

Анализ активности пользователей в чатах Пачки и Слак



-
- **Заказчик** - Яндекс практикум
 - **Проблема** - низкая обратная связь на важные посты, публикуемые в чатах.
 - **Цель исследования**
 - Выявить паттерны, динамику, цикличность в течение дня, недели, месяца, года. Визуализировать свои находки.
 - Понять, когда активность студентов в чатах наибольшая, и когда лучше публиковать посты/анонсы, чтобы получить больше откликов.
 - Проанализировать различные типы каналов, когорт, групп. оценить в каких общении больше, в каких меньше, и как они различаются от когорты к когорте, от канала к каналу



Анализ полученных данных

- Итого, в таблице 26533 записей, и 14 столбцов. 7 столбцов - текстовые, 4 - с датами и временем, 2-целочисленные, 1 - булевый. 7 столбцов не имеют пропусков. Остальные семь имеют пропуски от 10% до 91%. Больше всего пропусков в столбцах с реакциями и временем ответа 84 и 91 процент.

latest_reply	91.320243
reactions	84.287491
subtype	68.654129
team	40.236686
thread_ts	31.323258
client_msg_id	31.172502
user	10.892097
text_words	0.000000
text_len	0.000000
attachments	0.000000
file_date	0.000000
channel	0.000000
ts	0.000000
type	0.000000

Предобработка данных

- **channel** - выделены названия канала, профессии и номер когорты.
- **ts (дата поста)** - выделены год, месяц, день недели, и время дня.
- **reactions (реакции)** - выделено количество реакций на сообщения.
- **user_type** (тип юзера - студент или преподаватель)
- **subtype (метка действий пользователя)** - Не смотря на то , что служебные сообщения составляют 20% от всех сообщений, их придется удалить

student	22115	subtype	
teacher	4418	channel_leave	3265
		bot_message	2890
		channel_join	2026
		thread_broadcast	108
		tombstone	14
		channel_unarchive	10
monday	7919	channel_purpose	2
tuesday	4146	bot_remove	1
thursday	3814	channel_name	1
wednesday	3490		
friday	3380		
sunday	2182		
saturday	1602		
		день_после_обеда	9191
		утро	6187
		день_до_обеда	5512
		вечер	4208
		ночь	1435
	

