



Чтобы не терять деньги:

оповещения о падениях продуктовых метрик

Иван Еремеев
руководитель направления аналитики контента в Дзен

Давайте знакомиться



Иван Еремеев

Руководитель направления аналитики
контента в Дзен (VK)

ex Руководитель направления продуктовой аналитики
в рексистемах VK (Пuls, Relap, Медиапроекты)

Ведущий преподаватель и со-автор курсов по
прикладной аналитике данных, a/b тестированию,
веб-аналитике, статистическим методам анализа
данных и продуктовой аналитике



 **образование**

 **GeekBrains**

 **МФТИ**



Команда проекта



Иван Еремеев



Руководитель
направления
аналитики контента
в Дзен



Татьяна Гайченкова



Продуктовый
аналитик,
Медиапроекты
mail.ru(vk)



Сергей Нефедов



Продуктовый
аналитик Дзен



1 Проблематика, польза для бизнеса и почему мы не стали использовать готовые решения

2 Поиск наилучшего подхода

Сравнение текущего периода с предыдущим

- Скользящие границы
- ML-методы

3 Инфра

4 Интерфейс

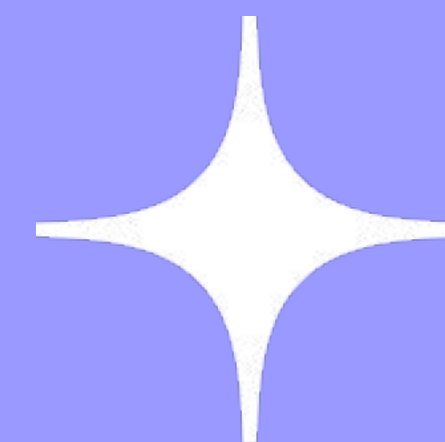
Redash

- Почтовые оповещения
- Telegram-бот

5 Эффект в деньгах

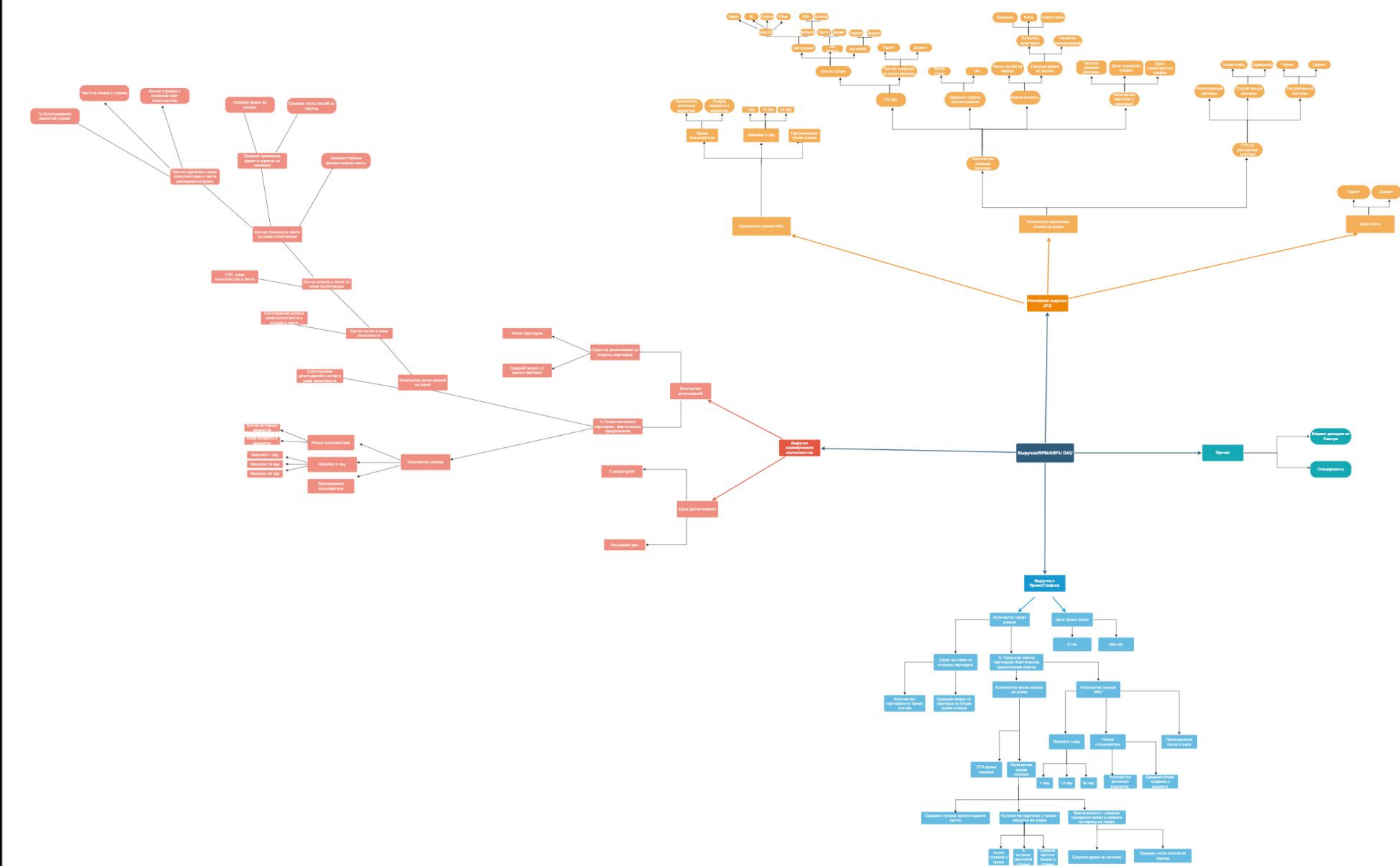


Проблематика, польза для бизнеса и почему
мы не стали использовать готовые решения



Почему мы задумались об алертах?

Десятки метрик и сотни срезов за которым невозможно постоянно следить вручную, можно легко упустить из виду важные изменения метрик, сигнализирующее о проблеме



Почему нам не подошли готовые решения?



У готовых инструментов есть ряд минусов:

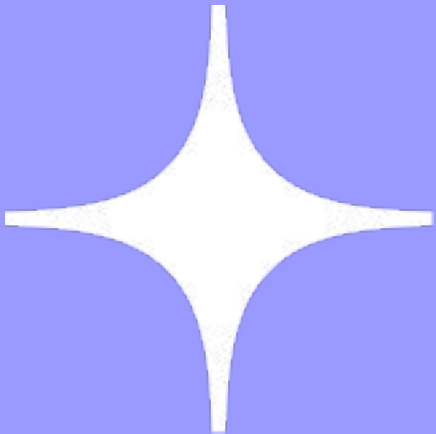
- Сложность кастомизировать под особенности бизнеса и потребности команды, недостаточная гибкость настройки
- Дороговизна внедрения некоторых инструментов
- Сложность перехода на новую инфру и ее поддержки
- Могут приостановить сотрудничество с российскими продуктами из-за санкций



Польза бизнесу / пользователям от внедрения проекта

- ✦ Какой импакт на бизнес? Позволяет сохранять несколько миллионов рублей в месяц за счет более быстрой реакции на инциденты. Автоматизация ручного труда, экономия времени для более творческих задач
- ✦ Кто пользователи? Внутренние команды рексистем (в частности продуктовые аналитики и менеджеры)
- ✦ В дальнейшем возможно развитие продукта для других бизнес-юнитов

Алгоритм



Дзен



Как мы подступались к проблеме?

