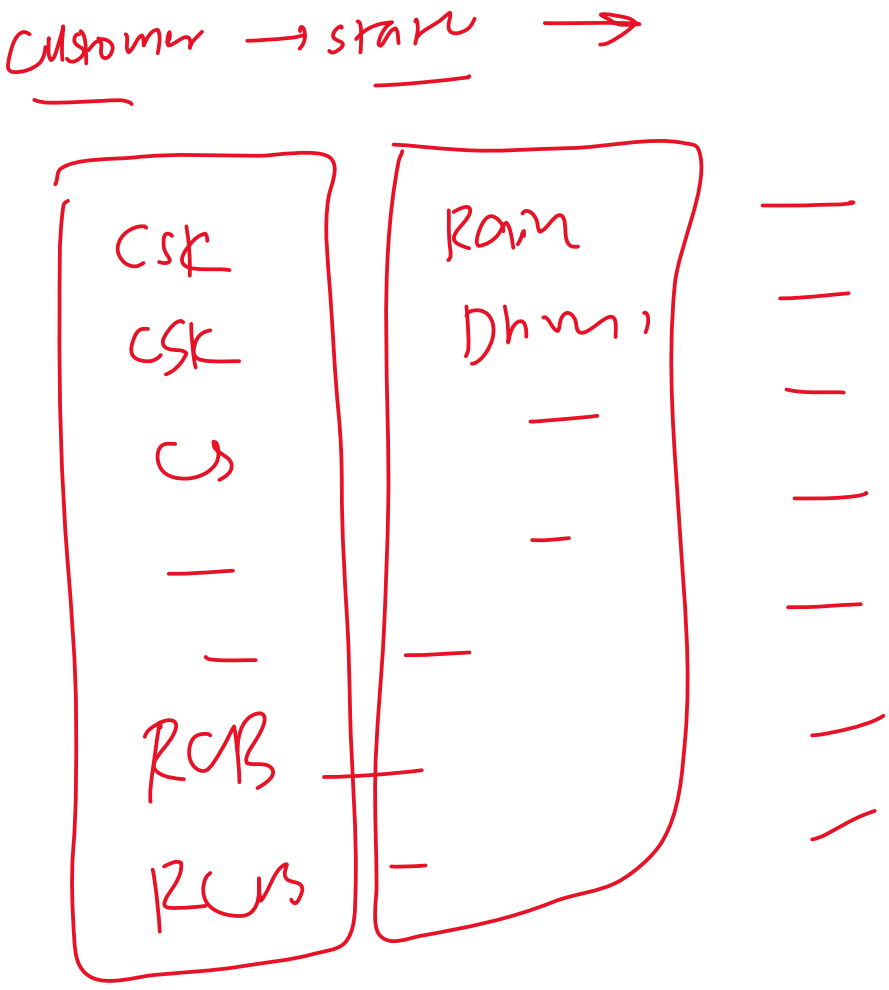
Ranking

01 March 2023 13:57

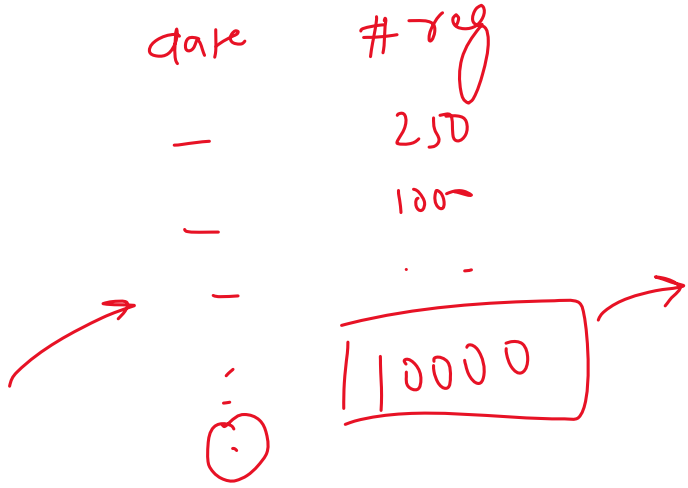


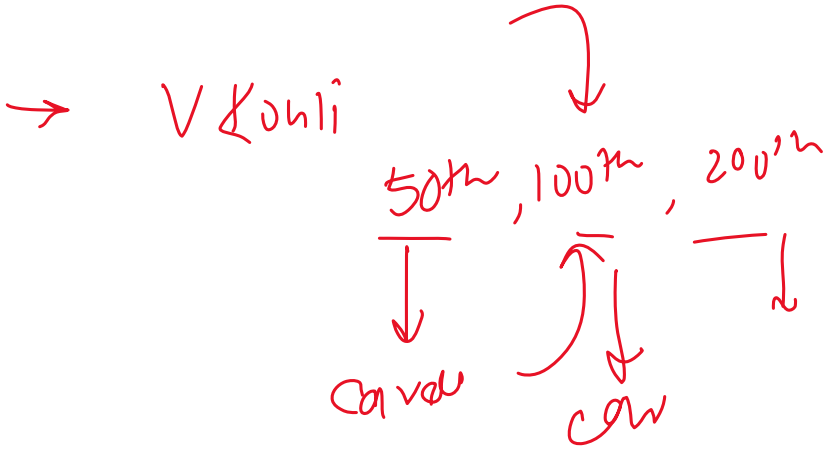
Cumulative Sum

01 March 2023 13:58

Cumulative sum is another type of calculation that can be performed using window functions. A cumulative sum calculates the sum of a set of values up to a given point in time, and includes all previous values in the calculation.

month	year	sales	cumulative_sum
1	2020	100	100
2	2020	150	250
3	2020	200	450
4	2020	175	625
5	2020	250	875
6	2020	300	1175
7	2020	275	1450
8	2020	200	1650





Cumulative Average

01 March 2023 13:58

Cumulative average is another type of average that can be calculated using window functions. A cumulative average calculates the average of a set of values up to a given point in time, and includes all previous values in the calculation.

Student_ID	Test_Number	Score	Cumulative_Avg
1	1	85	85.00
1	2	90	87.50
1	3	80	85.00
2	1	70	70.00
2	2	75	72.50
2	3	80	75.00
3	1	90	90.00
3	2	95	92.50
3	3	90	91.67

Running Average

01 March 2023 13:58

Running average (also known as moving average) is a statistical technique that calculates the average value of a dataset over a moving window of consecutive data points.

The window size determines the number of data points used to calculate the average, and as the window moves forward in time, the average is recalculated using the new data points and dropping the oldest one. This means that the running average is continuously updated and reflects the most recent trends in the data.

For example, a running average of a batsman's runs scored over a window of 10 matches will calculate the average runs scored in the last 10 matches, then move the window one match forward and recalculate the average for the new set of 10 matches, and so on.

Running averages are often used in finance, economics, and engineering to smooth out noisy or volatile data series, and to identify trends or patterns that may be obscured by random fluctuations in the data.