Исследовательский анализ данных

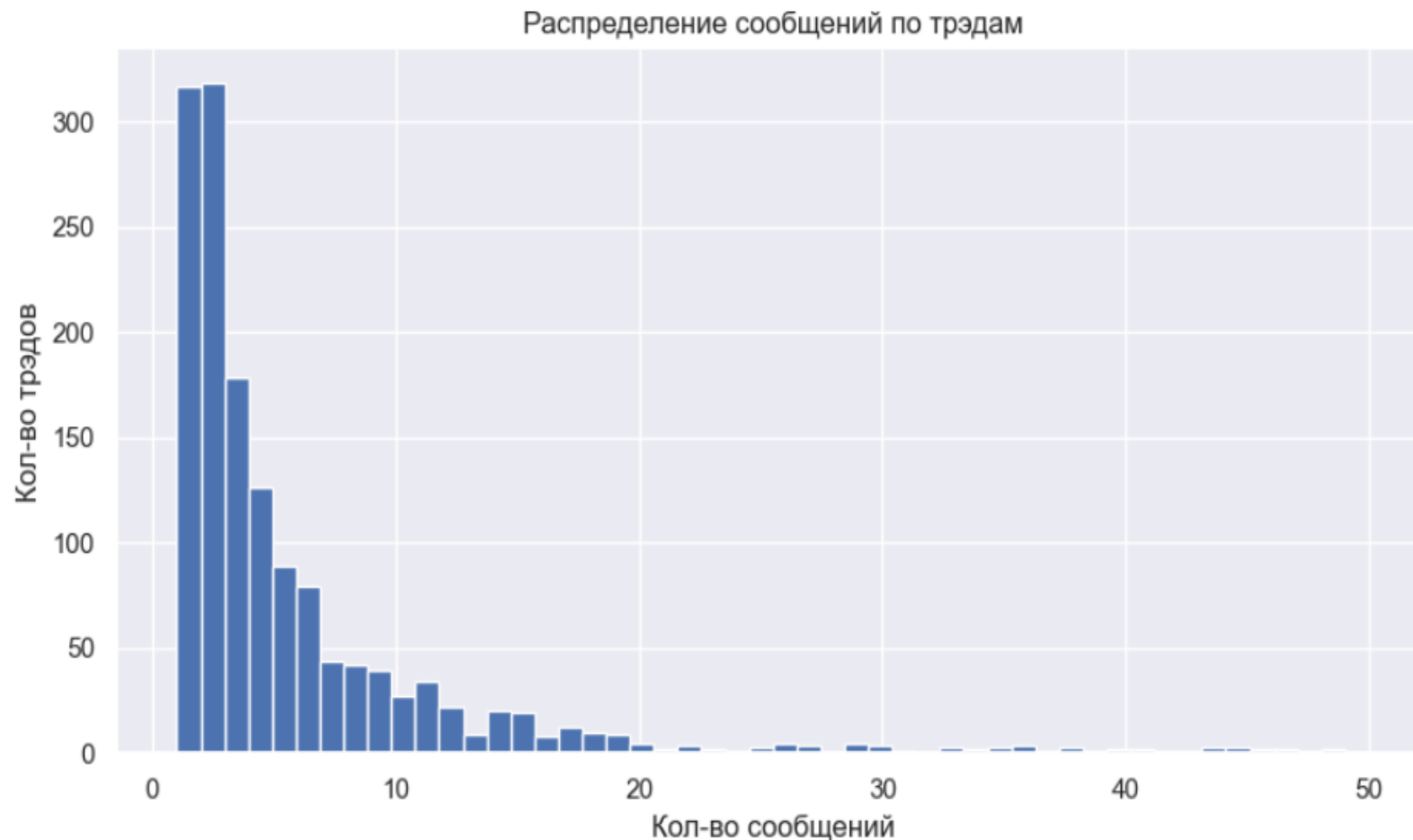
- За какой период представлены данные
- Наблюдается резкий всплеск активности первого декабря и такое же резкое падение 15ого декабря у студентов и преподавателей



Исследовательский анализ данных

Активность в трэдах

Всего в выгрузке 1540 трэдов. Большая часть из них состоит из одного - двух сообщений, по 300 штук. Медианное количество сообщений в трэдах - 3. Таких 150шт. Первая квартиль - 2 сообщения. Третья квартиль - 8 сообщений



Исследовательский анализ данных

Время до первого сообщения

Минимальное время до первого сообщения менее 1ой минуты. Таких трэдов около 140.

