Тестовое задание на вакансию «младший аналитик»

У компании есть CRM система, в которую приходят заявки от клиента. Дальше эти заявки обрабатываются операторами и могут менять свое состояние. Например, статус «Сбор информации» значит, что уточняется информация по заявке. А «Проверка товара на складе», что кладовщик в данный момент проверяет, есть ли нужный товар в наличии.

Соответственно, Вам предоставляется набор данных (csv файл) со следующими колонками:

* Id – идентификатор объекта (заявки)
* timestamp – момент времени, когда произошла смена статуса
* status – новый статус объекта

**Необходимо**: написать на языке SQL (любой удобный диалект) запрос, который бы показывал среднее время (в минутах) нахождения заявки в каждом из статусов. Дробное время округлять до 1 знака после запятой.

Для написания запросов можно либо экспортировать данные в удобную вам БД, либо воспользоваться онлайн сервисами ([например](https://sqliteonline.com/))

**Требования**: при добавлении новых статусов в систему запрос должен продолжать корректно работать. То есть новый статус должен появиться в результате исполнения запроса.

**Доп. вопросы:** корректны ли результаты, которые Вы получили, с точки зрения бизнеса / здравого смысла? Если да, то почему. Если нет, то какая причина и как изменить запрос, чтобы стало лучше.

**Пример**:

|  |  |  |
| --- | --- | --- |
| **Id** | **Timestamp** | **Status** |
| 1 | 123 | Сбор информации |
| 2 | 128 | Проверка товара на складе |
| 1 | 222 | Ожидание |
| 1 | 569 | Согласование нового товара |

**Ожидаемый результат**:

|  |  |
| --- | --- |
| Status | Time\_in\_minutes |
| Сбор информации | 1.7 |
| Ожидание | 5.8 |

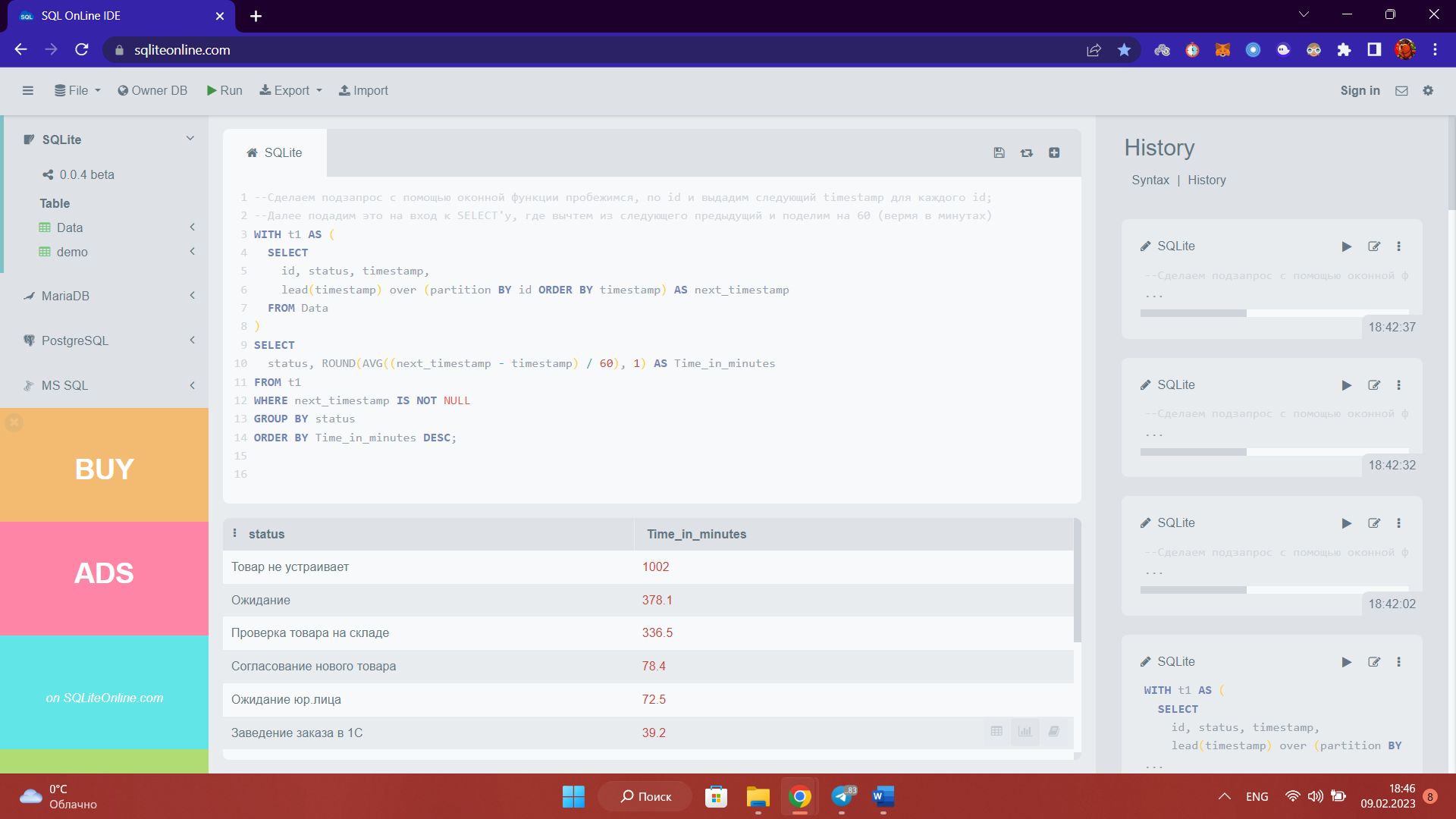
Объяснение: товар с id = 2 не менял свой статус, поэтому нельзя сказать, сколько товар находился в этом статусе. Аналогично для id = 1 статус «Согласование нового товара» на данный момент является текущим.