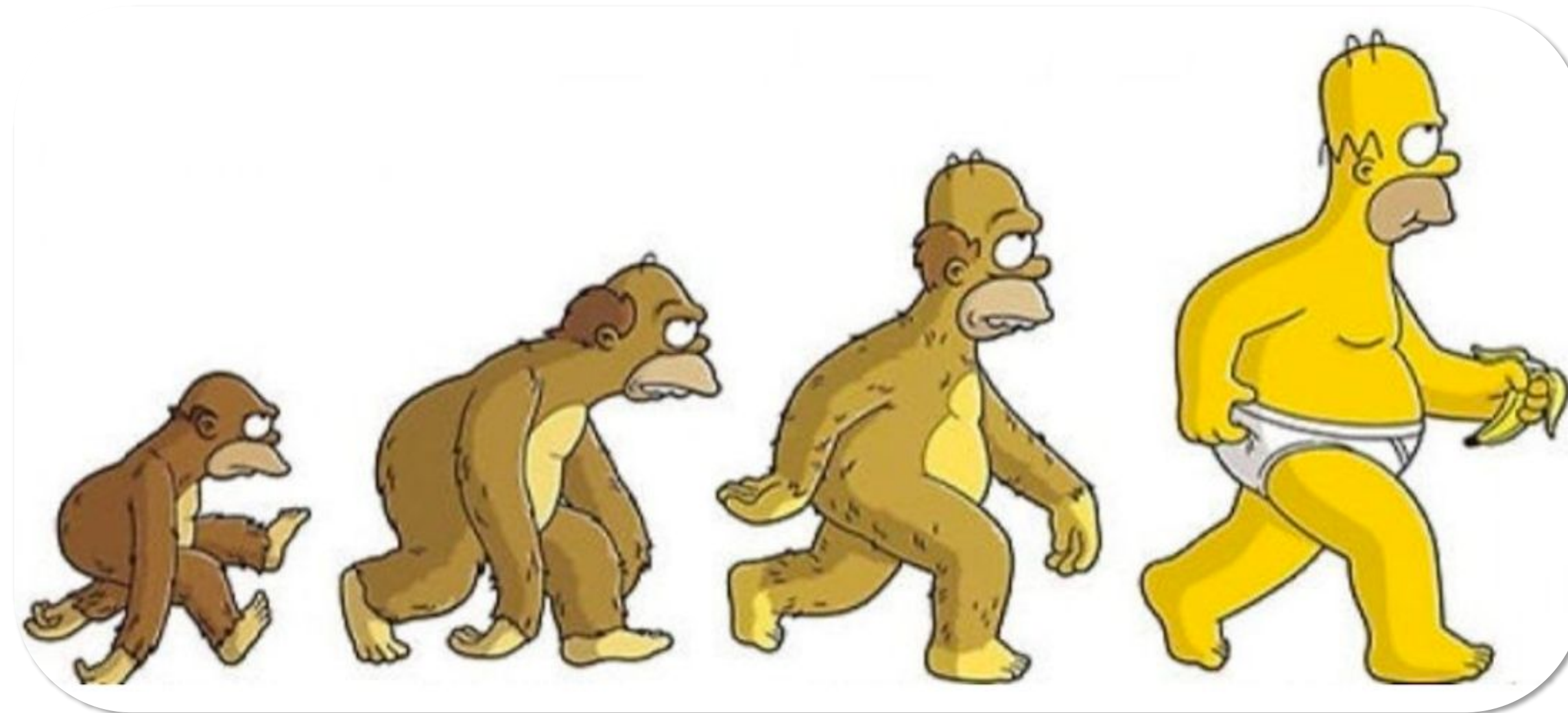


Мы задали себе вопросы:

- Что считать аномалиями в наших метриках?
- Когда присылать оповещения об их изменениях?
- Как найти оптимальную периодичность, чтобы не слишком надоедать коллегам, но при этом не упустить критичные отклонения?



Пробовали поочередно различные подходы,
пока не нашли оптимальный для нас



Сравнение
текущего периода
с предыдущим

Скользящее
среднее

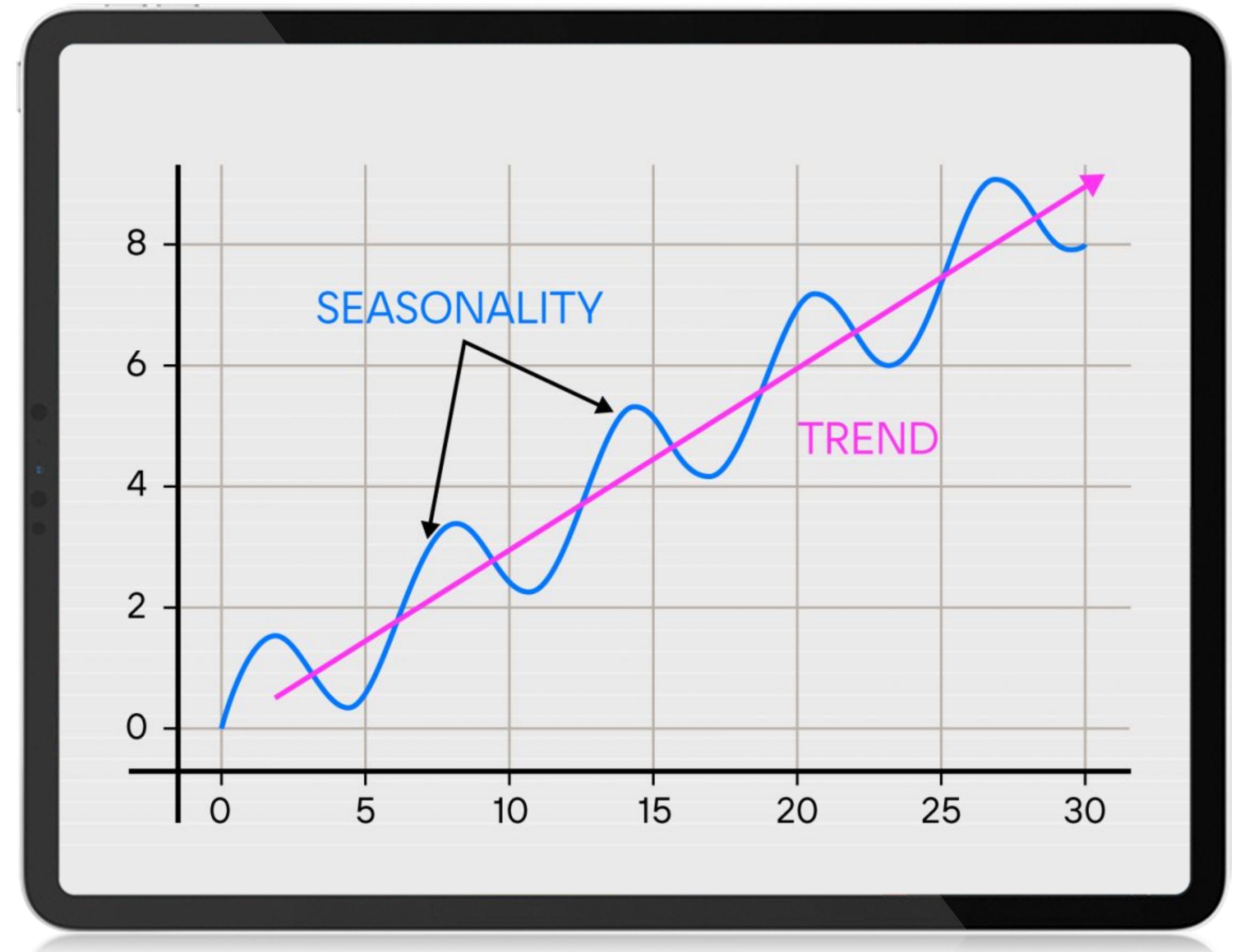
Скользящие
границы

ML-библиотеки

Что такое временной ряд?

