



> ПРОДУКТОВЫЕ МЕТРИКИ

> Зачем нужен Retention?

Мы можем много смотреть на метрики активности пользователей в разные временные промежутки — на MAU, DAU, WAU. Радоваться тому, какие они большие и как они (потенциально) растут со временем, понимая, что это означает большую вовлеченность наших пользователей... Или нет? Почему в интернете можно найти много постов на тему того, что всё это — «метрики тщеславия», пользы от которых мало?

Проблема в том, что эта «активность пользователей» может не означать **реального интереса** пользователей к нашему продукту. Рост этих метрик может быть искусственным и не отражать **долговременную вовлечённость**. Иными словами, это легко манипулируемые, «gameable» метрики.

Примеры того, как мы можем увеличить активность пользователей, при этом потенциально пустив бизнес под откос:

1. Завлекать пользователей в наш продукт деньгами («установи приложение и получи 1000 рублей в подарок»).
2. Завлекать их обманом или дешёвым эпатажем — по аналогии с clickbait и hate clicks
3. Накидать ботов в наш продукт, создавая видимую активность.
4. «Подписать» человека на наш продукт без его желания .

Все эти варианты могут привести к (видимой) увеличенной активности, но реально приводить к оттоку наших клиентов и даже гибели бизнеса. Можете также придумать свои варианты, когда такое может случиться! Подумайте в сторону технических проблем и product-market fit.

Если же мы хотим понять, сколько пользователей реально остаётся и продолжает пользоваться продуктом, мы должны посчитать **retention** (он же **retention rate**). Часто можно видеть следующую формулу:

$$RR = \frac{\text{количество клиентов на конец периода} - \text{количество новых клиентов за период}}{\text{количество клиентов в начале периода}} * 100$$

Но не все используют именно этот вариант формулы — можно встретить и другие. Например, часто встречается формулировка «количество активных пользователей за текущий период/общее число пользователей за прошлый период». Или «число пользователей за n-ый период/число пользователей в когорте». **Эти варианты больше соответствуют тому, как Толя считал retention в лекции :)**

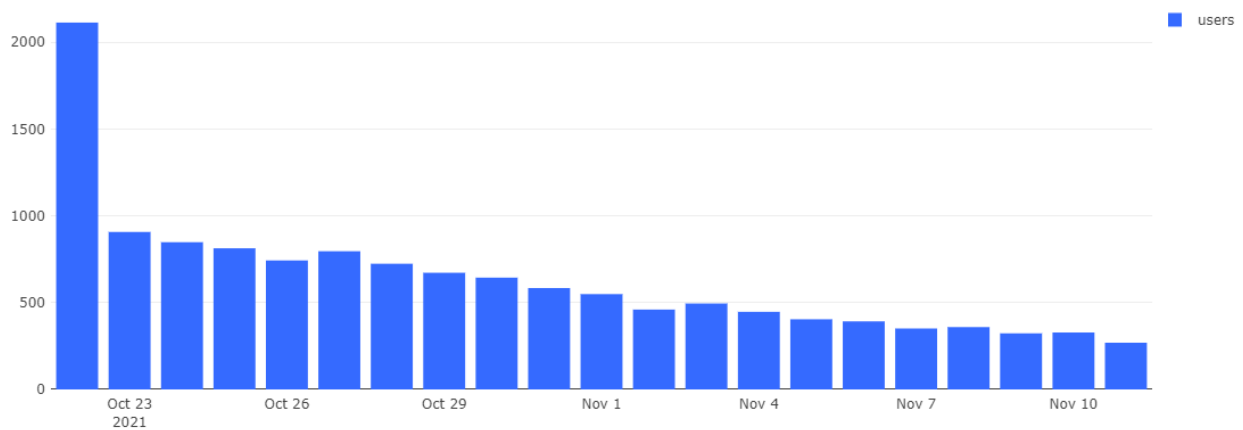
Есть и другие типы retention. Более того, можно считать не остающихся клиентов, а тех, которые уходят — это будет называться churn rate.

Внимание! Потенциально манипулировать можно любой метрикой, и retention не исключение. Помните про закон Гудхарта и будьте осторожны.

