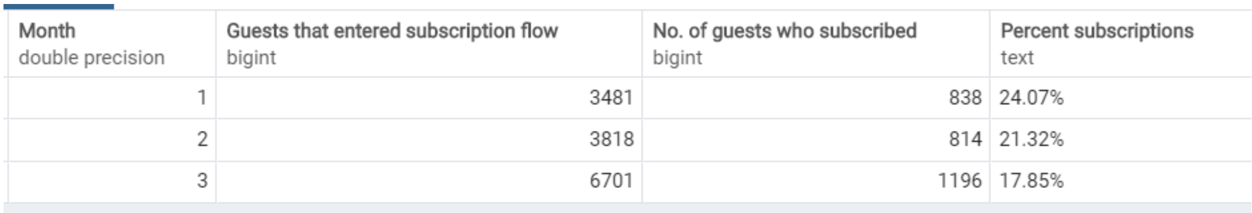
1. Please write a query for how many customers started the subscription flow each month as well as the number and % that completed.

The months are presented as number of each month(Jan - 1, Feb - 2, etc)
Guests that entered the subscription flow are the total number of rows in subscriptions table for that month.
Last 2 parameters are calculated as per logic given in exercise.

```
SELECT
extract(month from date_created) AS "Month",
COUNT(*) AS "Guests that entered subscription flow",-- total customers who started
subscription flow
SUM(
CASE WHEN date_initialized IS NOT NULL
        THEN 1
        ELSE 0
END) AS "No. of guests who subscribed",--total subscriptions
ROUND(SUM(
CASE WHEN date_initialized IS NOT NULL
        THEN 1
        ELSE 0
END)*100.0/COUNT(*),2)||'%' AS "Percent subscriptions"
FROM thistle_web.subscriptions_subscription
GROUP BY
extract(month from date_created)
```

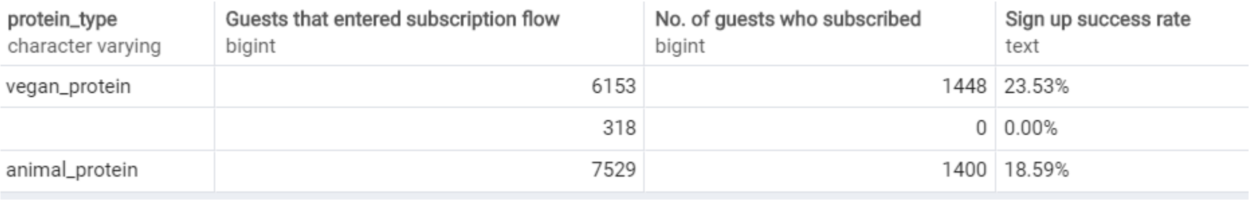


Month double precision	Guests that entered subscription flow bigint	No. of guests who subscribed bigint	Percent subscriptions text
1	3481	838	24.07%
2	3818	814	21.32%
3	6701	1196	17.85%

2. What is the signup success rate (# of people signing up for a subscription vs. all people who enter the checkout flow) for meat vs. veg plans?

A group by clause with protein type gives the answer for this question tweaking the SQL in 1st question.

```
SELECT
protein_type,
COUNT(*) AS "Guests that entered subscription flow",-- total customers who started
subscription flow
SUM(
CASE WHEN date_initialized IS NOT NULL
        THEN 1
        ELSE 0
END) AS "No. of guests who subscribed",--total subscriptions
ROUND(SUM(
CASE WHEN date_initialized IS NOT NULL
        THEN 1
        ELSE 0
END)*100.0/COUNT(*),2)||'%' AS "Sign up success rate"
FROM thistle_web.subscriptions_subscription
GROUP BY
protein_type
```

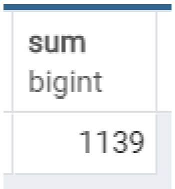


protein_type	Guests that entered subscription flow	No. of guests who subscribed	Sign up success rate
character varying	bigint	bigint	text
vegan_protein	6153	1448	23.53%
	318	0	0.00%
animal_protein	7529	1400	18.59%

3.Please calculate how many customers cancel within 14 days of signing up.

A simple count of total customers where difference in cancelled date and subscription date is less than or equal to 14.

```
SELECT
COUNT(DISTINCT(user_id))
FROM
thistle_web.subscriptions_subscription "s"
JOIN
thistle_web.subscriptions_subscriptioncancellation "c"
ON s.id = c.subscription_id
WHERE DATE_PART('day',c.date_cancelled::timestamp-s.date_initialized::timestamp) <= 14
```