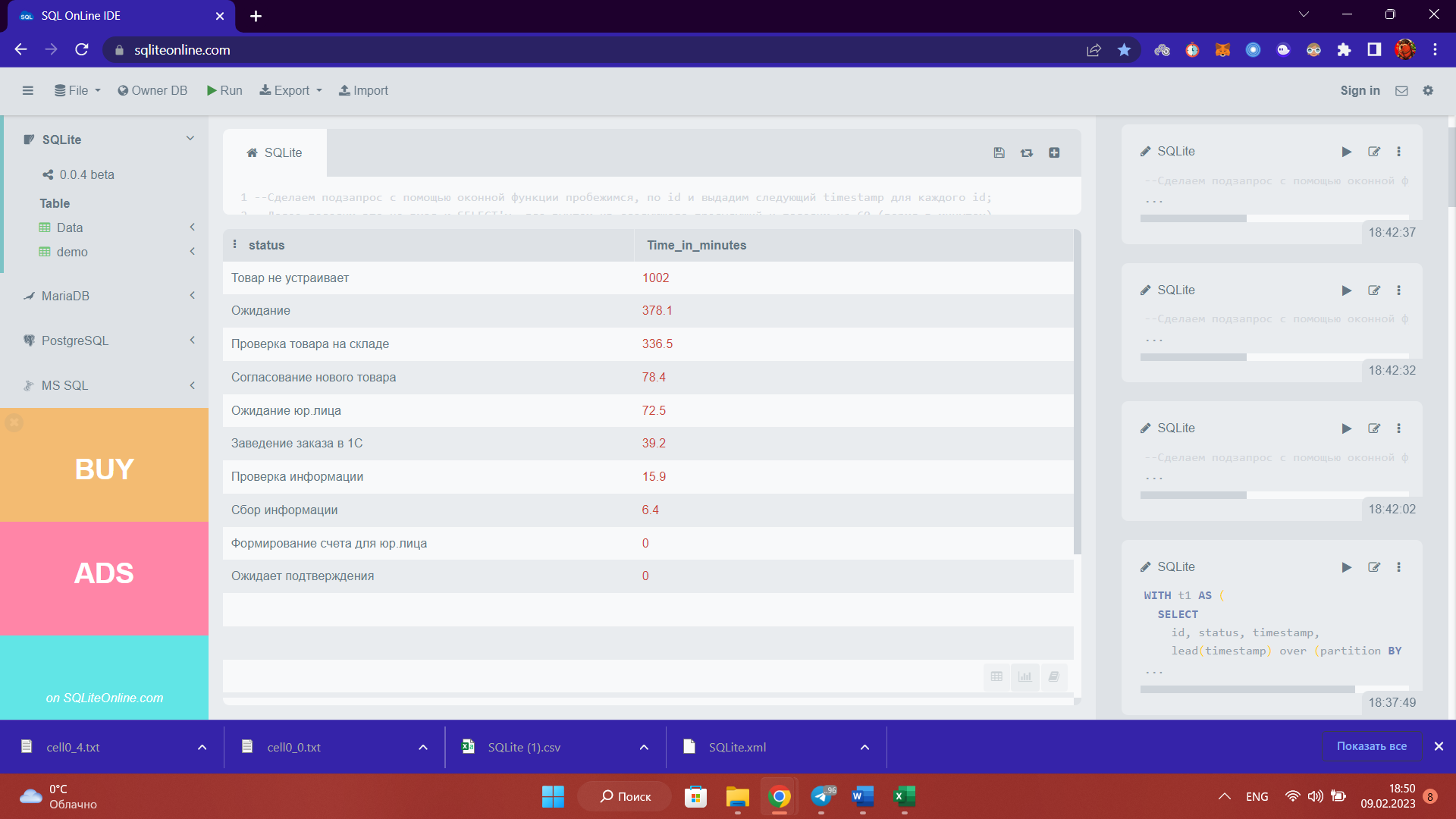
**Доп. задача**: отображать в результате статус, который является текущим. Подумать, корректно ли его отображать и почему.

