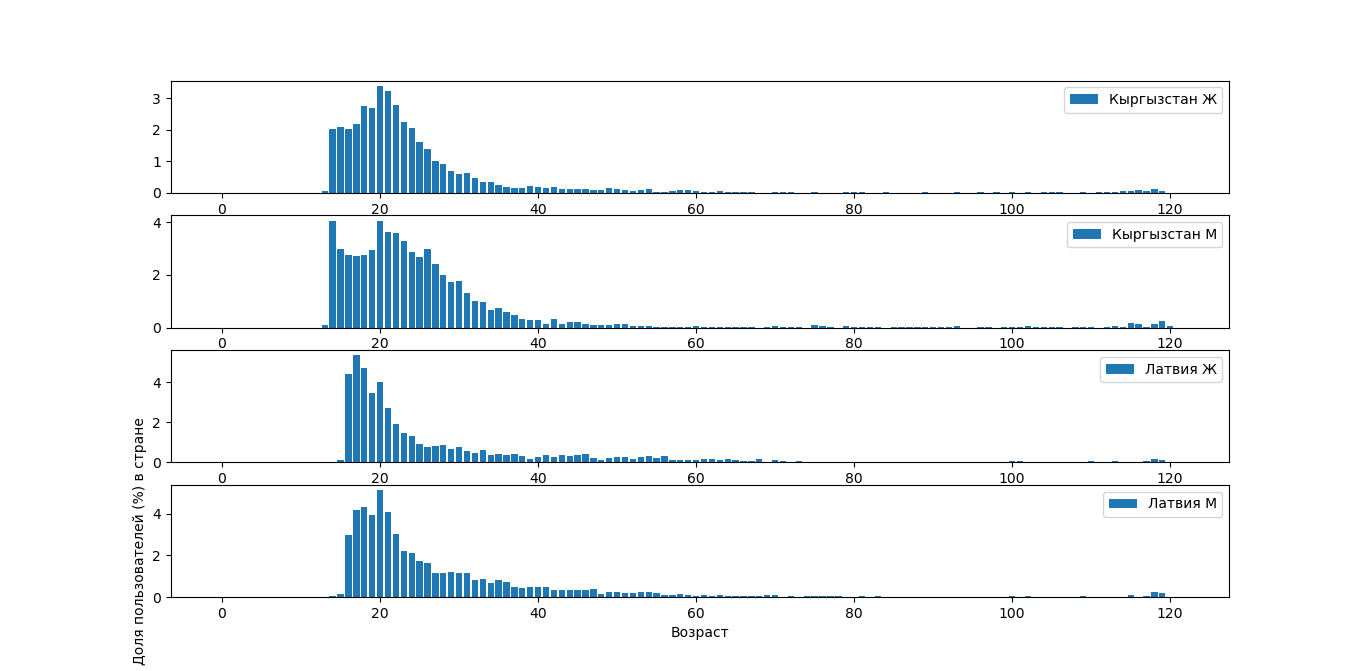
Во всем анализе учитывались пользователи, user\_botscore которых менее 0.3.

Была проведена оценка **состава** аудитории двух стран. В Киргизии общее количество пользователей превышает количество пользователей Латвии: в Киргизии 301526 пользователей мужского пола, 199800 женского; в Латвии 148632 мужского, 125204 женского. Поэтому сравнение шло в процентах от общего числа пользователей страны.



Основной аудиторией ВК в обеих странах являются люди в возрасте от 18 до 30 лет. Вне зависимости от пола пользователя, распределение пользователей примерно одинаковое для обеих стран.

Для выявления **критериев активности** использования сообщений были построены графики зависимости суммы количества личных сообщений и сообщений, отправленных в чаты (messaging\_total\_to\_users + messaging\_total\_to\_chats) относительно следующих потенциальных критериев:

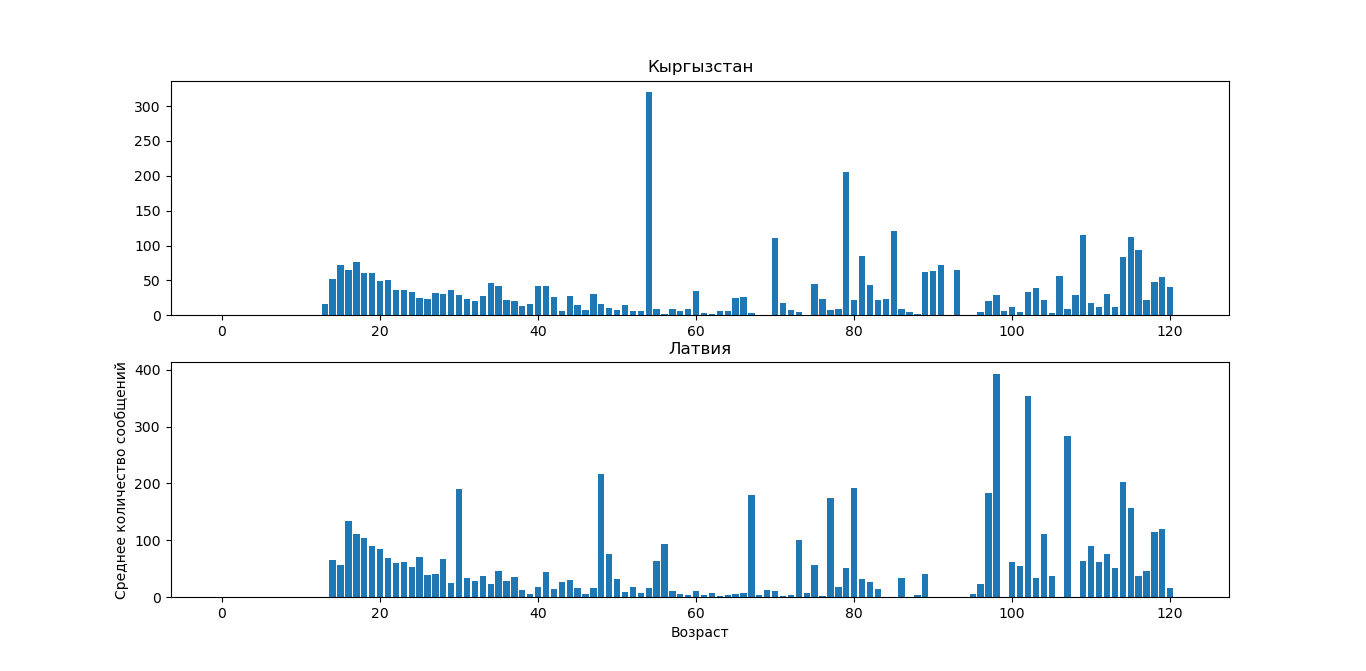
* user\_age
* user\_friends
* количество прочих установленных приложений из списка таблицы
* messaging\_inited\_direct\_with\_reply
* messaging\_inited\_direct\_first\_time\_with\_reply
* messaging\_inited\_chat\_conversations
* calls
* comments\_created
* timespent\_im

Не учитывались сообщения сообществам и от лица сообществ, так как такие сообщения являются “необходимыми”, “формальными”. Также не учитывались пользователи, не написавшие ни одного сообщения.

Далее приводятся графики (**визуализация данных**) зависимости для каждой из стран и **интерпретация** полученных результатов. Разделение по полу не велось, так как оценка состава аудитории показала, что состав аудитории не сильно зависит от пола.

Отдельно графики представлены в той же папке на GitHub.

