

Datameer Functions

Field Types Used in This Document

Field Type	Description	Internal representation
integer	64-Bit integer value	Java Long
big integer	Unlimited integer value	Java BigInteger
float	64-Bit float value	Java Double
big decimal	High-precision float value	Java BigDecimal
date	Date object	Java Date
string	String object	Java String
boolean	Boolean object	Java Boolean
list	a collection of multiple values of one data type	
number	float, big decimal, integer or big integer	
any	float, big decimal, integer, big integer, date, string or Boolean	

To learn about operators, see [Using Operators](#).

To learn about using functions, see [Using the Formula Builder](#).

To learn about dealing with date and time, see [Date and Time Parse Patterns](#)

Supported Functions - As Grouped in the Formula Builder

Comparison

- **COMPARE** Compares two values. Returns a negative integer, zero, or a positive integer as this object is less than, equal to, or greater than the specified object.
- **EQUALS** Checks to see if two values are identical.
- **GE** Greater equal.
- **GT** Greater than.
- **LE** Less equal.
- **LT** Less than.
- **NOT_EQUALS** Checks to see if two values aren't identical.

Date and Time

- **ADDTODATE** Adds (or subtracts) time to a date using an integer and one date constant.
- **AFTER** Tests if this date is after the specified date.
- **ASDATE** Converts a character sequence to a date using predefined (dd-MM-yyyy hh:mm:ss, dd.MM.yyyy hh:mm:ss, dd-MM-yyyy, dd.MM.yyyy) or an optional date parse pattern.
- **BEFORE** Tests if this date is before the specified date.
- **CEILINGDATE** Rounds a date argument up to the beginning of the next date interval.
- **CRON_MATCHER** Checks if the given date matches the given cron expression.
- **DAY** Returns only the day of the month of a date. The day is given as an integer, ranging from 1 to 31.

- **DAYOFYEAR** Returns the day of the year for the supplied date in a range from 1 to 366.
- **ENDOFMONTH** Returns the date of the last day of month (ending in the evening at 11:59:59.999 PM / 23:59:59.999) for the given date.
- **ExcelConverter** Converts integer dates from Excel to a date.
- **EXPAND_DATE_RANGE** Expands a date range into multiple rows, one for each interval, one day by default.
- **FLOORDATE** Rounds a date argument down to the beginning of the next date interval.
- **FORMATDATE** Converts date to string by applying a format pattern.
- **FORMATDATE** Converts date to string by applying a format pattern.
- **HOUR** Returns only the hour of a time value. The hour is given as an integer, ranging from 0 (12:00 A.M.) to 23 (11:00 P.M.).
- **MILLISECOND** Truncates a date to only display the milliseconds associate with the date.
- **MINUTE** Returns only the minute of a time value
- **MIDNIGHT** Returns the given date with a time of midnight (12:00:00 am).
- **MONTH** Returns only the month of a date represented by a serial number.
- **MSTODATE** Converts an integer representing a timestamp in milliseconds from January 1, 1970, 00:00:00 GMT to its corresponding date value.
- **NOW** Returns the current date and time.
- **QUARTER** Returns the quarter of the year for the supplied date in a range from 1 to 4.
- **ROUNDTIME** Rounds a date field down to the nearest (s)econd, (m)inute, (h)our, or (d)ay set by a time constant.
- **SECOND** Truncates a date to only display the seconds associate with the date.
- **SHIFTTIMEZONE** Shift the date from one timezone to another.
- **STARTOFMONTH** Returns the date of the first day of month (starting the morning at 00:00:00 / 12:00:00 AM) for the given date.
- **STARTOFWEEK** Returns the date of the first day of week (Sunday night/Monday morning 00:00:00) for the given date.
- **TIMESTAMP** Difference, measured in milliseconds, between the current time and midnight, January 1, 1970 UTC.
- **TIMESTAMPDIFF** Returns the number of whole date intervals between two date arguments.
- **TODAY** Returns the current day (time at midnight).
- **WEEKDAY** Returns an integer between (1-7) corresponding to the day of the week of the date.
- **WEEKOFYEAR** Returns the week of the year for the supplied date in a range from 1 to 53.
- **YEAR** Returns the year corresponding to a date. The year is returned as an integer, ranging from 1900-9999.