**Решение:**

Использовал SQLite:

--Сделаем подзапрос с помощью оконной функции пробежимся, по id и выдадим следующий timestamp для каждого id;

--Далее подадим это на вход к SELECT'у, где вычтем из следующего предыдущий и поделим на 60 (вермя в минутах)



WITH t1 AS (

SELECT

id, status, timestamp,

lead(timestamp) over (partition by id order by timestamp) as next\_timestamp

FROM Data

)

SELECT

status, ROUND(AVG((next\_timestamp - timestamp) / 60), 1) as Time\_in\_minutes

FROM t1

WHERE next\_timestamp IS NOT NULL

GROUP BY status

ORDER BY Time\_in\_minutes DESC;

|  |  |
| --- | --- |
| status | Time\_in\_minutes |
| "Товар не устраивает" | "1002" |
| "Ожидание" | "378.1" |
| "Проверка товара на складе" | "336.5" |
| "Согласование нового товара" | "78.4" |
| "Ожидание юр.лица" | "72.5" |
| "Заведение заказа в 1С" | "39.2" |
| "Проверка информации" | "15.9" |
| "Сбор информации" | "6.4" |
| "Формирование счета для юр.лица" | "0" |
| "Ожидает подтверждения" | "0" |

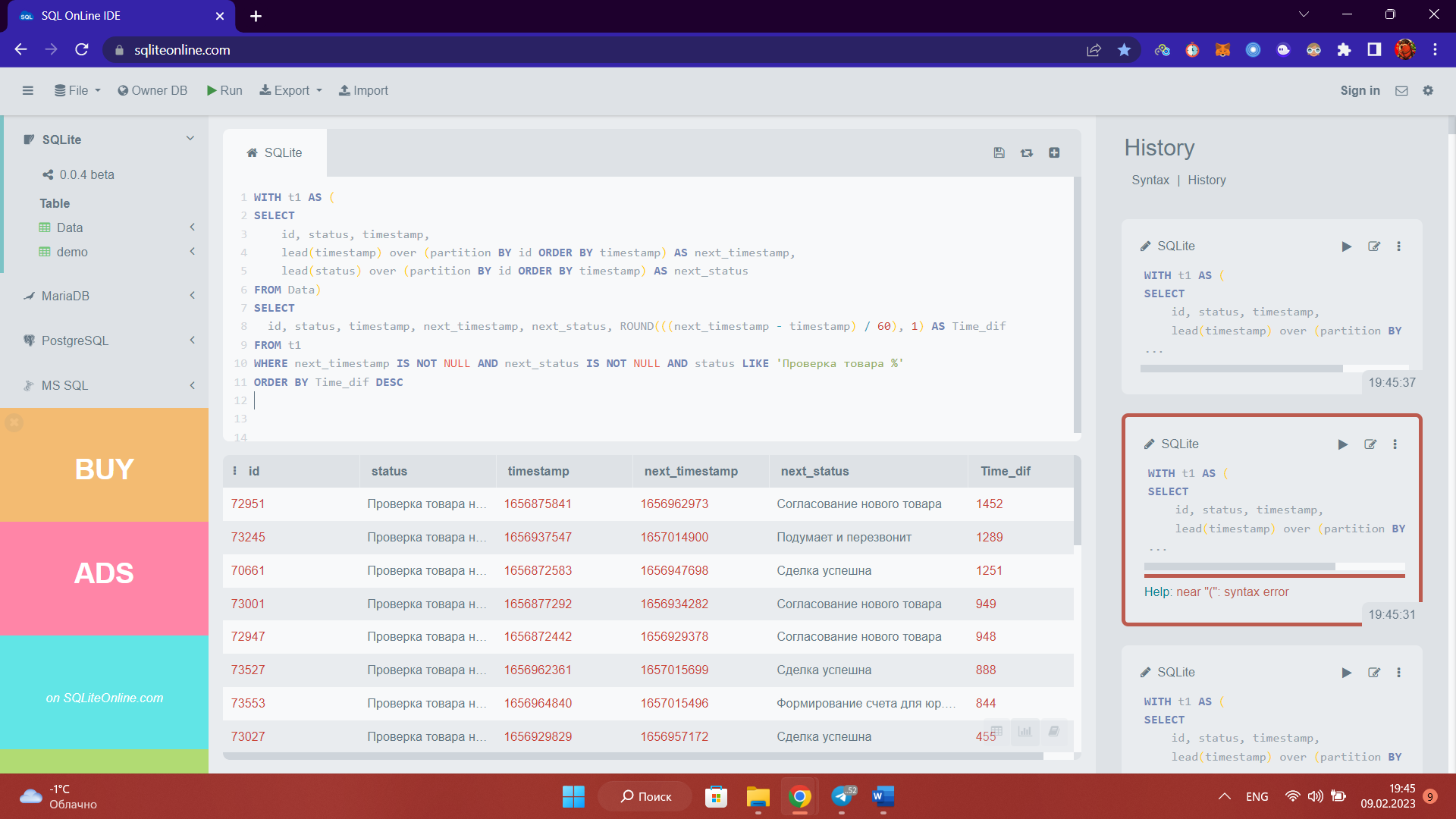
**Доп. вопросы:** корректны ли результаты, которые Вы получили, с точки зрения бизнеса / здравого смысла? Если да, то почему. Если нет, то какая причина и как изменить запрос, чтобы стало лучше.

**Ответ:**

Результаты данных запросов позволяют увидеть какие отделы работают быстрее, а какие медленнее. Не удивительно, что больше всего занимает категория, товар не устраивает.

