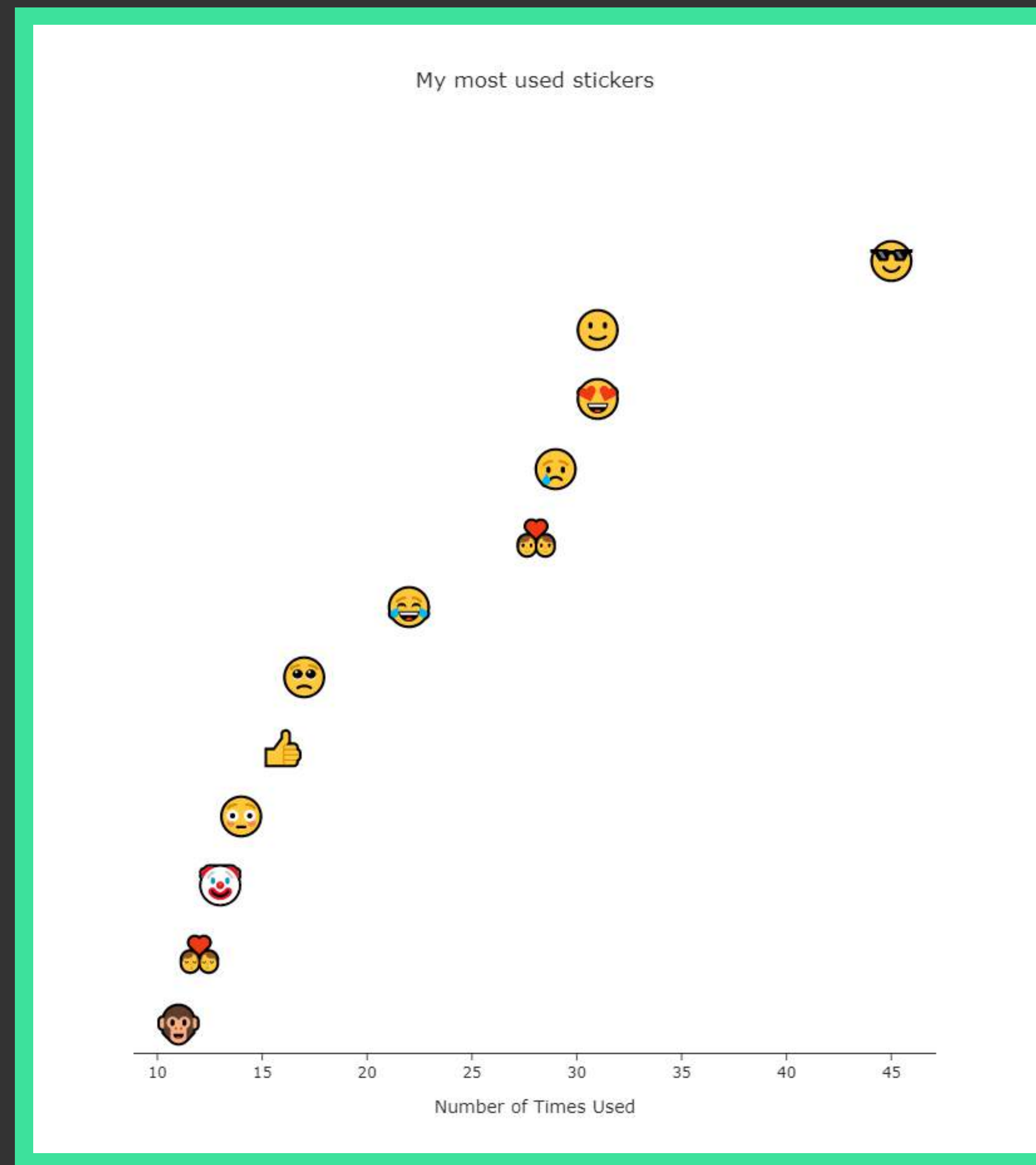
Correlation of sent/received

OBSERVATION

Here I've noticed an interesting regularity that almost all my female friends receive more messages from me than they send me.

Let's calculate the rate of **sent/received**.

