Steve Stedman
Freelance SQL Server Consultant
http://stevestedman.com

More TSQL 2012 Analytics CUME_DIST - Cumulative Distribution

Continuing on the TSQL 2012 Analytic Series now on to the CUME_DIST function

SQL Server 2012 introduces another new analytic function. The Cumulative Distribution Function CUME_DIST() refers to the probability that the value of a random variable falls within a specified range.

CUME_DIST is the function that maps values to their percentile rank in a distribution.
 CUME_DIST function calculates the possibility of another occurrence being of that value or lesser than that among a group of values.

Lets jump into the examples:

Which produces the following output:

In the first example with the years being the range looked at from 1999 to 2012, the difference between each cumulative distribution value is 7.1% which with an even calculation could be determined by dividing 100 by the number of values wich produces 7.1% (.0714285....).

That's the easy one. You don't need a TSQL Function to calculate 100 divided by the the number of rows.

Now on the the second distribution which calculates the distribution over the revenue numbers. I have added another column to the example above which rounds the Cumulative Distribution to a percentage with one decimal.

From here you can see that the lowest Revenue value of 10,000 was hit in both 2004 and 2012. Keeping in mind the number 7.14% from the previous example you can see that 14.3% is double 7.14% since there are two years with a $10,000 revenue.

This can be useful to determine the percentage of values that are less than or equal to the current value.

That's it for today. I hope you find CUME_DIST to be useful.