

# 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?

## DATA STRUCTURE

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**

7650	289519089	Діма	Private dialog	{'user_id': 289519089, 'first_name': 'Діма', 'last_name': 'Діма', 'phone': '0931234567', 'email': 'dima@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7651	290672065	Олег Кормушин телефонные опросы	Private dialog	{'user_id': 290672065, 'first_name': 'Олег', 'last_name': 'Кормушин', 'phone': '0931234567', 'email': 'oleg@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7652	292255630	Veronica	Private dialog	{'user_id': 292255630, 'first_name': 'Veronica', 'last_name': 'Veronica', 'phone': '0931234567', 'email': 'veronica@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7653	296614607	Наталя Слобода	Private dialog	{'user_id': 296614607, 'first_name': 'Наталя', 'last_name': 'Слобода', 'phone': '0931234567', 'email': 'natalia@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7654	304250770	Анжеліка	Private dialog	{'user_id': 304250770, 'first_name': 'Анжеліка', 'last_name': 'Анжеліка', 'phone': '0931234567', 'email': 'anhelina@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7655	309706694	Sanya Natalenko	Private dialog	{'user_id': 309706694, 'first_name': 'Sanya', 'last_name': 'Natalenko', 'phone': '0931234567', 'email': 'sanya@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7656	319031272	Ярославко Верхола	Private dialog	{'user_id': 319031272, 'first_name': 'Ярославко', 'last_name': 'Верхола', 'phone': '0931234567', 'email': 'yoslavko@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7657	319498185	anhelina	Private dialog	{'user_id': 319498185, 'first_name': 'anhelina', 'last_name': 'anhelina', 'phone': '0931234567', 'email': 'anhelina@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7658	324529903	Helen Gark	Private dialog	{'user_id': 324529903, 'first_name': 'Helen', 'last_name': 'Gark', 'phone': '0931234567', 'email': 'helen@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7659	327020439	Ivan Tkachenko	Private dialog	{'user_id': 327020439, 'first_name': 'Ivan', 'last_name': 'Tkachenko', 'phone': '0931234567', 'email': 'ivan@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7660	330644315	сандаша	Private dialog	{'user_id': 330644315, 'first_name': 'сандаша', 'last_name': 'сандаша', 'phone': '0931234567', 'email': 'sandasha@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7661	331012549	Олексій Кордуба!	Private dialog	{'user_id': 331012549, 'first_name': 'Олексій', 'last_name': 'Кордуба', 'phone': '0931234567', 'email': 'oleksii@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7662	332402450	Александр Севергин	Private dialog	{'user_id': 332402450, 'first_name': 'Александр', 'last_name': 'Севергин', 'phone': '0931234567', 'email': 'alexander@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7663	333095627	Cristina Cherevko	Private dialog	{'user_id': 333095627, 'first_name': 'Cristina', 'last_name': 'Cherevko', 'phone': '0931234567', 'email': 'cristina@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7664	333333918	Oleksandra Radzievska	Private dialog	{'user_id': 333333918, 'first_name': 'Oleksandra', 'last_name': 'Radzievska', 'phone': '0931234567', 'email': 'oleksandra@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7665	334976382	Dmytro Novykov	Private dialog	{'user_id': 334976382, 'first_name': 'Dmytro', 'last_name': 'Novykov', 'phone': '0931234567', 'email': 'dmytro@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7666	336314102	Лиза ☹	Private dialog	{'user_id': 336314102, 'first_name': 'Лиза', 'last_name': 'Лиза', 'phone': '0931234567', 'email': 'liza@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7667	340227209	Dmytro Miedviediev	Private dialog	{'user_id': 340227209, 'first_name': 'Dmytro', 'last_name': 'Miedviediev', 'phone': '0931234567', 'email': 'dmytro@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7668	342667961	Yullya Verbik	Private dialog	{'user_id': 342667961, 'first_name': 'Yullya', 'last_name': 'Verbik', 'phone': '0931234567', 'email': 'yullya@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7669	351412933	Vladyslav	Private dialog	{'user_id': 351412933, 'first_name': 'Vladyslav', 'last_name': 'Vladyslav', 'phone': '0931234567', 'email': 'vladyslav@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7670	353330283	Полковник Кольт	Private dialog	{'user_id': 353330283, 'first_name': 'Полковник', 'last_name': 'Кольт', 'phone': '0931234567', 'email': 'polkovnik@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7671	354402008	Настя ☹	Private dialog	{'user_id': 354402008, 'first_name': 'Настя', 'last_name': 'Настя', 'phone': '0931234567', 'email': 'nastia@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7672	358521785	Ira Horoshkevych	Private dialog	{'user_id': 358521785, 'first_name': 'Ira', 'last_name': 'Horoshkevych', 'phone': '0931234567', 'email': 'ira@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7673	360307970	Ulya	Private dialog	{'user_id': 360307970, 'first_name': 'Ulya', 'last_name': 'Ulya', 'phone': '0931234567', 'email': 'ulya@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7674	363384572	marusyabelle	Private dialog	{'user_id': 363384572, 'first_name': 'marusyabelle', 'last_name': 'marusyabelle', 'phone': '0931234567', 'email': 'marusyabelle@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7675	364743743	Загарт.Ю	Private dialog	{'user_id': 364743743, 'first_name': 'Загарт.Ю', 'last_name': 'Загарт.Ю', 'phone': '0931234567', 'email': 'zagart.yu@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7676	364750475	Максим Сидорчук	Private dialog	{'user_id': 364750475, 'first_name': 'Максим', 'last_name': 'Сидорчук', 'phone': '0931234567', 'email': 'maxim@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7677	370252186	Іра	Private dialog	{'user_id': 370252186, 'first_name': 'Іра', 'last_name': 'Іра', 'phone': '0931234567', 'email': 'ira@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7678	371105962	Тім	Private dialog	{'user_id': 371105962, 'first_name': 'Тім', 'last_name': 'Тім', 'phone': '0931234567', 'email': 'tim@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7679	372551595	Влад Шпигун	Private dialog	{'user_id': 372551595, 'first_name': 'Влад', 'last_name': 'Шпигун', 'phone': '0931234567', 'email': 'vlad@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7680	373321078	Ана.sta.sia	Private dialog	{'user_id': 373321078, 'first_name': 'Ана.sta.sia', 'last_name': 'Ана.sta.sia', 'phone': '0931234567', 'email': 'ana.sta.sia@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}
7681	376254220	Аня Вареник ☹	Private dialog	{'user_id': 376254220, 'first_name': 'Аня', 'last_name': 'Вареник', 'phone': '0931234567', 'email': 'anya@ua.com', 'city': 'Kyiv', 'country': 'Ukraine', 'status': 'active', 'created_at': '2018-01-01 12:00:00', 'updated_at': '2018-01-01 12:00:00'}

