Aggregate Functions

by Sophia

WHAT'S COVERED

In this lesson, you will explore the different types of aggregate functions that can be used for querying tables. Specifically, this lesson will cover:

1. The Value of Aggregating Data

Most of the queries we have been working with so far have produced a list of individual records in a table. However, in many cases, a database works better when using data sets rather than individual records. Many questions can be answered with a database by **aggregating** the data together as if it were a single unit. Being able to gather information like the company's top 10 customers or who bought the most expensive products opens the door to knowing more about the people or systems that the database is meant to track.

TERMS TO KNOW

Aggregate
To combine individual items or objects and treat the combined set as a single unit.

2. Aggregate Functions

In SQL, data is aggregated by using aggregate functions. Although different databases may have some specialized aggregate functions, most database systems share a common basic set.

SQL can perform various mathematical summaries for us within the database to help answer questions. For example, how many invoices were there in January? Of course, we could count this out by hand or by trying to count them in a large SELECT statement. But we also have functions like COUNT() that will do the work for us. Or what if we wanted to know the total revenue generated by sales in a certain week? We could use a SELECT query to filter the records in the Invoice table by a certain date range, and then use a SUM() function to sum the

© 2024 SOPHIA Learning, LLC. SOPHIA is a registered trademark of SOPHIA Learning, LLC.

values from the Total column in that result set. The functions that operate on multiple records to answer such questions are known as **aggregate functions**.

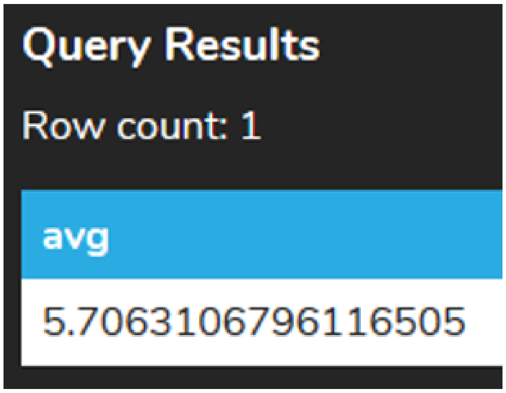
Aggregate functions have a wide variety of uses. For example, they could help us find the smallest invoice total in the prior year, the maximum invoice total of all time, or the average quantity ordered across all invoices. All of this is useful to managers, accountants, and people who need at-a-glance data. A large part of the job of a database administrator, data wrangler, or even a big data programmer is to use aggregate functions to write queries that will provide the information that managers and other decision-makers need to have.

Here is a list of the most common aggregate functions. We will get into more specifics on each in later lessons.

AVG

The AVG function calculates the average of non-null values.
As an example, if we were looking for the average of the total items purchased across all invoices, we could run:

```
SELECT AVG(total)<br>
FROM invoice;
```



Query Results

Row count: 1

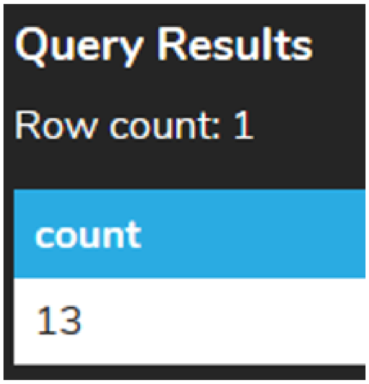
avg
5.7063106796116505

COUNT

The COUNT function returns the number of rows in a result set.
If we wanted to find out how many customers lived in the country USA, we would run:

```
SELECT COUNT(customer_id)<br>
FROM customer<br>
WHERE country = 'USA';
```

© 2024 SOPHIA Learning, LLC. SOPHIA is a registered trademark of SOPHIA Learning, LLC.



Query Results

Row count: 1

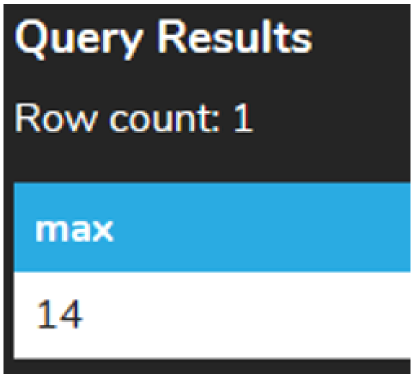
count
13

MAX

| The MAX function returns the largest of the non-null values.

If we wanted to find the largest order made between January 1, 2009, and January 1, 2010, we would run:

```
SELECT MAX(total)<br>
FROM invoice<br>
WHERE invoice_date between '2009-01-01' AND '2010-01-01';
```



Query Results

Row count: 1

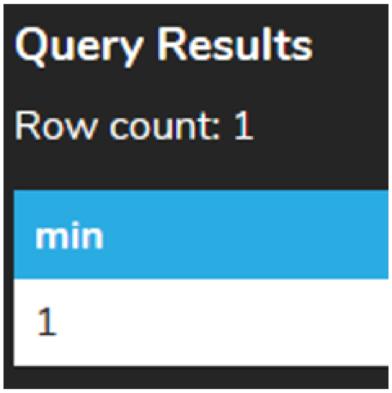
max
14

MIN

The MIN function returns the smallest of the non-null values.