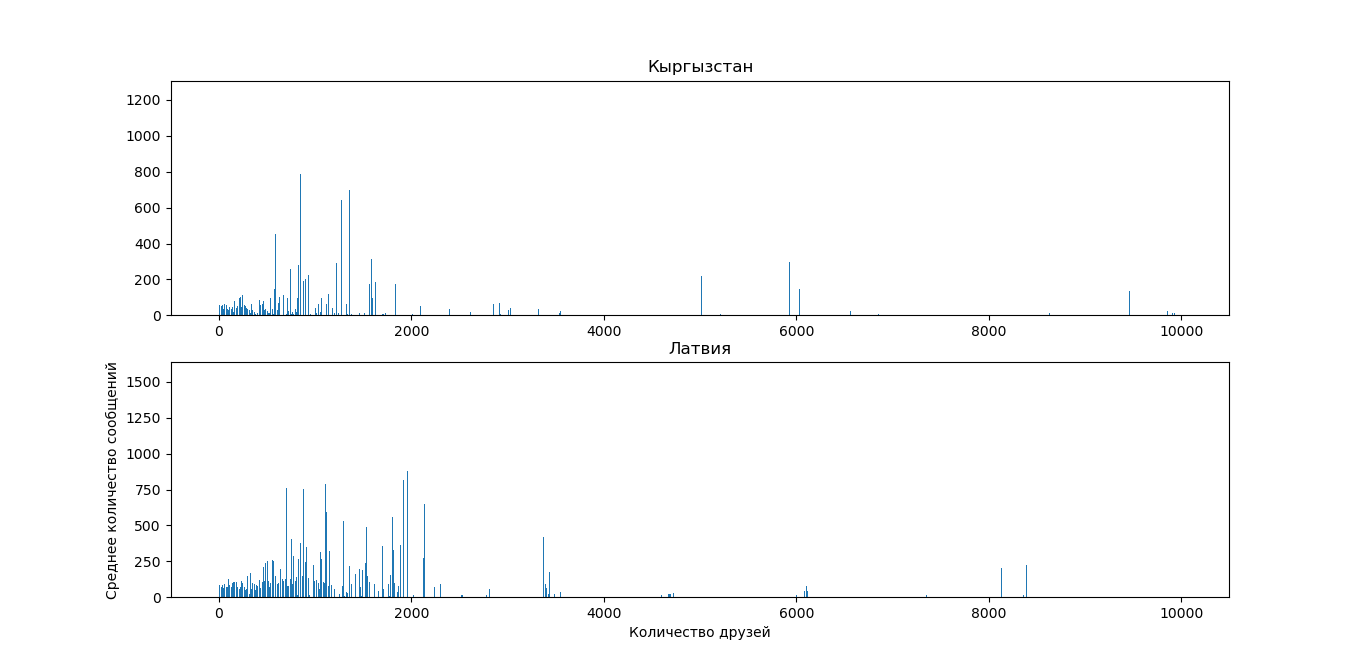
user\_age:



Ожидаемо количество сообщений постепенно падает с ~16 лет до 40-60. Большое количество сообщений на местах скачков графиков и после 100 лет скорее всего связано с указанием ненастоящего возраста или с необнаруженными ботами.

В Латвии в среднем отправляют больше сообщений, чем в Киргизии вне зависимости от возраста.

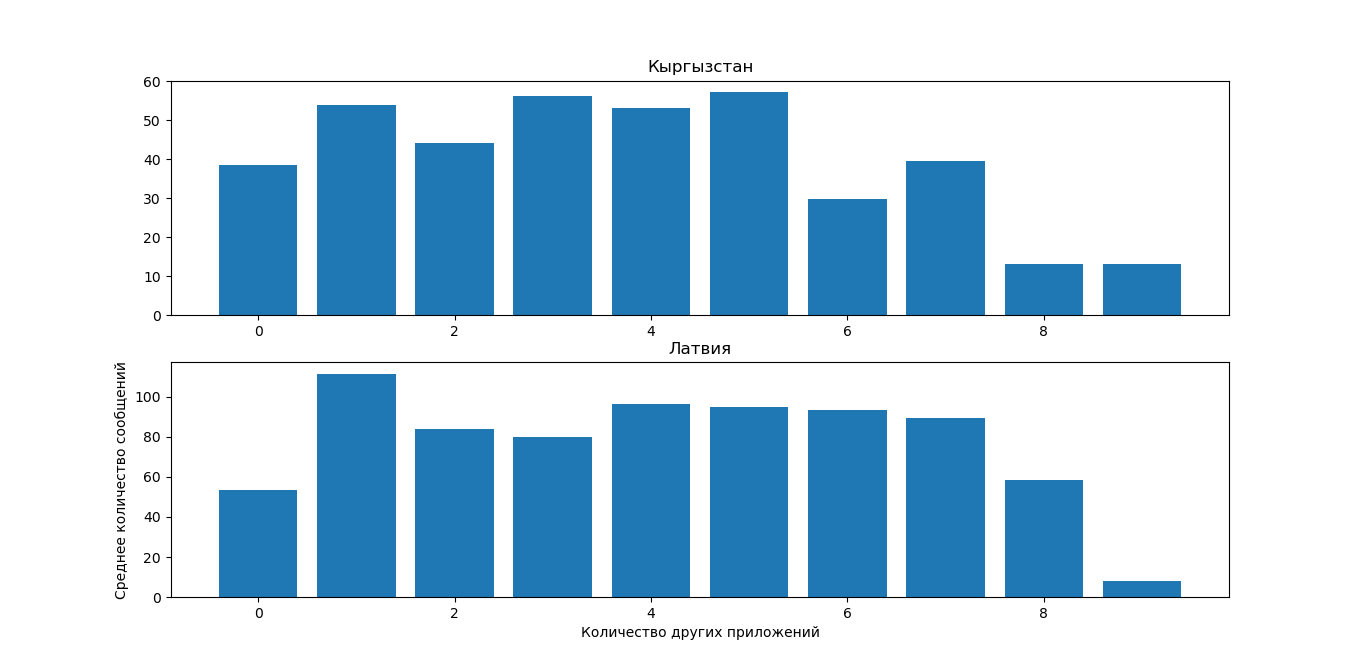
user\_friends:



Количество сообщений почти не зависит от количества друзей. До 1000 друзей количество сообщений не меняется. После 1000 количество резко возрастает, что, вероятно, связано с рекламными или рабочими рассылками.

Относительно данного критерия, в Латвии так же отправляют немного больше сообщений.

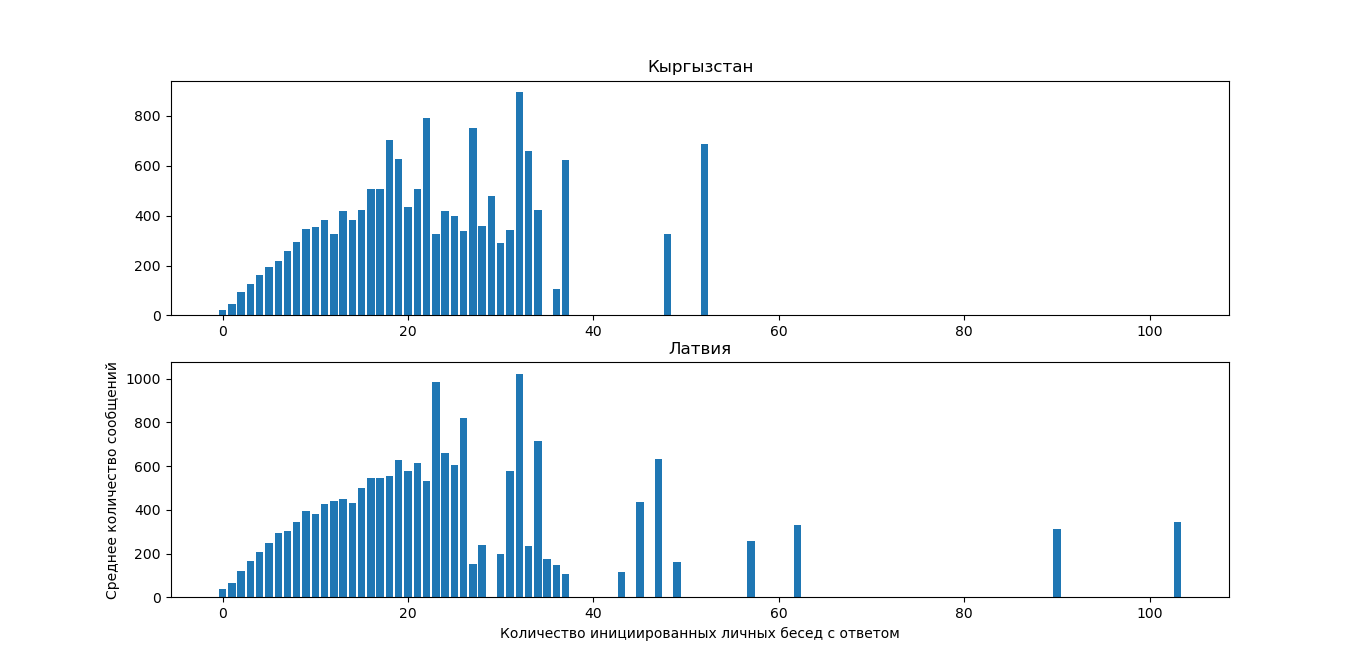
Количество других приложений:



Количество других приложений почти не влияет на количество сообщений. Только при большом количестве других мессенджеров количество сообщений в ВК начинает падать, что можно объяснить использованием других мессенджеров.