Временной ряд — это последовательность значений некоторой переменной (или переменных), регистрируемой через определённые промежутки времени (регулярные или нерегулярные). Любой временной ряд можно разложить на компоненты:

- **тренд** — функция, описывающая рост или падение;
- **сезонность** — отвечает за колебания в данных в течение года, недели, суток и т.д.;
- **праздники** или другие события, которые могут существенно влиять на поведение временного ряда;
- **шум** — случайные возмущения



Сравнение текущего периода с предыдущим

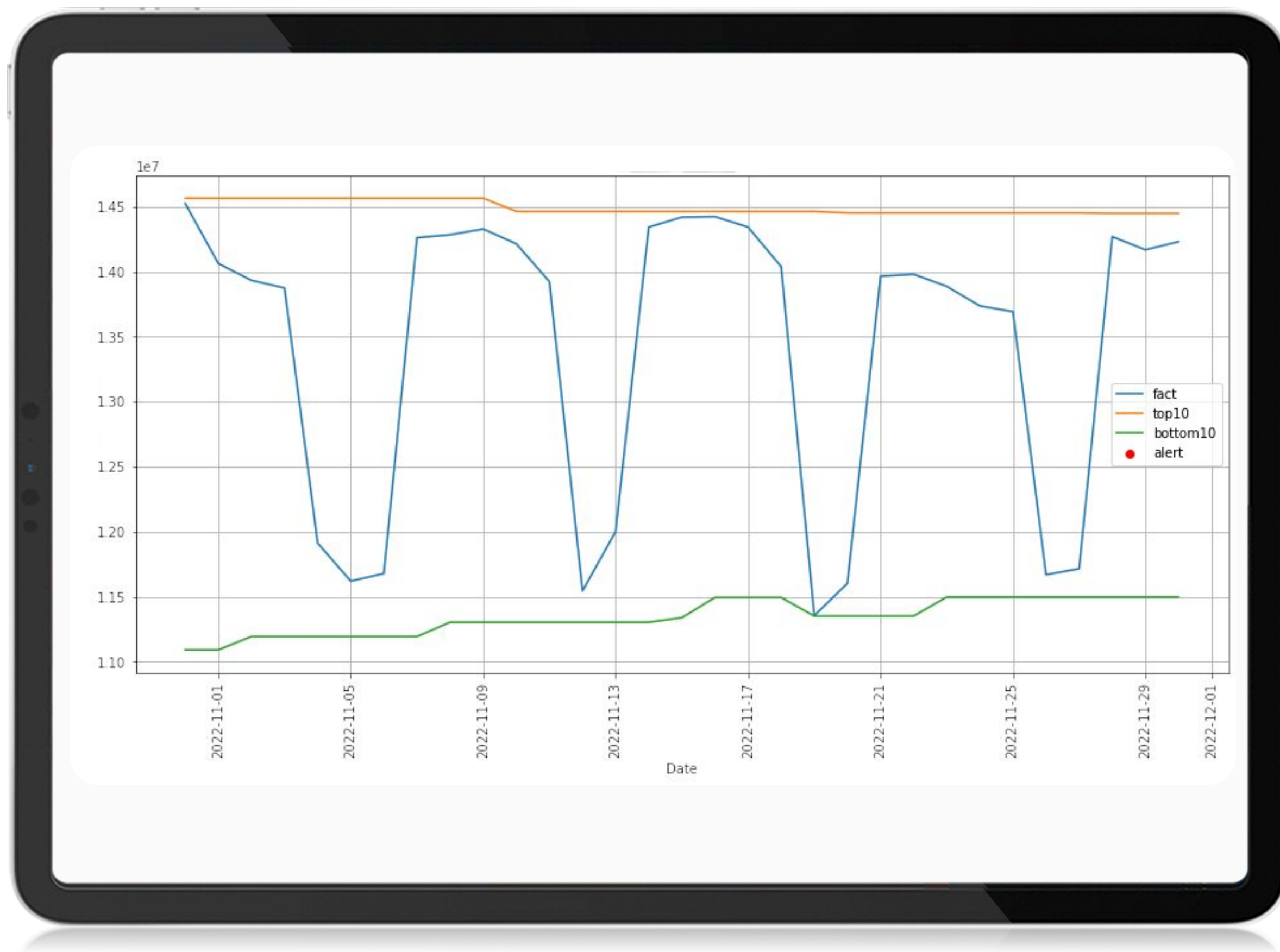


Простота — главное достоинство подхода, но:

- он не учитывает нормально весь спектр сезонности метрик и тренды(изменения за год)
- нам слишком часто присылались ложноположительные (false positive) оповещения, и мы решили пробовать дальше

Скользящие границы

Тогда мы решили модифицировать подход и рассчитывать для каждого показателя скользящие верхнюю и нижнюю границы (например, 90-й и 10-й перцентили распределения метрик), а также пробовали считать расхождения по стандартным отклонениям (например, брали $\pm 2\sigma$, $\pm 3\sigma$), чтобы при выходе метрики за эти пределы срабатывало оповещение

