ЦПУР

1. Изначально имеем ~200 млн. сообществ ВК (все живые сообщества). Фильтруем эти сообщества по двум критериям:
    a. Сообщество имеет 100 и больше подписчиков (остается ~10 млн. сообществ).
    b. Сообщество имеет хотя бы один пост за последний месяц. На этом этапе качаем первые 100 постов каждого сообщества и проверяем дату публикации (остается 1.83 млн сообществ).
2. Сокращаем выборку - случайно выбираем 18 тыс. сообществ из 1.83 млн.
3. Качаем у них стены за год (~9.5 миллиона постов).
4. Считаем статистику по публикациям:
    a. Сумма всех количественных показателей по месяцам (лайки, просмотры, репосты, комментарии).
    b. На основе месячной статистики считаем среднемесячные количественные показатели для каждого сообщества.
5. Убираем из выборки сообщества, у которых в среднем меньше 2 лайков (квартиль) в месяц (остается ~7 млн. постов от 12 тыс. сообществ).
6. Качаем лайки у 7 млн. постов (~5.4 млн. объектов).
7. Среди 7 млн. постов находим полнотекстовым поиском те, в которых присутствует тематика вакцинации (~14 тыс. постов).
8. Качаем всех подписчиков у 18 тыс. сообществ (~34 млн. валидных пользователей).
9. Считаем социально-демографическую статистику по пользователям для каждого сообщества (пол, возрастная группа, статус обучения в вузе, семейное положение, регион, факт указания университета).

В итоге получились следующие таблицы:

- Профили сообществ, имеющие более 100 подписчиков и хотя бы один пост за последний месяц - *1.83 млн. уникальных записей, период выгрузки - август 2021*
- Профили сообществ, выбранных случайно из 1.83 млн. - *18 тыс. уникальных записей*
- Посты 18 тыс. сообществ за год - *9.5 млн. уникальных записей, дата и время последнего поста - 2021-09-03 11:33:52*
- Лайки к постам 18 тыс. сообществ - *5.4 млн. уникальных записей, период выгрузки - сентябрь 2021*
- Профили подписчиков 18 тыс. сообществ - *34 млн. уникальных записей, период выгрузки - август 2021*

Формат извлечённых данных:
https://drive.google.com/drive/folders/15y5QwKcAIbQsELyJNz-oodvtW53znex8?usp=sharing