В Киргизии спад начинается с 6 приложений, а в Латвии с 8. Также можно обратить внимание, что пользователи, не имеющие прочих приложений, пишут реже.

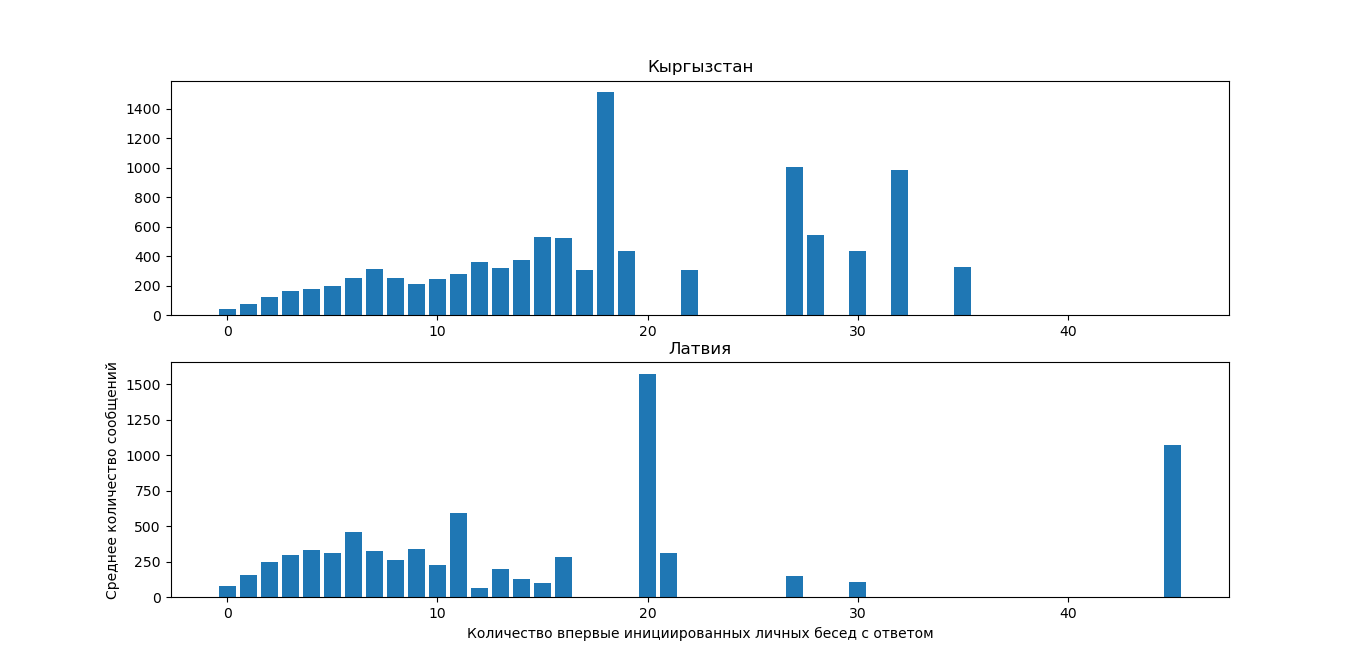
messaging\_inited\_direct\_with\_reply:



Более показательной зависимостью оказывается зависимость по данному параметру. Видно, что чем пользователь чаще пишет первым, тем больше он пишет сообщений в общем. Рост количества сообщений наблюдается примерно до 23 инициированных бесед. После 40 бесед количество сообщений почти нулевое, что может говорить о информационной рассылке.

Для Латвии и Киргизии показатели примерно одинаковы.

