**Задача** — смоделировать изменение балансов студентов.

Баланс — это количество уроков, которое есть у каждого студента.

Чтобы проверить, всё ли в порядке с нашими данными и составить список гипотез и вопросов, нам важно понимать:

* Сколько всего уроков было на балансе **всех учеников** за каждый календарный день
* Как это количество менялось под влиянием транзакций (оплат, начислений, корректирующих списаний) и уроков (списаний с баланса по мере прохождения уроков)

Также нам нужно создать таблицу, где будут балансы **каждого студента** за каждый день.

В результате должен получиться запрос, который собирает данные о балансах студентов за каждый прожитый ими день.

<https://metabase.sky.pro/question/60124>

with

----------- (шаг 1) первая дата успешной (со статусом success) транзакции

first\_payments as

(select distinct user\_id, min(transaction\_datetime::date) as first\_payment\_date

from skyeng\_db.payments

where status\_name = 'success'

group by user\_id

),

----------- (шаг 2) все дни 2016 года

all\_dates as

(select distinct(class\_start\_datetime::date) as dt

from skyeng\_db.classes

where date\_trunc('year', class\_start\_datetime) = '2016-01-01'

),

----------- (шаг 3) все даты жизни студента после того, как произошла его первая транзакция

all\_dates\_by\_user as

(select user\_id,

dt

from all\_dates s

join first\_payments t

on s.dt >= t.first\_payment\_date

),

----------- (шаг 4) список успешных транзакций

payments\_by\_dates as

(select

user\_id,

transaction\_datetime::date as payment\_date,

sum(classes) as transaction\_balance\_change

from skyeng\_db.payments

where id\_transaction is not null and status\_name = 'success'

group by user\_id, payment\_date

),

----------- (шаг 5) баланс студентов, который сформирован только транзакциями

payments\_by\_dates\_cumsum as

(select a.user\_id,

dt,

COALESCE(transaction\_balance\_change,0) as transaction\_balance\_change,

sum(transaction\_balance\_change) over (partition by a.user\_id order by dt) transaction\_balance\_change\_cs

from all\_dates\_by\_user as a

left join payments\_by\_dates b on b.user\_id = a.user\_id and b.payment\_date = a.dt

),

----------- (шаг 6) изменения балансов из-за прохождения уроков

classes\_by\_dates as

(select

user\_id

, class\_start\_datetime::date as class\_date

, count(id\_class)\*(-1) as classes

from SKYENG\_DB.classes

where (class\_status like 'failed\_by\_student' or class\_status like 'success')

and class\_type like 'regular'

group by user\_id, class\_date

),

----------- (шаг 7) кумулятивная сумма количества пройденных уроков

classes\_by\_dates\_dates\_cumsum as

(select a.user\_id,

dt,

COALESCE(classes,0) as classes,

sum(classes) over (partition by a.user\_id order by dt) classes\_cs

from all\_dates\_by\_user as a

left join classes\_by\_dates c on c.user\_id = a.user\_id and c.class\_date = a.dt

),

----------- (шаг 8) баланс каждого студента

balances as

(select d.user\_id,

d.dt,

transaction\_balance\_change,

d.transaction\_balance\_change\_cs,

classes,

classes\_cs,

classes\_cs+transaction\_balance\_change\_cs as balance

from payments\_by\_dates\_cumsum d

left join classes\_by\_dates\_dates\_cumsum e on e.user\_id = d.user\_id and e.dt = d.dt

)

select \*

from balances

order by user\_id, dt

limit 1000

**Задание 1**

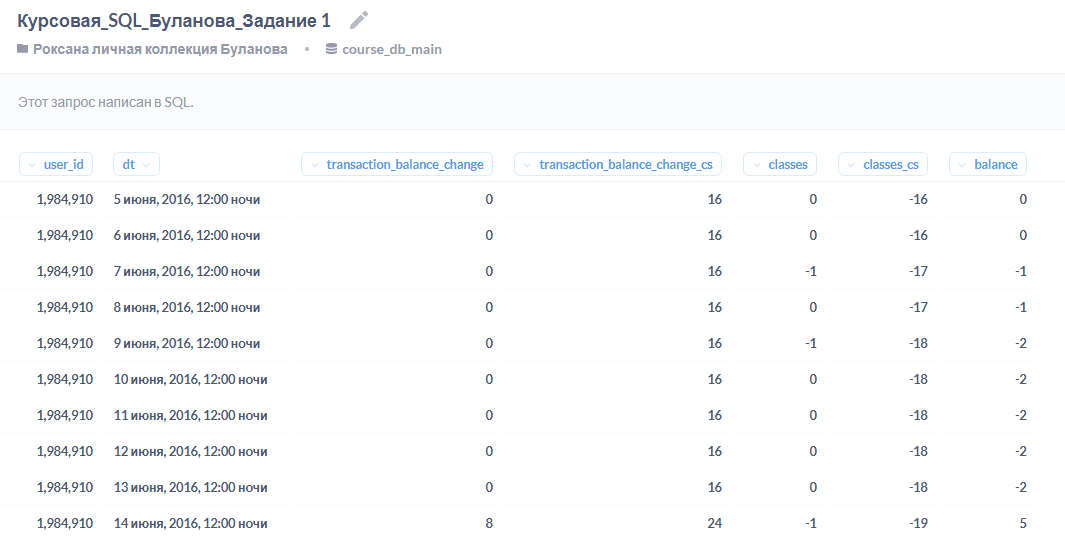
Выберите топ-1000 строк из CTE balances с сортировкой по user\_id и dt.