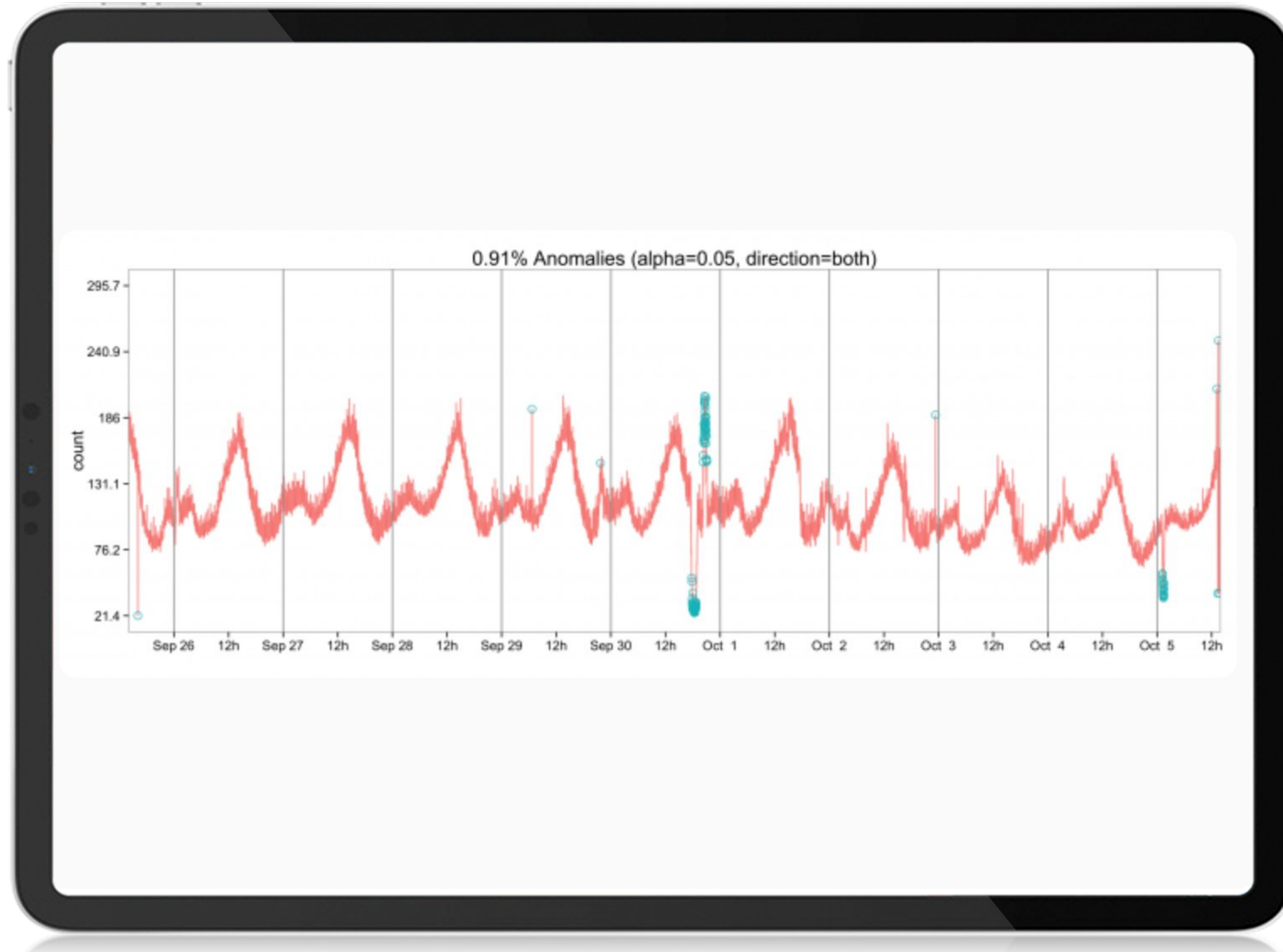


Простота — главное достоинство подхода

- Основной недостаток здесь — низкая чувствительность, то есть можно не заметить не только малые колебания метрики, но также и относительно большие.



Twitter Anomaly Detection



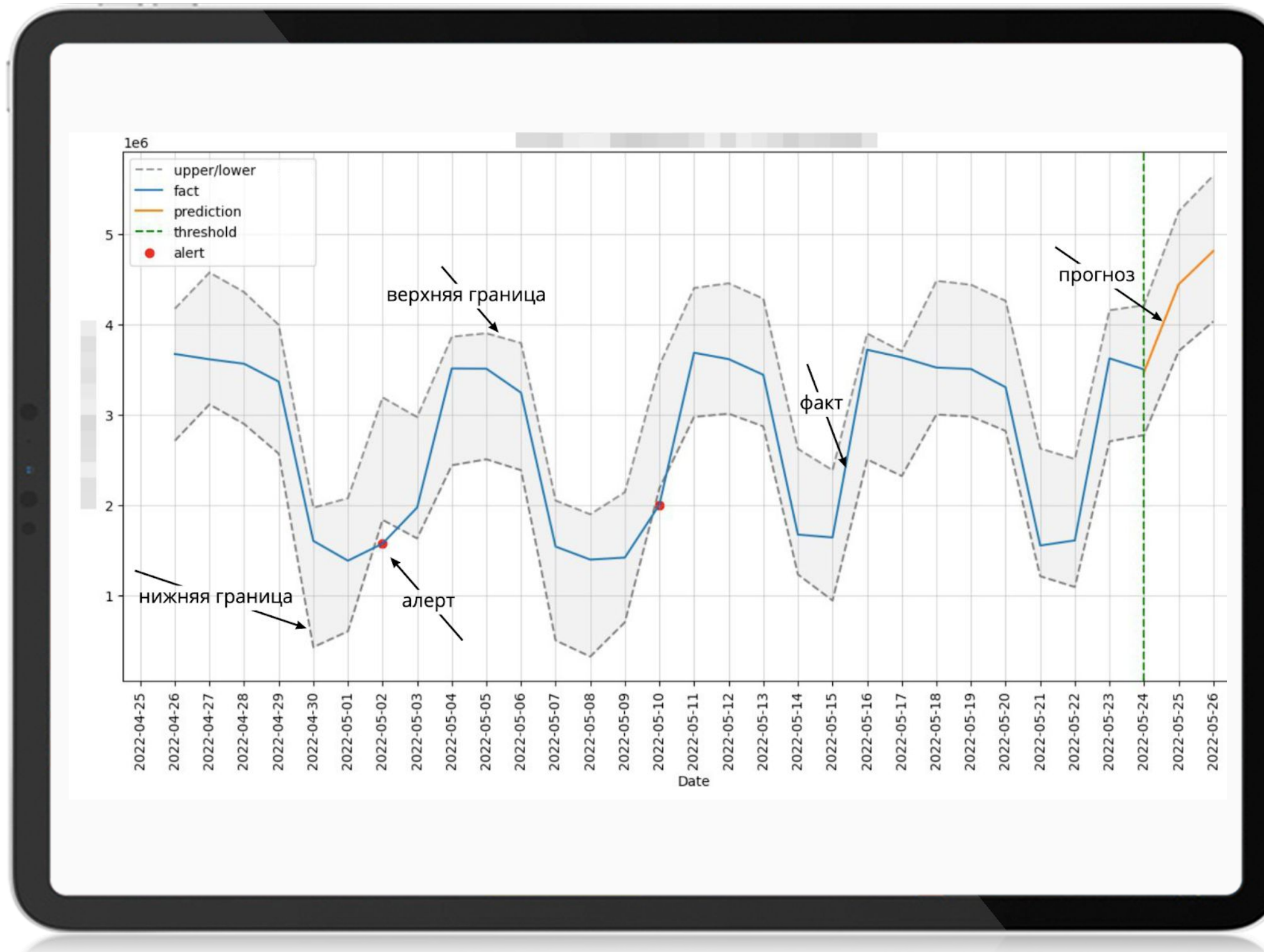
Для конечного пользователя работает по принципу black box

Излишне высокая чувствительность, пользователи устают от слишком частых уведомлений и перестают пользоваться системой

