Information Builders

iWay

Data Quality Workbench User's Guide

Version 3.2 and Higher

January 11, 2019

Active Technologies, EDA, EDA/SQL, FIDEL, FOCUS, Information Builders, the Information Builders logo, iWay, iWay Software, Parlay, PC/FOCUS, RStat, Table Talk, Web390, WebFOCUS, WebFOCUS Active Technologies, and WebFOCUS Magnify are registered trademarks, and DataMigrator and Hyperstage are trademarks of Information Builders, Inc.

Adobe, the Adobe logo, Acrobat, Adobe Reader, Flash, Adobe Flash Builder, Flex, and PostScript are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States and/or other countries.

Due to the nature of this material, this document refers to numerous hardware and software products by their trademarks. In most, if not all cases, these designations are claimed as trademarks or registered trademarks by their respective companies. It is not this publisher's intent to use any of these names generically. The reader is therefore cautioned to investigate all claimed trademark rights before using any of these names other than to refer to the product described.

Copyright © 2019, by Information Builders, Inc. and iWay Software. All rights reserved. Patent Pending. This manual, or parts thereof, may not be reproduced in any form without the written permission of Information Builders, Inc.

Contents

Preface 5

Documentation Conventions 5
Related Publications 6
Customer Support 6
Help Us to Serve You Better 7
User Feedback 8
iWay Software Training and Professional Services 8

1. Data Quality Workbench Introduction 11

General Structure 11
Profile tab 13
Data Grid tab 15

2. Data Quality Workbench Components Configuration Functionality Overview 17

Filtering: General Approach and Global Filters 17
Filtering by Patterns 21
Quick Filter 22
On-Screen Filter 23
Profiling 23
Profile Subject Page 23
Using the Add to Filter Option From the Profile Subject Page 24
Using the Show Data Option From the Profile Subject Page 27
Using the Add to Filter Option From the Profile Attribute Page 29
Using the Show Data Filter Option From the Profile Attribute Page 34
Data Tables 40
Table 43
Table Sorting 49
Showing and Hiding Columns 50
Compare to Original 51
Filter Panel 52
Using Filters 62
Using the Add to Filter Option 63
Rules and Operations 64

External rules. 68
Standard Rules. 69
Working With Operations. 77
Adding Operations from the Rules panel. 81
Adding Quick Operations From the Data Table. 85
Adding Complex Operations. 86
Processing and Executing Rules. 89
Reordering Rules and Operations 100
Dragging and Dropping Rules. 101
Dragging and Dropping Operations. 110
Content Types 112
Content Types in Omni Designer. 113
Content Types in the Workbench. 115
Content Types and Operations. 125
Detecting Errors (Highlight Defects) 126
Using the Highlight Defects Feature. 127
Defining Rules to Determine Whether a Value Should be Highlighted. 130
Profiling (Subject). 131
Profiling (Attribute). 135
Highlight Defects in On-Screen Filters. 145
Using Data Quality Workbench 145
Menu Items. 146
Reset Session. 147
Uncommitted Icon. 149
Saving State. 150
Loader. 150

Preface

This document describes how to configure and use Data Quality Workbench, a component of Omni-Gen that allows you to analyze data and create rules for data standardization.

How This Manual Is Organized

This manual includes the following chapters:

	Chapter/Appendix	Contents
1	Data Quality Workbench Introduction	Provides an overview and page structure information for Data Quality Workbench.
2	Data Quality Workbench Components Configuration Functionality Overview	Provides an overview of the various functions that are used in Data Quality Workbench.

Documentation Conventions

The following table lists and describes the documentation conventions that are used in this manual.

Convention	Description

`THIS TYPEFACE`
or
`this typeface`
	Denotes syntax that you must type exactly as shown.

this typeface	Represents a placeholder (or variable), a cross-reference, or an important term. It may also indicate a button, menu item, or dialog box option that you can click or select.
underscore	Indicates a default setting.
Key + Key	Indicates keys that you must press simultaneously.
{}	Indicates two or three choices. Type one of them, not the braces.
|	Separates mutually exclusive choices in syntax. Type one of them, not the symbol.

Convention	Description
...	Indicates that you can enter a parameter multiple times. Type only the parameter, not the ellipsis (...).
. . .	Indicates that there are (or could be) intervening or additional commands.

Related Publications

Visit our Technical Documentation Library at *http://documentation.informationbuilders.com*. You can also contact the Publications Order Department at (800) 969-4636.

Customer Support

Do you have questions about this product?

Join the Focal Point community. Focal Point is our online developer center and more than a message board. It is an interactive network of more than 3,000 developers from almost every profession and industry, collaborating on solutions and sharing every tips and techniques. Access Focal Point at *http://forums.informationbuilders.com/eve/forums*.

You can also access support services electronically, 24 hours a day, with InfoResponse Online. InfoResponse Online is accessible through our website, *http://www.informationbuilders.com*. It connects you to the tracking system and known-problem database at the Information Builders support center. Registered users can open, update, and view the status of cases in the tracking system and read descriptions of reported software issues. New users can register immediately for this service. The technical support section of *www.informationbuilders.com* also provides usage techniques, diagnostic tips, and answers to frequently asked questions.

Call Information Builders Customer Support Services (CSS) at (800) 736-6130 or (212) 736-6130. Customer Support Consultants are available Monday through Friday between 8:00 A.M. and 8:00 P.M. EST to address all your questions. Information Builders consultants can also give you general guidance regarding product capabilities. Be prepared to provide your six-digit site code (*xxxx.xx*) when you call.

To learn about the full range of available support services, ask your Information Builders representative about InfoResponse Online, or call (800) 969-INFO.

Help Us to Serve You Better

To help our consultants answer your questions effectively, be prepared to provide specifications and sample files and to answer questions about errors and problems.

The following table lists the environment information that our consultants require.

Platform	
Operating System	
OS Version	
JVM Vendor	
JVM Version	

The following table lists additional questions to help us serve you better.

Request/Question	Error/Problem Details or Information
Did the problem arise through a service or event?	
Provide usage scenarios or summarize the application that produces the problem.	
When did the problem start?	
Can you reproduce this problem consistently?	
Describe the problem.	
Describe the steps to reproduce the problem.	
Specify the error messages.	

Request/Question	Error/Problem Details or Information
Any change in the application environment: software configuration, EIS/database configuration, application, and so forth?	
Under what circumstance does the problem *not* occur?	

The following is a list of error and problem files that might be applicable.

- ❑ Input documents (XML instance, XML schema, non-XML documents)
- ❑ Transformation files
- ❑ Error screen shots
- ❑ Error output files
- ❑ Trace files
- ❑ Custom functions and agents in use
- ❑ Diagnostic Zip
- ❑ Transaction log

User Feedback

In an effort to produce effective documentation, the Technical Content Management staff welcomes your opinions regarding this document. Please use the Reader Comments form at the end of this document to communicate your feedback to us or to suggest changes that will support improvements to our documentation. You can also contact us through our website, *http://documentation.informationbuilders.com/connections.asp*.

Thank you, in advance, for your comments.

iWay Software Training and Professional Services

Interested in training? Our Education Department offers a wide variety of training courses for iWay Software and other Information Builders products.

For information on course descriptions, locations, and dates, or to register for classes, visit our website, *http://education.informationbuilders.com*, or call (800) 969-INFO to speak to an Education Representative.

Interested in technical assistance for your implementation? Our Professional Services department provides expert design, systems architecture, implementation, and project management services for all your business integration projects. For information, visit our website, *http://www.informationbuilders.com/consulting*.

Chapter 1

Data Quality Workbench Introduction

Data Quality Workbench is a component of Omni-Gen that allows you to analyze data and create rules for data standardization and enrichment without any previous knowledge of Data Quality (DQ) products. You can access the Workbench through the Omni Governance Console (OGC).

The Workbench provides an easy-to-use web-based interface that allows you to perform data mastering tasks, such as:

- Profiling data.
- Creating cleansing and validation rules.
- Reviewing rules results.
- Executing rules.
- Viewing summary and statistics for the original data set and cleanse/match results.

In this chapter:

- General Structure
- Profile tab
- Data Grid tab

General Structure

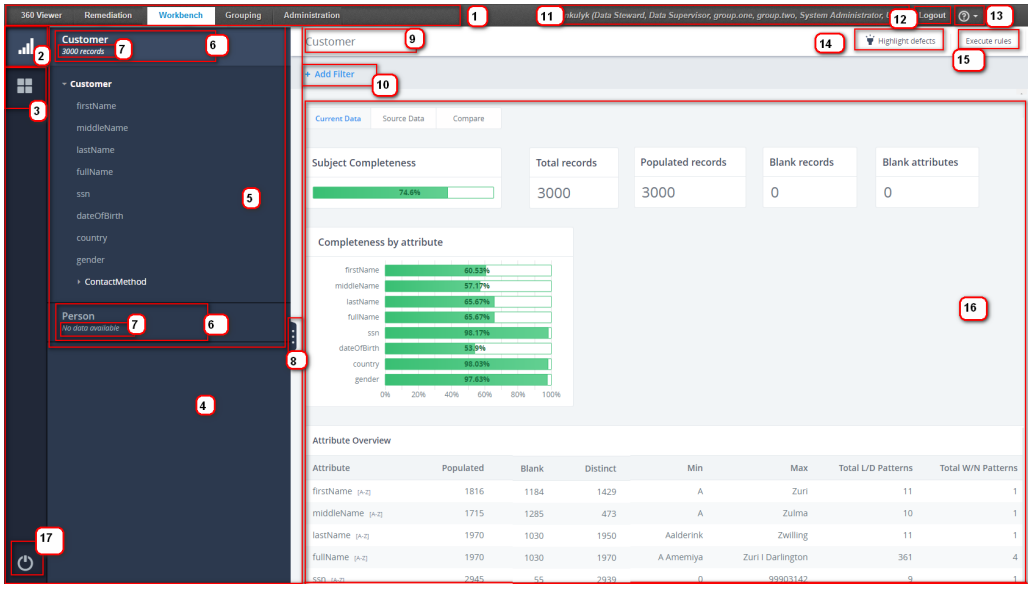
The data structure displayed in the Workbench is based on the data structure that was created using Omni Designer. The number of records displayed matches the number of records that were generated using Omni Console or DQS.

The Input Data Model is a specific data structure created for the database (for example, Employees unit or Customers). Subjects represent data tables that will be populated with corresponding data (for example, Addresses, Customers). Attributes represent data columns in the future data model, which are characteristics or important values for a predefined subject.