window = 5 matches

current form

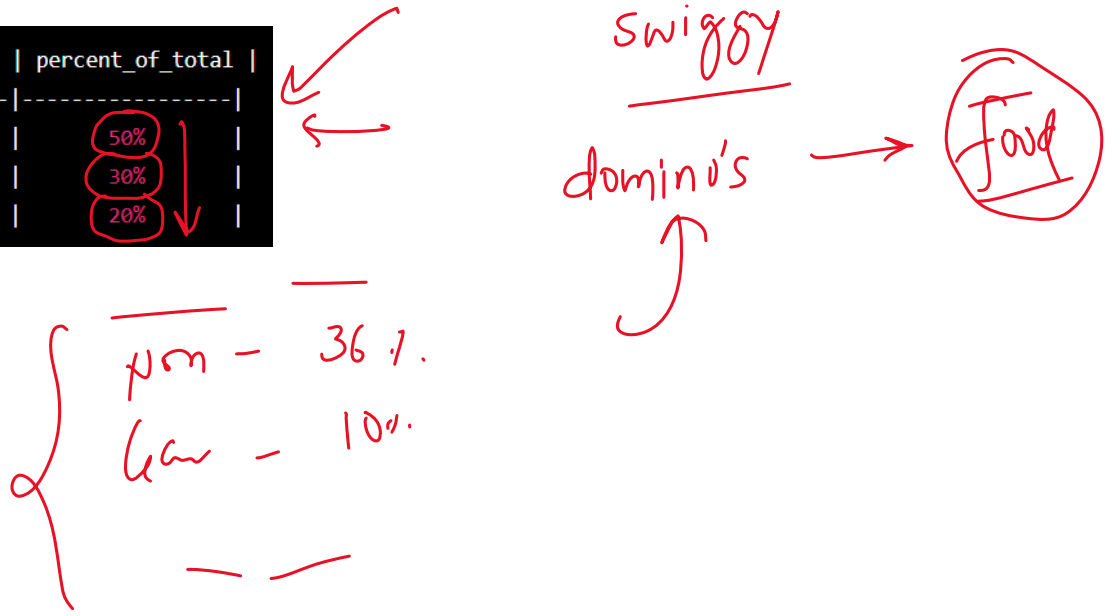
match_id	runs_scored	running_avg	cumulative_avg
1	52	52.0	52.0
2	41	46.5	46.5
3	17	36.7	36.7
4	68	44.5	44.5
5	36	42.8	42.8
6	91	49.2	50.0
7	22	44.0	45.1
8	55	44.9	45.6
9	81	51.2	48.9
10	13	41.6	45.6
11	29	41.5	45.3
12	44	42.3	45.2
13	36	41.4	44.8
14	72	47.9	45.8
15	87	56.0	48.7

Percent of total

01 March 2023 13:59

Percent of total refers to the percentage or proportion of a specific value in relation to the total value. It is a commonly used metric to represent the relative importance or contribution of a particular value within a larger group or population.

category	total_sales	percent_of_total
Category A	500	50%
Category B	300	30%
Category C	200	20%

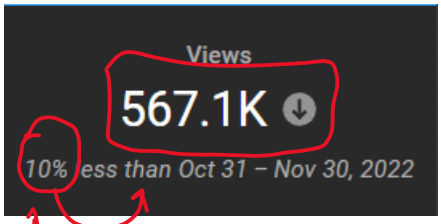


Percent Change

01 March 2023 13:59

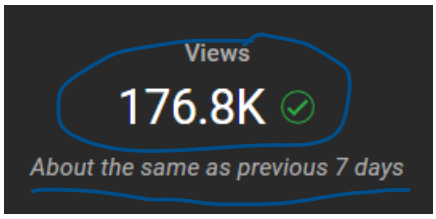
Percent change is a way of expressing the difference between two values as a percentage of the original value. It is often used to measure how much a value has increased or decreased over a given period of time, or to compare two different values.

percent change = ((new value - old value) / old value) x 100



Dec

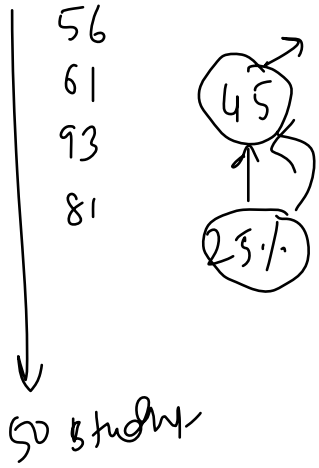
Jan



Percentiles & Quantiles

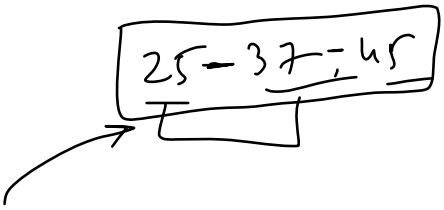
01 March 2023 13:59

A **Quantile** is a measure of the distribution of a dataset that divides the data into any number of equally sized intervals. For example, a dataset could be divided into **deciles** (ten equal parts), **quartiles** (four equal parts), **percentiles** (100 equal parts), or any other number of intervals.