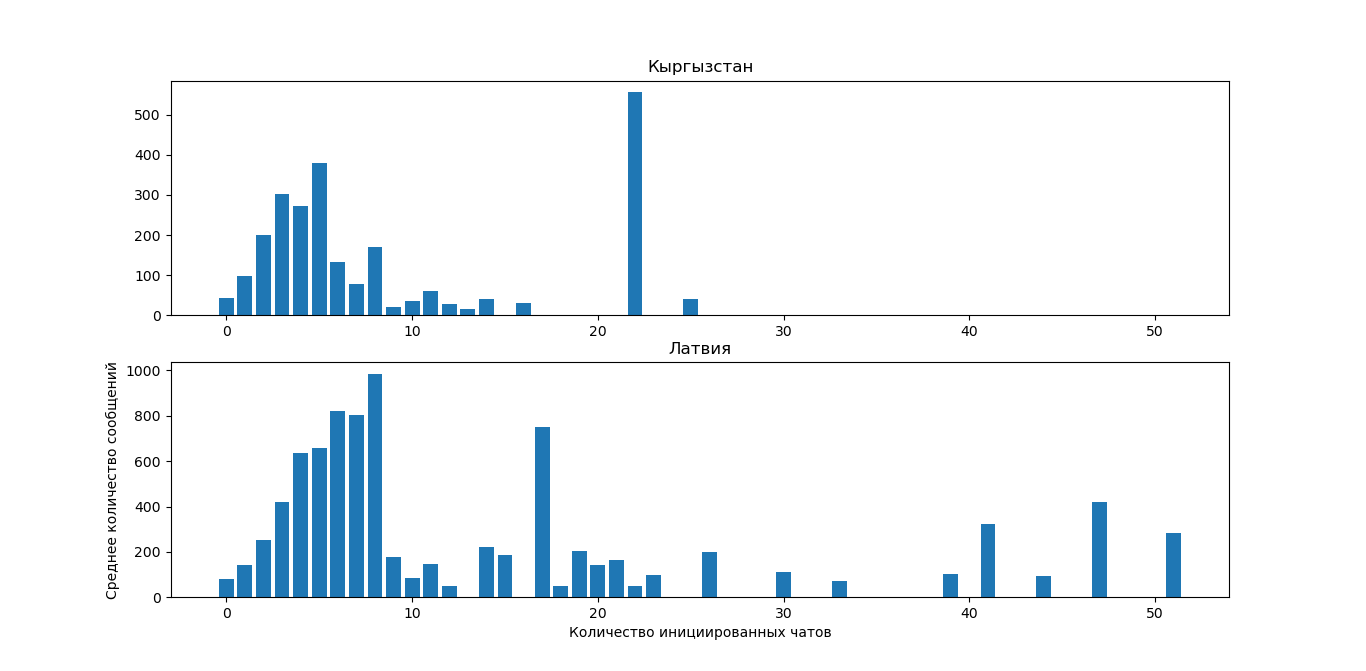
messaging\_inited\_direct\_first\_time\_with\_reply:



Количество впервые инициированных диалогов немного увеличивает количество сообщений.

Для Киргизии увеличение идет до 20 диалогов, для Латвии до 5, а потом идет на спад.

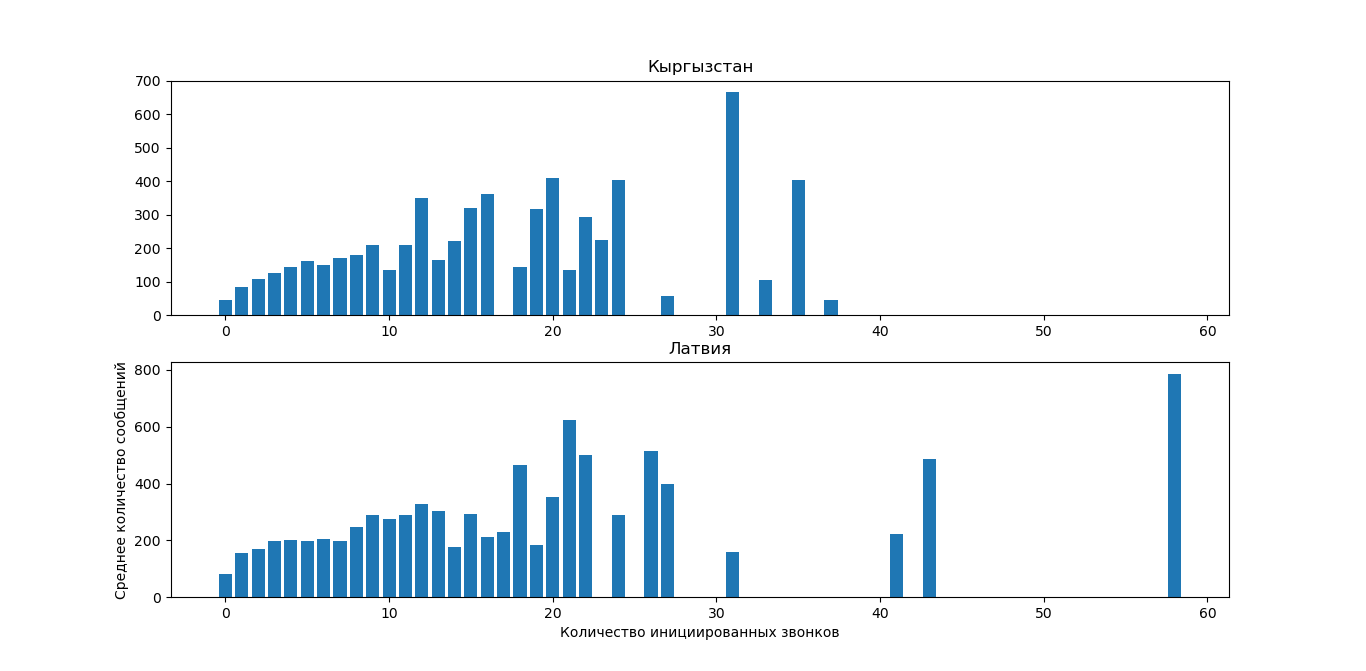
messaging\_inited\_chat\_conversations:



Количество инициированных чатов значительно повышает количество сообщений.

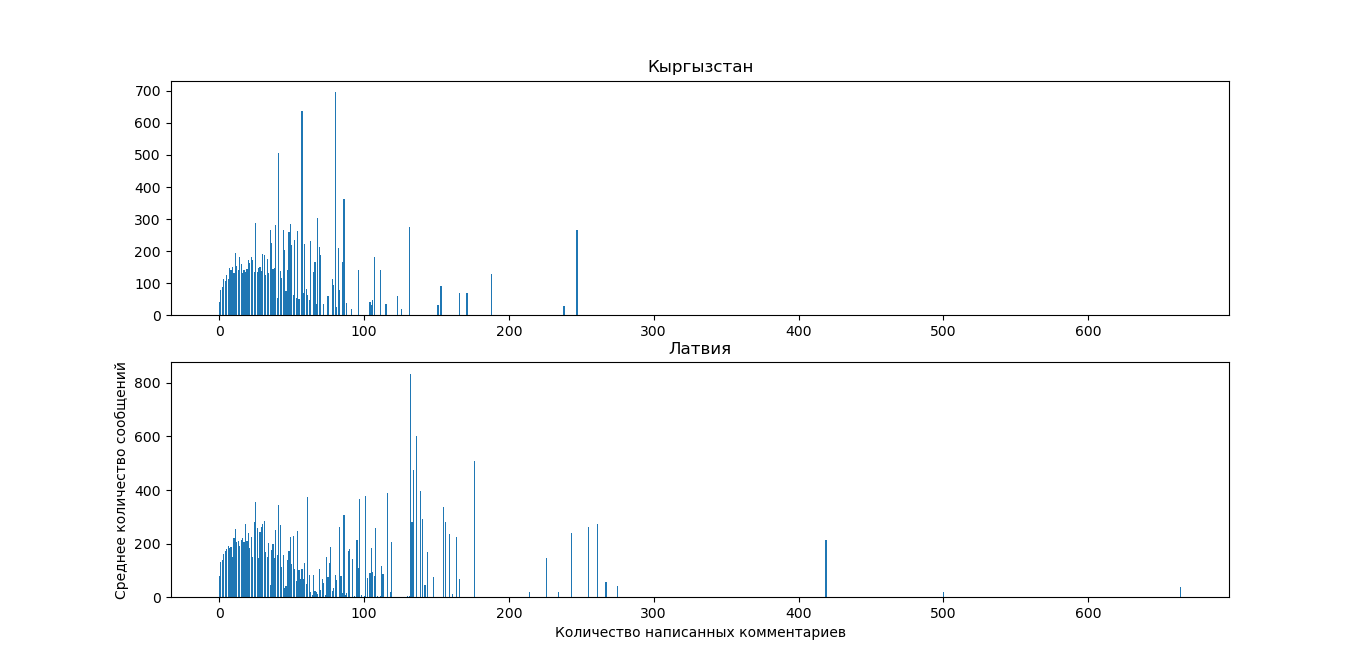
Для Киргизии повышение идет до 5 чатов, для Латвии до 8.

calls:



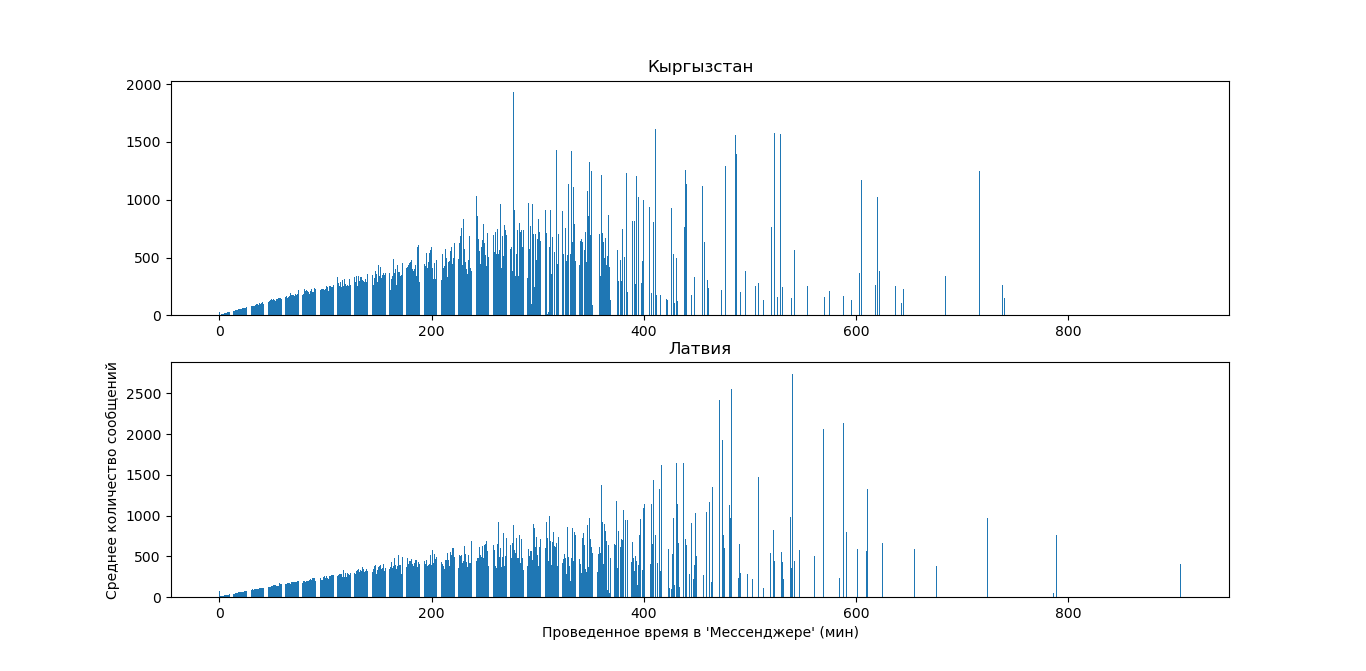
Количество инициированных звонков почти не влияет на количество сообщений.

comments\_created:



Количество комментариев так же несильно влияет на количество сообщений, кроме значений у нуля.

timespent\_im:



Время, проведенное в мессенджере показывает почти линейную зависимость количества сообщений от данного критерия, что для Латвии, что для Киргизии.

В общем можно подвести, что наиболее значимые критерии активности использования сообщений - это messaging\_inited\_direct\_with\_reply, messaging\_inited\_chat\_conversations и timespent\_im. В среднем в Латвии пишут больше сообщений, чем в Киргизии, но зависимость активности использования сообщений от критериев одинаковая.

В качестве **дальнейшего направления исследований** можно провести то же исследование с использованием нейронных сетей. На вход сети можно подавать сразу все потенциальные критерии активности, а на выходе ожидать messaging\_total\_to\_users и messaging\_total\_to\_chats. Такой подход позволит оценить зависимость количества сообщений не от одного параметра, а от всех сразу.

Подавать данные всех пользователей нельзя. Как было видно из графиков данного исследования, некоторые пользователи, не определенные как боты отправляли неестественно большое количество сообщений.

Также можно исследовать время, когда используется мессенджер, наиболее используемые слова и смайлы. Влияние параметров принятых сообщений (длина, количество, содержимое или тематика содержимого) на количество отправленных.