sum
bigint
1139

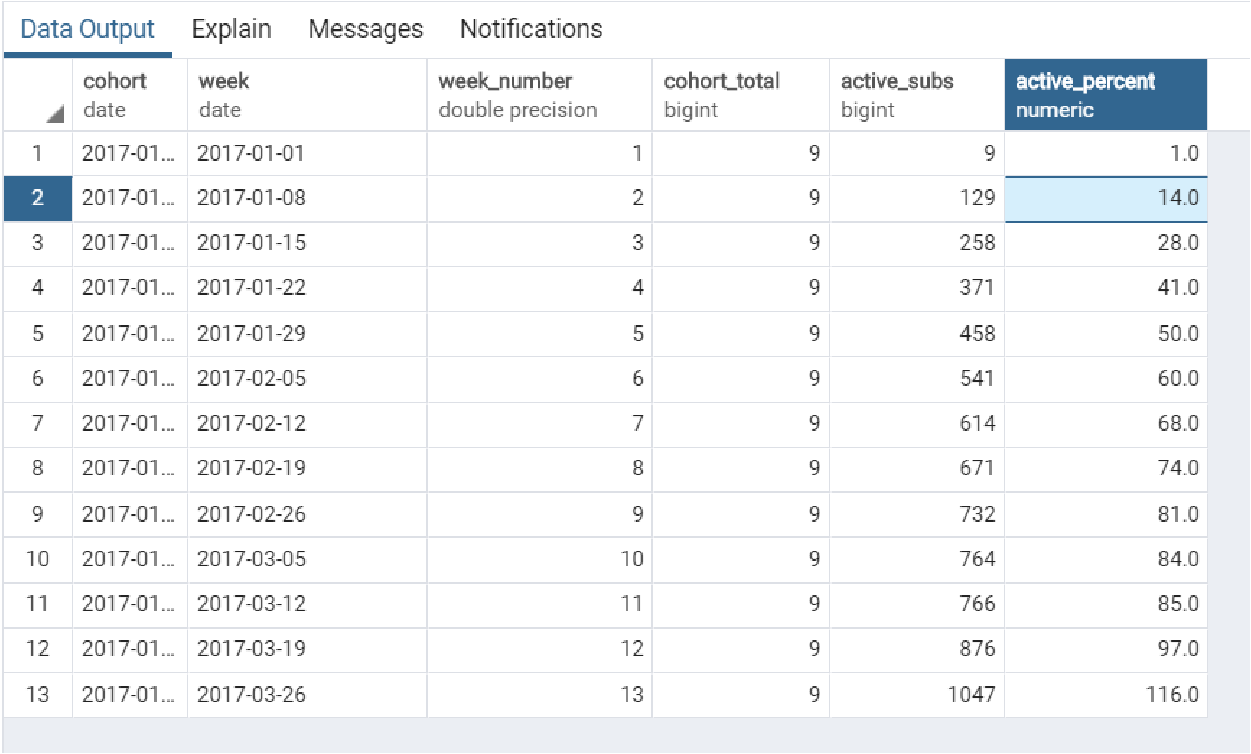
4. Please calculate retention by weekly cohort.

This report starts from day first subscription was added to the subscription table.
Cohort total is counted as the number of subscriptions on the first day.

‘Week’ column refers to the first day of each new week.
Active subscriptions are counted as ‘subscriptions started before or on a given day’ minus ‘subscription cancelled on or before that day’. (given day = first day of each week in table)

```
WITH first_day(first_day,last_day) AS
(
    SELECT
        MIN(date_initialized)::date,
        MAX(date_initialized)::date
    FROM
        thistle_web.subscriptions_subscription
),
first_dow (first_dow) AS
(
    SELECT
        dow
    FROM
        public.etl_calendar,first_day
    WHERE
        day=first_day.first_day
),
start_cohort (scount) AS
(
    SELECT
        COUNT(*)
    FROM thistle_web.subscriptions_subscription,first_day
    WHERE date_initialized::date <= first_day.first_day
)
SELECT
    first_day "cohort",
    day "week",
    DATE_PART('day',day::timestamp-first_day::timestamp)/7+1 "week_number",
    start_cohort.scount "cohort_total",
    start_cohort.scount-
    (SELECT count(*) FROM thistle_web.subscriptions_subscriptioncancellation
    WHERE date_cancelled <= day AND date_cancelled>=first_day)+
    (SELECT count(*) FROM thistle_web.subscriptions_subscription
    WHERE date_initialized <= day)"active_subs",
    (start_cohort.scount-
    (SELECT count(*) FROM thistle_web.subscriptions_subscriptioncancellation
    WHERE date_cancelled <= day AND date_cancelled>=first_day)+
    (SELECT count(*) FROM thistle_web.subscriptions_subscription
    WHERE date_initialized <= day))/start_cohort.scount*1.0 "active_percent"
```

```
FROM
        public.etl_calendar,first_dow,first_day,start_cohort
WHERE
        dow=first_dow.first_dow
AND day::date < last_day
```



	cohort date	week date	week_number double precision	cohort_total bigint	active_subs bigint	active_percent numeric
1	2017-01...	2017-01-01	1	9	9	1.0
2	2017-01...	2017-01-08	2	9	129	14.0
3	2017-01...	2017-01-15	3	9	258	28.0
4	2017-01...	2017-01-22	4	9	371	41.0
5	2017-01...	2017-01-29	5	9	458	50.0
6	2017-01...	2017-02-05	6	9	541	60.0
7	2017-01...	2017-02-12	7	9	614	68.0
8	2017-01...	2017-02-19	8	9	671	74.0
9	2017-01...	2017-02-26	9	9	732	81.0
10	2017-01...	2017-03-05	10	9	764	84.0
11	2017-01...	2017-03-12	11	9	766	85.0
12	2017-01...	2017-03-19	12	9	876	97.0
13	2017-01...	2017-03-26	13	9	1047	116.0