Encoding

- **DEC2HEX** Converts a decimal number to hexadecimal.
- **DEC_BASE64** Decodes a base64 encoded string.
- **ENC_BASE64** Encodes string as base64.
- **HEX2DEC** Converts a hexadecimal number given as a string to decimal number.
- **IP_TO_NUMBER** Converts IP4 address to decimal presentation.
- **NUMBER_TO_IP** Converts number to internet standard dot notation.
- **TEXT2HEX** Converts text into a hex string.

General

- **COALESCE** Returns the first non-null argument.
- **COPY** Copies values from a sheet column to a new sheet.
- **DENULLIFY** Converts null values to the non null default value.
- **FLOAT** Converts a text to a float value.
- **ISBLANK** Returns TRUE if the value is blank.
- **ISEVEN** Returns TRUE if the number is even.
- **ISNULL** Returns TRUE if the value is null.
- **ISODD** Returns TRUE if the number is odd.

Grouping

- **GROUPACCUMULATE** Returns the sum of all previous records in a group.
- **GROUPAND** Returns the Boolean value after checking if all the group's values are TRUE.
- **GROUPANOVA** Compares the means of three or more samples to see if they are significantly different at a certain significance level.
- **GROUPANY** Returns any random value of a group.
- **GROUPAVERAGE** Returns the average of its arguments.
- **GROUPBOTTOMN** Selects the bottom N values from a group. If this function is applied on a date column, bottom N means the N least recent dates.
- **GROUPBY** Groups records.
- **GROUPBYBIN** Groups values into bins.
- **GROUPBYCUSTOMBIN** Groups selected values into bins created at a custom sizes.
- **GROUPBYGAP** Groups records by maximum gap between two values in a timestamp or numerical value.
- **GROUPCOMBIN** Generates combinations of values in a group.
- **GROUPCONCAT** Creates a list of all non-null values seen in a group.
- **GROUPCONCATDISTINCT** Concatenates all non-null values seen in a group.
- **GROUPCOUNT** Counts the records in one group.
- **GROUPCOUNTDISTINCT** Counts the distinct values in one group.
- **GROUPFIRST** Returns the first value of a group.
- **GROUPJSONOBJECTMERGE** Merges all elements in a grouped series of JSON maps.
- **GROUPLAST** Returns the last value of a group.
- **GROUPMAP** Creates a JSON map based on two columns of all records within a group, keys and values.

- **GROUPMAX** Maximum of its arguments.
- **GROUPMEDIAN** Returns the median of all values in a group.
- **GROUPMIN** Minimum of its arguments.
- **GROUPOR** Returns the Boolean value after checking if any of the group's values are TRUE.
- **GROUPPERCENTILE** Return the number which is nth percentile in this distribution.
- **GROUPPREDICTIVEWINDOWS** This function assigns IDs to events. It does that based on an event that defines the center of a time-based window (center-event) and all events in that window get assigned the ID of the center-event.
- **GROUPROWNUMBER** Returns the row number within a sorted group.
- **GROUPSELECT** Create a column with selected values using a key constant from a specified grouped column in the workbook.
- **GROUPSESSIONS** Assigns IDs to events based on session start and session end. All events with the session ID of the start event between start and end event get assigned the ID of the start event.
- **GROUPSTDEVP** Estimates standard deviation on the entire population.
- **GROUPSTDEVS** Estimates standard deviation on a sample of the entire population.
- **GROUPSUM** Adds its arguments.
- **GROUPTOPN** Selects the top N values from a group.
- **GROUPTTEST** The Student's t-test compares the means of two samples (or treatments), even if they have a different numbers of replicates. It is used to determine if sets of data are significantly different from each other.
- **GROUPUNIQUES** Outputs unique column values for a group.
- **GROUP_DIFF** Computes the difference to the previous value in a group.
- **GROUP_JACCARD_DIST** Computes the jaccard distance of all values in a group.
- **GROUP_PAIR** Generates pairs of values in a group.
- **GROUP_PATH** Create paths from a field (e.g., for click stream analysis).
- **GROUP_PATH_CHANGES** Create paths from a field, but only if the value changes (e.g., for click stream analysis).
- **GROUP_PREVIOUS** Returns the value of the previous record in a group.
- **GROUP_SORT_ASC** Groups records in sorted ascending order.
- **GROUP_SORT_DESC** Groups records in sorted descending order.