6
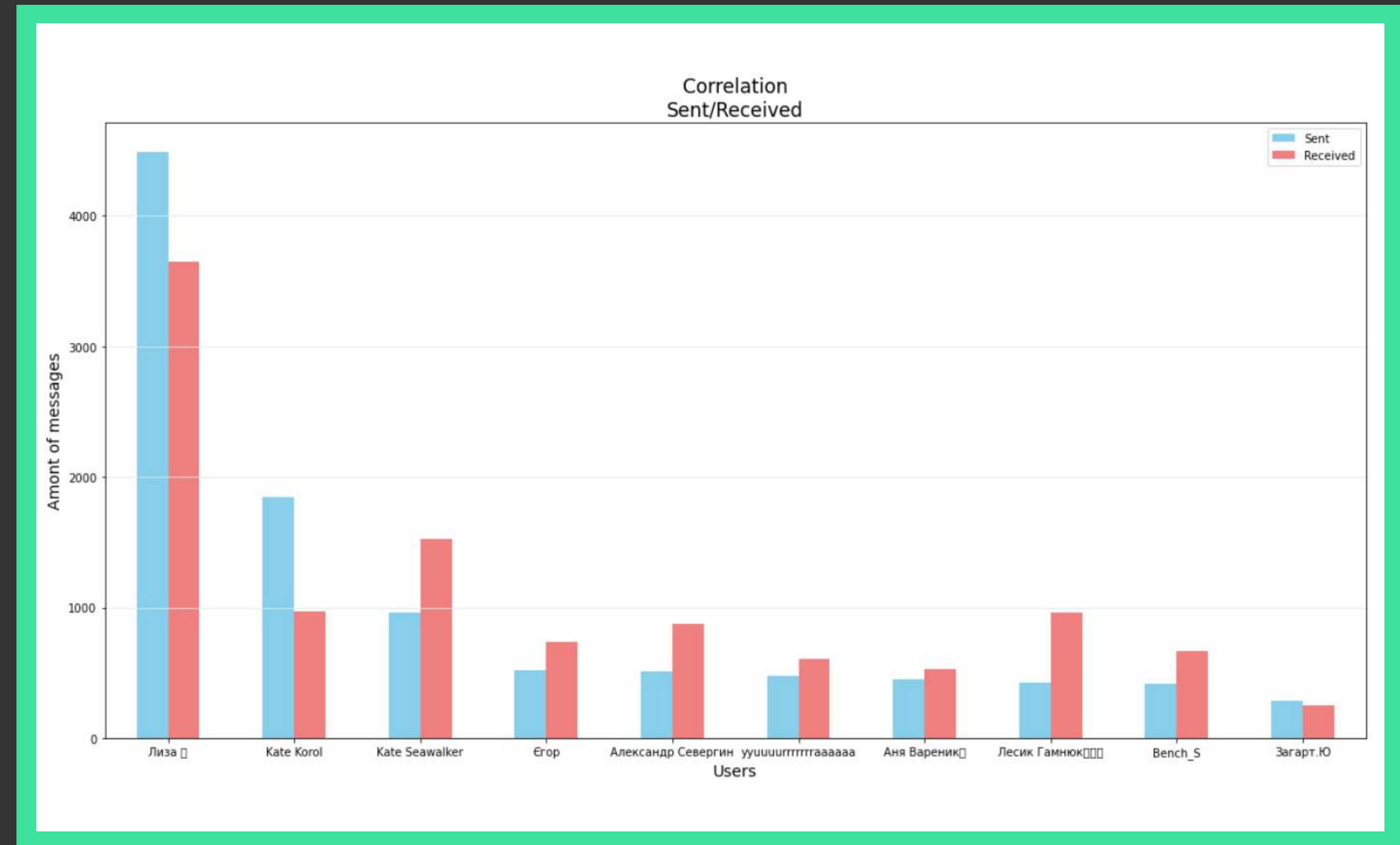
Correlation of sent/received

OBSERVATION

Among 10 of my biggest personal chats we can see very different pictures. User from 1st place receives **2** messages for **1**, while the 10th receives **0.5** messages for **1**.

So let's count the difference between my male and female collocutors. I call it **simp rate** 😎

