--------------------Курсовая работа------------------------------

-----------Моделирование изменения балансов студентов------------

-----------------Step 1----------------------

with first\_payments as

(

select user\_id

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

from

(select user\_id

, transaction\_datetime

, row\_number() over(partition by user\_id order by transaction\_datetime) rw\_payment

from skyeng\_db.payments as p

where status\_name = 'success'

group by user\_id, transaction\_datetime

order by date\_trunc('day', transaction\_datetime) desc) as row\_st\_lessons

where rw\_payment = 1

),

-----------------Step 2----------------------

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'

),

-----------------Step 3----------------------

all\_dates\_by\_user as

(

select user\_id

, dt

from all\_dates

join first\_payments

on all\_dates.dt >= first\_payments.first\_payment\_date

order by user\_id, dt

),

-----------------Step 4----------------------

payments\_by\_dates as

(

select user\_id

, date\_trunc('day', transaction\_datetime) as pay\_day

, sum(classes) as transaction\_balance\_change

from skyeng\_db.payments

where status\_name = 'success'

group by user\_id, pay\_day

order by user\_id, pay\_day

),

-----------------Step 5----------------------

payments\_by\_dates\_cumsum as

(

select a.user\_id

, dt

, transaction\_balance\_change

, case when transaction\_balance\_change is null then 0

else transaction\_balance\_change

end as transaction\_balance\_change\_no\_zero

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

, case when sum (transaction\_balance\_change) over (partition by a.user\_id order by dt) is null then 0

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

end as transaction\_balance\_change\_cs\_no\_zero

from all\_dates\_by\_user a

left join payments\_by\_dates b

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

and a.dt = b.pay\_day

),

-----------------Step 6----------------------

classes\_by\_dates as

(

select user\_id

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

, count(class\_start\_datetime) \* -1 as classes

from skyeng\_db.classes

where class\_type != 'trial'

and class\_status in ('success', 'failed\_by\_student')

group by user\_id, class\_date

),

-----------------Step 7----------------------

classes\_by\_dates\_cumsum as

(

select a.user\_id

, dt

, classes

, case when classes is null then 0

else classes

end as classes\_no\_zero

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

, case when sum (classes) over (partition by a.user\_id order by dt) is null then 0

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

end as classes\_cs\_no\_zero

from all\_dates\_by\_user a

left join classes\_by\_dates b

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

and a.dt = b.class\_date

),

-----------------Step 8----------------------

balances as

(

select a.user\_id

, a.dt

, transaction\_balance\_change\_no\_zero

, transaction\_balance\_change\_cs\_no\_zero

, classes\_no\_zero

, classes\_cs\_no\_zero

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

from payments\_by\_dates\_cumsum a

left join classes\_by\_dates\_cumsum b

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

and a.dt = b.dt

)

-----------------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`?

-- 1. Явно видно что в какие-то выходные и праздичные ученики не занимаются

-- 2. В начале недели ученики проявляют большую активность к урокам.

-- 2. Ученики занимаются без оплат. В балансе иногда появляются отрицательные числа. Почему?

-- 3. Почему 3 мая и 18 октября была оплата ровно 64 уроков у двух разных учеников?Какие-то нетепичный выбросы.

-- Может была какая-то акция?

-- select \*

-- from balances

-- order by user\_id, dt

-- limit 1000

-----------------Задание 2----------------------

-----------------Step 9 ----------------------

select dt

, sum (transaction\_balance\_change\_no\_zero) tr\_bal\_change\_sum

, sum (transaction\_balance\_change\_cs\_no\_zero) tr\_bal\_change\_cs\_sum

, sum (classes\_no\_zero) cl\_sum

, sum (classes\_cs\_no\_zero) cl\_cs\_sum

, sum (balance) balance\_sum

from balances

group by dt

order by dt

-- Создайте визуализацию (линейную диаграмму) итогового результата.

-- Какие выводы можно сделать из получившейся визуализации?

-- 1. Пройденные уроки и покупки показывают планомерный рост, причем транзакции растут большими темпами, что положительно сказывается на балансе (синяя шкала), который также планомерно растет.

-- 2. Из этого следует вывод об успешности кампаний либо по привлечению учеников либо по их удержанию (возможно и той и той)

-- 3. На графике видны всплески активности (баланса) учеников - это весна и осень. В течение лета хоть и наблюдается рост, но менее выраженный. По видимому это связано с сезонностью учебного года.

-- 4. п. 3. может говорить о том, что основные ученики - школьники и студенты, что проверяется по данным учеников. Соответственно, возможно расширение клиентской базы за счет других групп людей.

-- 5. Касательно прохожденя уроков (CL\_SUM) мы явно видим, что в выходные и праздичные дни ученики занимаются менее активно, что особенно заметно в мае - 1-го и 8-го числа (праздичные дни)

-- 6. При этом транзакции в выходные и праздничные дни также происходят значительно реже (например 30-го января, 7, 8-го мая оплат вообще не было)

-- 7. Также на графике транзакций мы наблюдаем выраженные пики по будням несколько раз за каждый месяц. Возможно это результат маркетинговой кампании.

-- 8. 31-го декабря было пройдено всего 7 уроков, однако приобретено 148. Можно предположить, что данные уроки были подарочными.