Идея синтетического контроля на основе прогноза

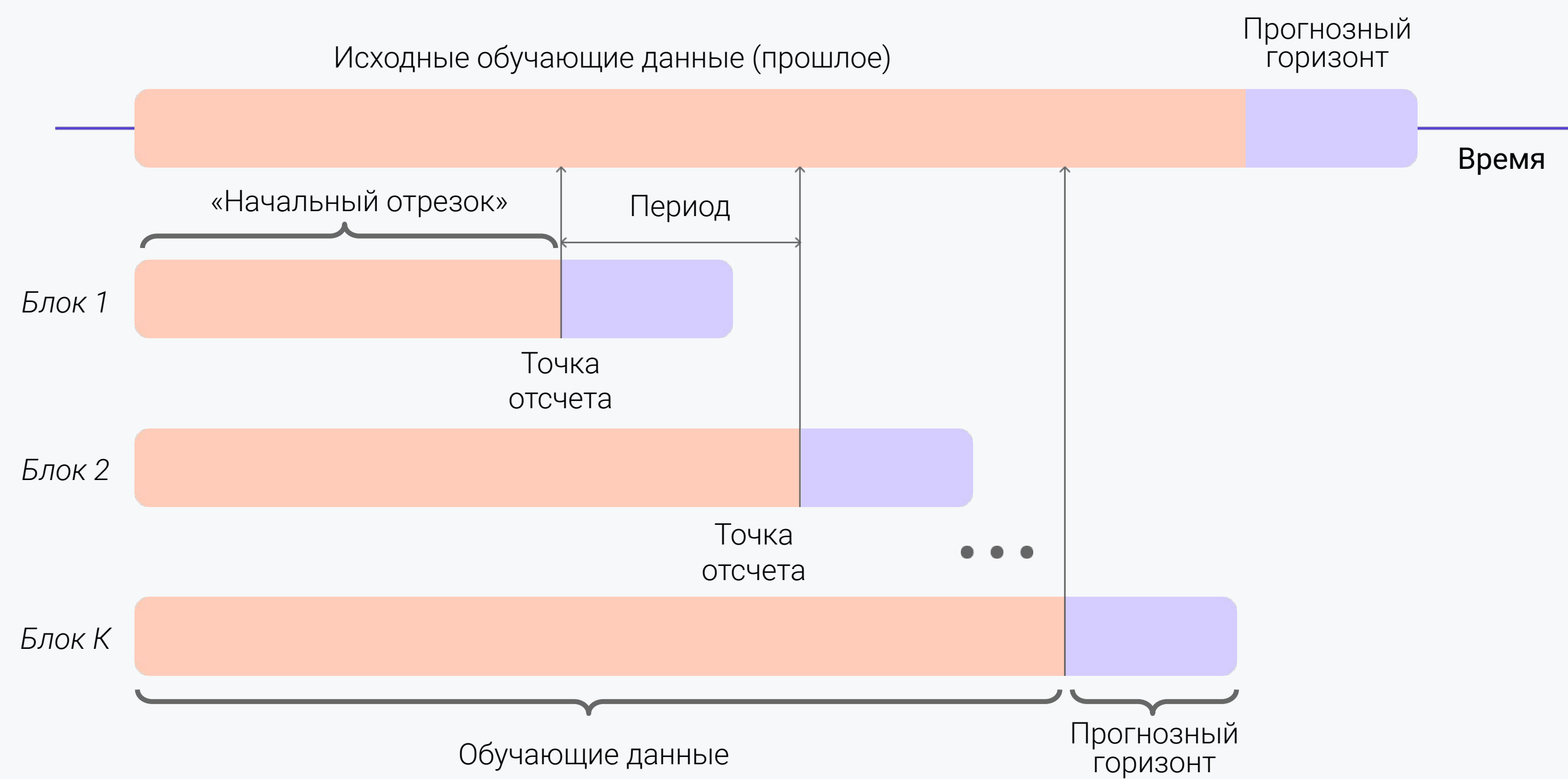
MAPE — относительная величина, что позволяет сравнивать между собой ошибки прогнозирования различных метрик (например, DAU и клики на уникального пользователя)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \cdot 100\%$$

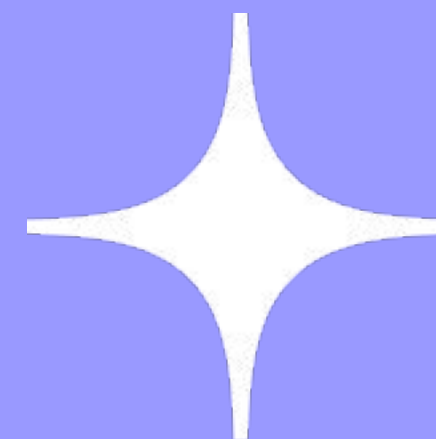


Выбрали библиотеку [Prophet](#)

Кросс-валидация с помощью метода “имитированных исторических прогнозов”

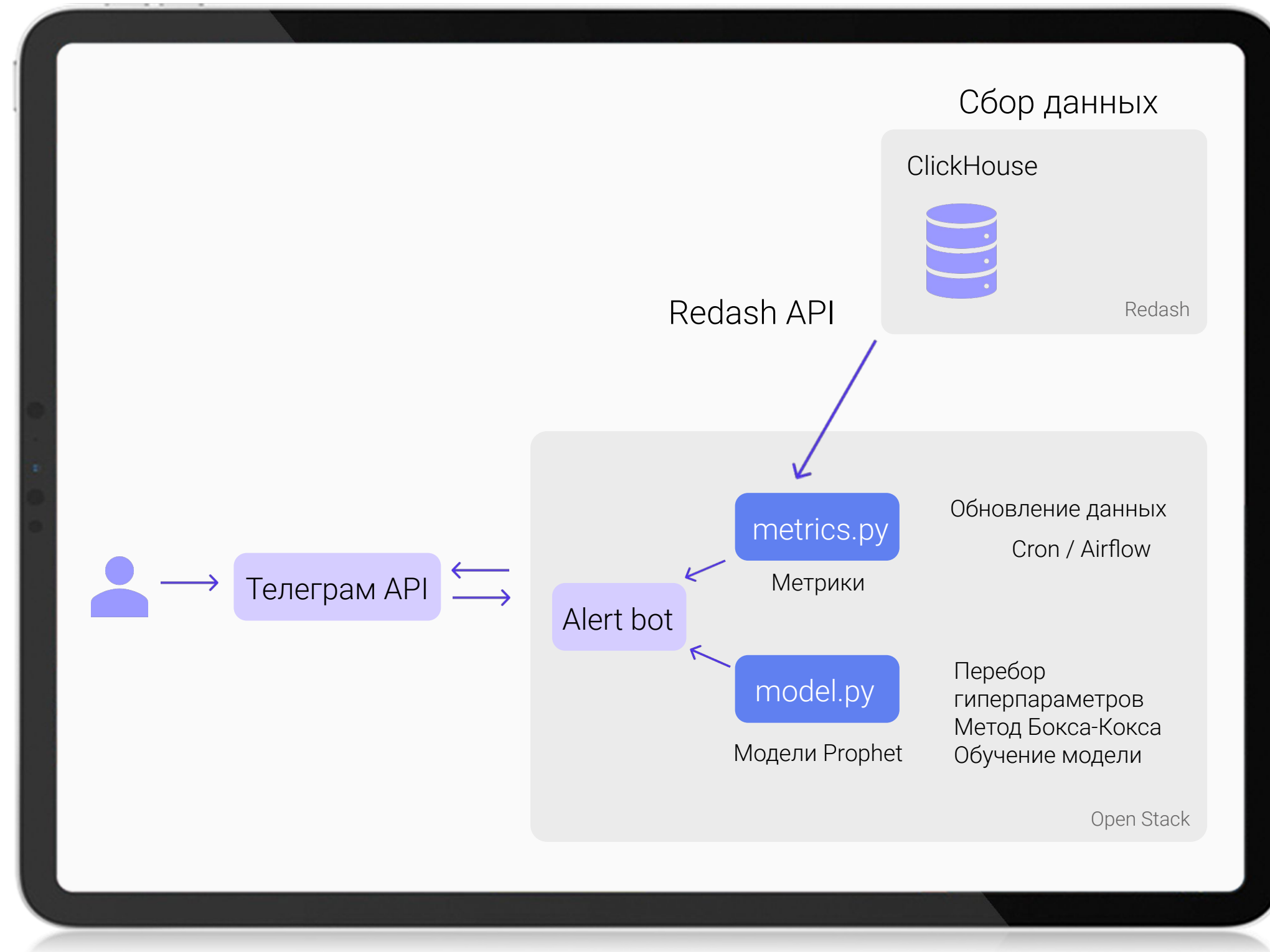


Инфра



Дзен



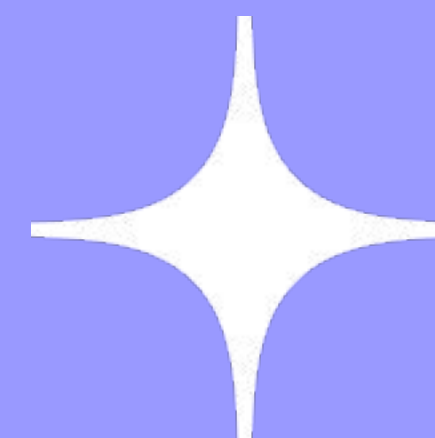


Наш алгоритм

Мы написали скрипт для сбора метрик через API [Redash](#)-а (Redash — это сервис, который позволяет работать с большими объёмами данных, создавать запросы к базам данных и строить визуализации).