Посмотрите на изменения балансов студентов.

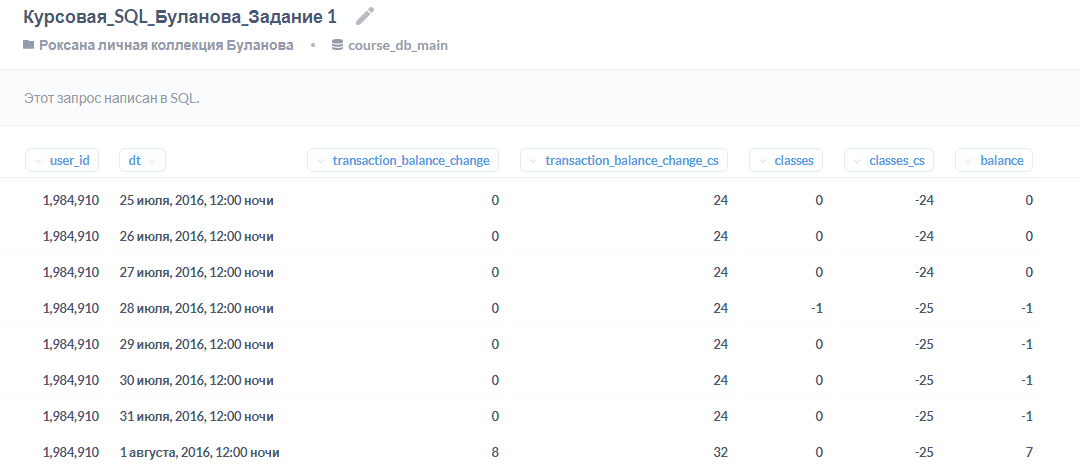
Какие вопросы стоит задавать дата-инженерам и владельцам таблицы payments?

[Курсовая\_SQL\_Буланова\_Задание 1 · Metabase (sky.pro)](https://metabase.sky.pro/question/60143)

Анализирую изменения баланса студентов первых топ-1000 строк, заметила, что у клиента с id 1,984,910 в 07.06.2016 и 09.06.2016 был нулевой баланс, но при этом урок был проведен и списан, т.е. списание было в долг? Либо клиент оплатил, но оплата прошла позже? И баланс обновился только 14.06.2016? Или ошибка в данных?



Аналогичная произошла с данным пользователем ситуация 28.07.2016

****

**Задание 2**

Создайте визуализацию (линейную диаграмму) итогового результата.

Какие выводы можно сделать из получившейся визуализации?

[Курсовая\_SQL\_Буланова\_Задание 2 · Metabase (sky.pro)](https://metabase.sky.pro/question/60139)