dialogs\_users\_all.csv contents

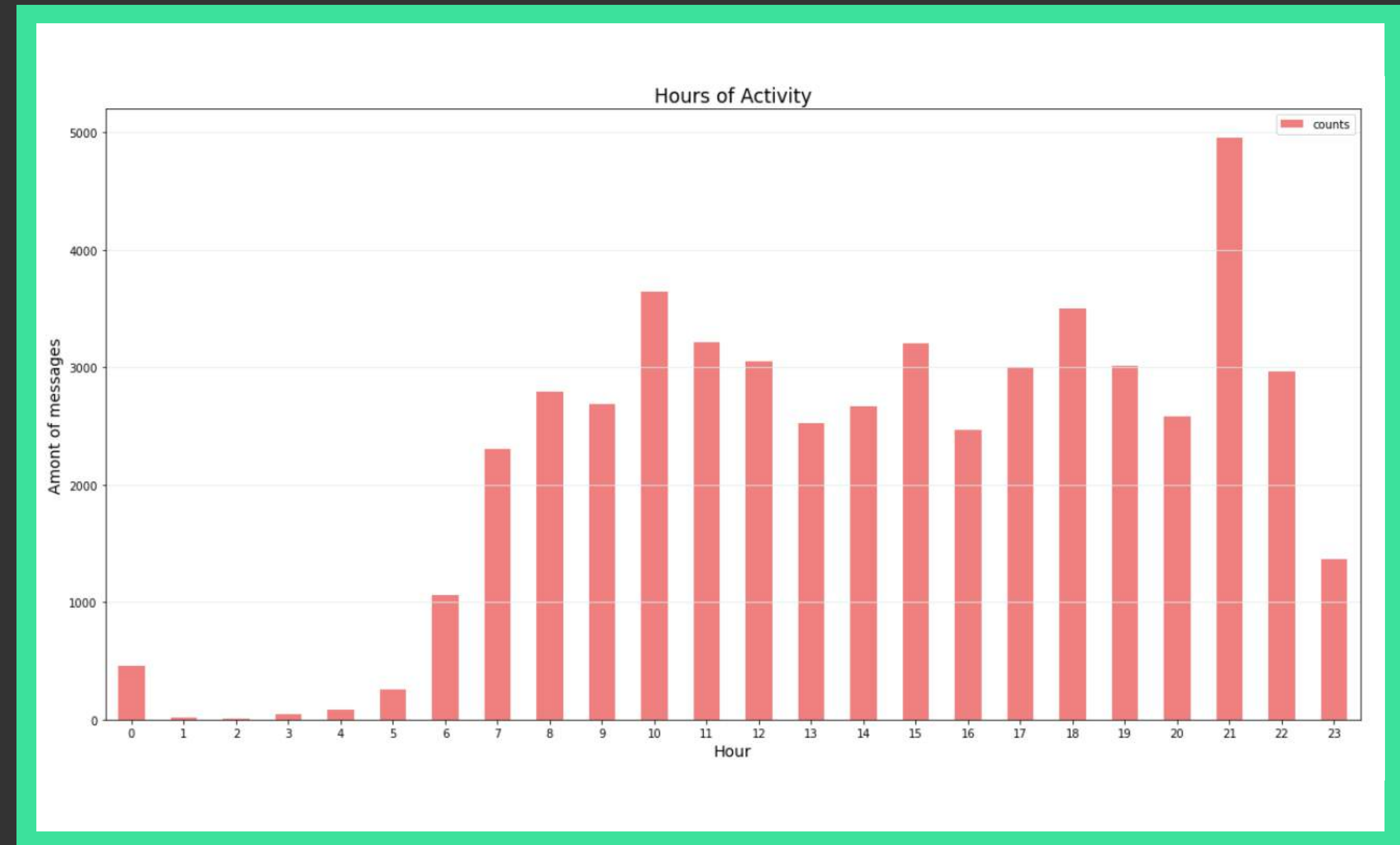


# My daily activity by hours

## 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.

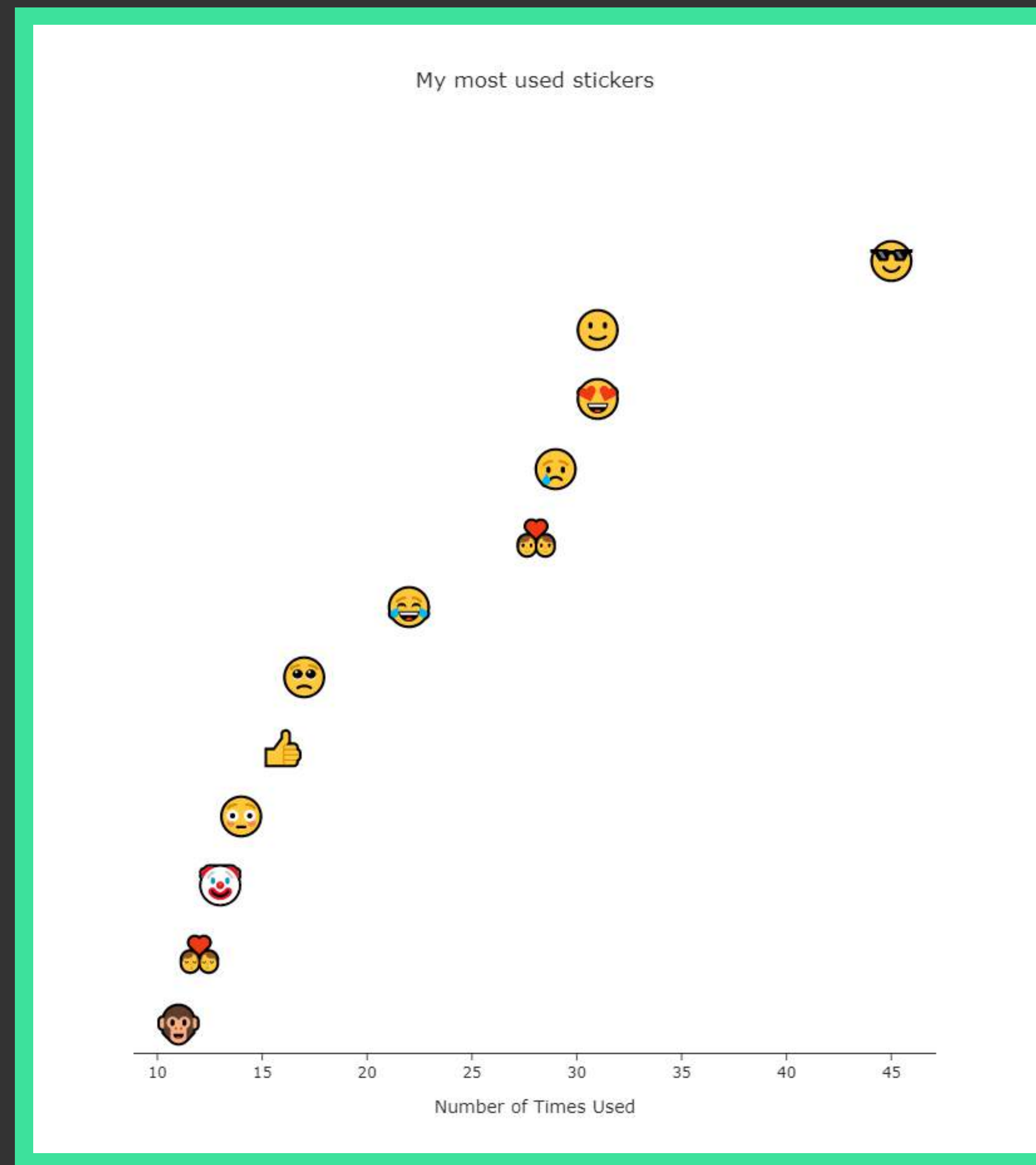
Obviously, the lowest activity is between **1-4am**. We can see the strong rise around **5-6am**, usually I schedule my day in my "saved messages".



# My favorite sticker?

## OBSERVATION

This plot is pretty simple.  
Just my favorite stickers.

