

SQL Group By

We have seen that Group Functions summarize values in an interim answer set to return values grouped across many rows.

The GROUP BY clause enables SQL to provide subtotals.

The GROUP BY tells SQL to perform the group function across a **subset** of rows in the interim answer set and provide a total for each subset of rows.

SQL Group By

Remember that this query triggers an error?

```
SELECT OrderID, ProductID, SUM(UnitPrice), SUM(Quantity)
      FROM "alanparadise/nw"."orderdetails"
      WHERE OrderID in (10248, 10249, 10250, 10251);
```

→ Single sums; many orders & products.

The level of the group function must match the level of detail in your select statement.

SQL Group By

By adding a GROUP BY, we avoid the mismatch, and get subtotals by OrderID.

```
SELECT OrderID, SUM(UnitPrice) as "Total Price",  
        SUM(Quantity) as "Total Quantity"  
FROM "alanparadise/nw"."orderdetails"  
WHERE OrderID in (10248, 10249, 10250, 10251)  
GROUP BY OrderID;
```

SQL Group By

To avoid this mismatch, remember the rule:

When using a GROUP BY, every column/expression in the SELECT statement must either be

- a group function

OR

- a column that you are grouping by

SQL Group By

Let's practice

1. How many customers does Northwinds have in each country?

```
SELECT country, count(customerid) as "Total"  
      FROM "alanparadise/nw"."customers"  
  
GROUP BY country;
```

2. What is the average product price for each supplier?

```
SELECT supplierid, avg(unitprice) as "Total"  
      FROM "alanparadise/nw"."products"  
  
GROUP BY supplierid;
```

(How can I make this output more useful?)

SQL Group By

NOTE:

Even though the GROUP FUNCTION is sorting the interim answer set in order to calculate totals, you must **STILL** use an ORDER BY to list the totals in group sequence.

SQL Group By

3. From which supplier does Northwinds carry the most inventory?

```
SELECT SupplierID, SUM(UnitsInStock) AS "Inventory"  
FROM "alanparadise/nw"."products"  
GROUP BY SupplierID  
ORDER BY 2 DESC LIMIT 1;
```


SQL Group By

HAVING

Defines a condition that selects a subset of rows from the answer set based on the group functions

8

When you use a GROUP BY, the HAVING is like a WHERE clause against the interim answer set

The syntax rules for condition in the HAVING are the same as for the WHERE

SQL Group By

4. In which countries does Northwinds have more than five customers?

```
SELECT country, count(customerid) as "Total"  
      FROM "alanparadise/nw"."customers"  
      GROUP BY country  
      HAVING count(customerid) > 5  
      ORDER BY 2 DESC;
```

SQL Group By

Remember:

You can select rows using the WHERE

You can select rows using the HAVING

So what's the difference?