

SQL

Задача 1. Какой запрос вернет самое большое число? Почему?

A.

```
select count(distinct first.counter_column)
from first
left join second on first.join_key = second.join_key
where second.filter_column >= 42;
```

B.

```
select count(distinct first.counter_column)
from first
left join second on first.join_key = second.join_key and second.filter_column >= 42;
```

C.

```
select count(distinct first.counter_column)
from first
right join second on first.join_key = second.join_key
where second.filter_column >= 42;
```

D.

```
select count(distinct first.counter_column)
from first
right join second on first.join_key = second.join_key and second.filter_column >= 42;
```

Мое решение:

B.

Нам нужен Left Join, потому что таким образом основной таблицей будет таблица first, на которую мы не накладываем фильтры. Соответственно, можно предположить, что уникальных пользователей там останется больше, чем в таблице second. Ограничение внутри джоина, а не после позволяет нам также оставить первую таблицу целиком.

Задача 2. Найти среднее время между первым и вторым заказом у пользователей. В запросе запрещается использование JOIN'ов. Тестовые заказы фильтровать не нужно. Для этого задания нам потребуются несколько дополнительных таблиц:

- **ClientOrder** – информация о заказах

Название столбца	Тип данных	Описание
ID	bigint	ID заказа
Date	datetime	Время оформления заказа
ClientID	bigint	ID пользователя
ClientOrderStateID	int	ID статуса заказа (1 – оформлен, 2 – получен, 3 – отменен)

- **ClientOrderItems** – информация о товарах в заказе

Название столбца	Тип данных	Описание
ClientOrderID	bigint	ID заказа
ItemID	bigint	ID товара
categoryLvl1	varchar	Коммерческая категория 1-го уровня
price	float	Цена за одну единицу товара
qty	int	Единиц товара в заказе

- **ClientOrderAdditionalInfo** – дополнительные атрибуты заказа

Название столбца	Тип данных	Описание
ClientOrderID	bigint	ID заказа
code	varchar	Название атрибута (OrderType – тип заказа, Platform – платформа, с которой совершен заказ)
value	varchar	Значение атрибута (OrderType – test или regular, Platform – ios, android, site или mobile)

Мое решение (синтаксис Кликхауса):

```

1  SELECT
2      avg(second_order_date - first_order_date)/86400 As avg_date_diff
3
4  FROM
5
6      (
7      SELECT DISTINCT ClientID,
8                      MIN(Date) As first_order_date,
9                      MAX(Date) As second_order_date
10     FROM
11         (
12         SELECT ClientID,
13                Date
14         FROM test.ClientOrder
15         ORDER BY Date ASC
16         LIMIT 2 BY ClientID
17         )
18     GROUP BY ClientID
19 )

```

Задача 3. Для каждой пары платформы и категории товара найти топ-3 пользователей, у которых наименьшее количество дней между первым и последним не тестовым заказом товаров из этой категории.

Мое решение (синтаксис Кликхауса):

```
1  SELECT
2      -- получаем все сочетания платформы и категории товара
3      concat(platform, ' ', categoryLv11) as platform_category,
4      c.ClientID as clientId,
5      -- получаем разницу между датой заказа и датой первого заказа этого клиента в днях
6      (c.Date - min_date)/86400 as date_diff
7
8  FROM
9      (
10         -- берем не тестовые заказы
11         SELECT DISTINCT ClientOrderID
12         FROM test.ClientOrderAdditionalInfo
13         WHERE code == 'OrderType' and value != 'test'
14     ) As a
15
16  LEFT JOIN
17      (
18         -- добавляем данные по платформе
19         SELECT ClientOrderID,
20                value as platform
21         FROM test.ClientOrderAdditionalInfo
22         WHERE code != 'OrderType'
23     ) As b on a.ClientOrderID = b.ClientOrderID
24
25  -- добавляем данные по заказам
26  LEFT JOIN test.ClientOrder as c on a.ClientOrderID = c.ID
27
28  -- добавляем данные по дате первого заказа
29  LEFT JOIN
30      (
31         SELECT ClientID,
32                min(Date) as min_date
33         FROM test.ClientOrder
34         GROUP BY ClientID
35     ) As d on c.ClientID = d.ClientID
36
37  -- добавляем данные по категории
38  LEFT JOIN test.ClientOrderItems as e on a.ClientOrderID = e.ClientOrderID
39
40  GROUP BY platform_category, date_diff, clientId
41  -- сортируем по возрастанию, так как нас интересуют клиенты,
42  -- у которых наименьшее количество дней между первым и текущим заказом
43  ORDER BY platform_category, date_diff ASC
```