----------- (шаг 9) как менялось общее количество уроков на балансах студентов

select dt,

sum(transaction\_balance\_change) as sum\_transaction\_balance\_change,

sum(transaction\_balance\_change\_cs) as sum\_transaction\_balance\_change\_cs,

sum(classes) as sum\_classes,

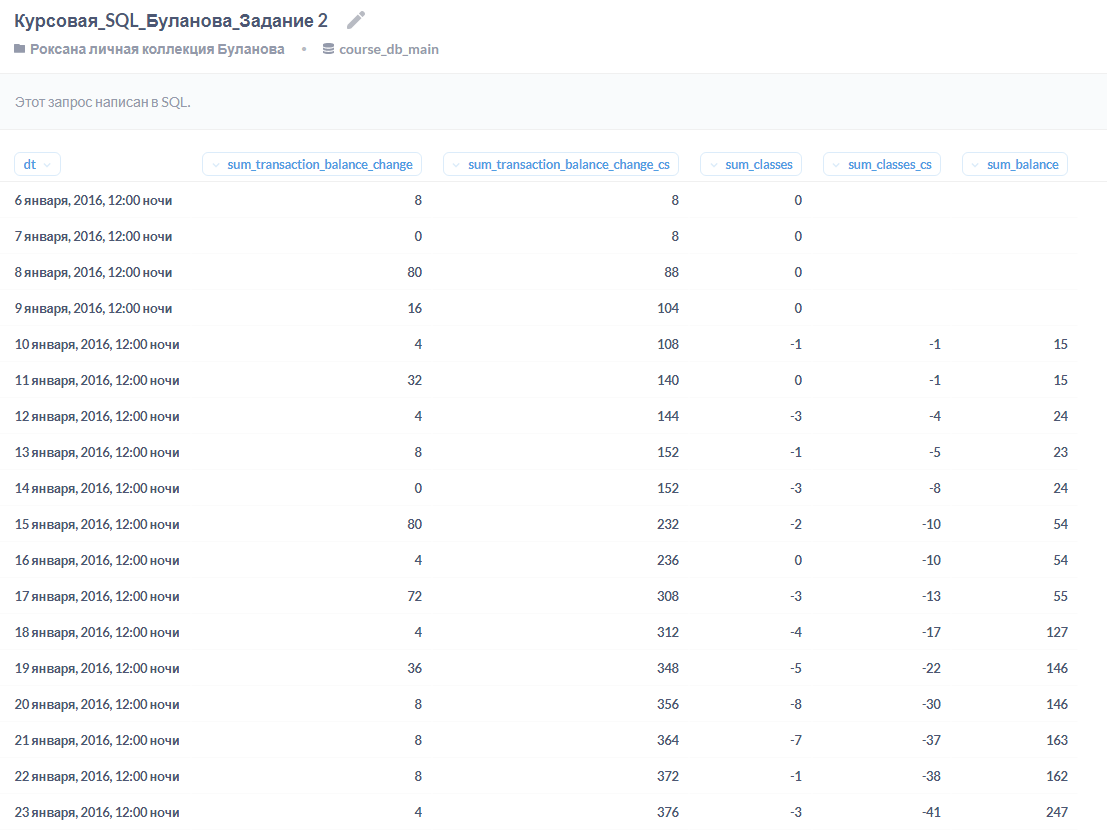
sum(classes\_cs) as sum\_classes\_cs,

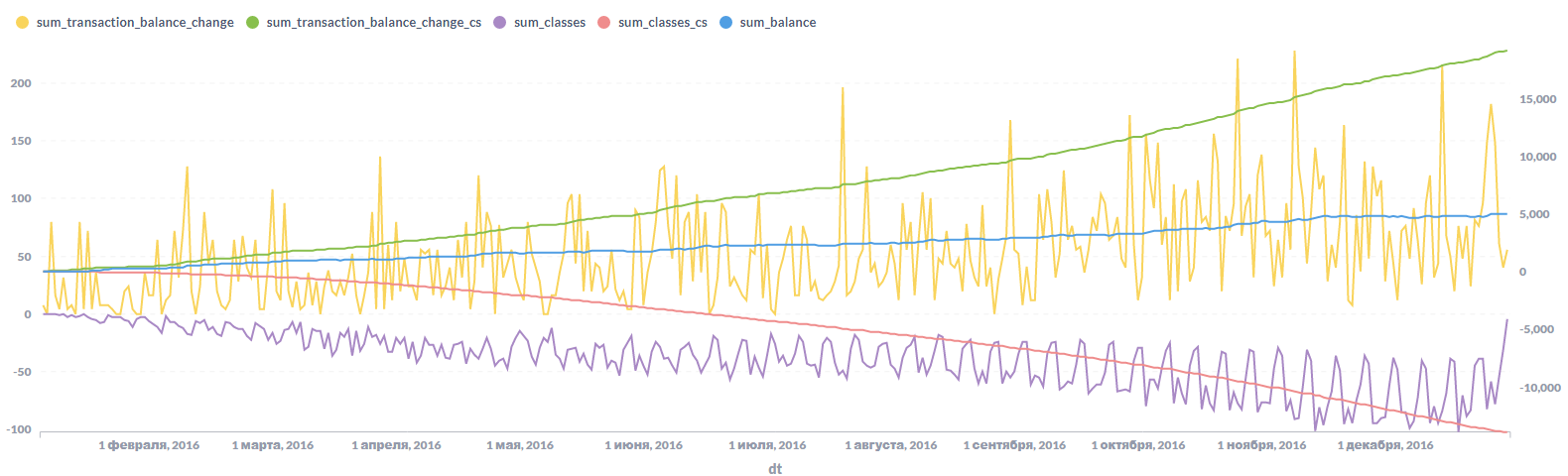
sum(balance) as sum\_balance

from balances

group by dt

order by dt





По итогам общего анализа видна динамика увеличения баланса c студентов.

На основании линии **sum\_classes** по выходным дням уроков проводят меньше, чаще всего уроки проходят по средам и четвергам.