The Workbench is a single-page application and has the following structure:

1. Top menu
2. Profile tab

3. Data grid tab
4. Navigation panel
5. Input Data Model tree
6. Input Data Model
7. Number of data records
8. Expand/collapse button
9. Navigation breadcrumbs
10. Global filter
11. User information
12. Logout button
13. About button
14. Highlight defects button
15. Execute rules button
16. Information about the selected data tree item
17. Reset session button

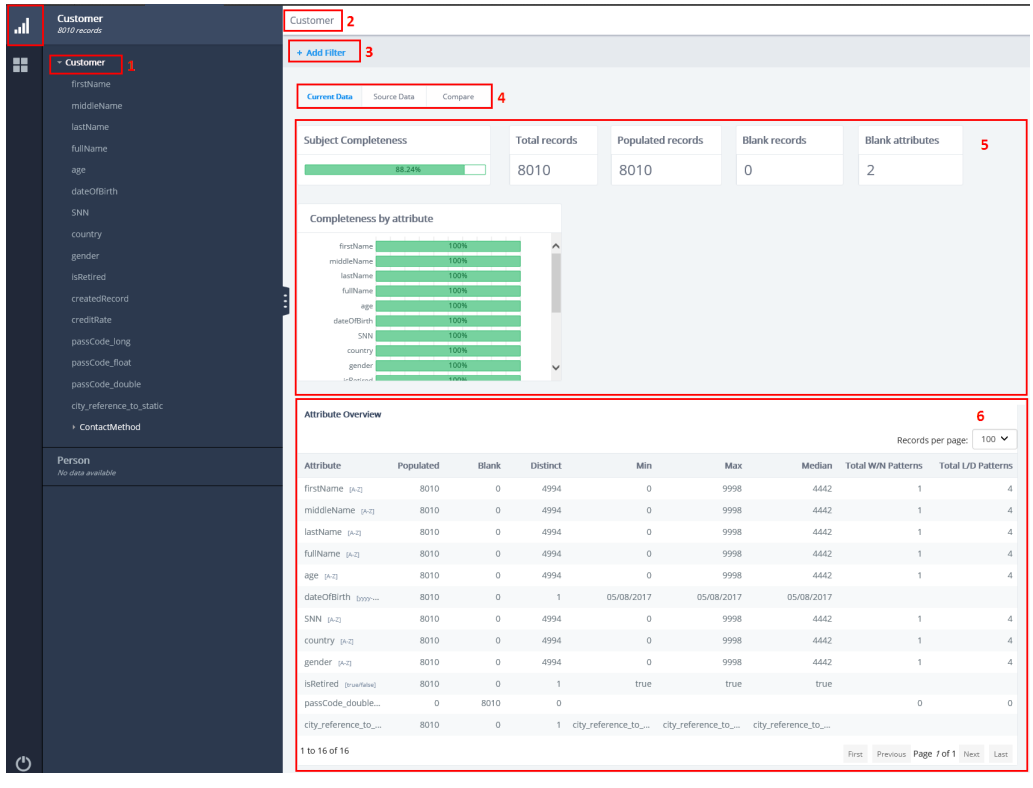


The tabs that appear are dependent on the user role and permissions.

Profile tab

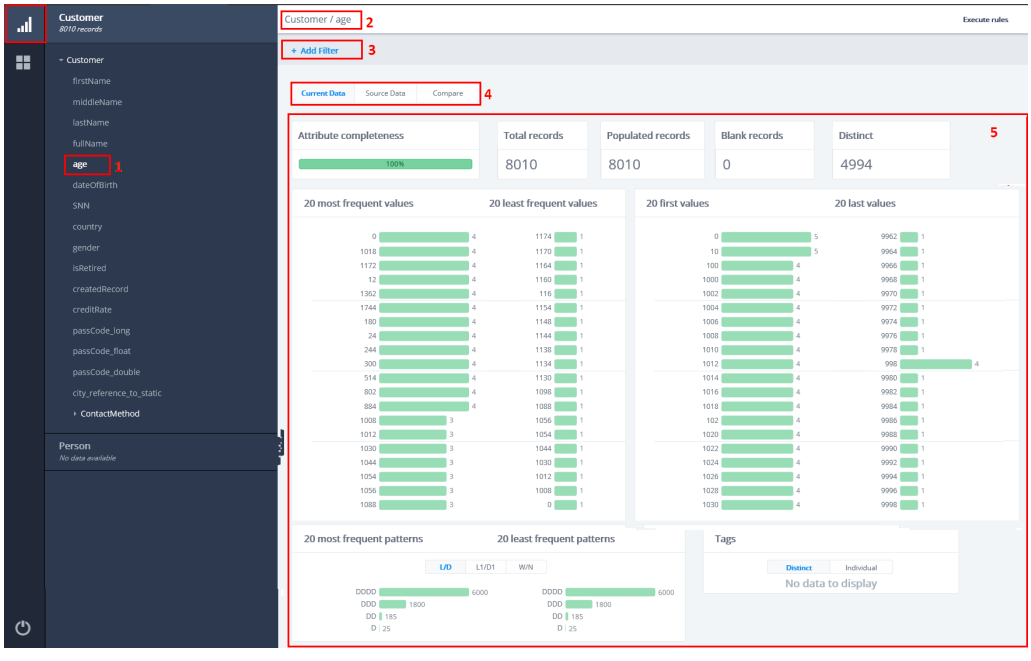
The Profile tab is displayed by default once a user has logged in. This tab has the following structure (it is similar for Subject and Attribute levels):

1. Subject/Attribute selected item
2. Navigation breadcrumbs
3. Add filter
4. Current/Source/Compare view mode
5. Subject/attribute charts/tiles
6. Attribute Overview grid for the selected subject



On the Subject level, a user can view information regarding Subject Completeness, Total records, Populated records, Blank records, and Distinct records not considering attributes influence. Data that describes general information regarding each attribute is displayed in the *Completeness by attribute* bar chart and *Attribute Overview* grid.

To view detailed information regarding each attribute, the user should navigate to the attribute in the navigation tree. Attribute Completeness, Total records, Populated records, Blank records, and Distinct records will show general information for the selected attribute and for the 20 most/least frequent values, 20 first/last values, 20 most/least frequent patterns, and Tags can show deep analysis based on each value.



When a user navigates to the Workbench, Subject, the Profile tab is displayed for a specific data model. the first data model is expanded and the focus is on the root subject line. All subjects are collapsed except for the root subject.

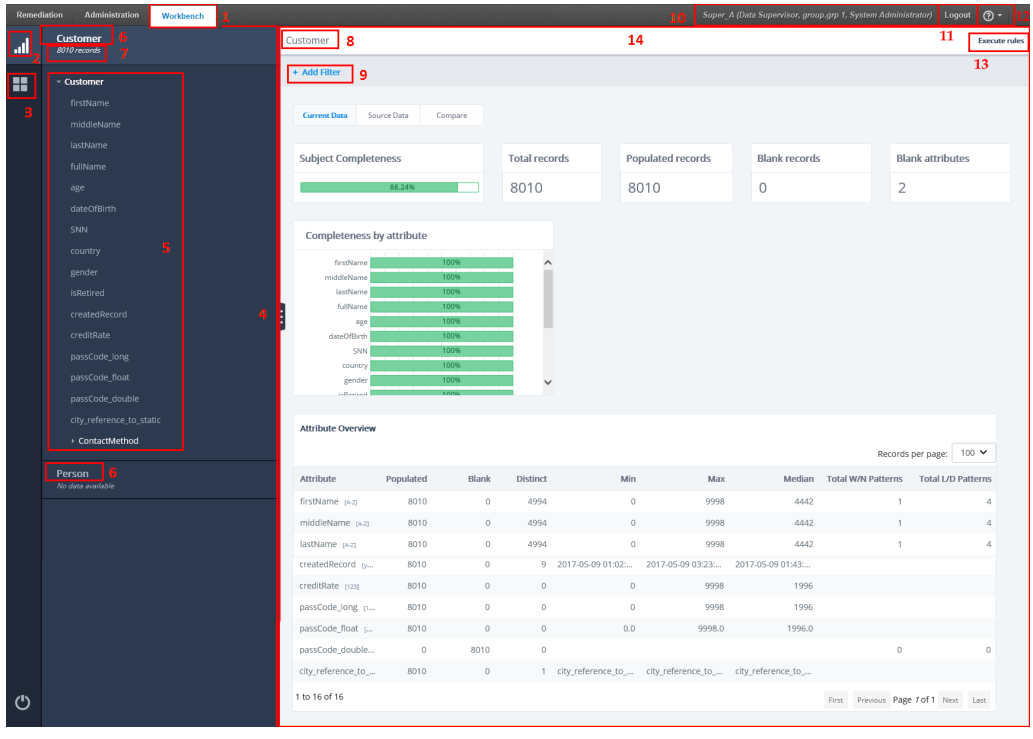
In the event that no data is present in the Workbench:

- A *No data available* message is displayed instead of the number of records for the input data model.
- All bar charts will be empty and a value of 0% is displayed.

If there is no data that satisfies some filter criteria, a *No data available* message is displayed for the grid/On-Screen filter.

Data Grid tab

The Data Grid tab provides a profiling view into the data. It includes patterns, record completeness, and other profiling features.



Chapter 2

Data Quality Workbench Components Configuration Functionality Overview

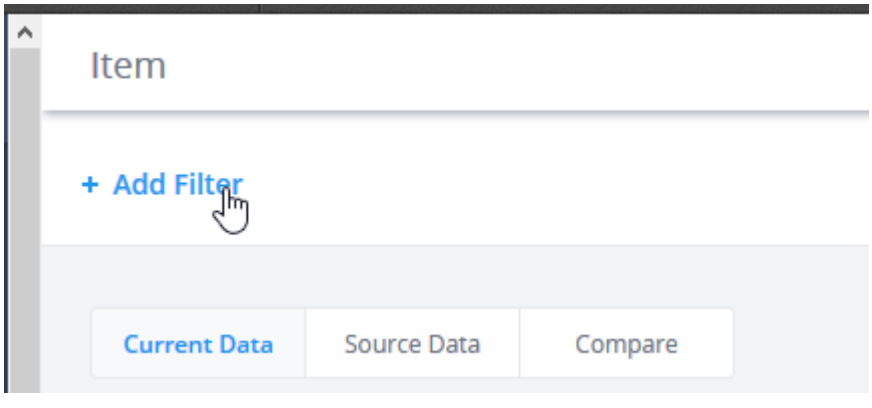
This section provides an overview of the various functions that are used in Data Quality Workbench.

In this chapter:

- Filtering: General Approach and Global Filters
- Profiling
- Profile Subject Page
- Data Tables
- Detecting Errors (Highlight Defects)
- Using Data Quality Workbench

Filtering: General Approach and Global Filters

You can add a global filter by clicking *Add Filter* on the upper pane of the screen and using the context menus.

