--Шаг 1 Узнаем, когда была первая транзакция для каждого студента. Начиная с этой даты, мы будем собирать его баланс уроков.

with first\_payments as (

select user\_id

, date\_trunc ('day', min(transaction\_datetime)) as first\_payment\_date

from skyeng\_db.payments

where status\_name = 'success'

group by user\_id

),

--Шаг 2 Соберем таблицу с датами за каждый календарный день 2016 года.

all\_dates as (

select distinct (date\_trunc ('day', class\_start\_datetime)) as dt

from skyeng\_db.classes

where class\_start\_datetime between '2016-01-01' and '2017-01-01'

),

--Шаг 3 Узнаем, за какие даты имеет смысл собирать баланс для каждого студента.

all\_dates\_by\_user as (

select user\_id

,dt

from first\_payments

join all\_dates

on dt >= first\_payment\_date

),

--Шаг 4 Найдем все изменения балансов, связанные с успешными транзакциями.

payments\_by\_dates as (

select user\_id

,date\_trunc ('day', transaction\_datetime) as payment\_date

,sum(classes) as transaction\_balance\_change

from skyeng\_db.payments

where status\_name = 'success'

group by user\_id , payment\_date

),

--Шаг 5 Найдем 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(coalesce(transaction\_balance\_change, 0)) over (partition by a.user\_id order by dt) as transaction\_balance\_change\_cs

from all\_dates\_by\_user a

left join payments\_by\_dates b

on a.user\_id = b.user\_id

and a.dt = b.payment\_date

),

-- Шаг 6 Найдем изменения балансов из-за прохождения уроков.

classes\_by\_dates as (

select user\_id

,date\_trunc ('day', class\_start\_datetime) as class\_date

,count(user\_id)\*-1 as classes

from skyeng\_db.classes

where class\_type != 'trial'

and (class\_status = 'success' or class\_status ='faild\_by\_student')

group by user\_id, class\_date

),

--Шаг 7 По аналогии с уже проделанным шагом для оплат создадим CTE для хранения кумулятивной суммы количества пройденных уроков.

classes\_by\_dates\_dates\_cumsum as(

select c.user\_id

,c.dt

,coalesce(classes , 0) as classes

,sum(coalesce(classes , 0)) over (partition by c.user\_id order by c.dt) as classes\_cs

from all\_dates\_by\_user c

left join classes\_by\_dates d

on c.user\_id = d.user\_id

and c.dt = d.class\_date

),

--Шаг 8 Создадим CTE balancesс вычисленными балансами каждого студента.

balances as(

select i.user\_id

,i.dt

,transaction\_balance\_change

,transaction\_balance\_change\_cs

,classes

,classes\_cs

,sum(classes\_cs)+sum(transaction\_balance\_change\_cs) as balance

from payments\_by\_dates\_cumsum i

left join classes\_by\_dates\_dates\_cumsum f

on i.user\_id = f.user\_id

and i.dt=f.dt

group by 1,2,3,4,5,6

),

--Задание 1

select \*

,dense\_rank () over ( order by user\_id) as rn

from balances

limit 1000

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

-- --Результат запроса - выборка по 7ми ученикам. На примере этих учеников видим, что они не поддерживают стабильный баланс доступных занятий:

-- --то есть закупают какое-то количество классов, а следующую оплату совершают уже после отработки доступных занятий и выхода баланса в ноль.

-- --У некоторых учеников наблюдается отрицательный баланс (например, 3335256 с 9го ноября) - то есть они не заплатили за пройденные занятия.

-- Шаг 9

select dt

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

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

, sum(classes) as sum\_c

, sum(classes\_cs) as sum\_cc

, sum(balance) as sum\_b

from balances

group by 1

order by 1

--Задание 2

--На графике 5 кривых, отражающих сумму значений метрики на каждую дату:

--sum\_tbc - оплачено занятий - в целом равномерная тенденция по году, максимальные значения достигаются в ноябре, то есть в этом месяце наиболее активные оплаты

--sum\_tbcc - оплачено занятий нарастающим итогом - активный рост наблюдаем с середины года

--sum\_c - пройдено уроков - в целом наблюдаем рост к концу года, но в сравнении с оплаченными занятиями не такой сильный

--sum\_cc - пройдено уроков нарастающим итогом - активный рост наблюдаем с середины года

--sum\_b - оплачено - пройдено уроков - равномерно растет на протяжении всего года, на конец года имеем положительный баланс, то есть оплачено уроков больше, чем пройдено.