7
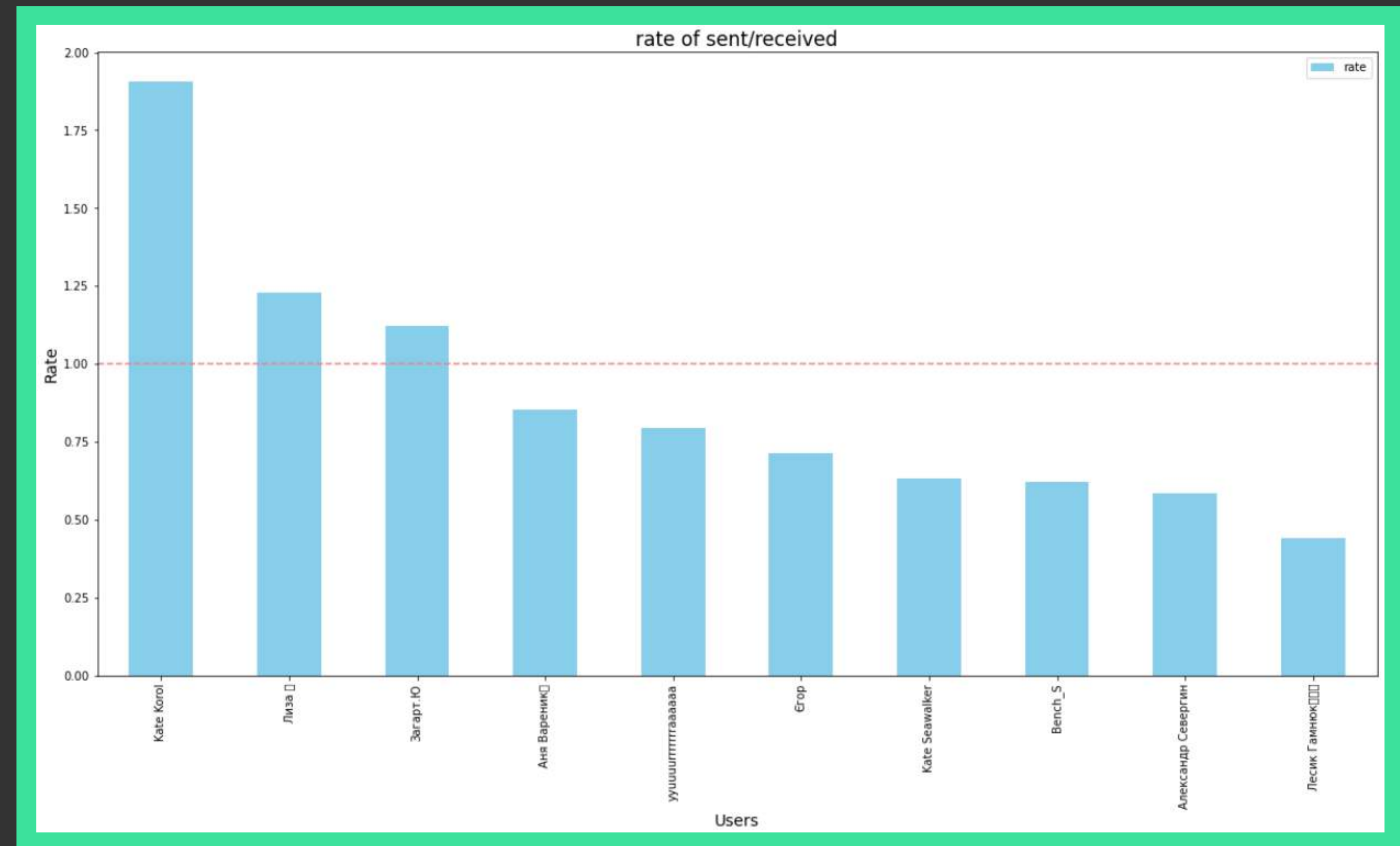
Correlation of sent/received

OBSERVATION

As result, girls usally receive from me more messages than they send, while boys receive less((((