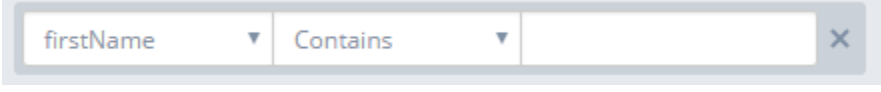


A filter can be added from the Profile and Data Table pages (from the table value and table header). If you added a filter from the table header, then the focus will be on the filter value, where you can enter the data immediately.

Clicking the *Add Filter* button on the screen creates a filter form where you can populate the Property and Value. User can also change the Operator, which is pre-populated with the first value in the drop-down list.

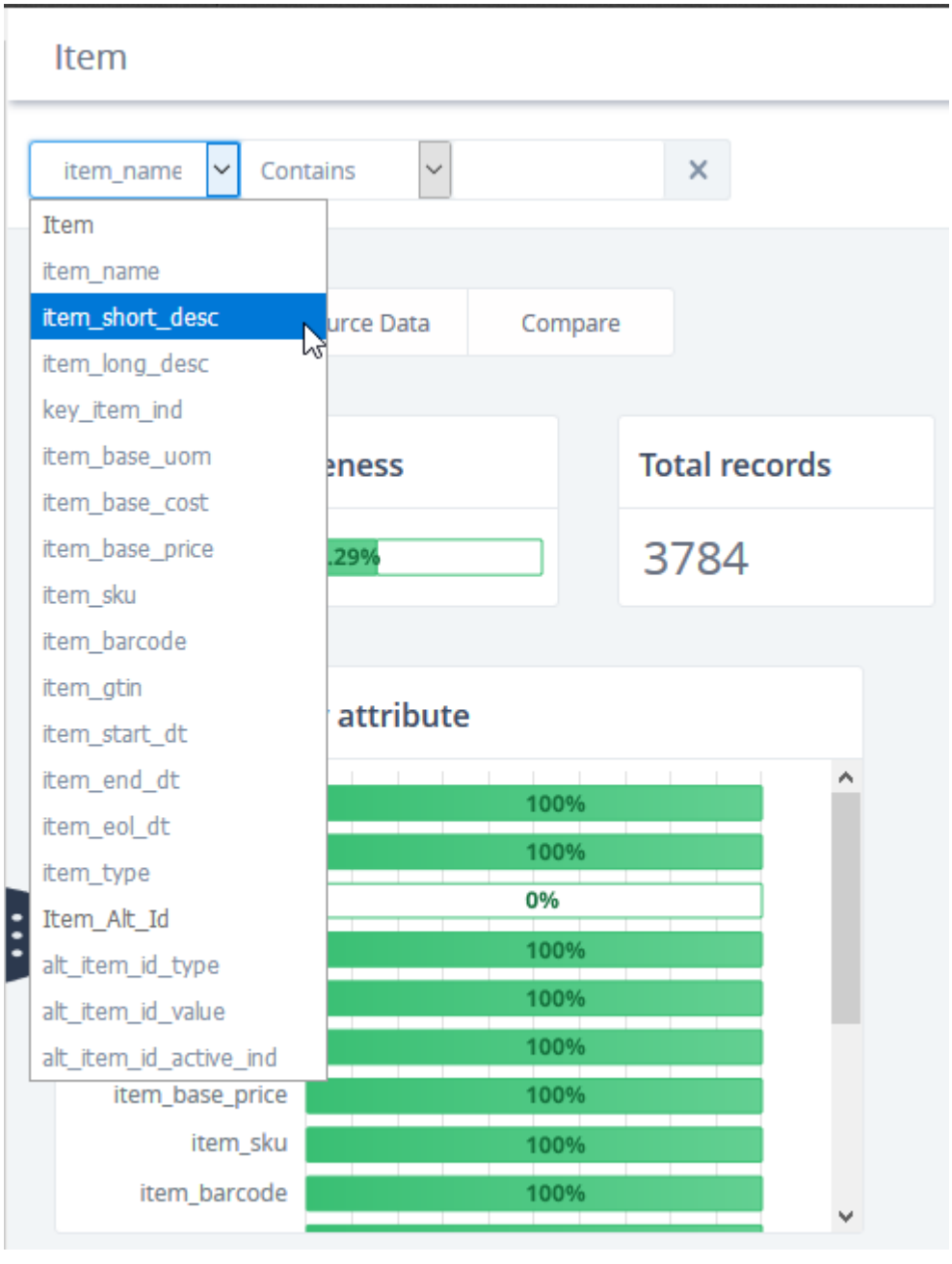
By default, the filter consists of the following fields:

- Property
- Operand
- Value



In the Properties drop-down list, the whole list of Model subjects and their attributes are available.

The first attribute and operator are set as default. For some properties, the filter will consist of two fields: Is empty, and Is not empty. However, some values will be prepopulated (for example, for the Boolean attribute type, the value will be set to *True*). Attributes and subjects are displayed in the drop-down list according to their location, but currently, only attributes can be selected.

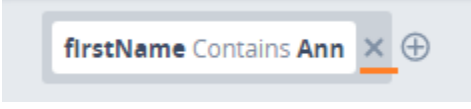

Item
item_name
Contains
Item
item_name
item_short_desc
item_long_desc
key_item_ind
item_base_uom
item_base_cost
item_base_price
item_sku
item_barcode
item_gtin
item_start_dt
item_end_dt
item_eol_dt
item_type
Item_Alt_Id
alt_item_id_type
alt_item_id_value
alt_item_id_active_ind
urce Data
Compare
eness
.29%
Total records
3784
attribute
100%
100%
0%
100%
100%
100%
100%
100%
100%
item_base_price
item_sku
item_barcode

Operators are different depending on the type of attribute selected:

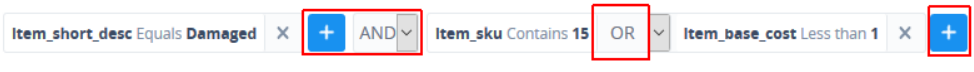
- *String type:* Contains, Not contains, Equals, Not equals, Is empty, Is not empty, Like, Not like, Tagged with, Not tagged with.
- *Integer, Date and DateTime types:* Equals, Not equals, Is empty, Is not empty, Greater than, Greater than or equals, Less than, Less than or equals.
- *Boolean type:* Equals, Not equals, Is empty, Is not empty.

Once you have added a filter with a criteria, the Profile page, data table, and onscreen filter will be recalculated and data records will be displayed according to the added filter.

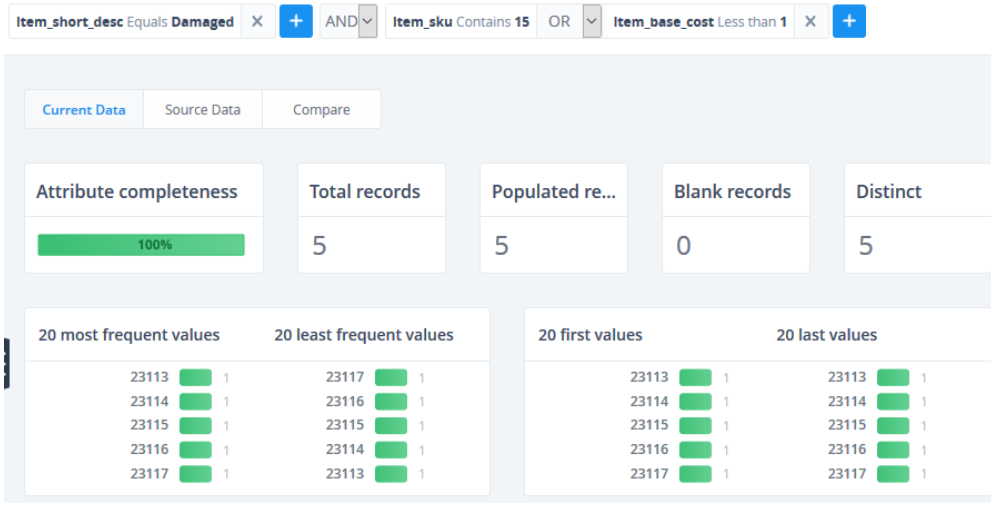
If you click on the X icon on the right side of the Filter pane, added filter is removed.



You can add multiple filter tags by combining them with the *AND/OR* operands. The first two added filter tags can use the *AND* operand. Filters separated with the *OR* operand are combined on the user interface and can not add new filters between them. Multiple filters can be added from the table or by the add (+) button in the global filter area. New filters can be inserted after each filter tag:



Values will be filtered according to the added filters and operands. If you combine filters for different attributes with the *AND* operand, then records that match all filter criteria will be displayed.



Filters separated with the *OR* operand are combined on the user interface where you can not add new filters between them. If you combine filters with the *OR* operand, then records that match at least one of the filtered criteria will be displayed.

Filters are saved only per one browser session, until you refresh the page or log out.

Filtering by Patterns

The Pattern (a regular expression – RegEx) is a special text string for describing a filter pattern. You can think of regular expressions as wildcards. Patterns are used when you need to filter by some general criteria: letters and digits, four letters and two digits, word and number. The general idea is to simplify filtering.

- If you are filtering data using the On-Screen filter/Patterns tab, you should select the bar chart that corresponds to what you need.

  For example, if you want to filter all names consisting of five letters like "Megan", you should select *LLLLL* as the filter criteria, and for filtering SSN "457-55-5462", you should select *DDD-DD-DDDD*.

- If you want to filter data using a global filter, you need to transform the filter criteria to the RegEx format since *Filtering by patterns* works only when the RegEx format is entered:

The *Like/Not like* operators are used for filtering by patterns, which are available only for string values.

Patterns will be transformed to the *RegEx* format in the following ways:

There are several types of data patterns: L/D, L1/D1, and W/N patterns.

L/D Patterns (Letters/Digits):

1. Change all L instances to [a-z]
2. Change all D instances to [0-9]

For example, if you want to filter some specific value Iris45 you should enter *[a-zA-Z][a-z][a-z][a-z][0-9][0-9]* as filter criteria.

L[1]/D[1] Patterns

1. Change all L[x] to [a-z]{x}
2. Change all D[x] to [0-9]{x}

{x} means the quantity of Letters or Digits. {x} should not be written if there is a single digit or letter.

For example, to filter the same value Iris45, you can also transform *[a-z]{4}[0-9]{2}* as filter criteria.

For example, you can transform Iris4 as *[a-zA-Z]{4}[0-9]*.

W/N Patterns

1. Change all W to [a-z]+
2. Change all N to [0-9]+ Patterns
3. Change all L to [a-z]
4. Change all D to [0-9]

For example, to transform Iris45, you should enter *[a-zA-Z]+[0-9]+*.

Transforming to RegEx should be used according to the case level (upper/lower).

If you want to filter records using wildcards or by using special characters (. ? + * | { } [ ] ( ) " \) – put the backslash (\) before the required special character.

For example, to filter patterns such as *ID123*a?*, you should enter *[A-Z]{2}[0-9]{3}*[a-z]\?*.