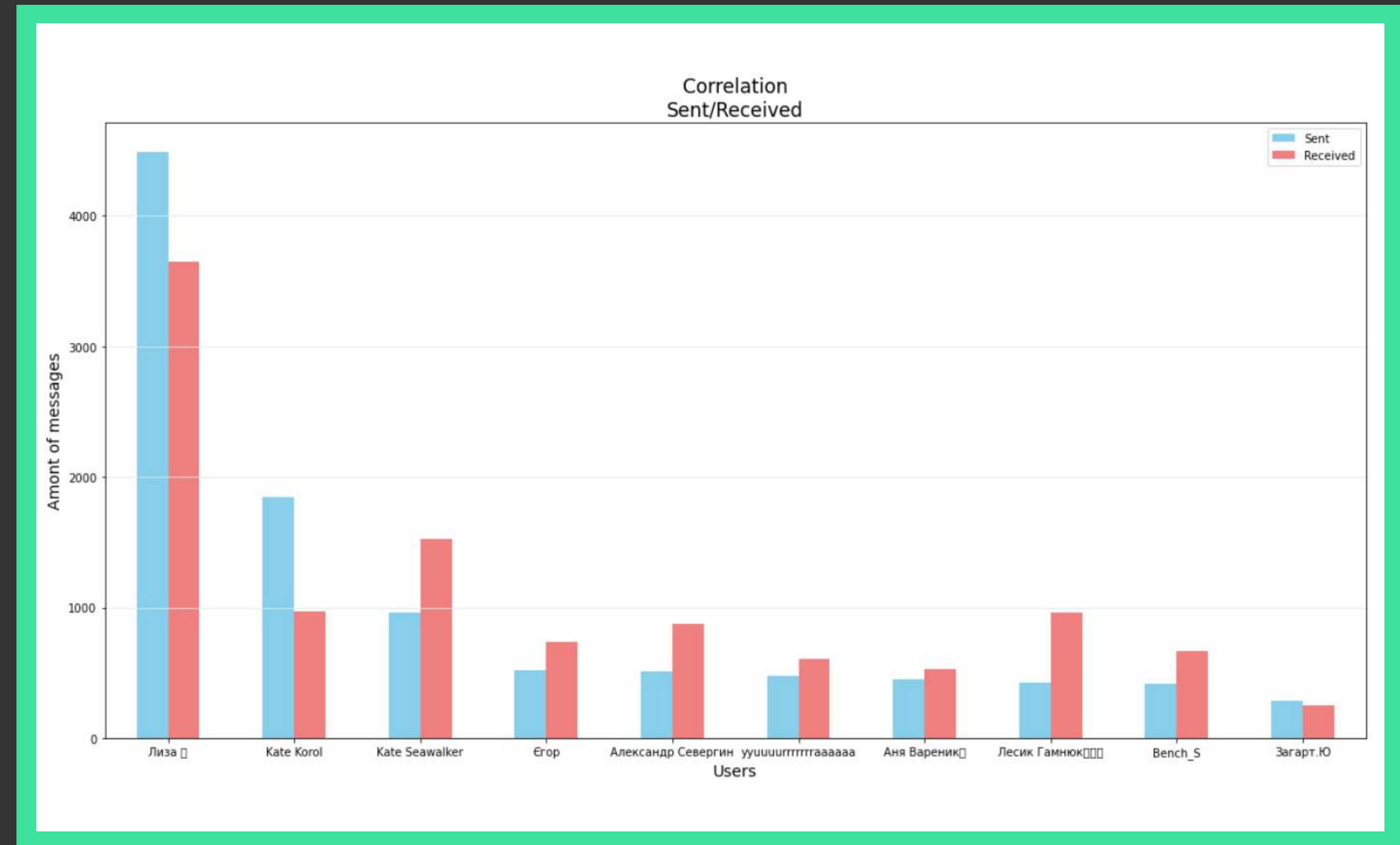


# Correlation of sent/received

## 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**.

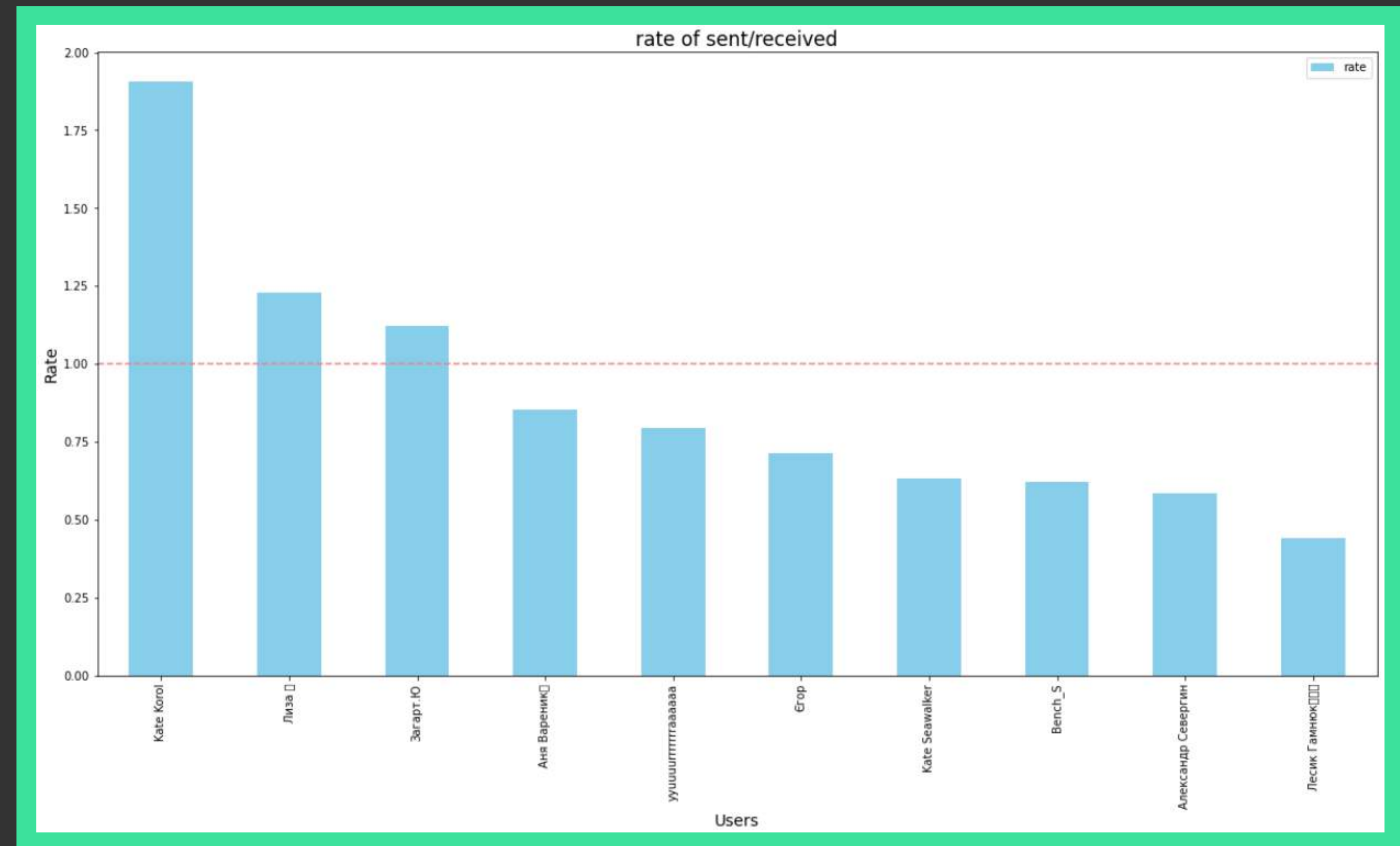


# Correlation of sent/received

## 