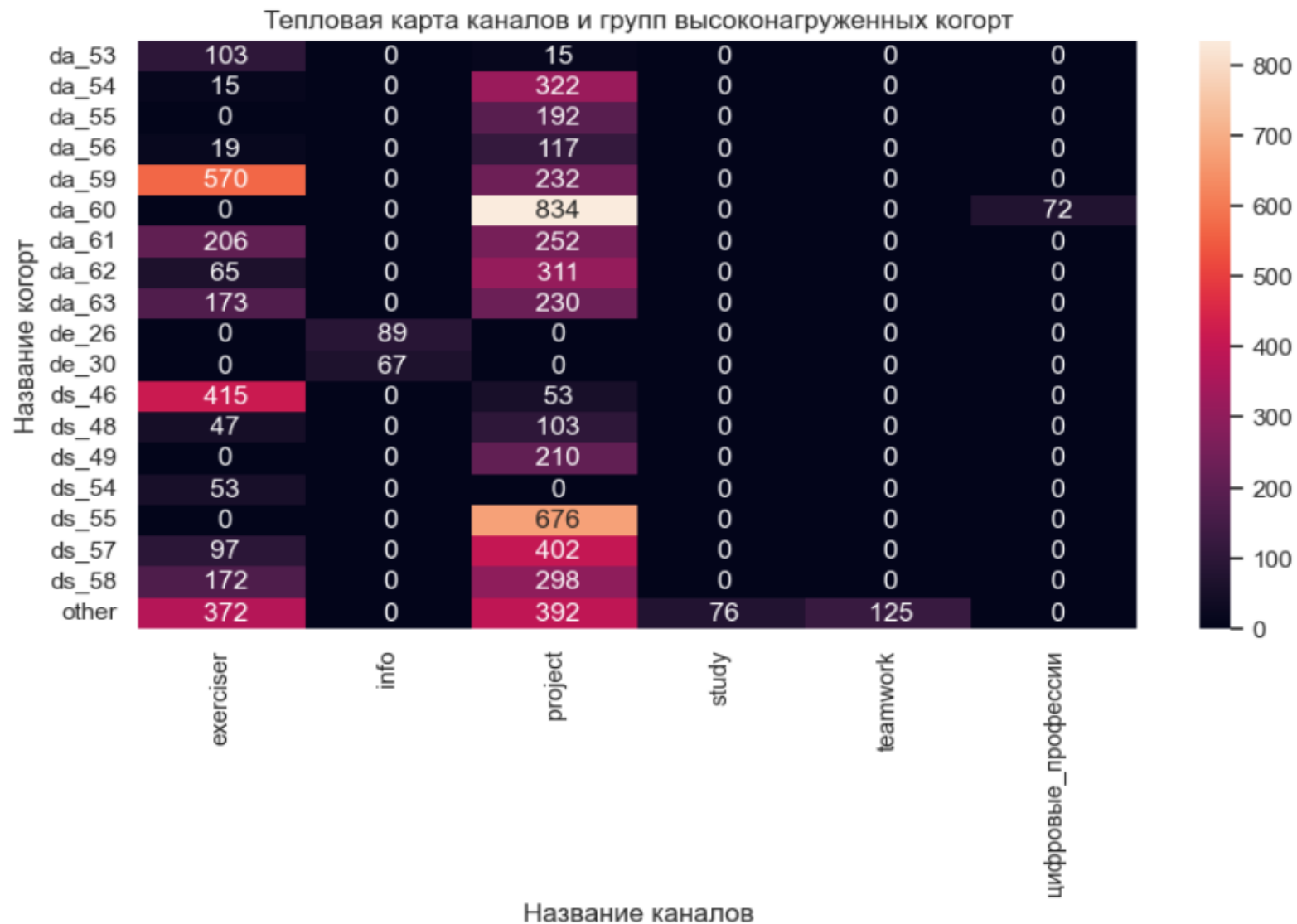
Медианное время до первого сообщения 75 мин. 1я квартиль - 9минут 3я квартиль - 679минут.



Исследовательский анализ данных

Распределение сообщений по группам и каналам

Самые нагруженные каналы это "exerciser" и "project". И самые активные профессии это da и ds.