Интерфейс



Дзен

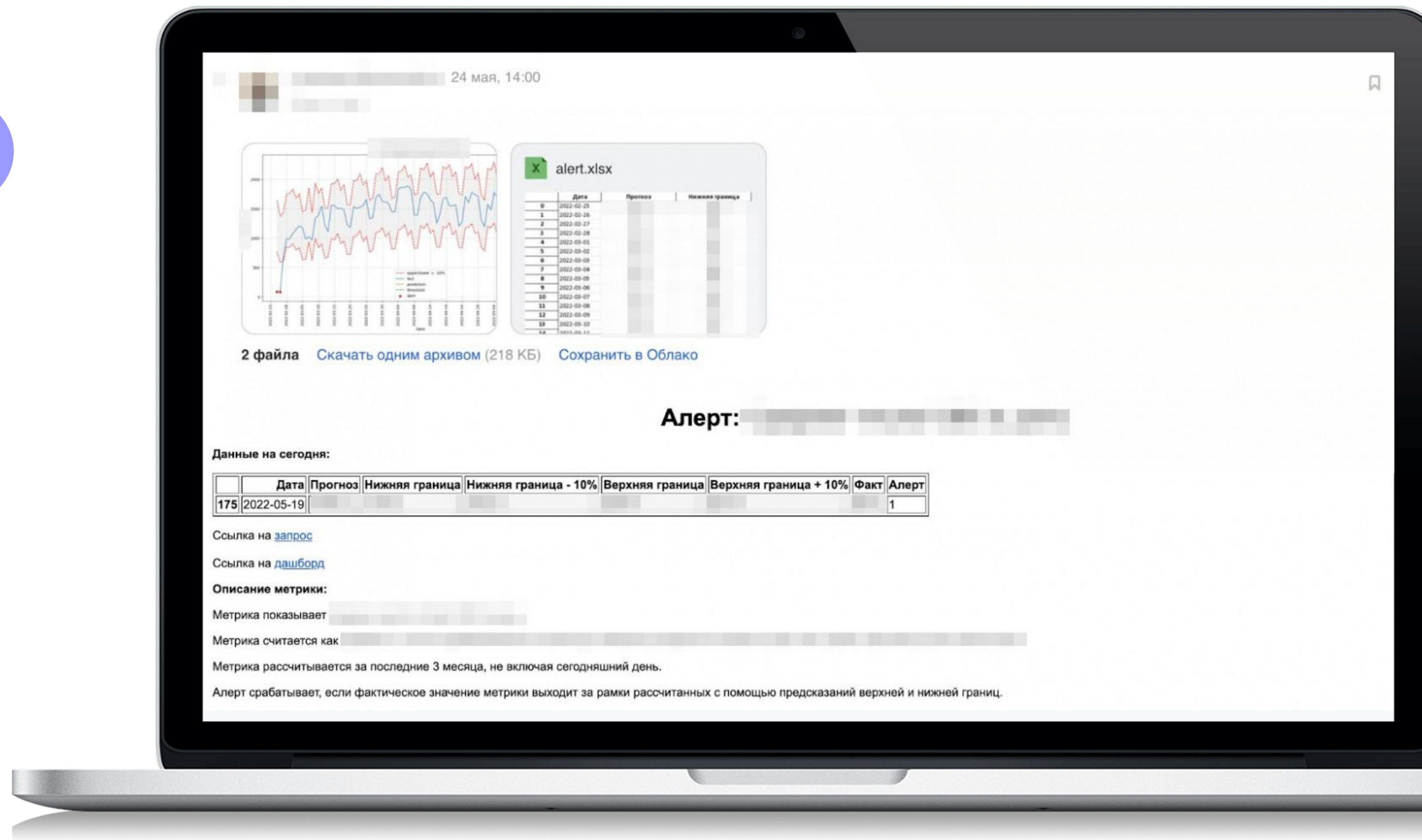


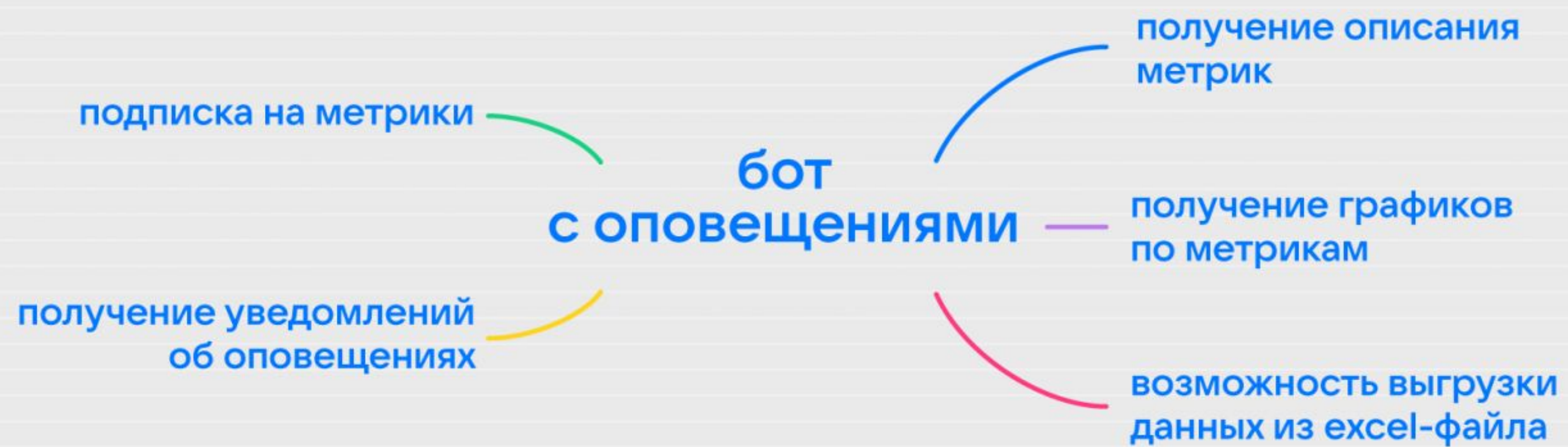


Redash

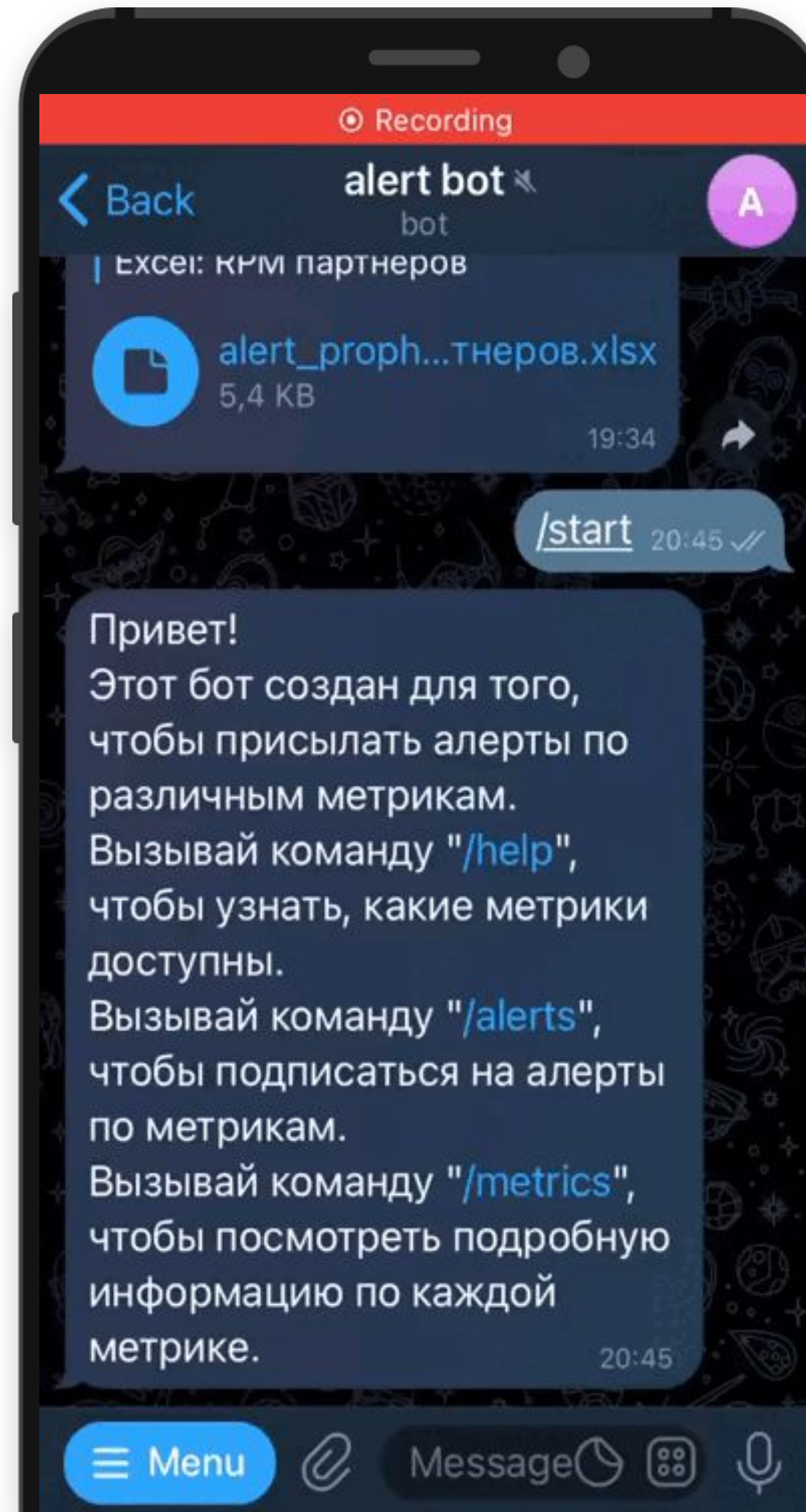
Alerts			
Name	State	Last Updated At	Created At
ARPU DAU xiaomi alert: alert equals 1	TRIGGERED	6 months ago	07/10/21 09:58
xiaomi clicks per uniq user alert: alert equals 1	OK	6 months ago	23/09/21 14:04
lenta clicks per uniq user alert: alert equals 1	OK	7 months ago	23/09/21 14:03
total revenue per stream alert: alert equals 1	OK	6 months ago	22/09/21 20:47
Fresh shows count alert	TRIGGERED	5 months ago	12/03/21 16:57
Traflo shows	OK	6 months ago	26/11/20 23:37
