

Аномалии, которые мы ищем:

- Активность по действиям пользователя

Проверяем активность пользователя за N секунд, если больше M - бот.

Проверяем IP адреса пользователей. Если больше M с одного адреса - подозрение на ботов.

- Сбросы нумерации просмотров

Проверяем, что в одной сессии нумерация не сбивается.

Проверяем пропуски в нумерации событий.

Надо проверять, начинается ли сессия с 1.

Проверяем соотношение `page_view_order_number` и `event_order_number`

Проверяем дубликаты событий с одинаковыми `ts`, `counter_id`, `event`

- Активность по странам

Проверяем активность в странах по сравнению с общей активностью. Если у нас в стране возрастает активность по сравнению с активностью до этого в этой стране и по сравнению с другими регионами - могут быть боты.

Проверяем несоответствие `geo_city_id` и `geo_location_timezone`

Проверяем резкие изменения локации для одного пользователя

- Активность по типам устройств

Если возрастает активность на одном конкретном типе (особенно если не `mobile` и не вечернее время) - могут быть боты.

- Активность в регистрации пользователей

Если резко возрастает число пользователей, которые начинают регистрироваться или попадают на страницу регистрации – что-то не так.

- Количество просмотров

Если количество просмотров за N минут резко снизилось по сравнению с остальными временными окнами, могут быть проблемы с сайтом. Необходимо учитывать время дня и наличие популярных трансляций.

- Время, проведенное на странице

Ищем нетипичное поведение пользователя по сравнению с другими пользователями. Например, если пользователь на каждой странице проводит по 1 секунде.

Надо проверять, что сессии не слишком короткие, если резко возросло кол-во коротких сессий - что-то не так с сайтом.