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** 😎

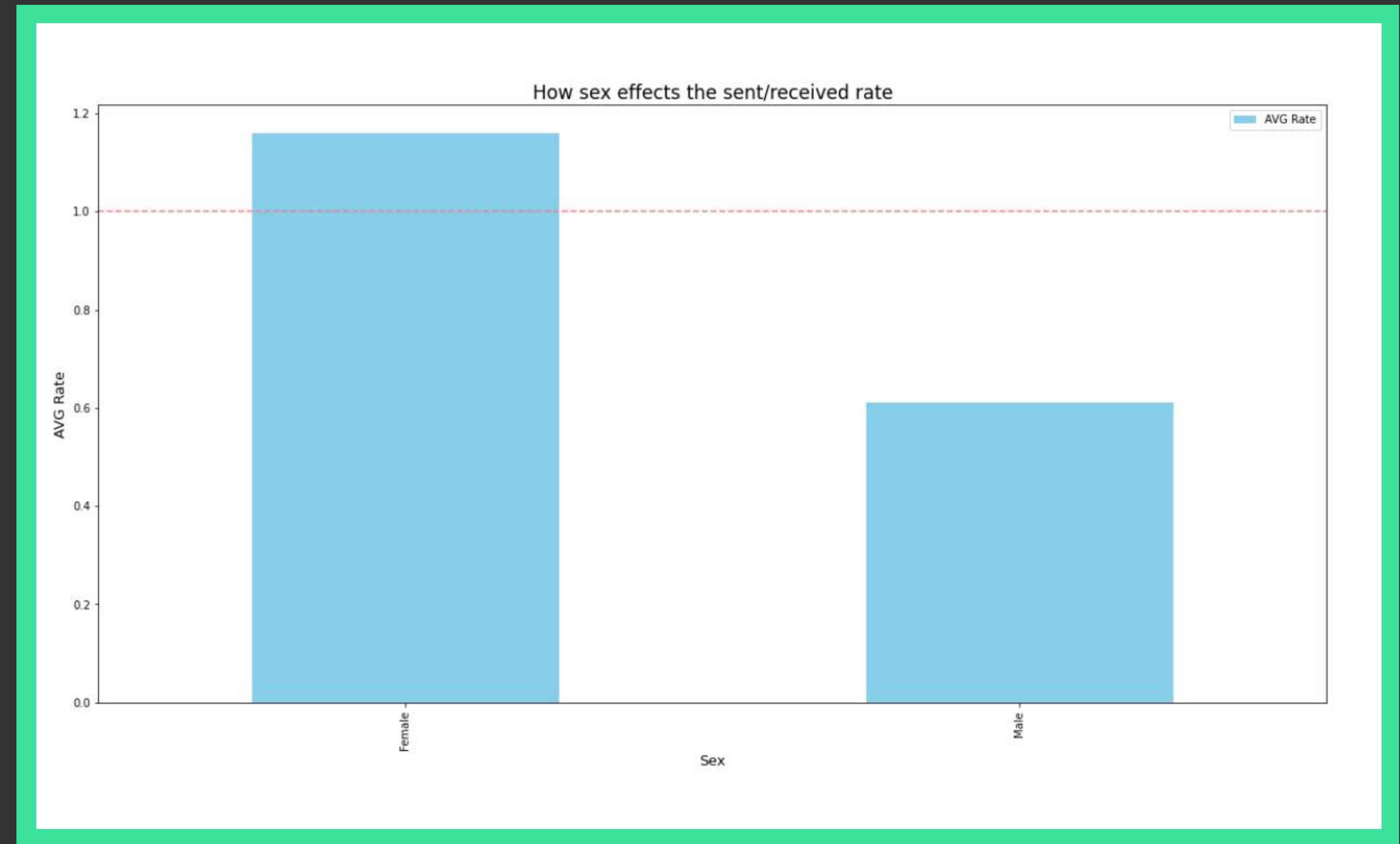


# Correlation of sent/received

## 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 😎



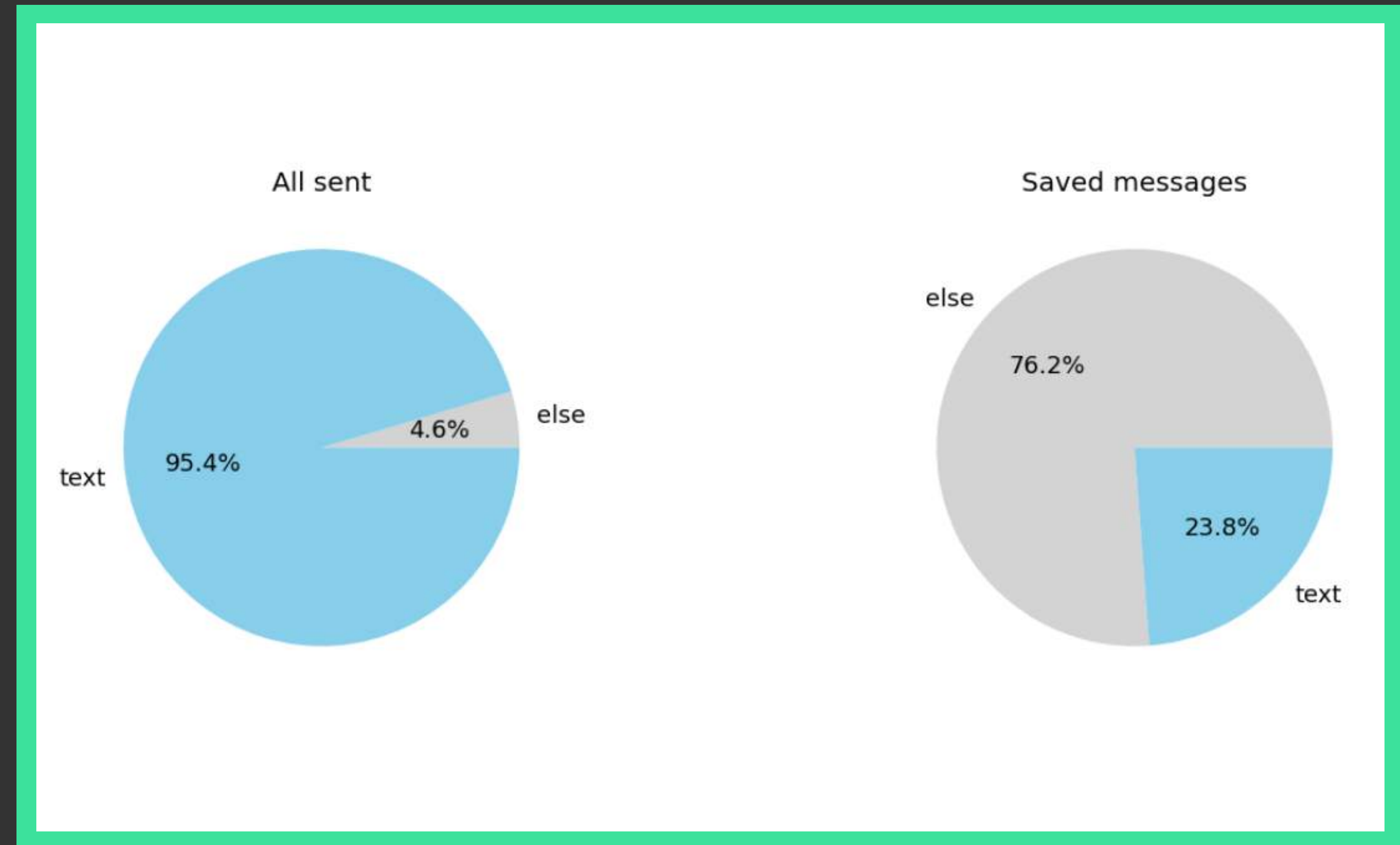


