DATA_PROFILE view.

A view containing the results of the column profiling. The view contains information from the DATA_PROFILING_COLUMN_STATISTICS table, which the DATA_PROFILE_COLUMNS stored procedure fills. The view also has information dynamically calculated in the DATA_PROFILING_INFER_TYPES user defined table function. The view also calculates columns internally.

Column	Type	Description
DATABASE_NAME	string	The name of the database containing the column.
SCHEMA_NAME	string	The name of the schema containing the column.
TABLE_NAME	string	The name of the table containing the column.
COLUMN_NAME	string	The column profiled for its statistics and contents.
HAS_INFORMATION	boolean	True if the sample found a single value that is not null or blank.
VALUE_CARDINALITY	integer	The cardinality of the samples in the column, (non-null + non-blank). In general, the higher the value cardinality, the more likely the column contains information usable for analytics.
INFORMATION_DENSITY	float	A ratio from 0 to 1 the value cardinality divided by the sample size. This measures how close the column is to containing unique values where 1.0 represents complete uniqueness and 0.0 represents all the same values.
DEFINED_TYPE	string	The data type defined for the column in the Snowflake information_schema.
INFERRED_TYPE	string	The data type the profiler infers the column should be.
CONFIDENCE_LEVEL	string	A qualitative assessment of how confident the profiler is that the inferred data type is correct.

HAS_TYPE_MISMATCH	boolean	A convenience column to filter where the defined type and inferred type do not match.
MOST_COMMON_VALUES	variant	A JSON containing the most common values in the sample, up to 16 total.
MIN_VALUE	string	The minimum value in the sample using the min() function on the defined type.
MAX_VALUE	string	The maximum value in the sample using the max() function on the defined type.
CHARACTER_MAX_LENGTH	integer	The maximum length of a text type column found in the sample. Returns null for non-text types.
NUMERIC_PRECISION	integer	The numeric precision defined in the Snowflake information_schema. Returns null for non-numeric types.
NUMERIC_SCALE	integer	The numeric scale defined in the Snowflake information_schema. Returns null for non-numeric types.
DATETIME_PRECISION	integer	The precision of datetime, timestamp, and time columns. Returns null for other data types.
PROFILE_TIMESTAMP	timestamp	The timestamp when the profiler last examined this column.
SAMPLE_COUNT	integer	The number of rows sampled on the last profile.
NULL_COUNT	integer	The number of null values found in the last profile’s sample.
CARDINALITY	integer	The database cardinality (including blank values) found in the last profile’s sample.
IS_UNIQUE	boolean	A flag indicating that all values in the last sample were unique.
FLOAT_COUNT	integer	The number of values that converted properly to the float type in the last sample.
INTEGER_COUNT	integer	The number of values that converted properly to the integer type in the last sample.
DECIMAL_COUNT	integer	The number of values that converted properly to the decimal type in the last sample.
BOOLEAN_COUNT	integer	The number of values that converted properly to the boolean type in the last sample.

TIMESTAMP_COUNT	integer	The number of values that converted properly to the timestamp type in the last sample. The conversion uses a user-defined function (UDF) named TRY_MULTI_TIMESTAMP that converts from several common timestamp formats. You can add new timestamp formats to this UDF.
DATE_COUNT	integer	The number of values in the sample that converted properly to the date type and retained the same value as the timestamp. This indicates that they convert to the date type and do not lose time information in the column when converting.
TIME_COUNT	integer	The number of values in the sample that converted properly to the time type and retained the same value as the timestamp. This indicates that they convert to the time type and do not lose date information in the column when converting.
BLANK_VALUES	integer	The number of non-null values that may represent blanks in the sample, such as zero-length strings, zeros, 1970-01-01 00:00:00 for timestamp, etc.
MIN_LENGTH	integer	The minimum length of the values in the sample. Useful mostly for string types.
AVG_LENGTH	float	The average length of the values in the sample. Useful mostly for string types.
MAX_LENGTH	integer	The maximum length of the values in the sample. Useful mostly for string types.
PUNCTUATION_COUNT	integer	The number of values in the sample containing one or more punctuation marks.
NON_ASCII_COUNT	integer	The number of values in the sample containing one or more non-ASCII Unicode characters.
NON_PRINTABLE_COUNT	integer	The number of values in the sample containing one or more non-printable characters such as tabs, ASCII or Unicode nulls, control characters, etc.
MULTI_LINE_COUNT	integer	The number of values in the sample containing one or more multi-line values.
IS_NUMERIC	boolean	A flag indicating that all sampled values successfully convert to a numeric type.
INFORMATION_ENTROPY	float	The Shannon Entropy of the sampled values in the column. The profiler calculates this only when the value_cardinality is between 2 and 16 inclusive.

ENTROPY_RATIO	float	The ratio of the column's information_entropy to the maximum possible entropy for the column's value_cardinality. Calculated only when the profiler calculates the information_entropy for a column. Data modelers can use this to infer whether a column containing between 2 and 16 distinct values has useful information or is mostly defaulting to a single value. This is especially useful for boolean columns. If all values are true or all false, then the entropy_ratio will be close to 0.0. If the column contains an equal distribution of true and false values, then the entropy_ratio will be 1.0. In general, columns with an entropy_ratio close to 1.0 are more likely to contain data interesting for analytics and those close to 0.0 are less likely.

DATA_PROFILE_COLUMN_NAMES view

A view containing the results of the column name profiling. The view contains information from the DATA_PROFILING_COLUMN_NAME_FINDINGS table, which the DATA_PROFILE_COLUMN_NAMES stored procedure fills. The view also calculates columns internally.

Column	Type	Description
DATABASE_NAME	string	The name of the database containing the column.
SCHEMA_NAME	string	The name of the schema containing the column.
TABLE_NAME	string	The name of the table containing the column.
COLUMN_NAME	string	The column profiled for its statistics and contents.
SEEMS_TO_HAVE	string	A plain-language description of what the column seems to have based only on the name of the column.

RULE_NAME	string	The rule in the DATA_PROFILING_COLUMN_NAME_REGEX that matched the column's name. Note that you can add new rules to this table.
CATEGORIES	string	The categories into which the matching rule belongs, such as location.
ORDINAL_POSITION	integer	The ordinal position of the column in the table.
FINDING_TIME	datetime	The time the profiler matched the finding.

DATA_PROFILING_COLUMN_NAMES_REGEX table

A fdfdlf

Column	Type	Description
IS_ENABLED	boolean	A flag indicating whether or not the rule is enabled for processing.
CONFIDENCE_RATIO	float	A subjective assessment of how likely the finding is accurate vs. a false positive.
RULE_NAME	string	The rule name. Since the rules work by matching column names, it is a best practice to prefix the rule name with a two-letter ISO code for the language used for the rule, e.g, en_FIRST_NAME for a rule matching common patterns for a column named in English for a first name..
DESCRIPTION	string	A description of the rule's findings. As a best practice, the description should be lowercase, not contain a period at the end, and fit into the sentence "This column may have ------."
CATEGORIES	string	A comma-separated list of categories in which the rule fits.
REGEX	string	The regular expression that needs to match for the profiler to add a finding.

REGEX_NEGATIVE	string	An optional negative regular expression to suppress false positives generated by the regex. For example, the regex_negative for street addresses filters out false positives for column names like ‘email_address’ or ‘web_address’.