If we wanted to find the smallest total before January 1, 2011, we would run:

```
SELECT MIN(total)<br>
FROM invoice<br>
WHERE invoice_date < '2011-01-01';
```



Query Results

Row count: 1

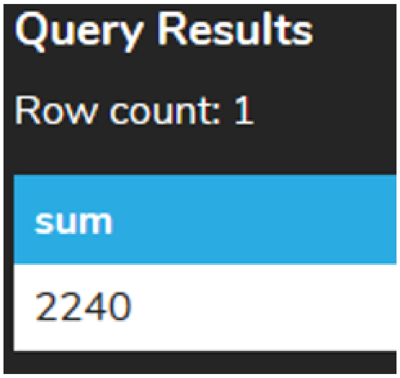
min
1

SUM

The SUM function returns the sum of all of the non-null values.

If we wanted to find out the number of all the tracks ordered for all time, we would run:

```
SELECT SUM(quantity)<br>
FROM invoice_line;)
```



Query Results

Row count: 1

sum
2240

|}

There are also modifiers you can use to find more specific data in the database.

STDDEV	This function returns the standard deviation based on the non-null values.
VARIANCE	This function returns the variance of the non-null values.
RANK	This function returns the rank of the row based on the value. A row with the same value as the prior row will return the same rank.

Each of these has its own special use.

STDDEV can be used for finding the standard deviation in a table. VARIANCE, RANK, and others are all part of data management.

For example, suppose you have a table in your database called netflix_titles, and it has a column called duration_minutes. If you wanted to find the standard deviation of the run times for all movies, you could use the following statement:

```
SELECT
ROUND(STDDEV(duration_minutes), 2) AS standard_deviation
FROM netflix_titles;
```

You could also find the VARIANCE of all the movies times in the database by using the following statement:

```
SELECT
ROUND(SQRT(VARIANCE(duration_minutes)), 2) AS stddev_using_variance
FROM netflix_titles;
```

You could use RANK to put all the items in a products table in descending order by price:

© 2024 SOPHIA Learning, LLC. SOPHIA is a registered trademark of SOPHIA Learning, LLC.

```
SELECT
product_id,
product_name,
price,
RANK () OVER (
           ORDER BY price DESC
) price_rank
FROM
          products;
```

In upcoming lessons, you will learn much more about different aggregate functions.

▶ WATCH

✎ TRY IT

Your turn! Open the SQL tool by clicking on the LAUNCH DATABASE button below. Then, enter in one of the examples above and see how it works. Next, try your own choices for which columns you want the query to provide.

🗎 TERMS TO KNOW

Aggregate Function

A function that operates on multiple records and returns a single value that summarizes them, such as producing a count, sum, or average.

AVG

The function that calculates the average of non-null values.

COUNT

The function that returns the number of rows in a result set.

MAX

The function that returns the largest of the non-null values.

MIN

The function that returns the smallest of the non-null values.

RANK

This function returns the rank of the row based on the value. A row with the same value as the prior row will return the same rank.

STDDEV

This function returns the standard deviation based on the non-null values.

SUM

The SUM function returns the sum of all non-null values.

© 2024 SOPHIA Learning, LLC. SOPHIA is a registered trademark of SOPHIA Learning, LLC.

VARIANCE

This function returns the variance of non-null values.

SUMMARY

In this lesson, you learned **the value of aggregating data** in deriving insights and summaries from large datasets. You also learned that PostgreSQL's **aggregate functions** are powerful tools for calculating across groups of rows and producing single values. SUM, AVG, COUNT, MIN, and MAX are functions that can operate on a column or an expression across multiple rows. These functions enable you to calculate averages, sum values, count occurrences, find minimums and maximums, and more.

Source: THIS TUTORIAL WAS AUTHORED BY DR. VINCENT TRAN, PHD (2020) AND Faithe Wempen (2024) FOR SOPHIA LEARNING. PLEASE SEE OUR **TERMS OF USE**.

TERMS TO KNOW

AVG

The function that calculates the average of non-null values.

Aggregate

To combine individual items or objects and treat the combined set as a single unit.

Aggregate Function

A function that operates on multiple records and returns a single value that summarizes them, such as producing a count, sum, or average.

COUNT

The function that returns the number of rows in a result set.

MAX

The function that returns the largest of the non-null values.

MIN

The function that returns the smallest of the non-null values.

RANK

This function returns the rank of the row based on the value. A row with the same value as the prior row will return the same rank.

STDDEV

This function returns the standard deviation based on the non-null values.

SUM

The SUM function returns the sum of all non-null values.

© 2024 SOPHIA Learning, LLC. SOPHIA is a registered trademark of SOPHIA Learning, LLC.

VARIANCE

This function returns the variance of non-null values.

© 2024 SOPHIA Learning, LLC. SOPHIA is a registered trademark of SOPHIA Learning, LLC.