Quick Filter

You can add any item from the data table to the global filter by right-clicking and selecting *Add to filter*. The global filter is enhanced with a new Filter tag where the property = <column in which the cell user clicked on lies>, operand = "Equals", and value = <actual value of the cell user clicked on>.

On-Screen Filter

The On-Screen filters panel is a panel with filters combined in groups. The On-Screen filters panel is located on Data Tables page and is displayed only for the selected table column.

Profiling

The Profile Subject and Attribute pages can be displayed in three modes:

- Current
- Source
- Compare

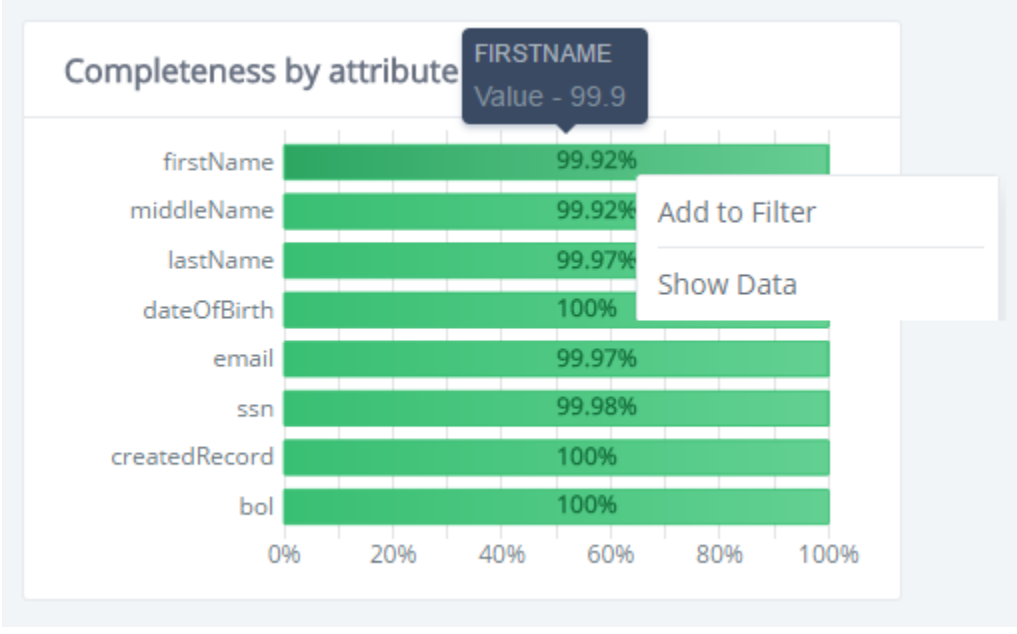
In general, all of the modes display the same page elements, but the Current mode displays records that you are currently working with. The Source mode displays source records and Compare mode displays Current and Source records to be compared.

Profile Subject Page

The Subject Completeness bar chart displays the percentage of Subject data completeness.

Total records, Populated records, Blank records, and Blank attributes tiles display the number of records corresponding to the name of the tile (Blank/Populated, and so on).

The Completeness by attribute bar chart displays the percentage of completeness per attribute. It shows you a visual representation of the subject you are working with and helps identify where the possible data problem might be.



The Attribute Overview table displays attribute data characteristics from different categories:

Populated, Blank, Distinct, Min, Max, Total W/N Patterns and Total L/D Patters.

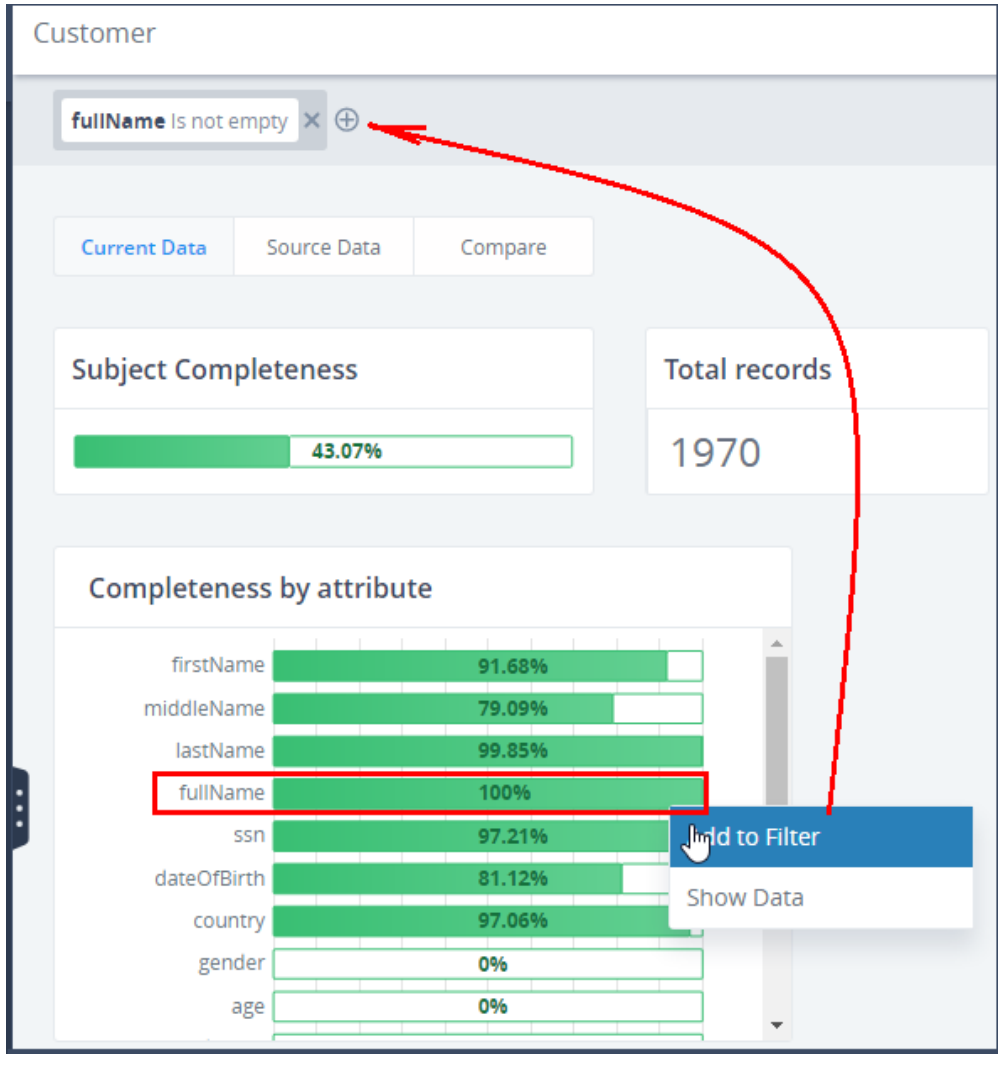
Attribute Overview

Attribute	Populated	Blank	Distinct	Min	Max	Median	Total W/N Patterns	Total L/D Patterns
firstName [A-Z]	5996	5	6046	0	999	18097	1	5
middleName [A-Z]	5996	5	6046	0	999	18097	1	5
lastName [A-Z]	5999	2	6048	0	9997	18097	1	5
dateOfBirth [MM/dd...	6001	0	2	05/22/2017	05/24/2017	05/24/2017		
email [A-Z]	5999	2	6048	10000	9997	18100	1	5
ssn [A-Z]	6000	1	6049	0	9997	18097	1	5
createdRecord [yyy...	6001	0	917	2017-05-23 07:05:4...	2017-05-25 06:33:5...	2017-05-25 06:07:0...		
bol [true/false]	6001	0	1	true	true	true		

Using the Add to Filter Option From the Profile Subject Page

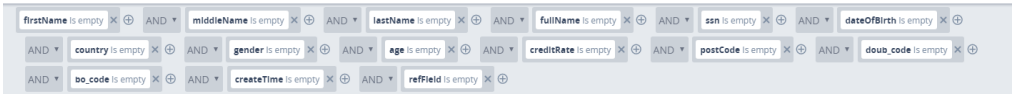
The Add to Filter functionality allows you to add a filter tag from the bar chart, tile, or grid. The Add to Filter option is present in the Context menus which are available by right-clicking on most of the following components: tile, bar chart, and table cell. There is no context menu available for the Blank Attributes tile.

After selecting the *Add to Filter* option, a global filter will be added containing the appropriate criteria.



The following lists describes the specifics for adding a filter from the different context menus:

- When adding a filter for the Populated records tile or Subject completeness bar chart, the global filter will contain every attribute in that subject, and the operator indicated that it is not empty, as shown in the image below. Filter tags will be separated with the *OR* operand.
- You can right-click on the Blank records tile and select the *Add to Filter* option. A global filter will be added containing a series of Filter tags for every column of the subject and each tag will have property = <column name> and operand *is empty*. Filter tags will be separated with the AND operand.