HTML

- **CONTENTS_BY_TAG_NAME** Returns all contents of elements with the specified tag name in a list.
- **ELEMENTS_BY_SELECTOR_QUERY** Returns all elements that match the selector query.
- **ELEMENTS_BY_TAG_NAME** Returns all elements with a specified tag name in a list.
- **PROPERTY_VALUE_BY_TAG_NAME** Returns the value of the specified property by tag name in a list.
- **REMOVE_ELEMENTS_BY_TAG_NAME** Removes elements by tag name from the document object model.
- **REMOVE_HTML_TAGS** Removes all HTML tags.

List

- **ADD** Adds a value or values into a list.
- **COMPRESS** Removes all null values from a list.
- **EXPAND** Creates a new row per element of the argument list.
- **INDEXOF** Returns (as an integer) the index of the first occurrence of an element in a list or -1 if the element isn't found.
- **INTERSECTION** Returns the intersection of all list arguments. The resulting list contains only the elements found in all the list arguments and doesn't contain any duplicate elements.
- **IEMPTY** Checks to see if a list is empty and returns a Boolean value.
- **LIST** Build a list value out of its arguments.
- **LISTELEMENT** Returns the given element of a list.
- **RANGE** Returns in the form of a list, integers ranging in numerical order from the first value (FROM) and ending with last value (TO).
- **REMOVE** Remove a value from a list.
- **REMOVEALL** Removes all elements from one list that exist in a specified list or lists.
- **REVERSE** Reverses a list.
- **SIZE** Returns the size of a list.
- **SORT** Sorts a list.
- **SUBLIST** Returns a sublist of a list.
- **TOJSON** Converts a list to a JSON array.
- **TOKENIZELIST** Tokenizes text returning a list of values or a specified element as a string.
- **UNIQUES** Removes all duplicate values from a list.

Logical

- **AND** Returns TRUE if all of its arguments are TRUE.
- **IF** Specifies a logical test to perform.
- **NOT** Reverses the logic of its argument.
- **OR** Returns TRUE if any argument is TRUE.
- **SELECT** Create columns with selected values using a key constant from a specified column in the workbook.

Math

- **ABS** Returns the absolute value of a number.
- **BIGDECIMAL** Converts the values of a column with a number data type to the big decimal data type.

- **BIGINTEGER** Converts the values of a column with a number data type to the big integer data type.
- **BITAND** Bitwise AND operation on given arguments.
- **BITOR** Bitwise OR operation on given arguments.
- **CEILING** Rounds a number to the nearest integer or to the nearest multiple of significance.
- **COMBIN** Returns the number of combinations for a given number of objects.
- **COMBIND** Returns the double representation of combinations for a given number of objects.
- **CONVERT** Converts a number from one measurement to another measurement.
- **DIFF** Returns the difference of the numbers (First argument value can't be null).
- **DIV** Returns the quotient of the numbers.
- **E** Returns the value of e.
- **EVEN** Rounds a number up to the nearest even integer.
- **EXP** Returns e raised to the power of a given number.
- **EXPLODE_RANGE** Explodes a integer or ip address range into multiple rows. A given ip address range is converted into a range of integer before.
- **FACT** Returns the factorial of a number.
- **FACTD** Returns the double factorial of a number.
- **FLOOR** Rounds a number down, toward 0 (zero).
- **GCD** Returns the greatest common divisor.
- **GEOHASH_DEC_LAT** Decodes the given geohash value into a latitude with a precision of 6 after the decimal place.
- **GEOHASH_DEC_LONG** Decodes the given geohash value into a longitude with a precision of 6 after the decimal place.
- **GEOHASH_ENC** Encodes the given latitude and longitude into a geohash value with precision of 12 digits.
- **INT** Rounds a number down to the nearest integer.
- **LCM** Returns the least common multiple.
- **LN** Returns the natural logarithm of a number.
- **LOG** Returns the logarithm of a number to a specified base.
- **LOG10** Returns the base-10 logarithm of a number.
- **LSHIFT** Shifts bits left by the given number of bits.
- **MOD** Returns the modula of the arguments.
- **MROUND** Returns a number rounded to the desired multiple
- **ODD** Rounds a number up to the nearest odd integer.
- **PI** Returns the value of pi.
- **PLUS** Adds its arguments (or concats string values).
- **POWER** Returns the result of a number raised to a power.
- **PRODUCT** Multiplies the arguments.
- **QUOTIENT** Returns the integer portion of a division.
- **RAND** Returns a random number between 0 (zero) and 1.
- **RANDBETWEEN** Returns a random number between the numbers that you specify.
- **RANDGAUSSIAN** Generates normal distributed random values with mean of 0 and standard deviation of 1.
- **ROUND** Rounds a number to a specified number of digits.
- **ROUNDDOWN** Rounds a number down, toward zero.
- **ROUNDUP** Rounds a number up, away from zero.
- **RSHIFT** Shifts bits right by the given number of bits.
- **SIGN** Returns the sign of a number.
- **SQRT** Returns the square root of a number.
- **SQRTPI** Returns the square root of (number * pi).
- **SUBTRACT** Subtracts all values from the first argument.
- **SUM** Adds its arguments.
- **TRUNC** Truncates a number to an integer or truncates its decimal places.
- **URSHIFT** The unsigned right shift operator shifts a zero into the leftmost position.