# What messages do I save?

## 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.

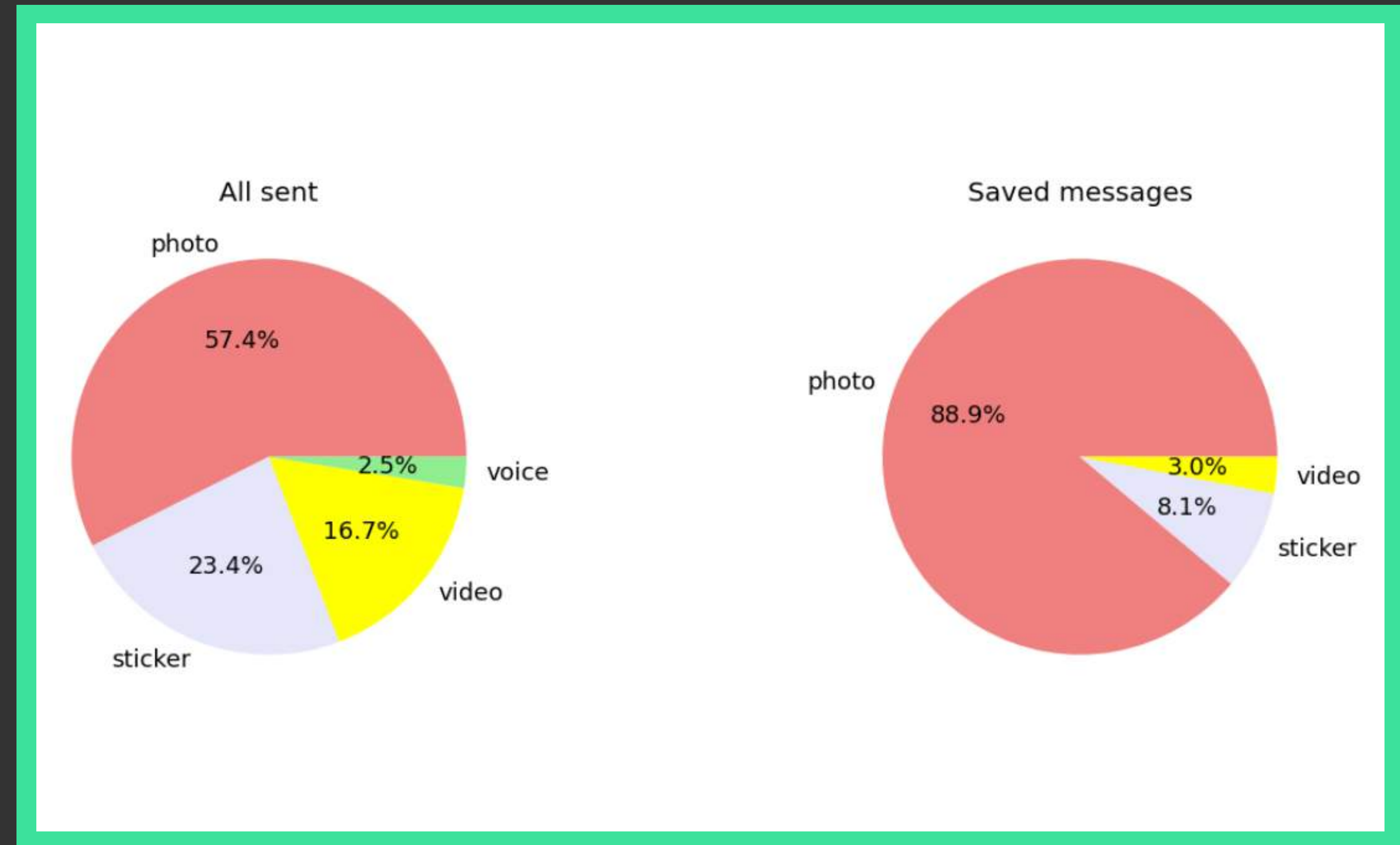


# What messages do I save?

## OBSERVATION

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

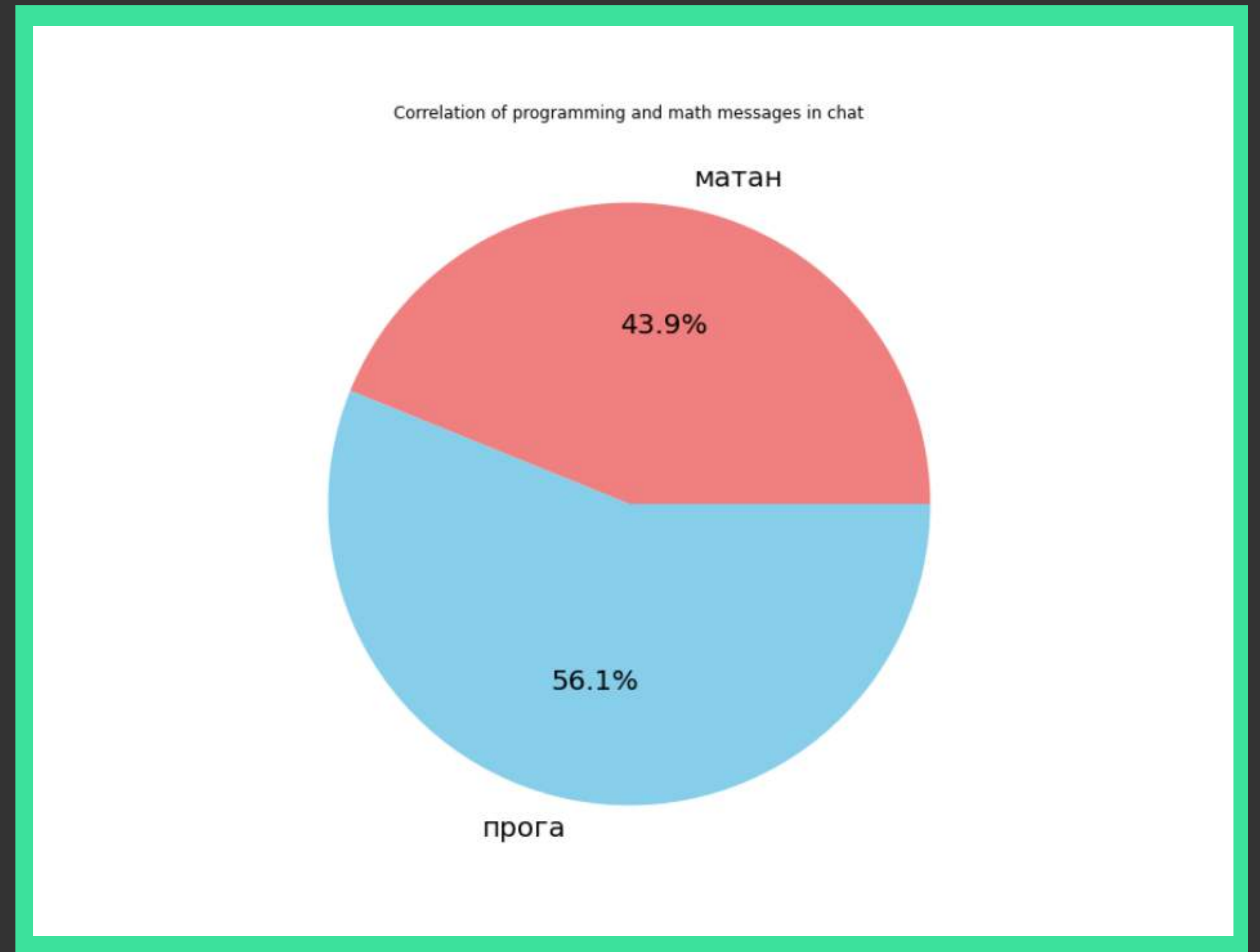
Firstly, I don't like listening to my voice, so there's **0** voice messages in my "saved". Also there's big difference between the percentage of photos. Most of those photos in "saved" are just screenshots of some lectures.