So, we've found that I'm
a huge simp 😎

How sex effects the sent/received rate
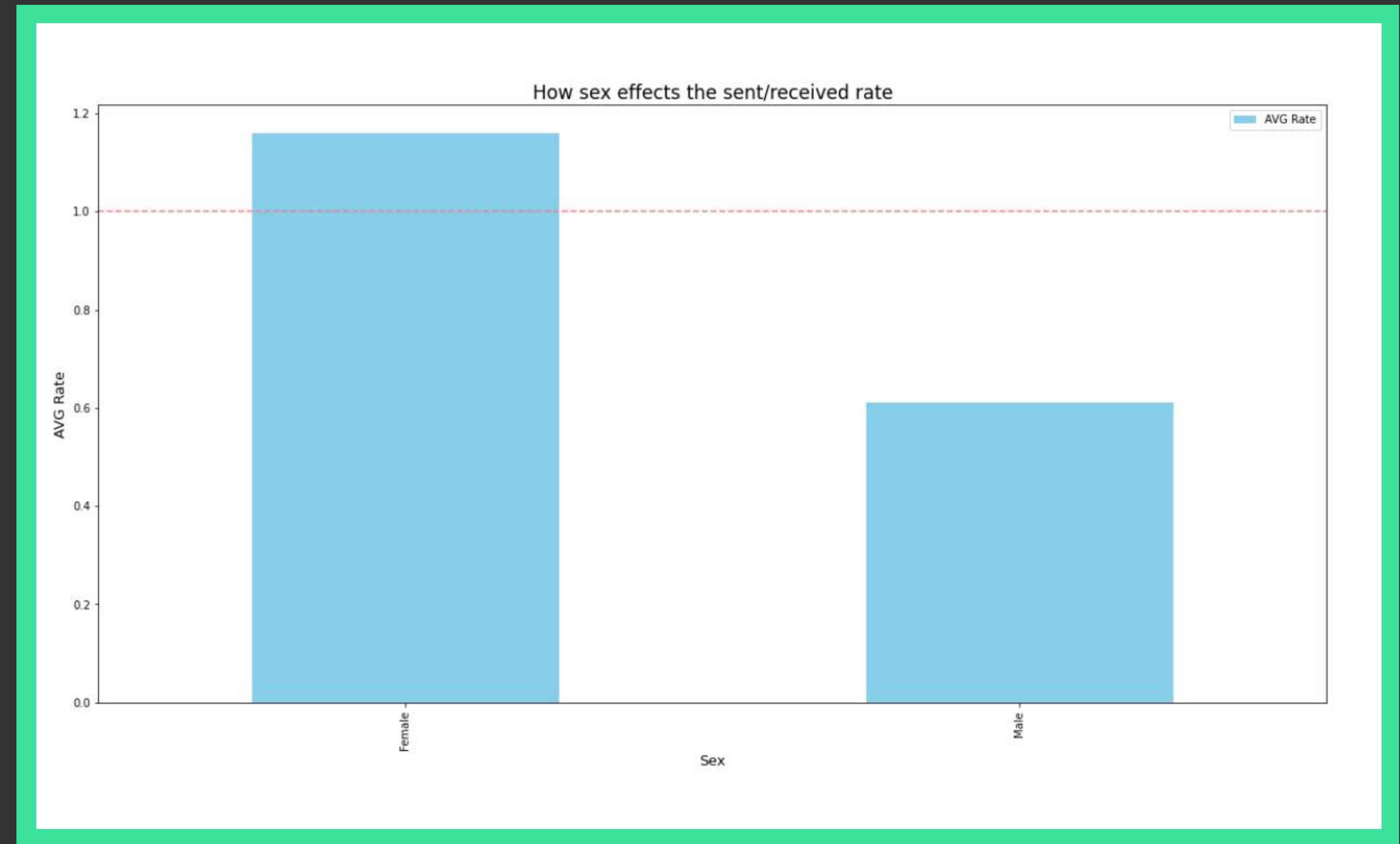
What messages do I save?

OBSERVATION

Here I compare all my messages and my **"saved messages"**

As we see, text messages are **95%** of all. While in **"saved"** they're only **23%**. Most of those texts are some reminders or day plans, because I use **"saved"** as a notebook.

9
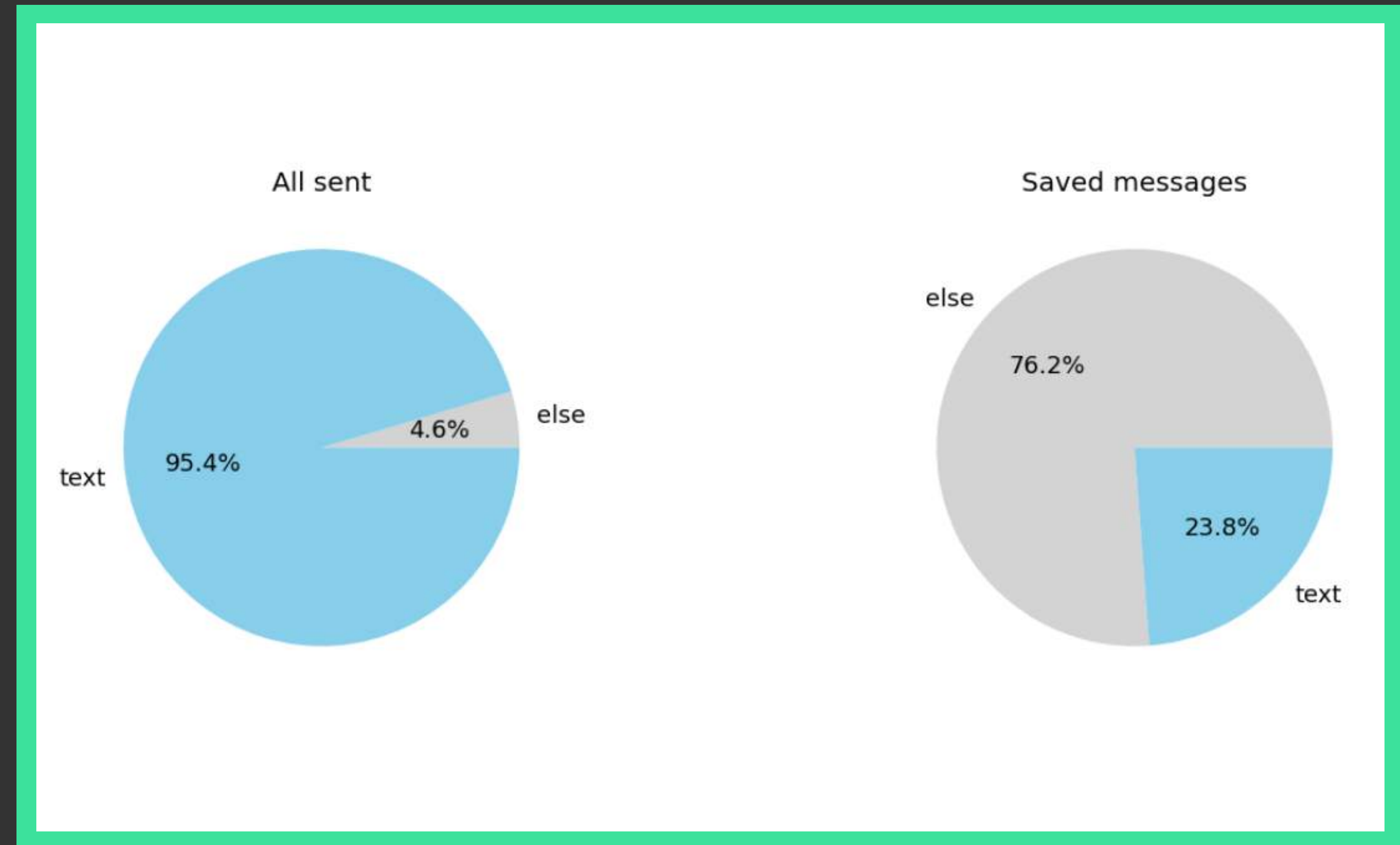
What messages do I save?