Statistical

- **AVERAGE** Returns the average of its arguments.
- **CHISQDIST** Returns a chi square distribution at a given value x for a certain number of degrees of freedom.
- **EXPONDIST** Returns the exponential distribution for a certain value with given lambda.
- **FDIST** Returns the F probability distribution for a certain value with given degrees of freedom for the numerator and degrees of freedom for the denominator.
- **GAMMADIST** Returns the gamma distribution for a certain value with given parameters alpha and beta.
- **LOGNORMDIST** Returns the log-normal distribution for a certain value with given mean and standard deviation .
- **MAX** Returns maximum of arguments.
- **MIN** Returns minimum of arguments.
- **NORMDIST** Returns the normal distribution for a certain value with given mean and standard deviation.
- **STDEVS** Estimates standard deviation on a sample of the entire population.
- **STDEV** Estimates standard deviation on the entire population.
- **TDIST** Returns student t-distribution at a given value x for a certain number of degrees of freedom.
- **WEIBULLDIST** Returns the weibull distribution for a certain value with given alpha and beta.

Text

- **CHAR** Returns the character specified by the codepoint number.
- **CODE** Returns a numeric code for the first character in a text string
- **CONCAT** Concatenates several text items into one text item.

- **CONTAINS** Returns whether text contains the second argument.
- **CONTAINS_IC** Returns whether text contains the second argument. (Not case sensitive)
- **ENDSWITH** Tests if this string ends with the specified suffix.
- **INDEX** Finds the first occurrence of the given regular expression.
- **JSONTOLIST** Converts the selected JSON data type column into a list data type column.
- **JSON_ELEMENT** Returns the given element of a json array.
- **JSON_ELEMENTS** Returns all elements of a json array.
- **JSON_KEYS** Returns all keys from a json string.
- **JSON_MAP** Creates a JSON map from two JSON arrays.
- **JSON_VALUE** Returns a value for a given key from a json object.
- **LEFT** Returns the leftmost characters from a text value.
- **LEN** Returns the number of characters in a text string.
- **LOWER** Converts text to lowercase.
- **LTRIM** Removes leading whitespace or other characters from text.
- **MATCHES** Return true if a string matches a regular expression.
- **NGRAM** Returns subsequences of N characters of a given text.
- **NGRAM_OLD** Returns subsequences of N words of a given text.
- **PROPER** Capitalizes the first letter in each word of a text value.
- **REGEX** Replaces characters within a text string which matches a regular expression with a different text string..
- **REGEXTRACT** Extracts tokens that match a regular expression.
- **REPLACE** Replaces characters within a text string with a different text string.
- **REPLACEALL** Replaces each substring of this string that matches the given regular expression with the given replacement.
- **REPT** Repeats text a given number of times.
- **RIGHT** Returns the rightmost characters from a text value.
- **RTRIM** Removes trailing whitespace or other characters from text.
- **STARTSWITH** Tests if this string starts with the specified prefix.
- **SUBSTITUTEALL** Substitutes new text for all occurrences of old text in a text string.
- **SUBSTITUTEFIRST** Substitutes new text for first occurrence of old text in a text string.
- **SUBSTR** Returns a specific number of characters from a text string starting at the position that you specify.
- **T** Converts its arguments to text.
- **TOKENIZE** Tokenizes text.
- **TRIM** Removes spaces from text.
- **UPPER** Converts text to uppercase.