Obmenka shows	OK	6 months ago	26/11/20 23:27
clicks trigger	TRIGGERED	2 years ago	23/07/20 12:24
shows trigger	TRIGGERED	8 months ago	23/07/20 12:24
DTD	TRIGGERED	2 years ago	22/07/20 18:00

Почтовые оповещения

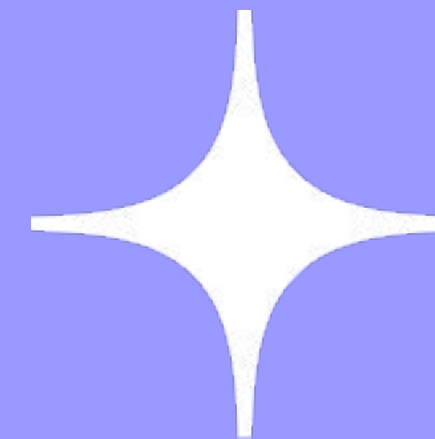




TG-bot



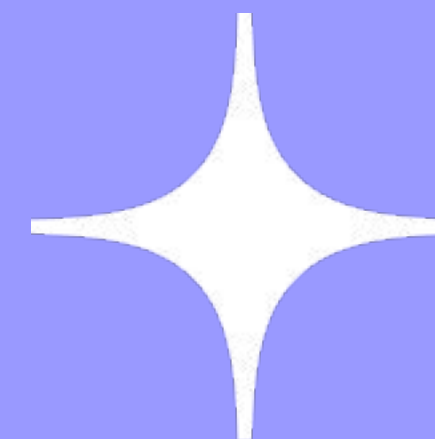
Приоритизация



Эффект в деньгах

$$\left(\begin{array}{l} \text{RPM fact} - \text{RPM predict} \\ \text{DAU fact} - \text{DAU predict} \end{array} \right) \begin{array}{l} * \text{ hits} \\ * \text{ ARPU DAU} \end{array} = \begin{array}{l} \text{revenue 1} \\ \text{revenue 2} \end{array}$$

