

SQL

Занятие 2

Агрегатные функции. Подзапросы.

Бояр Владислав

Группировка данных.
Агрегатные функции.



Что такое группировка данных?



Что такое агрегатная функция?



Можно ли применять группировку данных без агрегатной функции?



Можно ли применять агрегатную функцию без группировки?

Группировка данных

- Выполняется с помощью оператора **GROUP BY**

```
SELECT ...  
FROM ...  
WHERE ...  
GROUP BY column_1 [, column_2, ..., column_N] -- группировка данных
```

- При использовании группировки, в выводе могут участвовать только те колонки, которые указаны после GROUP BY, а также агрегационные функции над любыми колонками
- Группировка сокращает количество строк выборки до количества уникальных значений группируемых полей;

Пример группировки данных

Employee ID	Name	Department
1001	Kundan	Sales
1002	Virat	Marketing
1003	Santosh	Education
1004	Veer	Marketing
1005	Shivani	Sales
1006	Yogesh	Education

GROUP BY Department

Department
Sales
Marketing
Education

Что такое агрегатная функция?

Агрегатная функция вычисляет единственное значение в результате обработки множества строк. Наиболее часто применяется в сочетании с группировкой данных для вычисления значений по каждой группе.


Наиболее популярные агрегатные функции:

- **count**(количество), **sum**(сумма), **max**(максимум), **min**(минимум)
- **avg**(среднее), **mode** (мода)
- **string_agg**, **array_agg**, **json_agg** – объединение данных
- **stddev** (стандартное отклонение), **variance** (дисперсия), **percentile_cont** (перцентиль)

Агрегатная функция без группировки применяется ко всему набору данных (ко всей колонке).

Место агрегатных функции в запросе

SELECT можно здесь -- вычисление и вывод агрегированных значений
FROM ...
WHERE ...
GROUP BY ...
HAVING здесь -- фильтрация данных после группировки
ORDER BY и здесь -- сортировка на основе агрегированных значений



Синтаксис агрегатных функции

SELECT

- `count(*)` -- кол-во всех строк в таблице
- `count(column)` -- кол-во строк колонки, где значение не равно NULL
- `sum(column)` -- сумма
- `max(column)` -- максимальное значение в колонке
- `min(column)` -- минимальное значение в колонке
- `avg(column)` -- среднее значение в колонке
- `percentile_cont(0.5) WITHIN GROUP (ORDER BY column)` -- медиана
- `mode() WITHIN GROUP (ORDER BY column)` -- мода
- `stddev(column)` -- стандартное отклонение
- `variance(column)` -- дисперсия

FROM table



Можно ли применять группировку данных без агрегатной функции?



Можно ли применять агрегатную функцию без группировки?

Подзапросы

Что такое подзапрос?

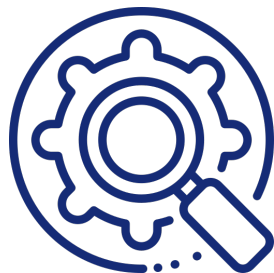
Подзапросы (subqueries) - это запросы, которые вложены в другие запросы.

Результатом подзапроса может быть одно значение или целая таблица.

Подзапросы используются для выполнения фильтрации, сравнения и вычисления значений.



Практика



Практика

1. Вывести уникальные направления перелётов (аэропорт отправления – аэропорт прибытия) двумя способами;
2. Найти аэропорт с наибольшей географией отправления;
3. Посчитать для таблицы с билетами (ticket_flights):
 1. Количество билетов каждого класса;
 2. Сумму, максимум, минимум, среднее, медиану, моду, стандартное отклонение и дисперсию стоимости билетов каждого класса;
4. Найти минимальную, среднюю и максимальную продолжительность полета для каждой модели самолёта. Формат вывода: «часы:минуты»;
5. Найти аэропорты (departure_airport), где количество рейсов превышает среднее количество рейсов по всем аэропортам;
6. Вывести номер бронирования, количество и список пассажиров в бронировании в виде массива. Отсортировать по количеству пассажиров в обратном порядке