Text Mining

- **ANALYZE_POLARITY** Analyze the polarity of a text.
- **CORRECT_SPELLING** Corrects spelling of tokenized text.
- **EXTRACT_ADJECTIVE** Extracts adjectives from text and returns the results in a list.
- **EXTRACT_DATE** Extracts dates from text and returns the results in a list.
- **EXTRACT_LOCATION** Extracts locations from text and returns the results in a list.
- **EXTRACT_MONEY** Extracts money from text and returns the results in a list.
- **EXTRACT_NOUN** Extracts nouns from text and returns the results in a list.
- **EXTRACT_ORGANIZATION** Extracts organizations from text and returns the results in a list.
- **EXTRACT_PERSON** Extracts people from text and returns the results in a list.
- **EXTRACT_SMILEYS** Extracts emoticons from text and returns the results as a list.
- **EXTRACT_VERB** Extracts verbs from text and returns the results in a list.
- **EXTRACT_WORDS** Converts text to a list of words using a preconfigured maximum entropy tokenizer.
- **REMOVE_STOP_WORDS** Removes StopWords (common words) from a list of tokens.
- **STEM_WORDS** Performs stemming of tokenized text using the Porter algorithm.

Trigonometry

- **ACOS** Returns the arccosine of a number.
- **ACOSH** Returns the inverse hyperbolic cosine of a number.
- **ASIN** Returns the arc sine of a number.
- **ASINH** Returns the inverse hyperbolic sine of a number.
- **ATAN** Returns the arc tangent of a number.
- **ATAN2** Returns the arctangent from x- and y-coordinates.
- **ATANH** Returns the inverse hyperbolic tangent of a number.
- **COS** Returns the cosine of a number.
- **COSH** Returns the hyperbolic cosine of a number.
- **COT** Returns the cotangent of a number.
- **DEGREES** Converts radians to degrees.
- **RADIANS** Converts degrees to radians.
- **SIN** Returns the sine of a number.
- **SINH** Returns the hyperbolic sine of a number.
- **TAN** Returns the tangent of a number.
- **TANH** Returns the hyperbolic tangent of a number.

URL

- **URL_AUTHORITY** Gets the authority part of the URL provided as string.
- **URL_DECODE** Decodes an encoded string
- **URL_DEFAULT_PORT** Gets the default port number of the protocol associated with this URL.
- **URL_ENCODE** Encodes a string.
- **URL_FILE** Gets the file name of the URL provided as string.
- **URL_HOST** Gets the host name of the URL provided as string.
- **URL_PARAM** Gets a value addressed by a param of the URL provided as string.
- **URL_PARAMS** Gets the values addressed by a param of the URL provided as list.
- **URL_PATH** Gets the path part of the URL provided as string.
- **URL_PLD** Extract the PLD (paid-level domain, as per the IRLbot paper) from a url.
- **URL_PORT** Gets the port number of the URL provided as long.
- **URL_PROTOCOL** Gets the protocol name of the URL provided as string.
- **URL_QUERY** Gets the query part of the URL provided as string.
- **URL_REF** Gets the anchor ('reference') of the URL provided as string.
- **URL_TLD** Gets the TLD of an URL.
- **URL_USERINFO** Gets the userinfo part of the URL provided as string.

XML

- **XMLVALUE** Returns a value selected by the given XPath expression out of an XML document.
- **XMLVALUES** Returns the values selected by the given XPath expression out of an XML document.