Однако проверка на складе занимает больше 5 часов. Если склад большой, это ещё имеет какой-то смысл, но скорее всего, дело в некорректном отображении статуса.

Сделав вот такой запрос можно убедиться, что timestamp иногда не меняет статус, и смена статуса с Проверка на складе на следующий переносится на несколько дней вперёд. Возможно, работники забывают изменить статус? Это можно проконтролировать.



Код запроса:

WITH t1 as (

SELECT

id, status, timestamp,

lead(timestamp) over (partition by id order by timestamp) as next\_timestamp,

lead(status) over (partition by id order by timestamp) as next\_status

FROM Data)

SELECT

id, status, timestamp, next\_timestamp, next\_status, ROUND(((next\_timestamp - timestamp) / 60), 1) as Time\_dif

FROM t1

WHERE next\_timestamp IS NOT NULL and next\_status IS NOT NULL AND status LIKE 'Проверка товара %'

ORDER BY Time\_dif DESC

Удалив такие выбросы среднее время в статусе “Проверка на складе” уменьшится.

Также можно добавить Id сотрудника, который производит смену статуса, для оптимизации работы и оценки продуктивности и выявление возможных ошибок

**Доп. задача**: отображать в результате статус, который является текущим. Подумать, корректно ли его отображать и почему.

**Ответ:**

Отображение текущего статуса запроса в этом запросе нецелесообразно, так как информация неполная и статус мог измениться с момента записи данных. Цель этого запроса — показать среднее время, проведенное запросами в каждом статусе.