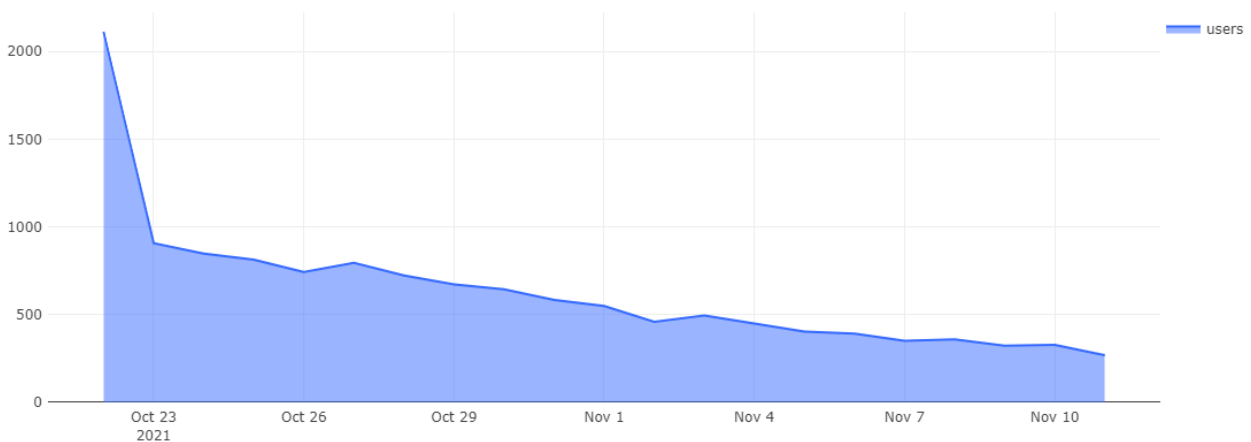
Внимание (2)! Retention не всегда бывает полезной метрикой. Например, если мы продаём **товары для маленьких детей**, то подавляющее большинство покупателей к нам не будет возвращаться, и это нормально :)

> Как изобразить Retention?

На лекции Толя изображал его через столбиковую диаграмму, меняющуюся во времени:



Естественно, вместо столбиков можно рисовать и более традиционные линии. При желании можно даже закрасить пространство под этой линией и сделать area chart:

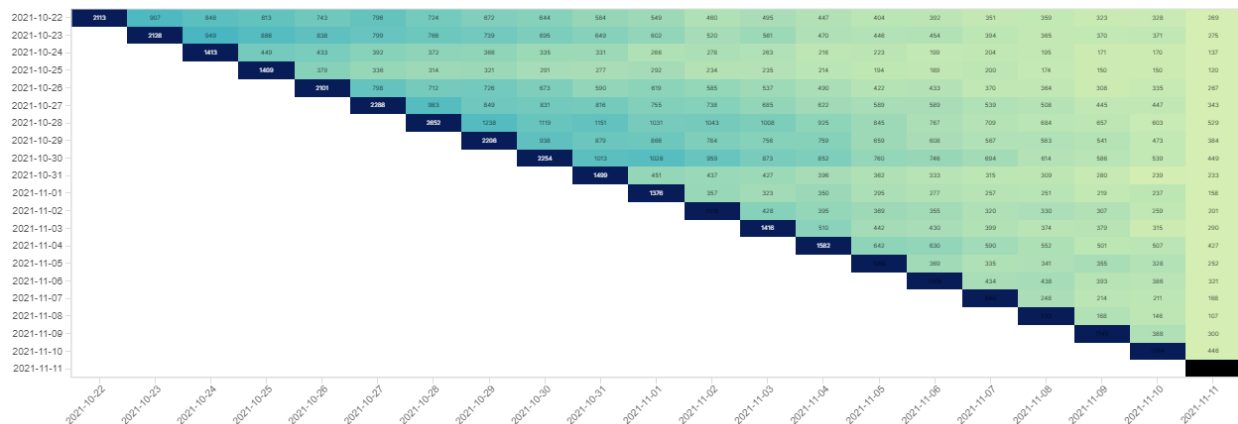


Но отдельные линии без закрашивания удобнее, если мы хотим визуализировать несколько **когорт**. Когорты — это группы наших пользователей, которые совершили определённое действие в определённый отрезок времени.