OBSERVATION

Lets get deeper and divide that "**else**" sector. So we receive the media distribution.

Firstly, I don't like listening to my voice, so there's **0** voice messages in my "**saved**". Also there's big difference between the percentage of photos. Most of those photos in "**saved**" are just screenshots of some lectures.

10
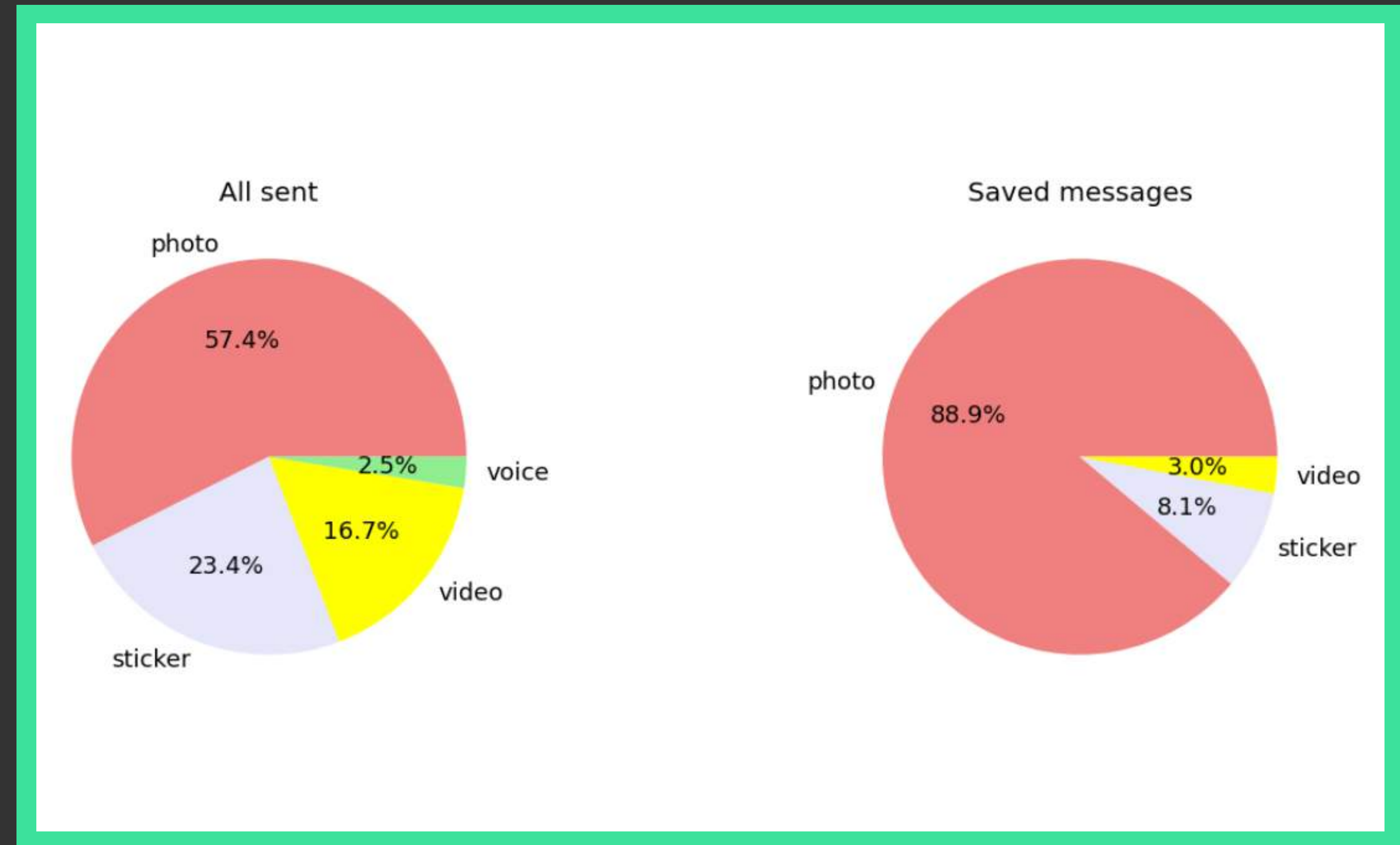
Words usage in "Primates" chat

OBSERVATION

All our program consists of two major studies: "**Math**" and "**Programming**". So let's study frequency of messages containing these words.