Исследовательский анализ данных

Распределение сообщений по дням недели и времени суток

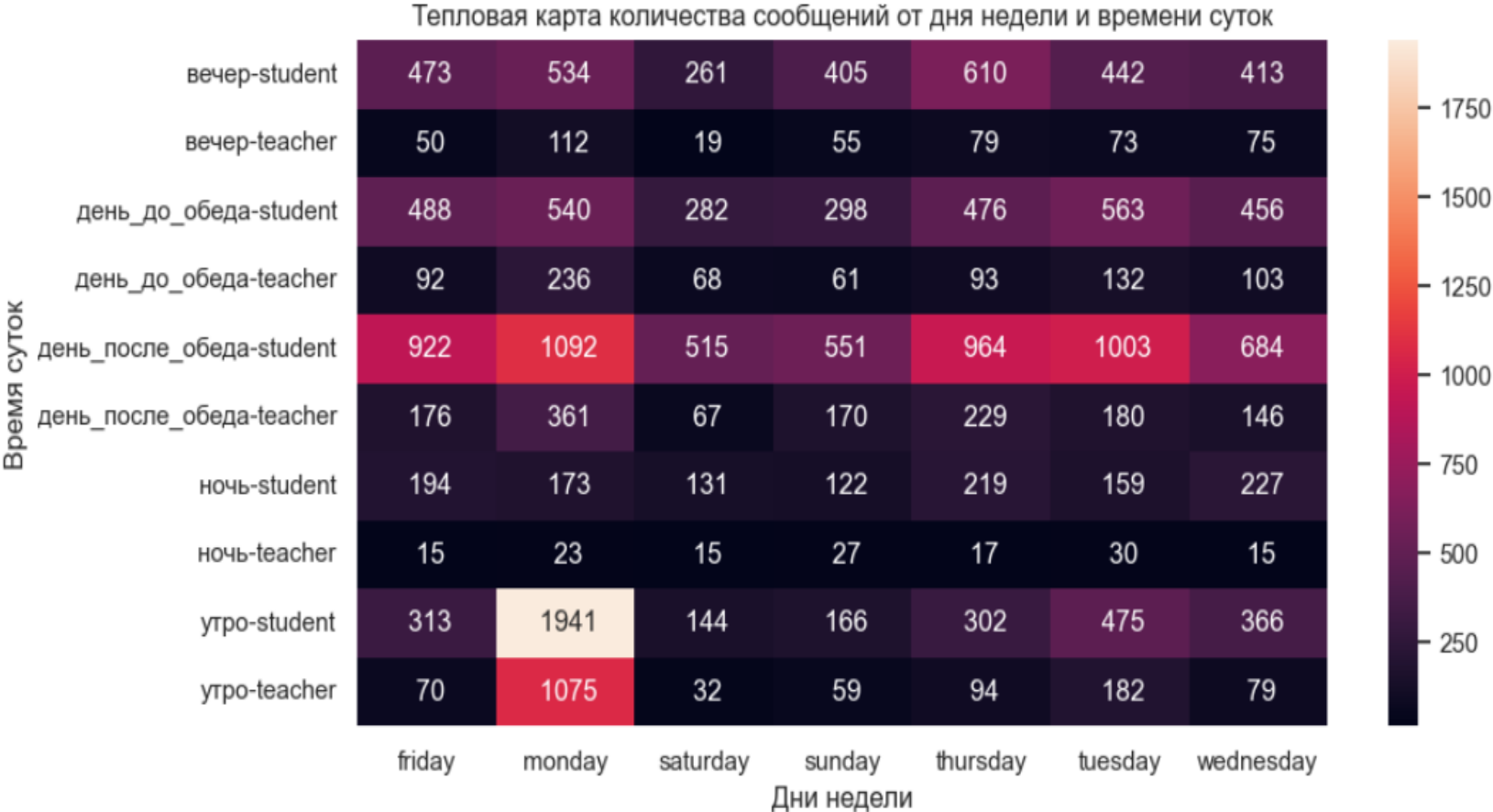
Понедельник утро - максимальная активность студентов и преподавателей.

Понедельник после обеда - высокая активность студентов. Активность преподавателей средняя.

Любой другой день после обеда - Высокая активность у студентов, низкая у преподавателей.

В остальное время активность у студентов и преподавателей - низкая.

Суббота и Воскресенье - низкая активность весь день.

