ГЛЕБ ЗАЙЦЕВ

DISTINCT В SQL

18 задач про дубли и уникальность

```
SELECT DISTINCT
       client_id, sku
FROM purchases
ORDER BY client_id, sku;

SELECT COUNT(DISTINCT sku)
FROM items;
```

уникальные клиенты · пары · ключи

COUNT DISTINCT · дубли · аудит

все запросы запускаются в SQLite

Глеб Зайцев

DISTINCT в SQL. 18 задач про дубли и уникальность

«Автор»
2026

Зайцев Г.

DISTINCT в SQL. 18 задач про дубли и уникальность /
Г. Зайцев — «Автор», 2026

Практическая книга по DISTINCT и дублям в SQL для аналитиков и начинающих разработчиков. Внутри 18 задач на маленьких таблицах: уникальные клиенты, COUNT DISTINCT, дубли справочника, повторы событий, уникальные пары, отличие DISTINCT от GROUP BY, дедупликация перед JOIN и финальный аудит каталога. Все примеры запускаются в SQLite.

Содержание

Глеб Зайцев
DISTINCT в SQL. 18 задач про дубли и уникальность

Перед первой задачей

Почему DISTINCT не должен быть кнопкой паники

Когда отчёт внезапно раздулся, рука тянется добавить DISTINCT. Иногда это правильно, но часто DISTINCT просто прячет проблему зерна или неправильного JOIN.

Уникальность всегда нужно формулировать явно: уникальные клиенты, уникальные SKU, уникальные пары клиент–товар или уникальные события по ключу.

Эта книга тренирует не только синтаксис DISTINCT, но и мышление вокруг дублей: где они допустимы, где опасны и когда их нельзя удалять без расследования.

Как читать задачи

Перед каждым запросом спросите себя: по каким колонкам мы считаем строки одинаковыми. DISTINCT сравнивает весь выбранный набор колонок.

Если нужно посчитать уникальные значения, используйте COUNT(DISTINCT ...). Если нужно выбрать представителя дублей, часто понадобится GROUP BY или отдельное правило.

Все запросы рассчитаны на SQLite. Базовые идеи DISTINCT, GROUP BY и проверки дублей полезны почти в любой SQL-системе.

Маршрут по задачам

Практические задачи

Задача 1. Получить список клиентов с заказами

Рабочий вопрос

Нужно вывести каждого клиента один раз, даже если он сделал несколько заказов.

Данные

```
CREATE TABLE orders(id INTEGER, client_id INTEGER);
INSERT INTO orders VALUES (1,10),(2,10),(3,20),(4,30),(5,20);
```

Запрос

```
SELECT DISTINCT client_id
FROM orders
ORDER BY client_id;
```

Ожидаемый результат

```
[[10], [20], [30]]
```

Почему работает

DISTINCT оставляет уникальные значения client_id.
Повторные заказы клиента 10 превращаются в одну строку.
ORDER BY делает список удобным для проверки.

Где легко ошибиться

Если добавить id заказа в SELECT, строки снова станут уникальными, потому что DISTINCT сравнивает весь выбранный набор колонок.

Самостоятельная проверка

Почему клиент 20 появляется один раз, хотя заказов у него два?
Ответ: Потому что в SELECT выбран только client_id, и DISTINCT убирает повтор этого значения.

Задача 2. Посчитать число уникальных клиентов

Рабочий вопрос

Нужно получить одну цифру: сколько разных клиентов сделали хотя бы один заказ.

Данные

```
CREATE TABLE orders(id INTEGER, client_id INTEGER);
INSERT INTO orders VALUES (1,10),(2,10),(3,20),(4,30),(5,20);
```

Запрос

```
SELECT COUNT(DISTINCT client_id) AS unique_clients
FROM orders;
```

Ожидаемый результат

```
[[3]]
```

Почему работает

COUNT(DISTINCT client_id) считает разные значения client_id.
Клиенты 10 и 20 имеют повторы, но учитываются по одному разу.
Итоговая цифра равна 3: 10, 20 и 30.

Где легко ошибиться

COUNT(*) здесь ответил бы на другой вопрос: сколько всего заказов, а не сколько уникальных клиентов.

Самостоятельная проверка

Какой результат дал бы COUNT(*) на этих данных?
Ответ: COUNT(*) дал бы 5, потому что в таблице пять строк заказов.

Задача 3. Уникальные пары клиент и товар

Рабочий вопрос

Нужно понять, какие пары клиент–товар встречались хотя бы раз, без повторов покупок.

Данные

```
    CREATE TABLE purchases(id INTEGER, client_id INTEGER, sku
TEXT);
    INSERT INTO purchases VALUES (1,10,'A1'),(2,10,'A1'),
(3,10,'B2'),(4,20,'A1');
```

Запрос

```
    SELECT DISTINCT client_id, sku
    FROM purchases
    ORDER BY client_id, sku;
```

Ожидаемый результат

```
    [[10, "A1"], [10, "B2"], [20, "A1"]]
```

Почему работает

DISTINCT смотрит на пару выбранных колонок.
Две покупки 10-A1 превращаются в одну пару.

Конец ознакомительного фрагмента.