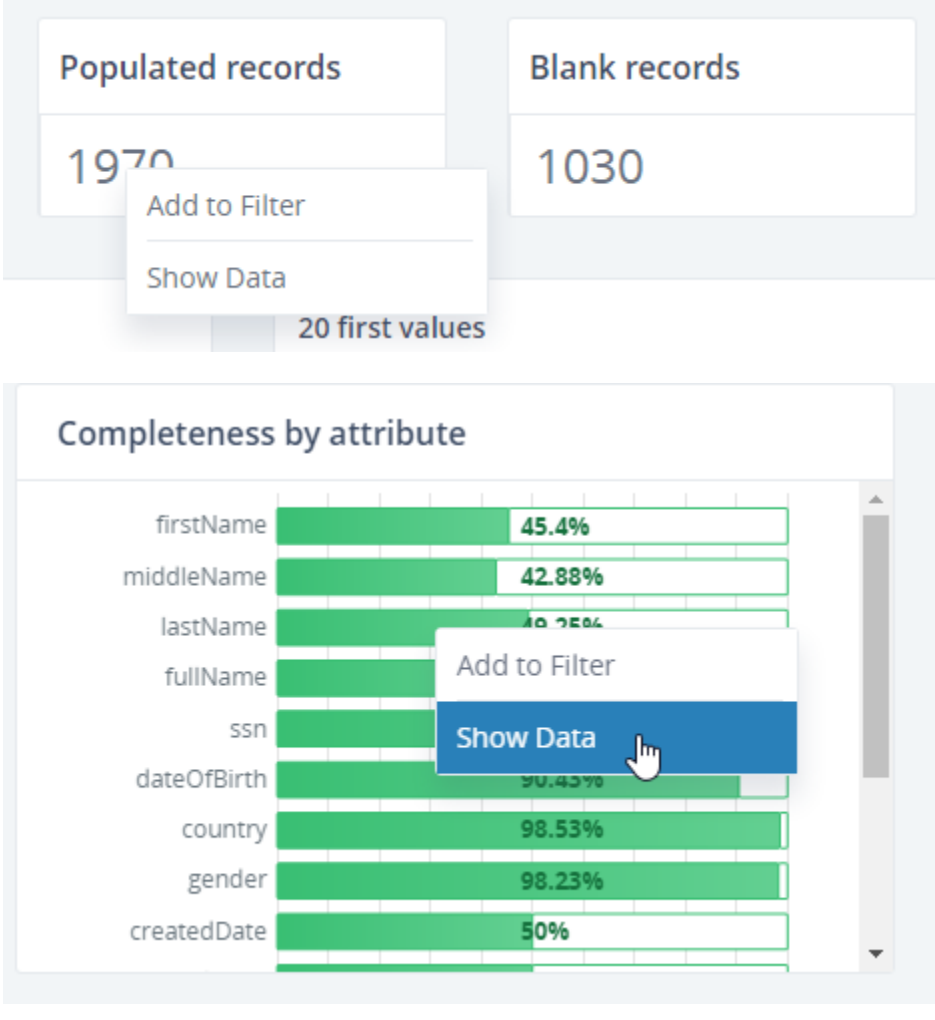


- There is no *Add to Filter* option for Total records because adding that filter makes no sense since it will show all the records in the system, and they are already displayed on the page.
- You can right-click on the Completeness by Attribute bar charts and select *Add to Filter*. As a result, the Filter tag will be added to the global filter having property = <column name> and operand *is not empty* (or *is empty* if subject is empty).
- You can right-click the Attribute name for a specific column in the Attributes Overview table and select the *Add to Filter* option from the context menu. The filter tag will be added with property = <column name you clicked on>, operand *Equals*, and value = <empty>. You will need to enter values for them.
- You can right-click the Populated value for a specific column in the Attributes overview table and select the *Add to Filter* option from the context menu. The filter tag will be added with property = <column name of the bar you clicked on> and operand *is not empty*.
- You can right-click the Blank value for a specific column in the Attributes overview table and select the *Add to Quick Filter* option from the context menu. The filter tag will be added with property = <column name of the bar you clicked on> and operand *Is empty*.
- You can right-click the Min value for a specific column in the Attributes overview table and select the *Add to Quick Filter* option from the context menu. The filter tag will be added with property = <column name of the bar you clicked on>, operand *Equals*, and value = <the min value you clicked on>. The Attribute Overview table will be filtered according to the filter criteria.
- You can right-click the Max value for a specific column in the Attributes overview table and select the *Add to Quick Filter* option from the context menu. The filter tag will be added with property = <column name of the bar you clicked on>, operand *Equals*, and value = <the max value you clicked on>. The Attribute Overview table will be filtered according to the filter criteria.
- There is no *Add to Filter* option for the Distinct, Total W/N Patterns, Total L/D Patterns columns, and the Attribute Overview table header.

Using the Show Data Option From the Profile Subject Page

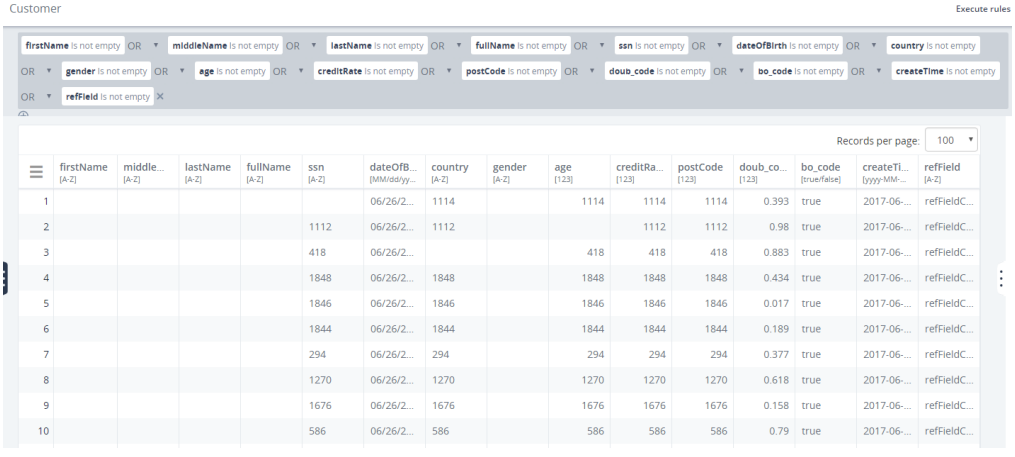
You can navigate to the Data table and see data according to the selected item in the on-screen filter or global filter data by selecting *Show data* from the context menu. The presence of an On-Screen filter and global filter on the Data table page depends on the selected bar chart on the Profile page.

You can right-click on any tile, table cell, or bar chart and select the *Show Data* option, as shown in the following image.



The following list shows how the Show Data option applies towards each different component.

- For the Populated records tile, the same filter will be set for the *Add to filter* option, and the Data table page will open with Populated records. None of the table columns will be selected.

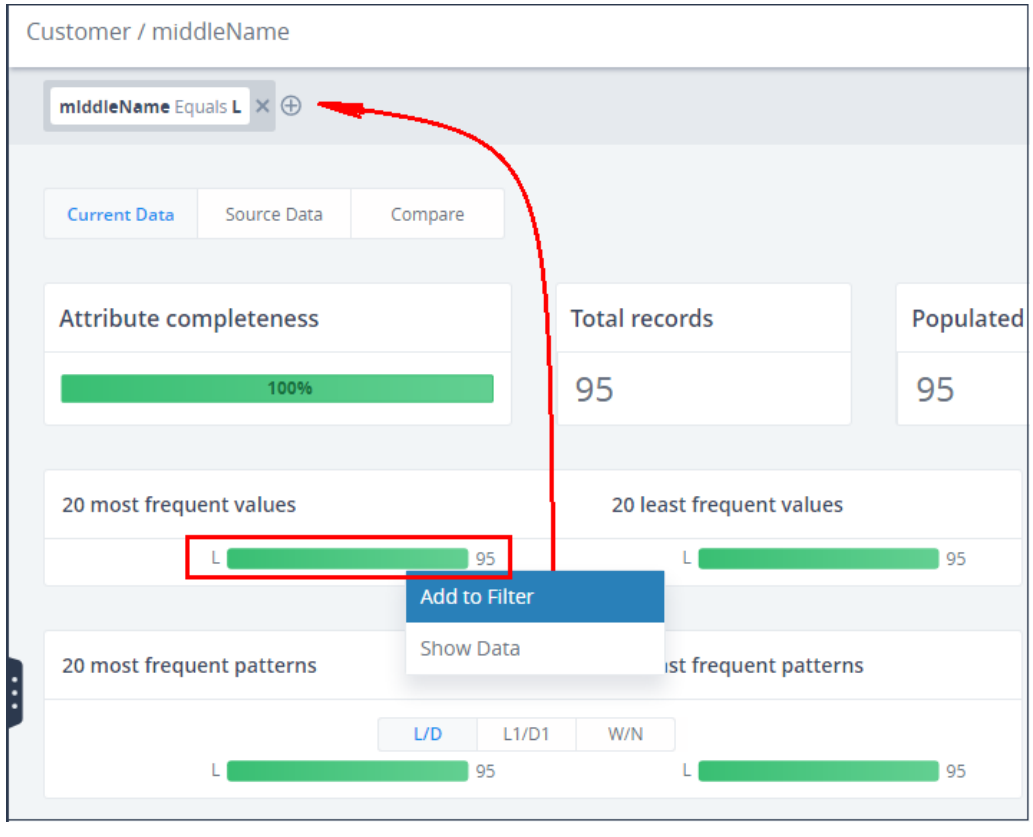


- The behavior will be the same for Blank records, but the Data table page will open with Blank records.
- For Total tile, the Data table page will open with all records shown.
- There is no *Show Data* option for the Blank attributes tile.
- For the Subject Completeness bar charts, results will be the same as the Populated records tile. None of the table columns will be selected.
- For the Completeness by attribute bar chart:
  - If you right-click the bar representing the Populated values, the Data table page for the selected column will open. The Column Profile pane has *Populated* selected.
  - If you right-click the bar part representing the Empty values, then the Data table page for the selected column will open. The Column Profile pane has *Empty* selected.
- You can right-click and select *Show Data* from every column at the Attribute Overview table. For example:
  - **Populated/Blank.** The Data table page opens with Populated/Blank records displayed for the selected column and with Populated/Blank tab in On-Screen filter selected.

- **Min column.** The Data table page opens with the selected column. The On-Screen Column Profile is switched to the First/Last values tab and the first value in the First tab is selected.
- **Max column.** The Data table page opens with the selected column. The On-Screen Column Profile is switched to the First/Last values tab and the last value in the Last tab is selected.
- When you right-click anywhere else in the Attributes overview table row and select *Show Data* from the context menu, the Data table page opens but none of the grid columns will be selected.

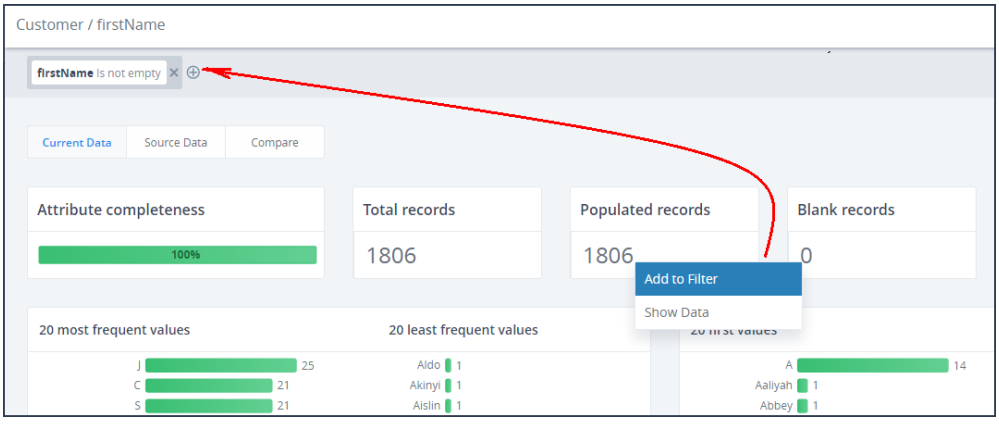
Using the Add to Filter Option From the Profile Attribute Page

The *Add to Filter* option can be selected in the Profile/Attribute page from the context menus by right-clicking the tile and bar chart components. After clicking *Add to Filter*, the filter tag will be added to the global filter, as shown in the following image.