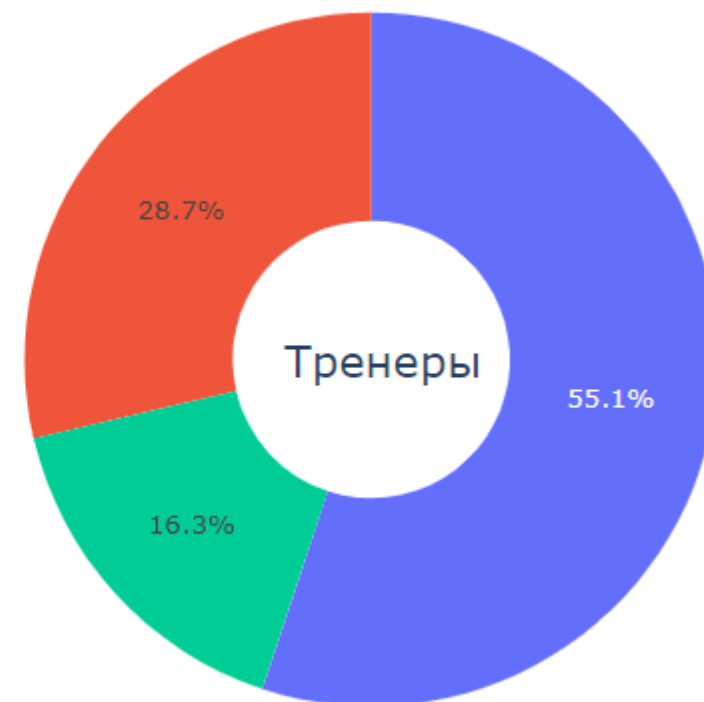


Исследовательский анализ данных

Распределение сообщений по количеству слов

По данным графикам становится видно что и студенты и преподаватели в половине случаев пишут средней длины сообщения (от 10 до 50 слов) или короткие - менее 10 слов (менее 30% случаев). Длинные сообщения преподаватели пишут чаще в 16% случаев против 10% случаев у студентов.

