В этой таблице безымянная частичная выгрузка перемещения пользователей по сайту. Один user\_id — один пользователь. Расскажите нам всё, что сможете понять по такой выгрузке

Проведите когортный анализ (за основу для когорты можно взять неделю регистрации), подсчитайте ориентировочный срок жизни клиента/когорты оцените популярность отдельных материалов и форматов

Попробуйте определить общие паттерны у тех пользователей, кто перестал учиться на платформе, и у тех, кто продолжает учиться. Вероятно, на основе этих данных вы сможете сделать предложения по адаптации платформы? Каких данных вам не хватает, чтобы провести более детальный анализ?

Выгрузка неполная, поэтому результаты могут быть искажены. Нам важны не итоговые значения, а ход решения, который вы выберите

**Решение**.

Судя по датасету, очень много пользователей уходит сразу после дня использования. Мне кажется, это связано с том, что платформа построена так, что требуется регистрация для просмотра контента платформы. Люди регистрируются, а потом понимают, что это не то, что они искали и больше мы их не встречаем. Подход с регистрацией хорош для увеличения общего числа пользователей на платформе, но он слабо помогает в увеличении активных пользователей.

При работе с новыми пользователями, было бы неплохо увеличить количество информации о нем для лучшего построения профиля пользователя. Если подход с регистрацией остается, то его можно улучшить предложив авторизацию через соцсети (из них же можно запросить больше демографических данных о пользователе), что позволит лучше предсказывать отток пользователей. В целом пользователь получает более быстрой доступ к платформе. Также, если мы понимаем, что пользователь не заходит на платформу больше недели (его средний срок жизни), то он ушел в отток и его можно попытаться вернуть рассылкой на почту.

Активные же пользователи характеризуются большим вовлечением в просмотр материалов сайта на регулярной основе. То есть видно по retention графику, что отток немного увеличивается, но не существенно по прошествую времени. Их можно попытаться еще больше вовлечь в использование платформы с помощью коллаборативной фильтрации пользователь-пользователь. Мне кажется, что коллаборативная фильтрация типа товар-товар не будет очень эффективной потому что врядли люди склонны смотреть похожие образовательные материалы. Мне кажется, на начальном этапе будет оптимально строить отдельные рекомендации для блогов, курсов, траекторий и видео, что я сделала для примера. В то же время, должно быть пересечение между тематикой разных форматов, поэтому есть смысл строить общие рекомендации, скорее всего используя контентный подход в будущем.

Я пыталась простроить диаграмму потока для анализа основных путей на сайте, но мне не хватило более детализированных данных. Было бы неплохо иметь более точные временные данные, которые бы позволили восстановить цепь событий. Исходя их этого, можно будет понять основные пути по сайту и также приоритизировать оптимизацию, простраивать воронку конверсии в целевые продукты (Эвотор, например, или переход на нужные сайты Сбера).

Для анализа оттока пользователей я вытащила такие показатели как: количество дней до регистрации на платформе, первая неделя посещения, последняя неделя посещения, количество пройденных курсов, траектории, количество просмотренных блогов и видео. Считала, что человек ушел в отток, если не проводил больше недели на платформе. Модель переобучается, потому что на мой взгляд недостаточно данных. Я посмотрела что люди включают в анализ, и судя по всему было бы неплохо подключить больше данных о профиле пользователя, также данные об устройстве с которого он заходит.

В целом я хотела бы отметить, что система платформы пока не предполагает какого-либо повышения ценности продукта с увеличением активности его использования, то есть по продукту следует проанализировать механики геймификации, применяемые на других образовательных площадках и попробовать их применить.