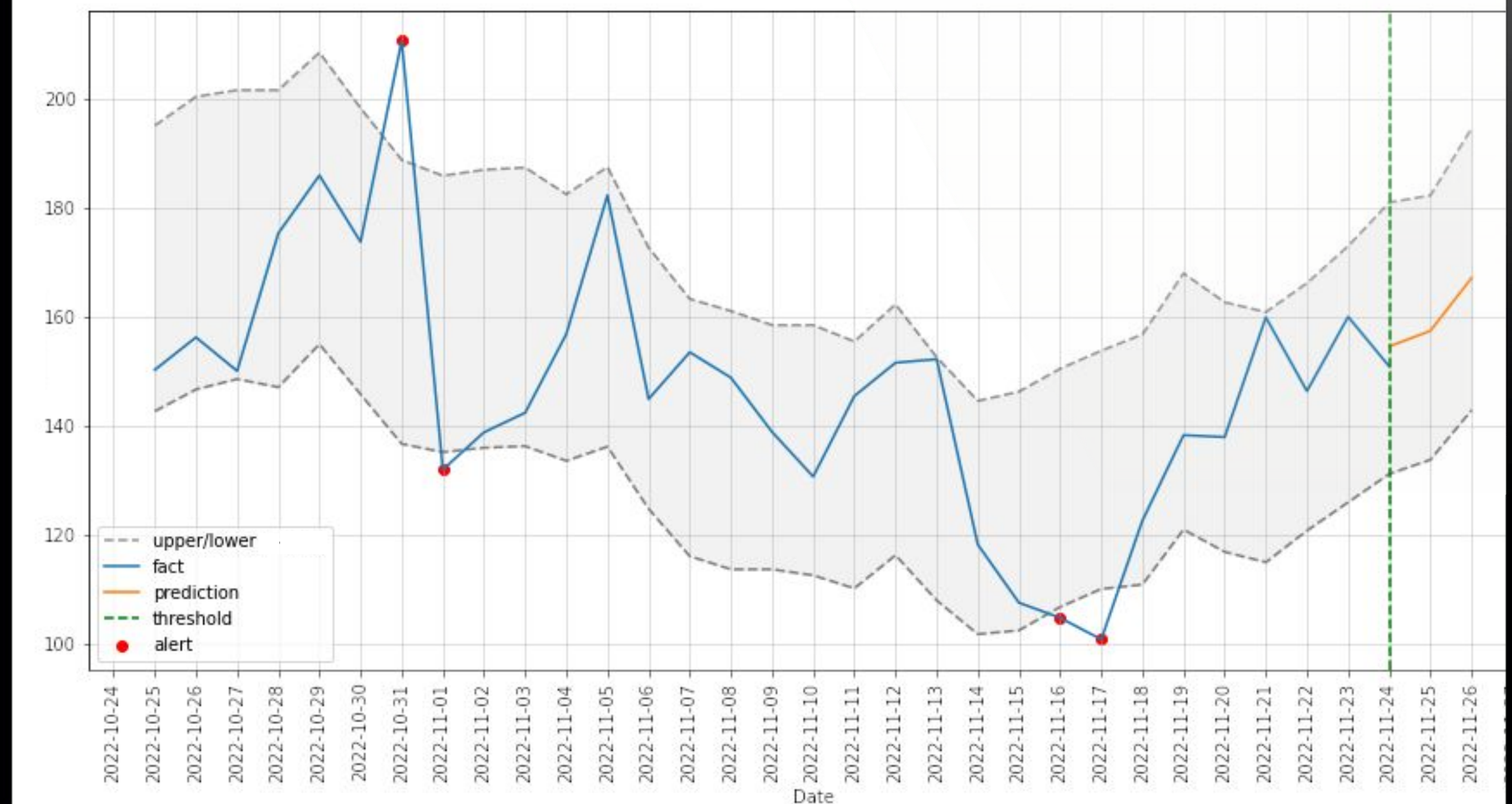
Кейсы



Дзен

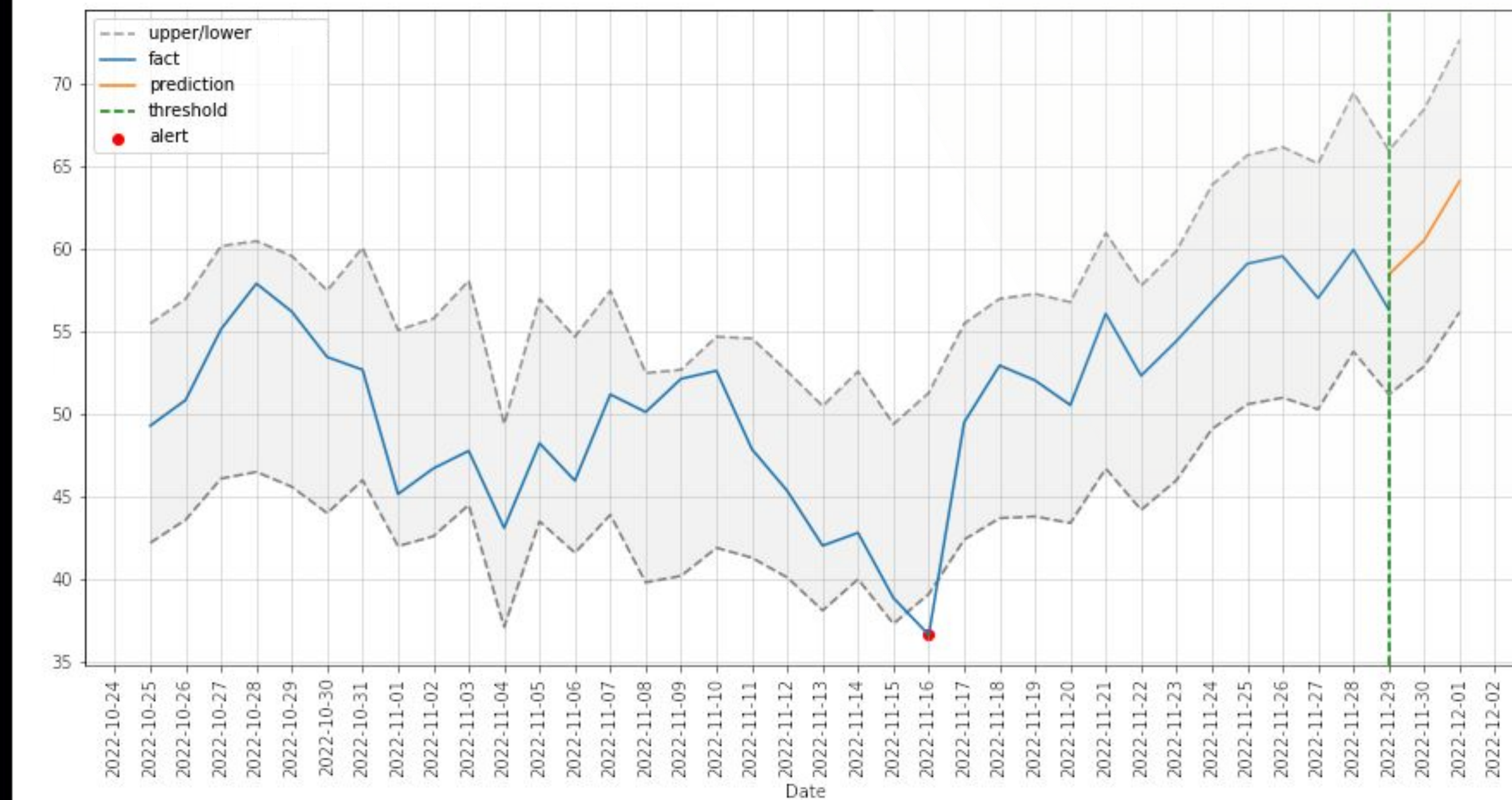
Заметили и вовремя
среагировали на падения
выручки на
1000 заходов у авторов

Если бы на проблему среагировали
спустя неделю мы потеряли бы
несколько миллионов рублей

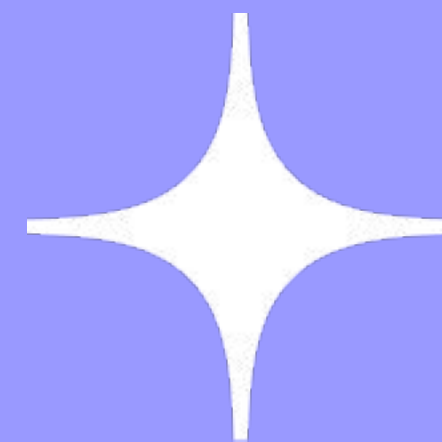


Заметили и вовремя
среагировали на
падения срт

Если бы на проблему среагировали
спустя неделю мы потеряли бы
более 400к рублей



Итоги



Дзен

- ✦ Таким образом, мы можем **экономить** как **время**, затраченное на обнаружение проблем и оценку потерянного дохода, так и сами **деньги**
- ✦ Сегодня наша система оповещения прогнозирует изменения десятков продуктовых метрик и каждый день сообщает об аномалиях, позволяет отслеживать критичные отклонения
- ✦ Это по сути цельный продукт : интерфейс с подписками /алгоритм



Дзен



Вакансии



Контакты



tg : @Vaness7



Q&A секция



Спасибо за внимание!