# 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.



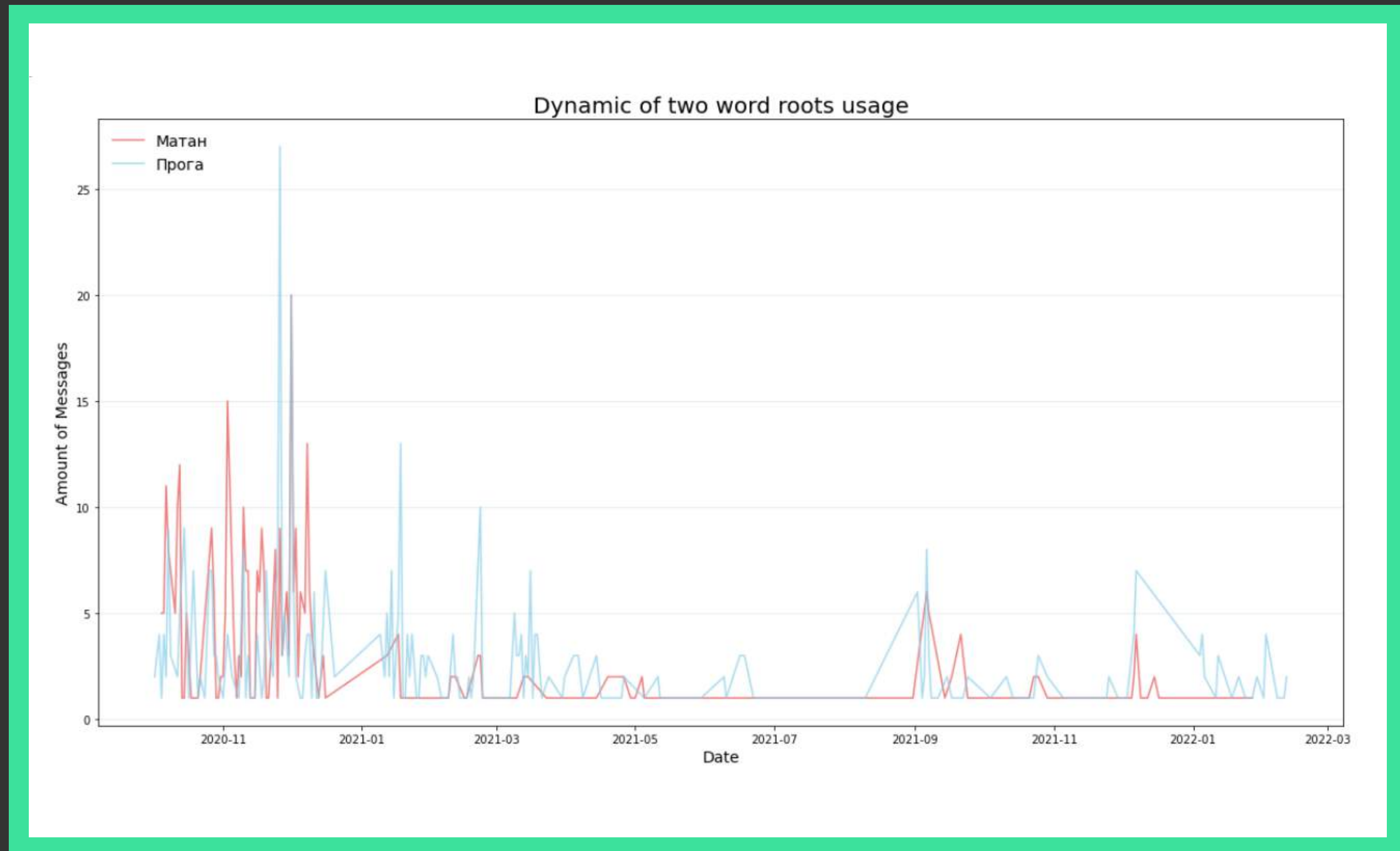
Programming: blue  
Math: red

# 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

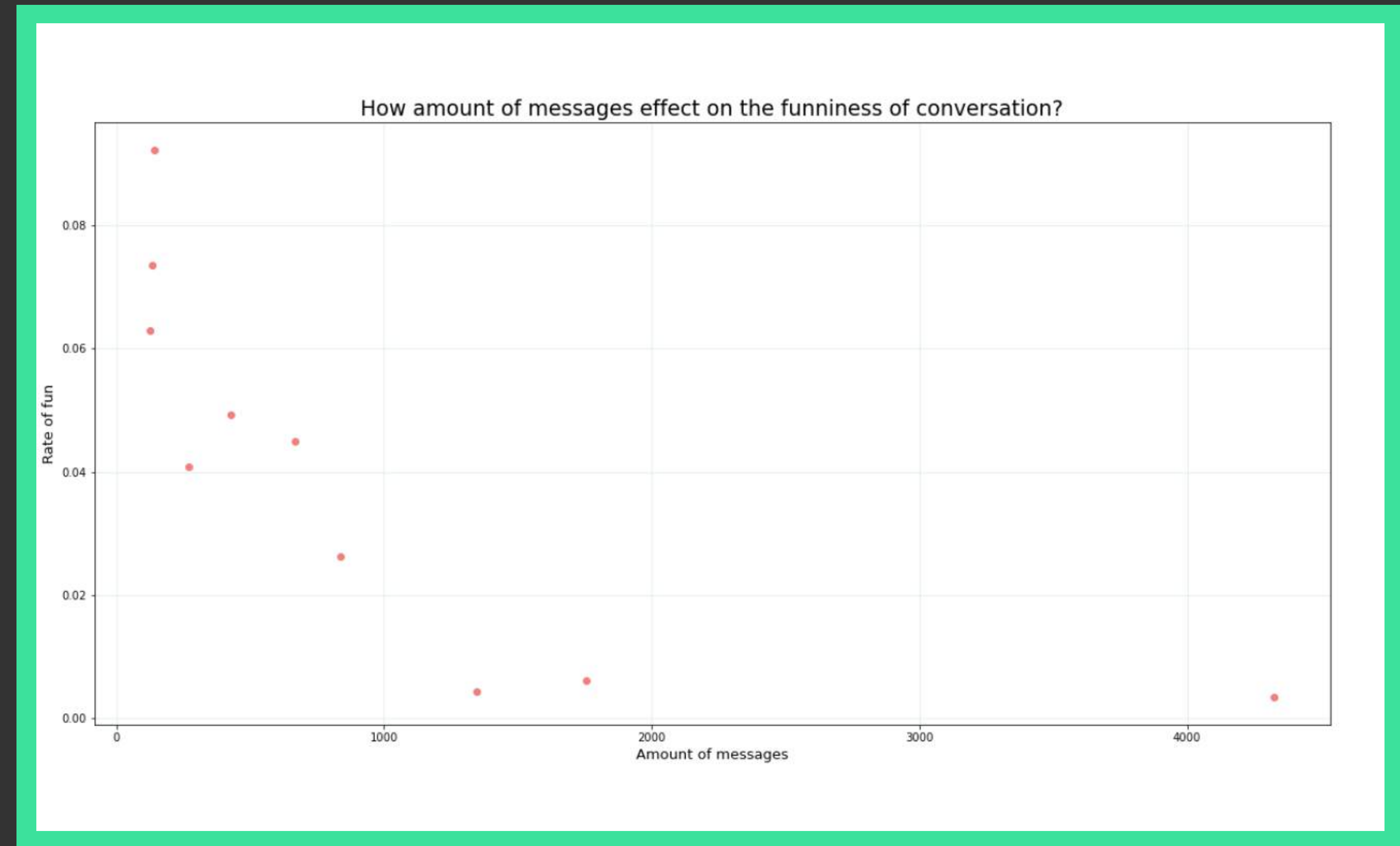


# Amount of fun/ Amount of messages

## 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 😞



# 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>