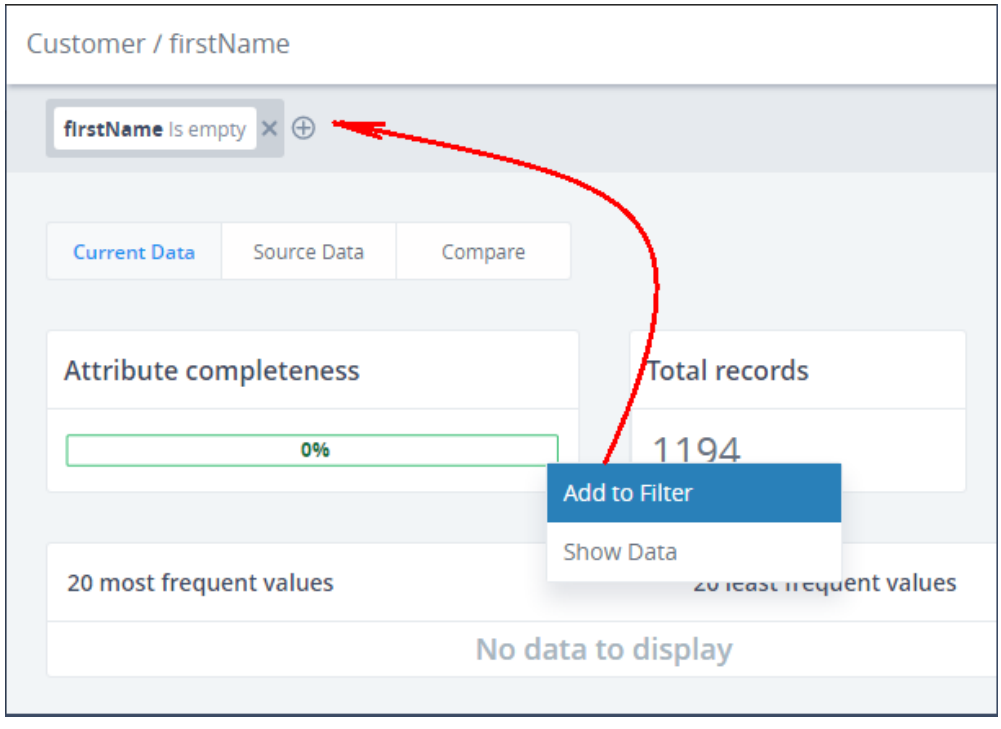


The following list describes the results of adding filters from different context menus.

- When adding a filter for the Populated records tile or Completeness by Attribute(green bar), the behavior will be the same: the global filter will be added for the selected attribute in the tree and contains the *is not empty* operator, as shown in the following image.



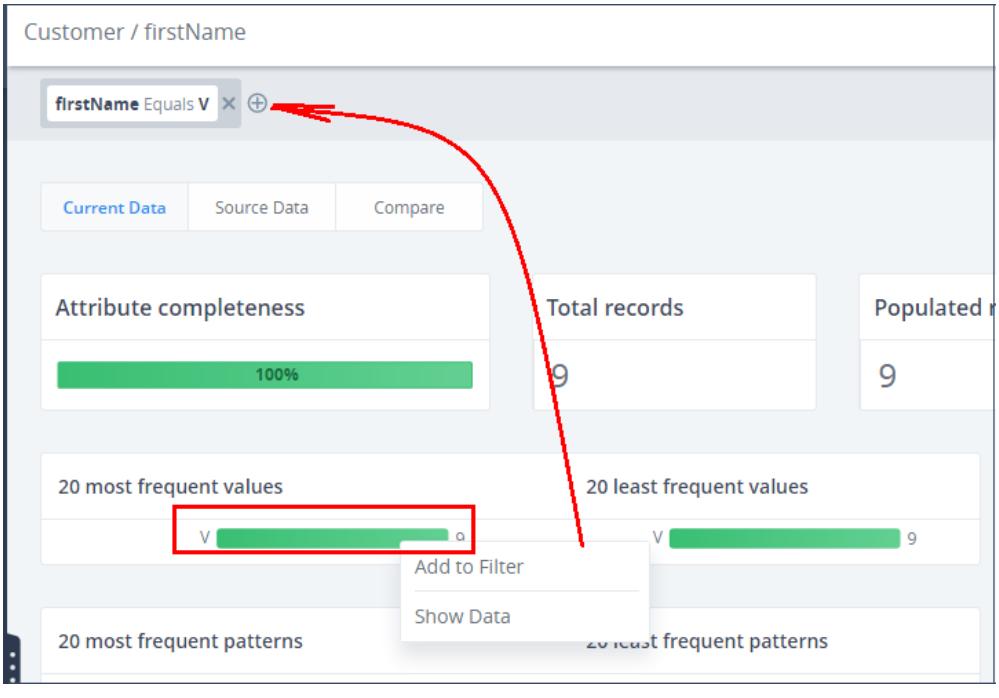
- When adding a filter for the Blank records tile or Completeness by Attribute (the un-filled white bar), the global filter will be added for the selected attribute in the tree and contains operator *is empty*, as shown in the following image.



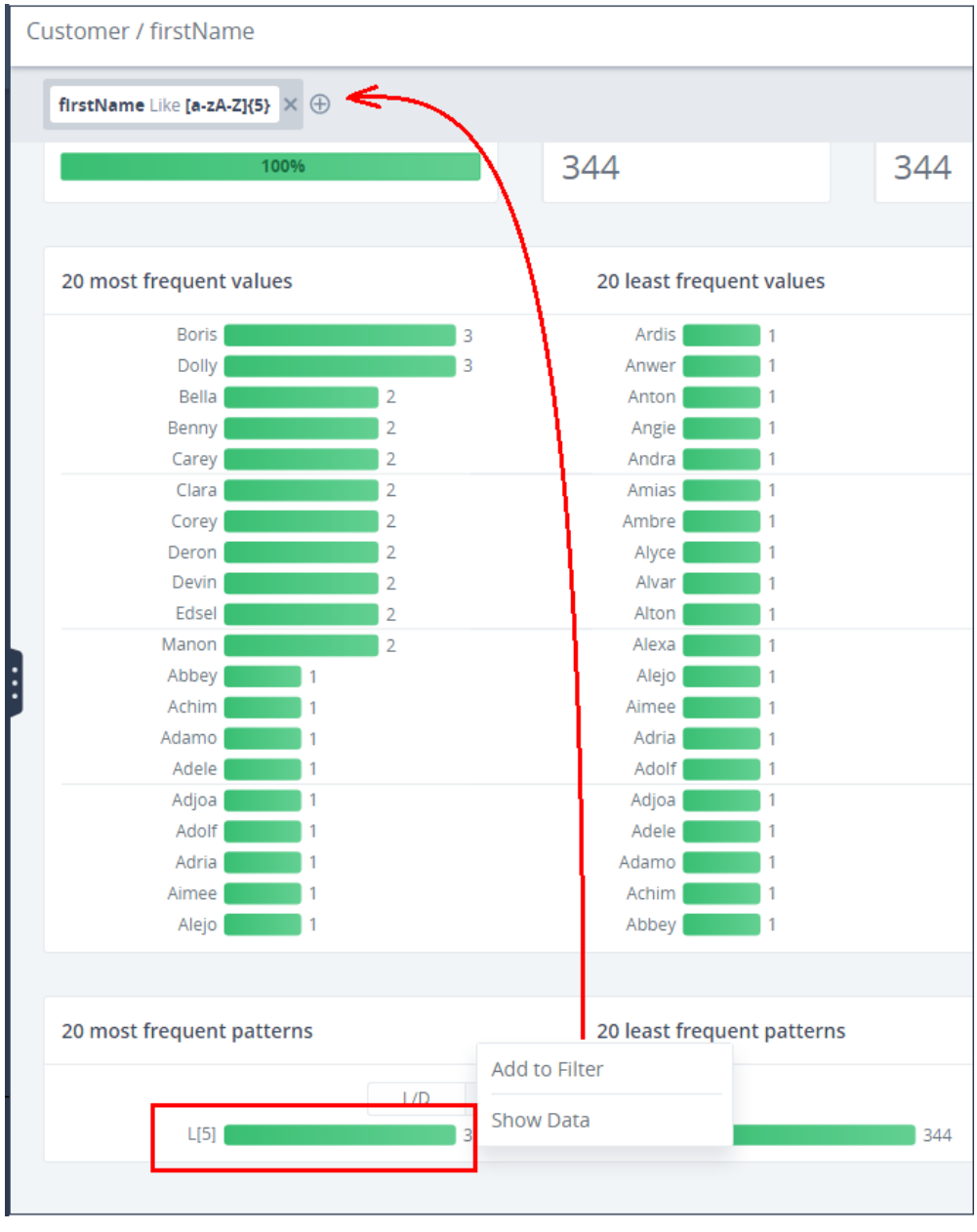
- There is no *Add to filter* option for Total records and Distinct records tiles.

  You can also open the context menu by right-clicking either the bars, name of the bars, or the value in the 20 most/least frequent values, 10 first/last values, 20 most/least frequent patterns, or tags.

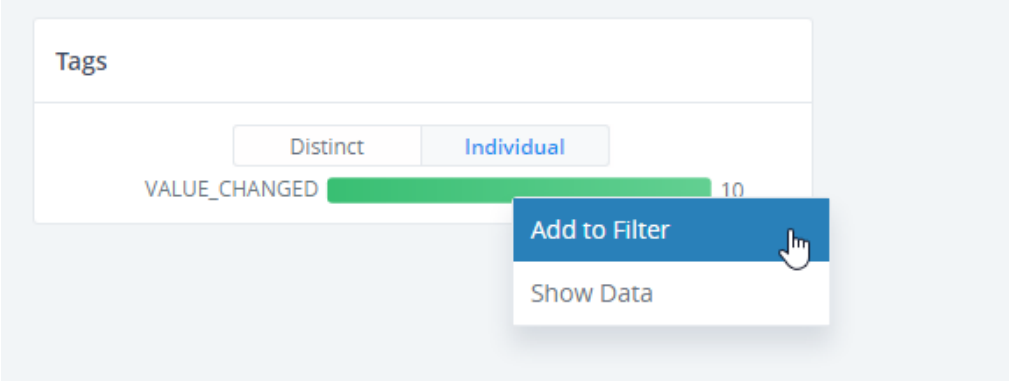
- You can right-click on the 20 most frequent values, 20 least frequent values, 10 first values, and 10 last values charts, and select *Add to Filter*. The filter tag will be added with property = <column name of the bar chart you clicked on>, operand *Equals* and value = <the value you clicked on>. The attribute page will be filtered according to the filter criteria, as shown in the following image.