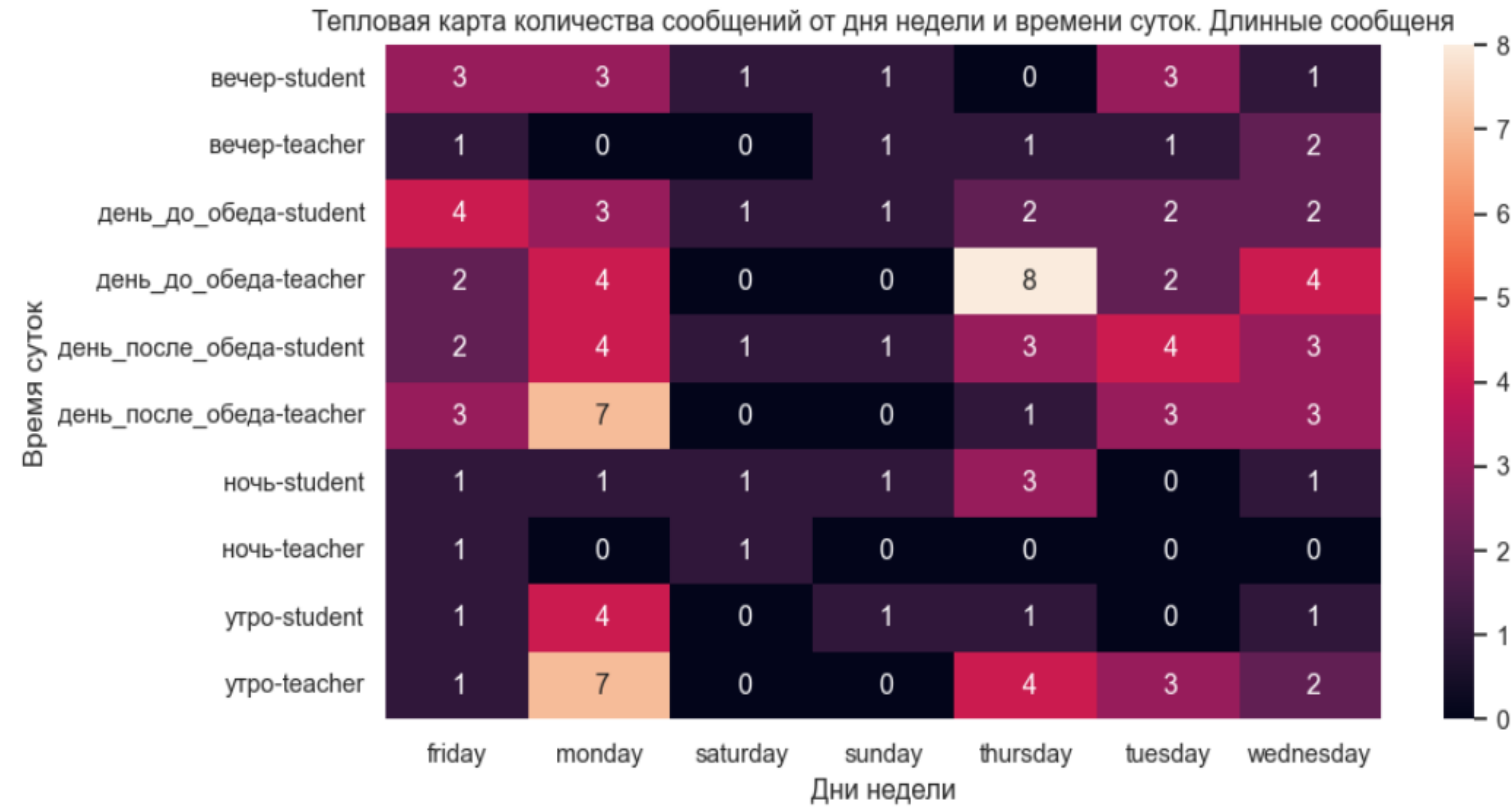
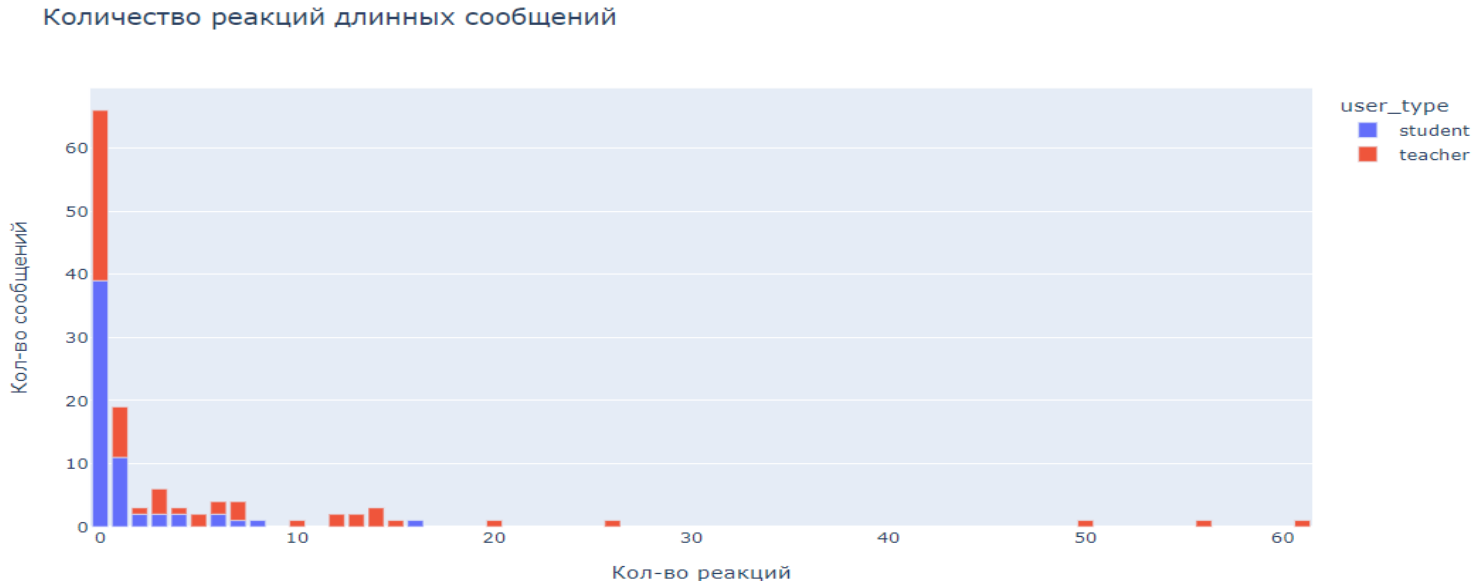
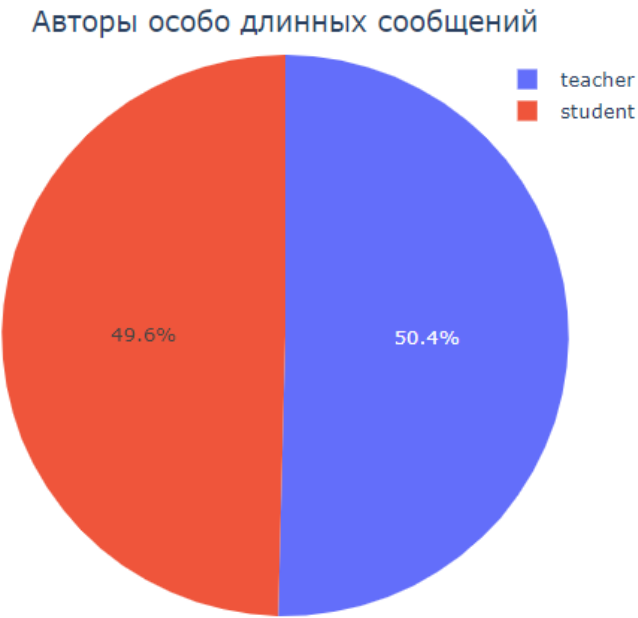
Соотношение кол-ва сообщений к длине сообщений