Correlation of programming and math messages in chat

матан

43.9%

56.1%

прога
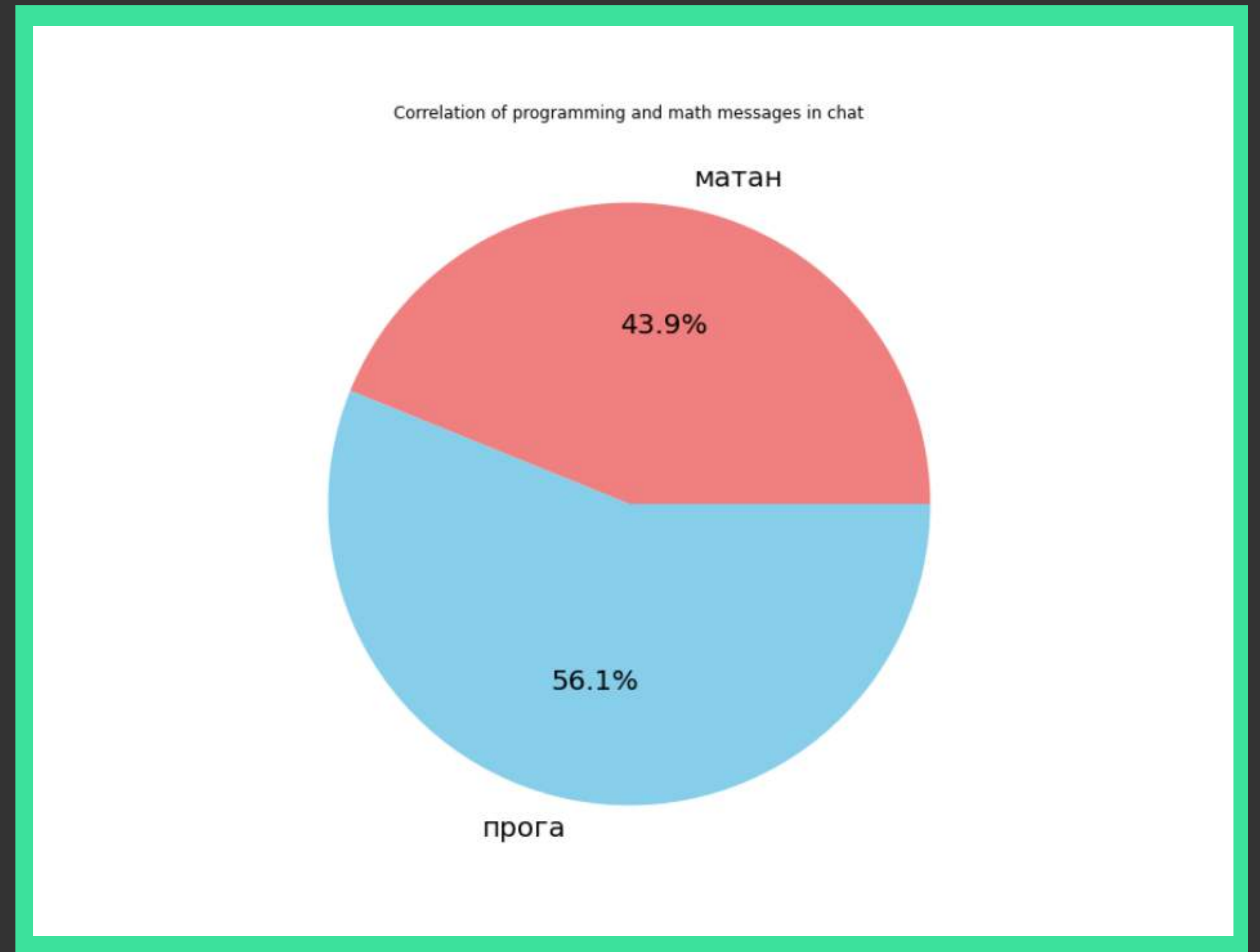
Programming: blue
Math: red

Words usage in "Primates" chat

OBSERVATION

Let's plot the messages by usage during time.

Firstly, we see that interest in study significantly fell on the 2nd year. Also the summer and winter vacations are very clearly seen.

Programming: blue
Math: red

12

Amount of fun/ Amount of messages

OBSERVATION

I've counted amount of **laugh** messages in every chat and created the rate of **fun messages/all**.