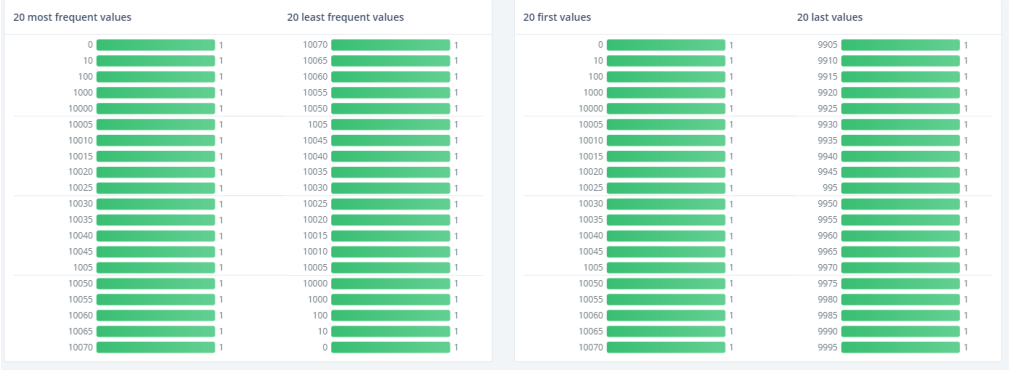
- When you right-click the (L/D, L1/D1, W/N) bar charts pattern and select *Add to Filter*, the filter tag will be added with property = <column name of the bar chart you clicked on>, operand *Like*, and value = <the value you clicked on transformed to RegEx format>. The Attribute page will be filtered according to the filter criteria, as shown in the following image.



- You can select *Add to Filter* for the Tags section, and select either the Distinct or Individual tab. The filter tag will be added with property = <column name of the bar chart you clicked on>, operand *Tagged with*, and value = <tag label>. The Attribute page will be filtered according to the filter criteria, as shown in the following image.



There is a delimiter line between each of the five bar charts for each bar chart section.

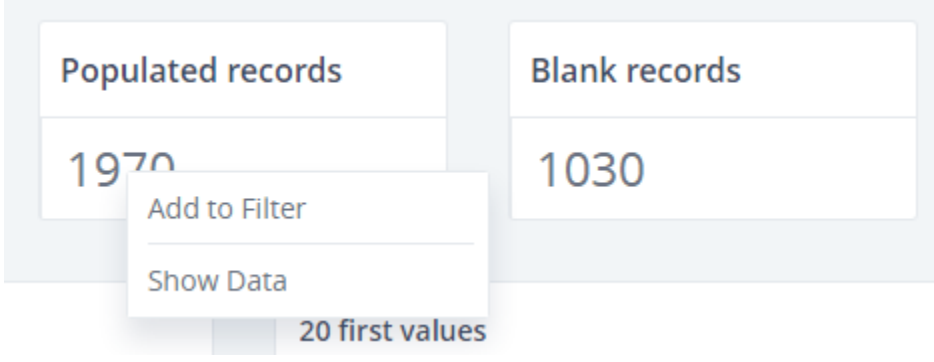


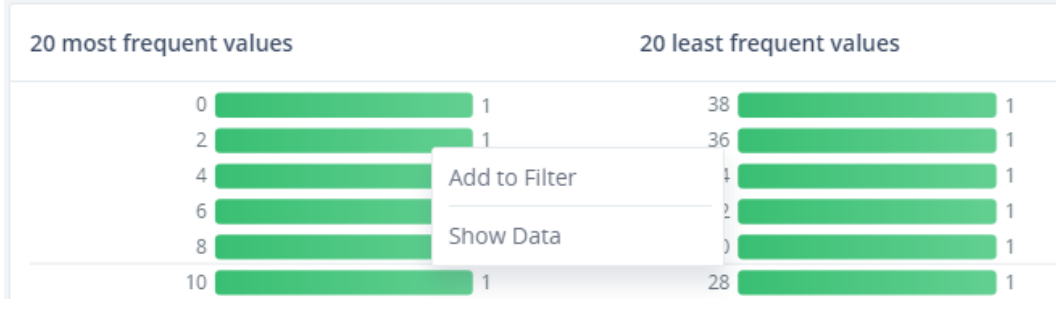
Once a filter is added, the Attribute page will be filtered according to the filter criteria.

Using the Show Data Filter Option From the Profile Attribute Page

You can navigate to the Data table and see data according to the selected item in the filter or global filter data by selecting the *Show Data* from the context menu. The presence of a filter and global filter on the Data table page depends on the selected bar chart on the Profile page.

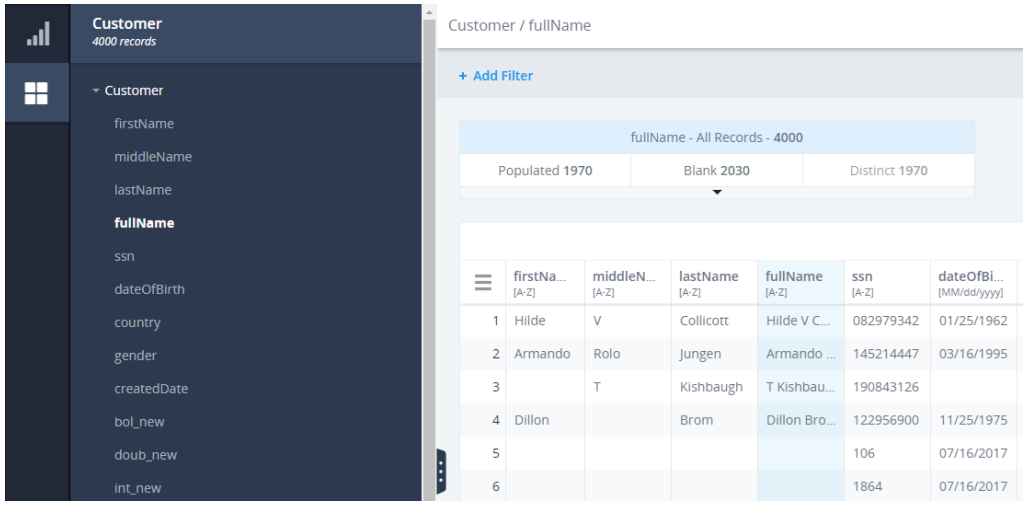
You can right-click on any tile, table cell, or bar chart and select *Show Data,* as shown in the following image.



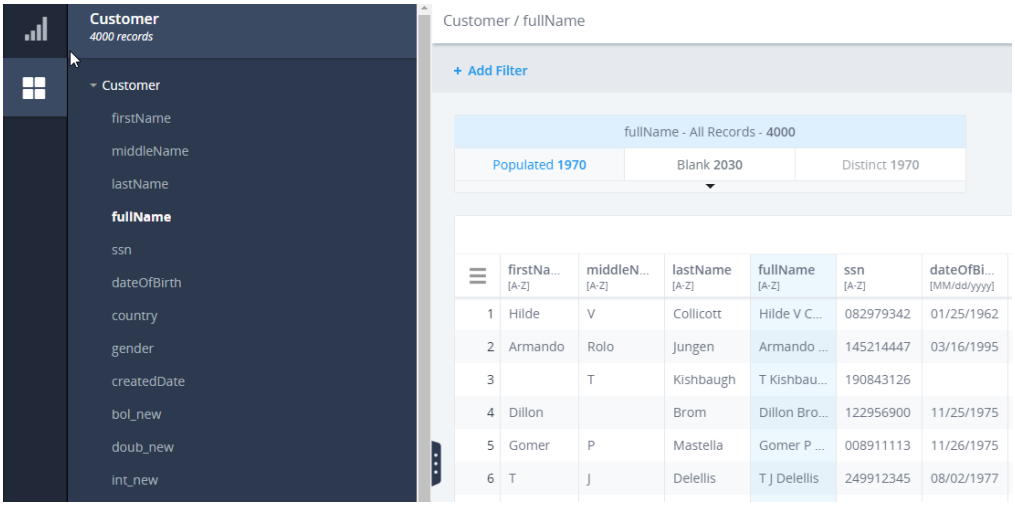


The following list describes the different functionalities of Show Data for each of the various components.

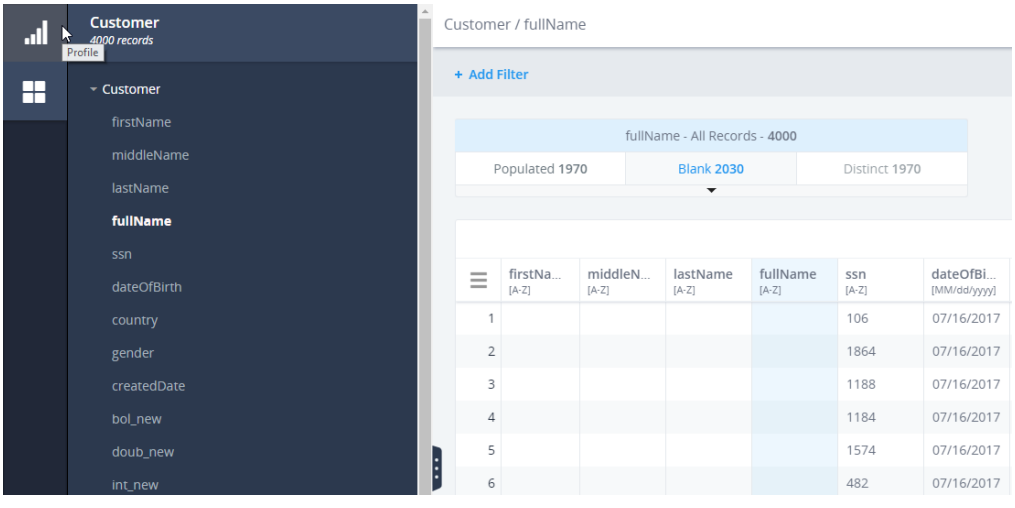
- For the Total records tile, the Data table page opens with no tab selected in the filter. The corresponding attributes are selected in the table, and all records are displayed.



- For the Populated records tile, the Data table page opens with the Populated records tab selected in the filter. The corresponding attributes are selected in the table, and populated records are displayed.



- For the Blank records tile, the Data table page opens with the blank records tab selected in filter. The corresponding attributes are selected in the table, and blank records are displayed.

