String Functions

Introduction

When dealing with string data, you may find that you need to remove unnecessary characters, standardize the case of text, or identify important information. The Formula tool offers a variety of powerful and dynamic functions to alter string data in Designer.

A dataset contains information on employee names, phone numbers, emails and industries. Before continuing with data analysis, cleanse this data using string functions in the Formula tool to remove unwanted spaces and commas from the column [Name] and remove all formatting from the column [Phone Number].

Remove Whitespace

First, remove the leading and trailing whitespace from the values in the column [Name].

Removing characters, including whitespace, can be accomplished by using a Trim function. Characters at the beginning of a string can be removed with a TrimLeft function, while characters at the end of a string can be removed with a TrimRight function. To remove characters from both ends of a string, use the Trim function.

Click the Function Library. Select the String function category, and select Trim from the list of options to insert this function into the Expression Editor.

Trim functions require two parameters; the string column to trim, and the characters to remove from the string, indicated by the placeholder "y". Trim characters from the column [Name]. When trimming non-whitespace characters, list the characters to remove in quotes. To remove whitespace, delete the "y" placeholder to leverage the function's default behavior: trimming whitespace.

After running the workflow and removing the leading and trailing whitespace, the quality of the column [Name] is 100% OK.

Replacing Characters

The String category of the function library includes functions for replacing instances of a character or sequences of characters. The Replace function will replace a specified target string with a replacement string anytime it is found, while the ReplaceFirst function will replace only the first occurrence of the target string. The ReplaceChar function will replace any listed character with a single character specified as the replacement value.

Use the function ReplaceChar to remove punctuation from the column [Phone Number] so that the column only contains digits.

Drag the punctuation to remove from the column [Phone Number] into the Expression Editor in the placeholder for "y".

Any character listed in the placeholder x will be replaced by the character specified in the placeholder "z". The placeholder "z" can be left empty to remove any listed characters from the string.

After running the workflow, the column [Phone Number] contains only digits.

Case
String functions can be used to standardize string values to upper, lower or title case, which can be important for data blending or reporting purposes.

The column [Email] contains values that include a variety of upper and lower cased letters. Choose a function to standardize the case of these values.

The Uppercase function requires a single parameter: the string column to modify. After running the workflow, the values in the column [Email] are comprised of capital letters.

The Lowercase function converts to the letters in a specified column to lowercase.

When using Title Case, the first character of a string and any character following a non-word character is capitalized. All other letters are lowercase.