Иными словами, мы можем визуализировать когорты людей, впервые проявившие активность в определённом месяце, и сравнить их динамику между собой:



Однако наибольшую популярность в индустрии получила визуализация когорт через тепловые карты:

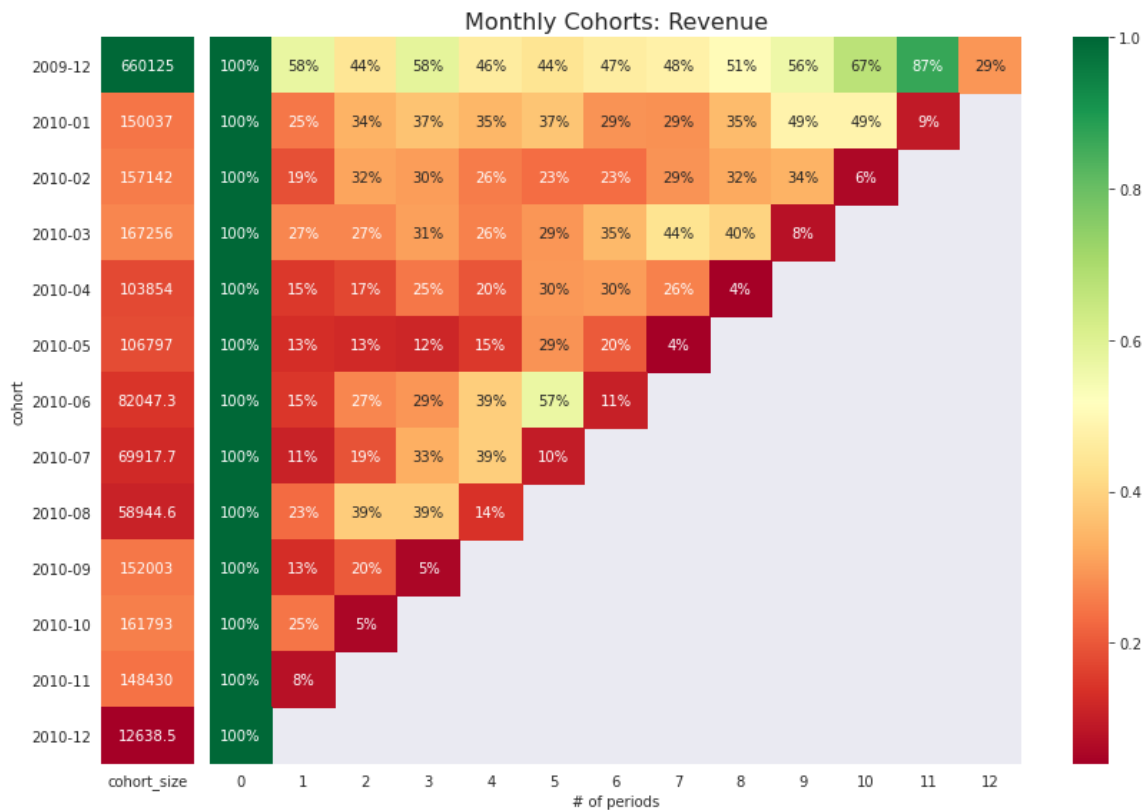


Как это интерпретировать?

1. По оси x — **шкала времени**. Здесь мы видим каждый новый день на определённом промежутке.
2. По оси y — **день первого появления**. По ней мы можем отследить, с какого дня начинается конкретная когорта.
3. Этому же соответствуют темно-синие ячейки на диагонали. Так, в верхнем левом углу мы видим пользователей, впервые начавших пользоваться нашим сервисом **2021-10-22**. Ниже по диагонали — те, кто появился **2021-10-23**. И так вплоть до **2021-11-11**, где наш график заканчивается.