- Нормальное сообщение
- Короткое сообщение
- Длинное сообщение

Исследовательский анализ данных

Анализ особо длинных сообщений

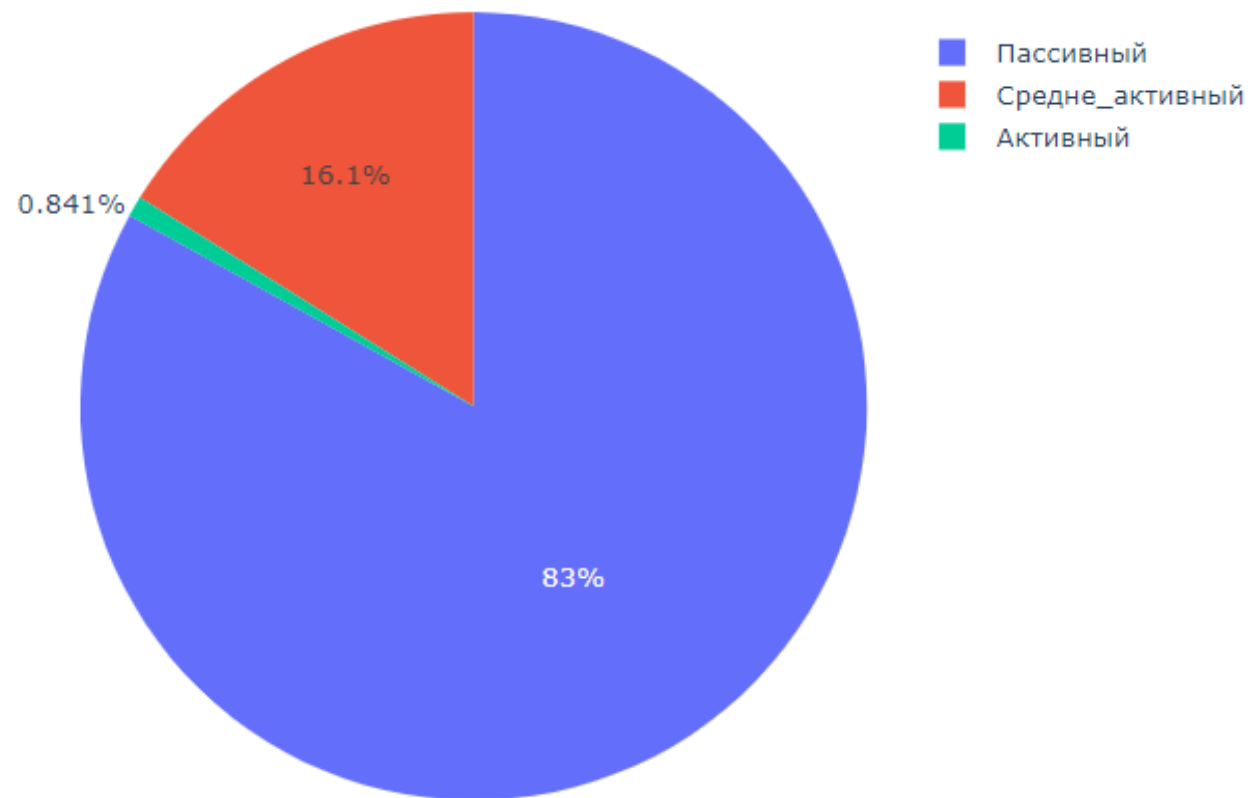


Исследовательский анализ данных

Проверка активности студентов

Выявилась неожиданная особенность. 83% всех студентов пассивно ведут себя в чатах (то есть, за все время исследования написали менее 10 сообщений и при этом поставили менее 10 реакций) А активных студентов, которые более 100 сообщений или поставили более 50 реакций, менее 1%.