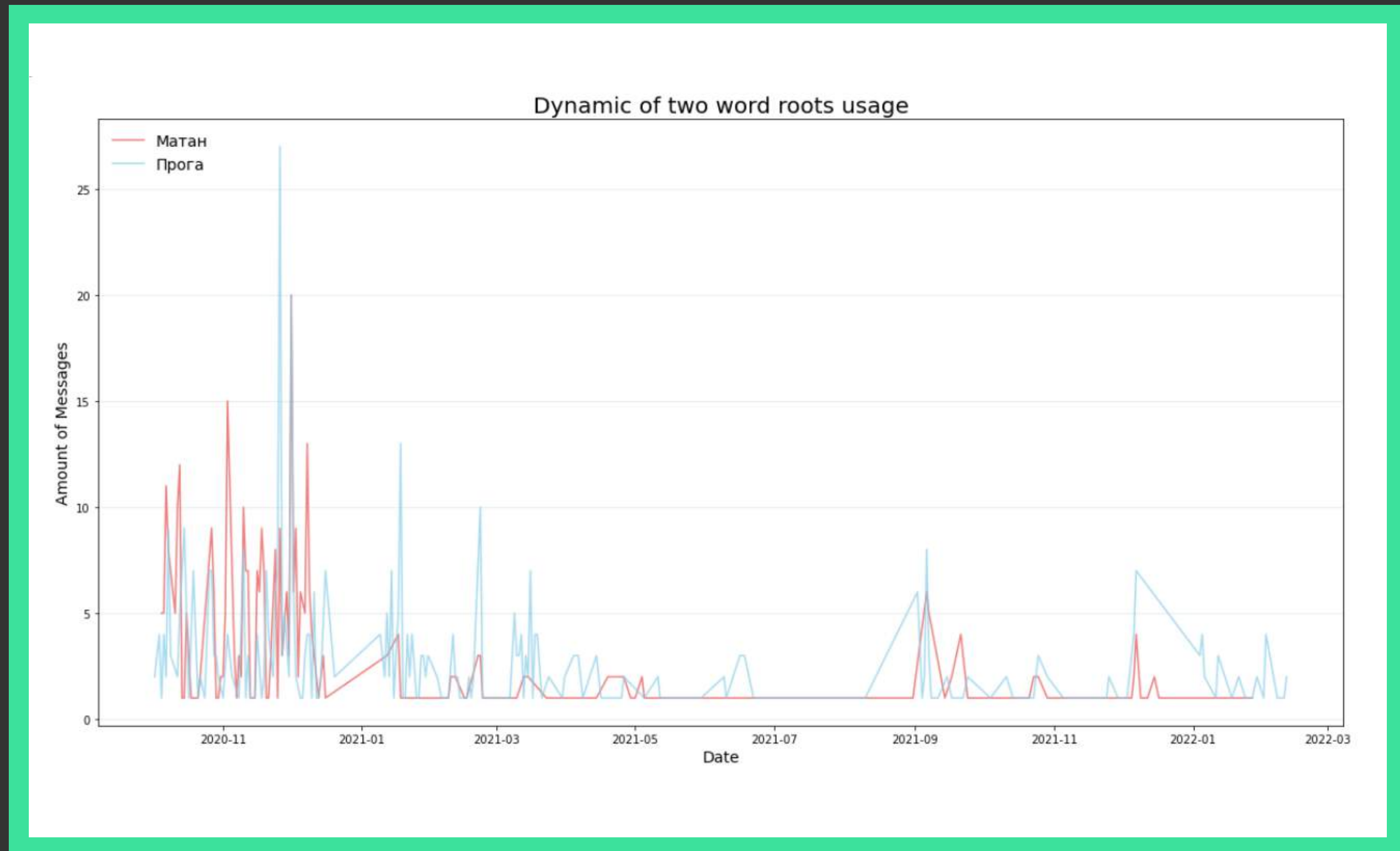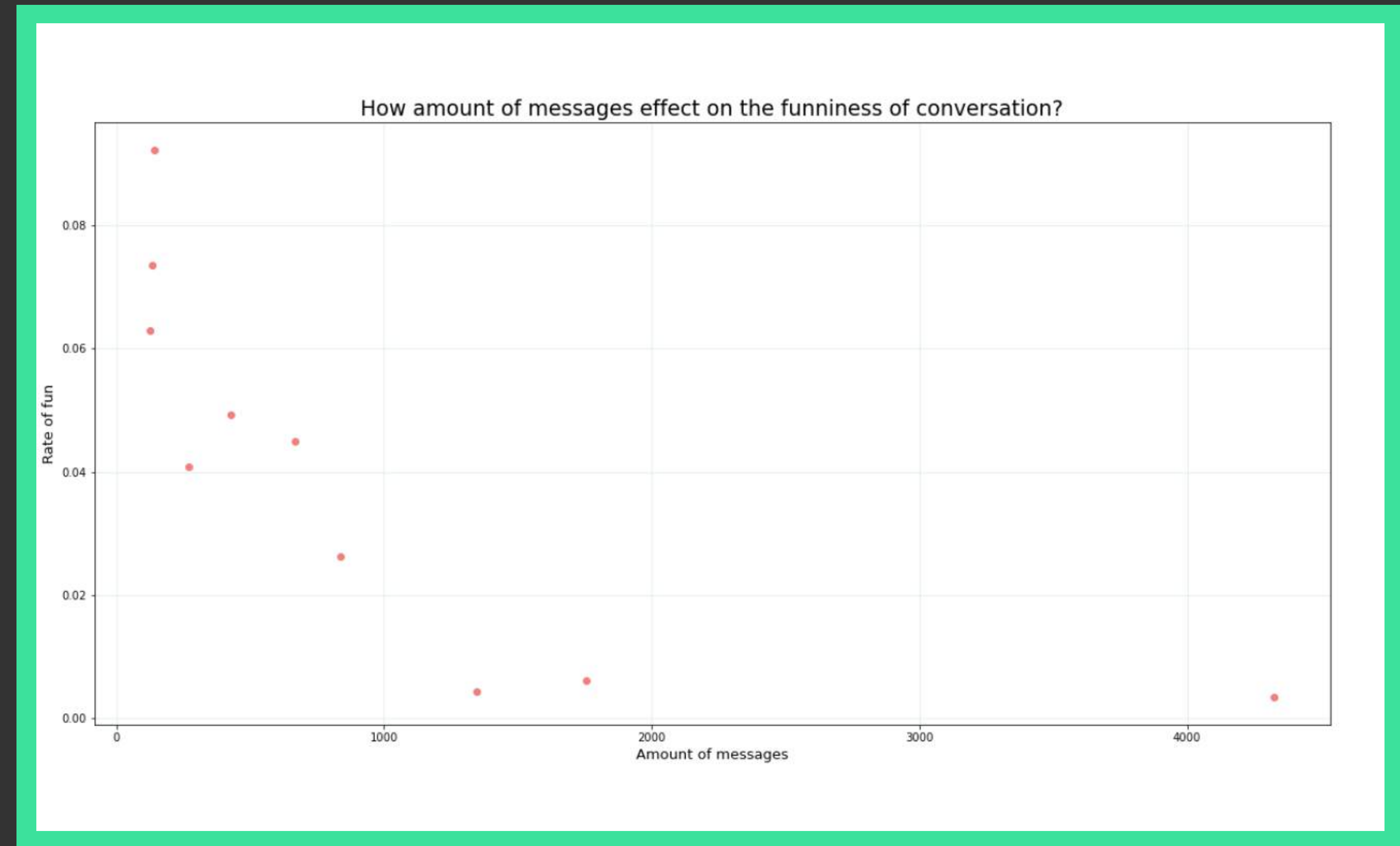It's a scatter plot of **fun rate** and amount of messages. It is clearly seen that more you chat with me, the less **fun** it becomes😢

How amount of messages effect on the funniness of conversation?

13

What to do next?

- NLP studies
- News analysis
- Punctuation mistakes
- Attitude towards countries/leaders
- Citation index (fwd messages)

Thanks for attention!

github link:
https://github.com/BHurskyy1/Telegram-data-analysis