### percentile()

그룹에 속한 값들 중 특정 분위에 해당하는 값을 구하는 함수입니다. 백분위 기준으로 분위 값을 입력하며, 100은 최상위, 0은 최하위를 의미합니다.

percentile(x,0.0) 결과값은 [min(x)](https://docs.logpresso.com/ko/query/min-aggregate-function)와 동일하고 percentile(x,100.0) 결과값은 [max(x)](https://docs.logpresso.com/ko/query/max-aggregate-function)와 동일합니다.

#### 문법

percentile(EXPR, PERCENTILE, [INTERPOLATION])

필수 매개변수

**EXPR**

그룹값을 반환하는 표현식

**PERCENTILE**

구하려는 분위 값

PERCENTILE 값은 0.0 ~ 100.0 사이 값이어야 하고, 범위를 벗어날 경우 쿼리 파싱 에러가 발생합니다.

**INTERPOLATION**

두 값 사이값을 계산할 방법

* interpolation 값은 선택 매개변수로, 원하는 분위에 데이터가 없는 경우 양쪽 값을 어떻게 처리하여 분위 값을 구할지를 결정합니다. 예를 들어 다음과 같이 [1,2,3,4] 4개 값이 있는 경우, percentile(x,50.0)을 구하려고 하면 2와 3 사이값을 구해야 합니다. 단순히 생각하면 두 값의 평균인 2.5를 결과로 받아오면 될 것 같지만 데이터 유형에 따라 다른 방법으로 계산할 필요가 있습니다.기본 값은 "linear"입니다.
* linear: 두 값의 PERCENTILE에 따른 가중치를 계산
* midpoint: 두 값의 평균을 계산
* lower: 두 값 중 더 작은 값을 선택
* higher: 두 값 중 더 큰 값을 선택
* nearest: 두 값 중 더 PERCENTILE에 가까운 값을 선택
* 예시) 다음 11개 값에 대해 37% PERCENTILE을 구할 경우, 각 INTERPOLATION 유형에 따른 결과는 다음과 같습니다.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 구분 | 1번 | 2번 | 3번 | 4번 | 5번 | 6번 | 7번 | 8번 | 9번 | 10번 | 11번 |
| 원본 | 97 | 18 | 2 | 85 | 81 | 85 | 65 | 14 | 96 | 22 | 42 |
| 정렬 | 2 | 14 | 18 | 22 | 42 | 65 | 81 | 85 | 85 | 96 | 97 |
| percentile | 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |

* 값이 11개이므로 1번째는 0%, 2번째는 10%, 3번째는 20%, ... 11번째는 100%입니다. 그러므로 37%는 4번째 값 22와 5번째 값 42 기준으로 계산합니다.

|  |  |  |
| --- | --- | --- |
| 옵션 | 계산방식 | 결과 |
| linear | 소수점이 0.7이므로 42의 비중을 70%, 22의 비중을 30%로 하여 계산 (22 \* 0.3) + (42 \* 0.7) | 36 |
| midpoint | (22+42)/2 | 32 |
| lower | min(22, 42) | 22 |
| higher | max(22, 42) | 42 |
| neareast | 3.7은 3과 4 중 4에 가까우므로 해당 값을 선택 | 42 |

#### 사용 예

percentile(x,80.0)| # 상위 80% 값을 계산percentile(x,15.5, "midpoint")| # 상위 15.5% 값을 계산. 정확히 15.5%에 해당하는 값이 없을 경우 양 사이값의 평균값을 계산.percentile(x,100.0)| # max(x)percentile(x,0.0)| # min(x)percentile(x,50.0)| # median(x)