График соотношения активности студентов



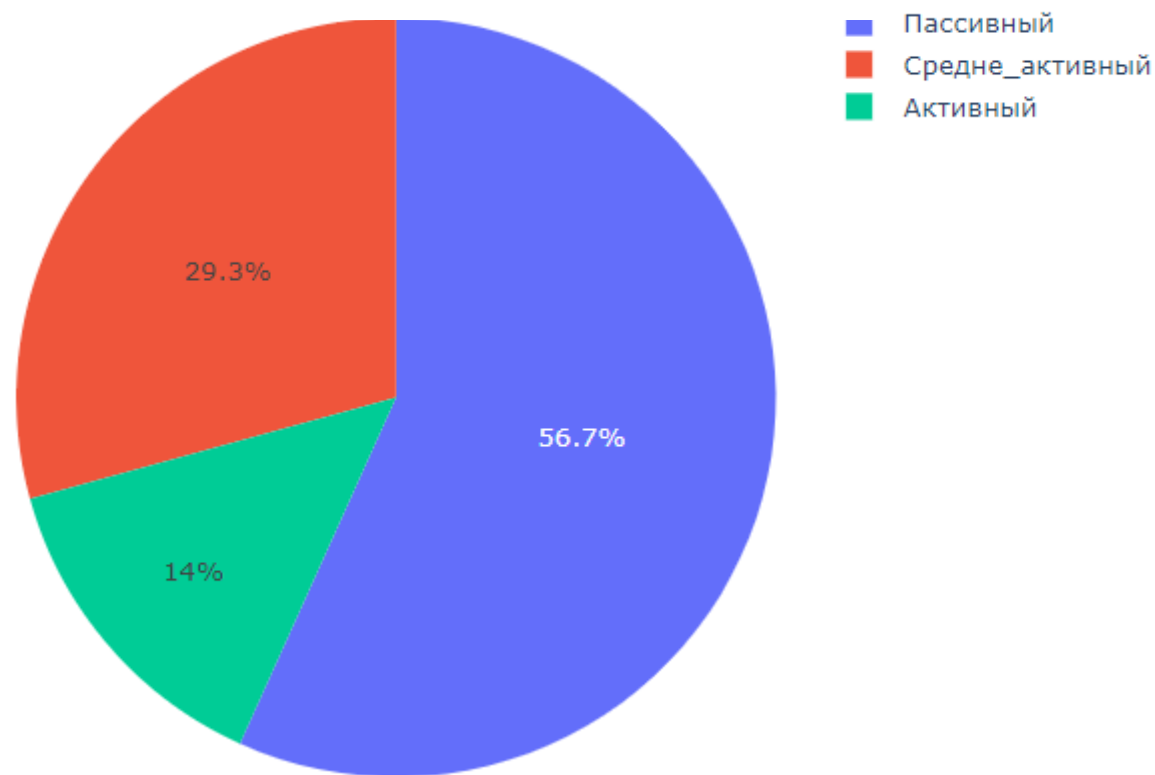
Активный - написал > 100 сообщений или поставил более 50 реакций
Пассивный - написал < 10 сообщений и поставил менее 10 реакций

Исследовательский анализ данных

Проверка активности студентов

У преподавателей ситуация выглядит лучше, чем у студентов. Половина преподавателей - активны в чатах. Но тем не менее 56% преподавателей написали менее 10 сообщений.

График соотношения активности преподавателей



Активный - написал > 100 сообщений или поставил более 50 реакций
Пассивный - написал < 10 сообщений и поставил менее 10 реакций

Рекомендации

- В связи с тем, что исследование показало **отсутствие активности в чатах 50% преподавателей**, необходимо провести отдельное исследование какие преподаватели должны быть активными в чате и разработать систему мотивации активности в чатах.
- Такой же подход можно применить к студентам. 80% студентов не активны в чатах. Система мотивации значительно увеличит данный показатель. **Например поставь 100 реакций или собери 100 лайков и получи скидку на следующий месяц.**
- время рассылки важных сообщений - **Понедельник утро и будние дни после обеда** для этого тот самый момент.





Спасибо за
внимание !!!