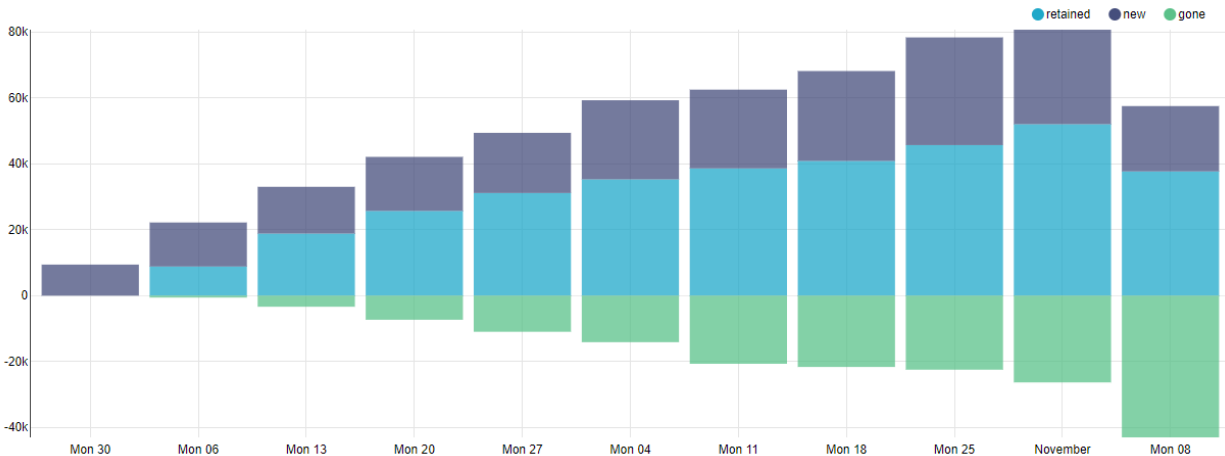
- По цвету каждой ячейки можно понять, насколько много пользователей «удержалось» по сравнению с первым днём. В данном случае чем **светлее** — тем **меньше** человек осталось в когорте. Естественно, это зависит от выбранной вами **цветовой палитры**.
- Если вам нужны конкретные значения, то в каждой ячейке можно посмотреть точное количество человек в конкретный день + их процент от изначального количества. Само собой, в первый день когорты этот процент равен 100 %)
- Очень важный момент — чтобы всё это было наглядно, нужна **нормировка тепловой карты** по колонкам. Различие между нормировкой по рядам и колонкам можно видеть тут.

Можно также встретить варианты тепловой карты вроде такой:



Из наиболее важных особенностей — карта «отзеркалена» по сравнению с прошлым вариантом, а по оси x указаны **номера периодов**, а не конкретные даты. Можете попробовать сделать что-то похожее в Superset :)

Также у Толи был вот такой график, нарисовать который вам предстоит самостоятельно:



На что стоит обратить внимание:

- Здесь указана **недельная** динамика, рисуются не все дни (так компактнее);
- В оригинальном запросе использовалось среднее количество юзеров за каждую неделю и каждую **группу**, однако вполне можно считать и просто **количество уникальных пользователей**;
- **Группы** пользователей (оставшиеся, новые, ушедшие) выделялись до рисования графика — вам придётся поработать в SQL;
- «Ушедшие» пользователи имеют отрицательные значения на этой шкале.

> Дополнительно про анализ

Анализировать данные об «удержании» (retention) и «оттоке» (churn) можно множеством разных способов в зависимости от поставленной гипотезы. Среди более продвинутых вариантов отдельно стоит выделить анализ выживаемости.

Изначально этот метод применялся в медицине для анализа факторов, влияющих на продолжительность жизни пациентов. При этом он учитывает тот факт, что пациенты могут выпадать из исследования ввиду своей смерти — отсюда и название.

Никто не заставляет использовать именно смерть в качестве возможного исхода! Мы можем анализировать **время до любого события, которое имеет бинарный исход** (произошло или нет).

Например, прекратил ли человек пользоваться сервисом или нет.

Пример подобного анализа можно видеть по ссылке вверху, а также в нашем открытом вебинаре по **обобщённым линейным моделям**. Модели такого рода начинают обсуждаться примерно с 1:05:00 :)

Использованный пакет: lifelines