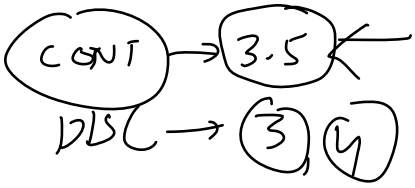
Each quantile represents a value below which a certain percentage of the data falls. For example, the 25th percentile (also known as the first quartile, or Q1) represents the value below which 25% of the data falls. The 50th percentile (also known as the median) represents the value below which 50% of the data falls, and so on.

Q1. Find the median marks of all the students
Q2. Find branch wise median of student marks.



PERCENTILE_CONT calculates the continuous percentile value, which returns the interpolated value between adjacent data points. In other words, it estimates the percentile value by assuming that the values between data points are distributed uniformly. This function returns a value that may not be present in the original dataset.
PERCENTILE_DISC, on the other hand, calculates the discrete percentile value, which returns the value of the nearest data point. This function returns a value that is present in the original dataset.

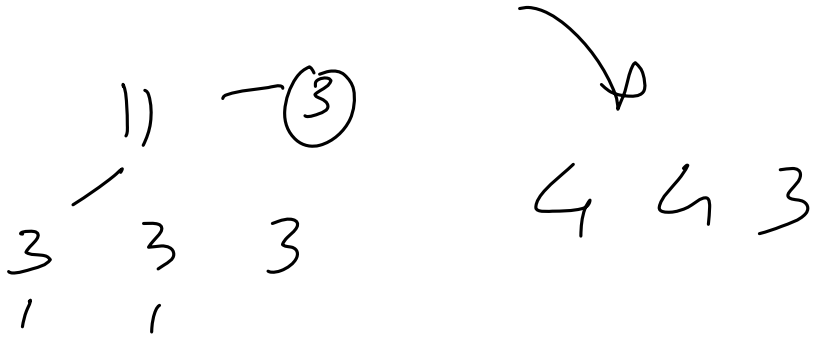
For example if we have 1,2,3,4,4,5



Segmentation

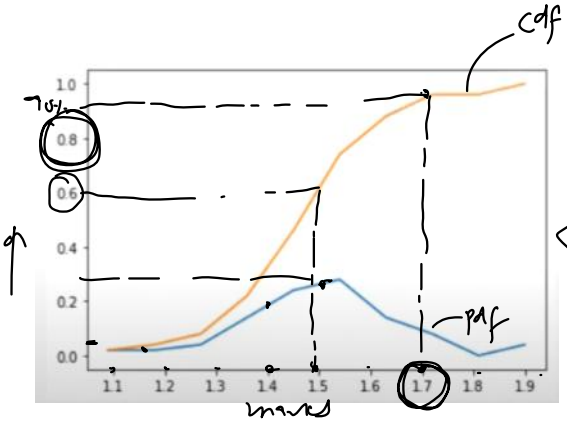
02 March 2023 09:30

Segmentation using NTILE is a technique in SQL for dividing a dataset into equal-sized groups based on some criteria or conditions, and then performing calculations or analysis on each group separately using window functions.



Cumulative Distribution

02 March 2023 09:03



"What percentage of the rows in the data set have a value less than or equal to the current row?"

CUME_DIST = (number of rows with a value less than or equal to the current row) / (total number of rows)

The cumulative distribution function is used to describe the probability distribution of random variables. It can be used to describe the probability for a discrete, continuous or mixed variable. It is obtained by summing up the probability density function and getting the cumulative probability for a random variable

Partition By multiple columns

01 March 2023 14:00