

Вопрос.

Выберите топ-1000 строк из CTE `balances` с сортировкой по `user_id` и `dt`. Посмотрите на изменения балансов студентов. Какие вопросы стоит задать дата-инженерам и владельцам таблицы `payments`?

Рассуждение.

На первый взгляд результат запроса по вышеуказанным параметрам выглядит логично, без явных отклонений. Присутствует и приобретенные уроки и проведенные, а также накопительные данные.

Попробуем порассуждать что первично, а что следует за первым шагом.

Исходя из ранее задаваемых вводных к запросам, мы понимаем, что оплата уроков является первичной, а только затем происходит списание в виде проводимых занятий. Исходя из этого данные по `transaction_balance_change`, `transaction_balance_change_cs` должны иметь значения от 0 до $+\infty$, а `classes`, `classes_cs` от $-\infty$ до 0.

Проведем эксперименты путем поочередной сортировки по столбцам .

`transaction_balance_change`, `transaction_balance_change_cs` отклонений не выявили.

Но при выстраивании `classes user_id 751,900,094` показал отрицательный `balance`, что является странным.

Попробуем отсортировать колонку `balance`. И вот мы видим, что у первого `user_id 59,995,492` проявился отрицательный баланс, который мог появиться при проведении уроков в кредит, что противоречит условиям школы. Пролистав ниже, мы видим, что таких `user_id` не единичное количество.

Что это значит? Это значит, что в витрине `payments` возможен сбой и не все данные по оплатам вносятся в расчетные формулы, что даёт некорректный результат.

Мы со своей стороны должны порекомендовать data science проверить таблицу, а также ввести алгоритм предупреждения данных разрывов.

Вопрос.

Создайте визуализацию (линейную диаграмму) итогового результата. Какие выводы вы можете сделать из получившейся визуализации?

Выводы-размышления.

Глядя на визуализацию, видим пять параметров из которых три являются накопительными. Их то мы и возьмем за основу рассуждений, а если быть точнее сосредоточимся на `balance`, тем более что он является конечным результатом.

Мы наблюдаем, что кривая `balance` постепенно поднимается и к концу года останавливается на отметке 3462. Это значит, что количество купленных уроков превышает количество проведенных. Это значит, что 18% купленных уроков оказались нереализованными. А с учетом починки `payments` этот показатель может быть даже выше.

Требуется ли данная ситуация каких-либо изменений и чем она может быть потенциально опасной?

Во-первых, становится понятно, что не хватает возможностей с нашей стороны проводить необходимое количество уроков (нехватка учителей, неверно выстроенный образовательный процесс).

Это может привести к оттоку покупателей уроков, которые вероятно почувствуют диссонанс между ожиданиями и реальностью, что возможно снизит нашу прибыль.

Вывод. Оптимизировать balance можно, снизив количество продаваемых уроков, но я считаю это неверное направление. С точки зрения прибыльности нужно увеличить интенсивность проводимых уроков, разобравшись в негативных причинах