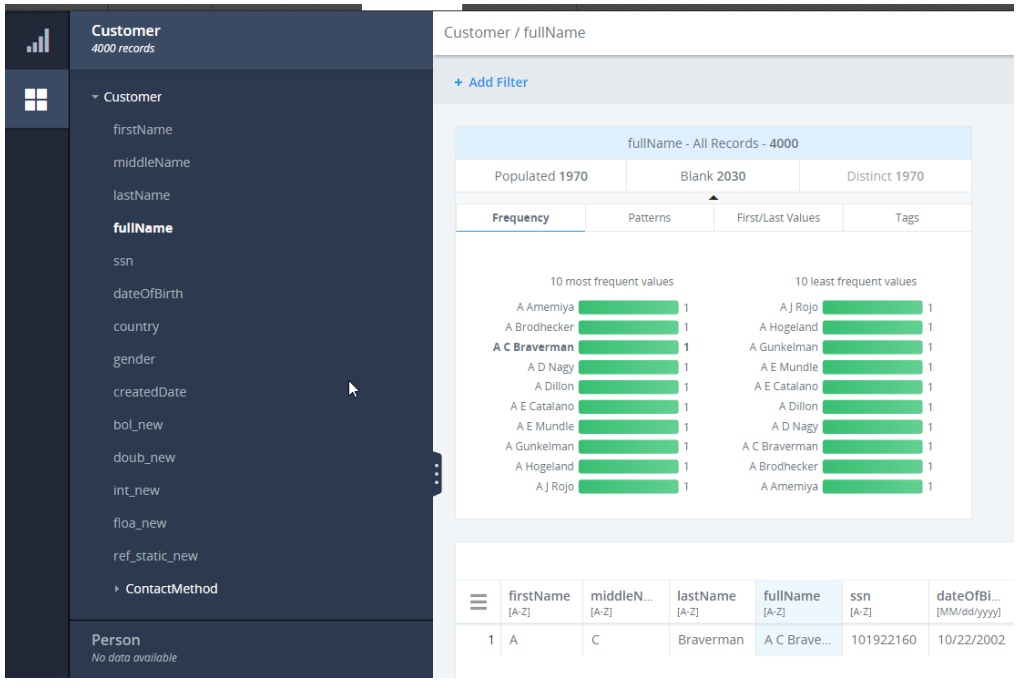


- There is no *Show Data* option for the Blank records tile.

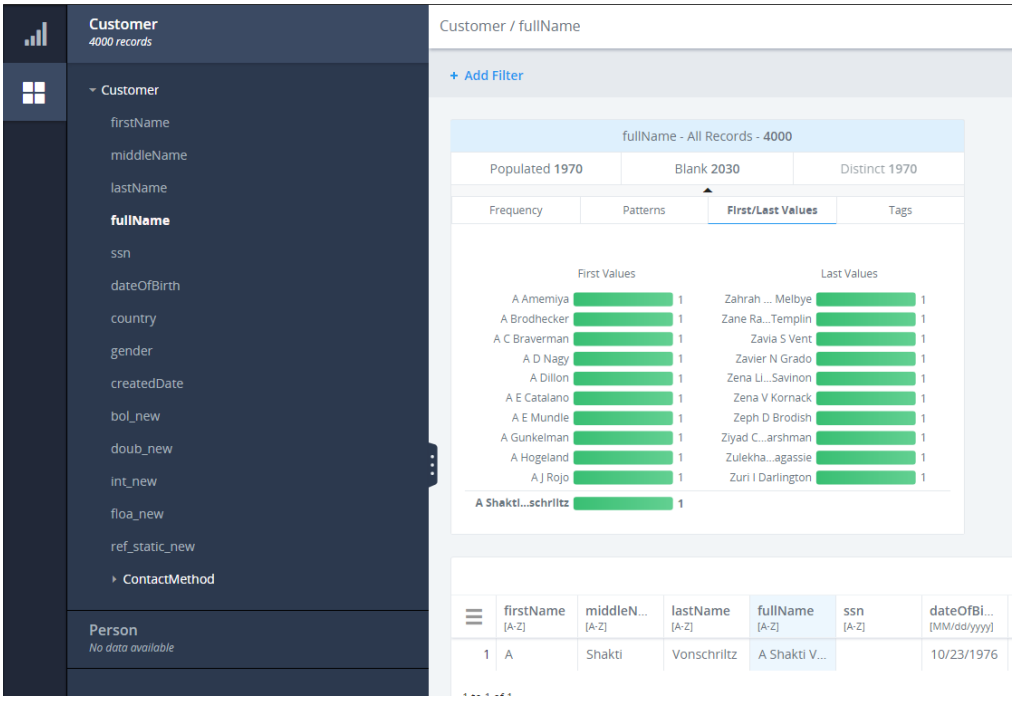
  You can right-click on the Attribute Completeness bar charts and select *Show Data*. The result will be the same for the Populated records tile.

- If you right-click on the green (filled) bars in the Attribute Completeness bar charts, the results will be the same as the Populated records tile. If you right click on the white (unfilled) part in the Attribute Completeness bar charts, the results will be the same as the Blank records tile.

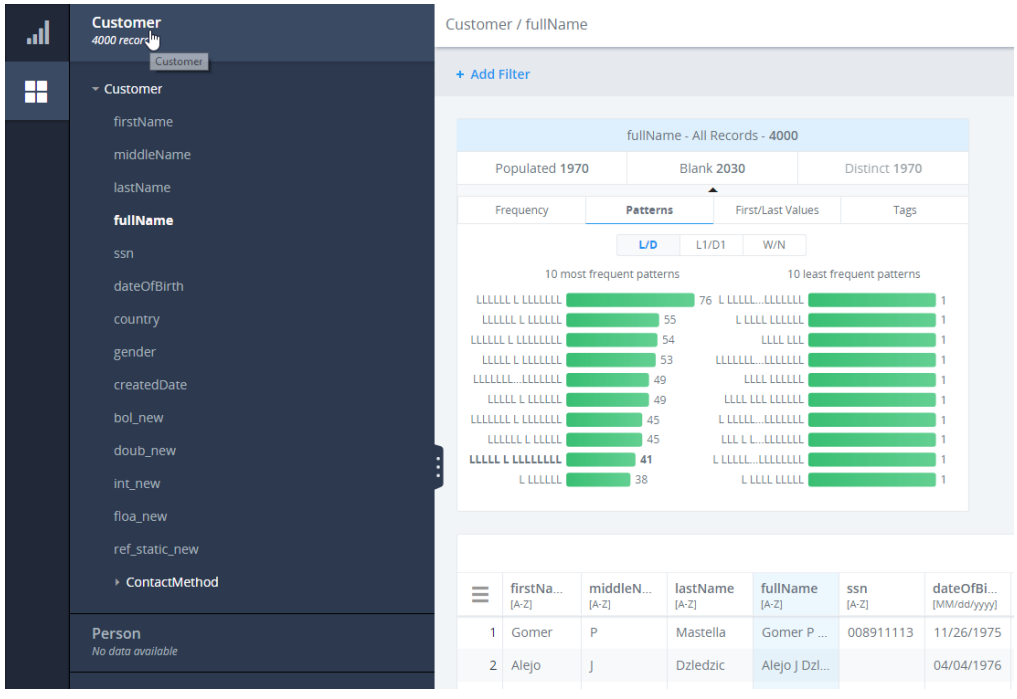
- You can right-click either of the bars or the value names in the 20 Most Frequent values, 20 Least Frequent values, 10 First values, and 10 Last values charts and select *Show Data* from the context menu. The Data table page opens with the column you selected. The Column Profile has the appropriate tab selected (displaying the same values you right-clicked on in the Profile) and the actual value used that was selected.



If the value is NOT in the default Column Profile due to space reasons (for example, if you clicked on the 19th Most Frequent value while the Column Profile displayed only the 10 Most Frequent values), the value appears as a special value in the correct place separated by a visual separator.

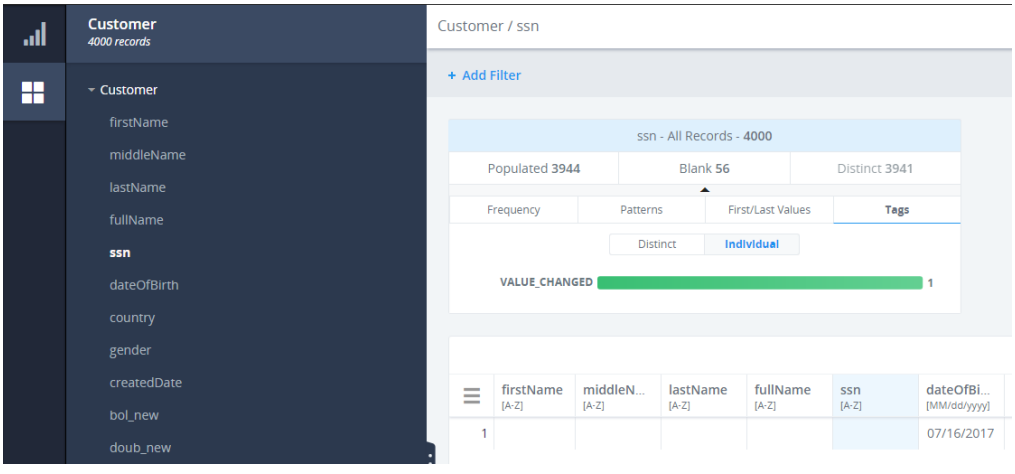


- You can right-click either the bar or the name of the pattern in the Letter/Digit, Word/Number patterns bar chart and select *Show Data* from the context menu. The Data table page opens with the column you selected. The Column Profile is switched to the Patterns tab, and the pattern you clicked on is selected.



If it does not appear, it is added as a new value at the end of the list separated by a visual separator.

- You can right-click either the bar or the name of the Tags in the Distinct or Individual tab bar chart and select *Show Data* from the context menu. The Data table page opens with the column you selected. The Column Profile is switched to the Tags tab, and the tag you clicked on is selected.

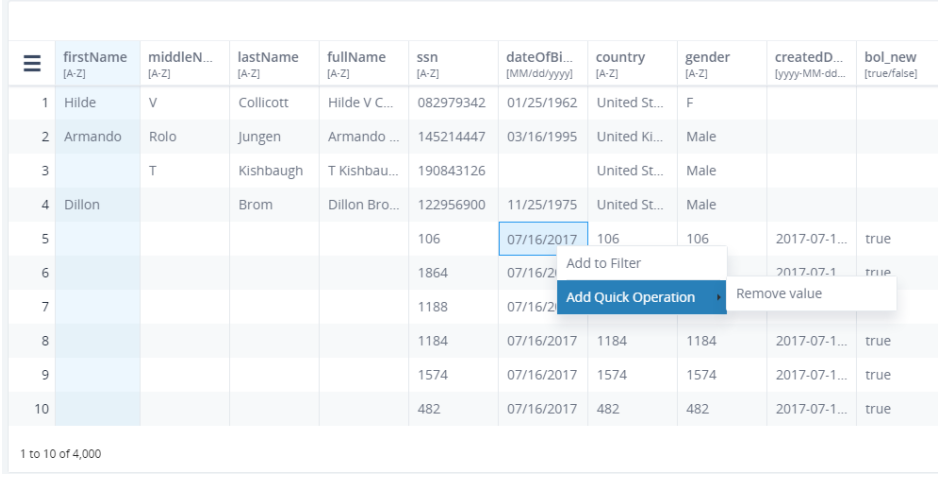


When you right-click anywhere else in the Attributes Page, the *Show Data* option will not be present.

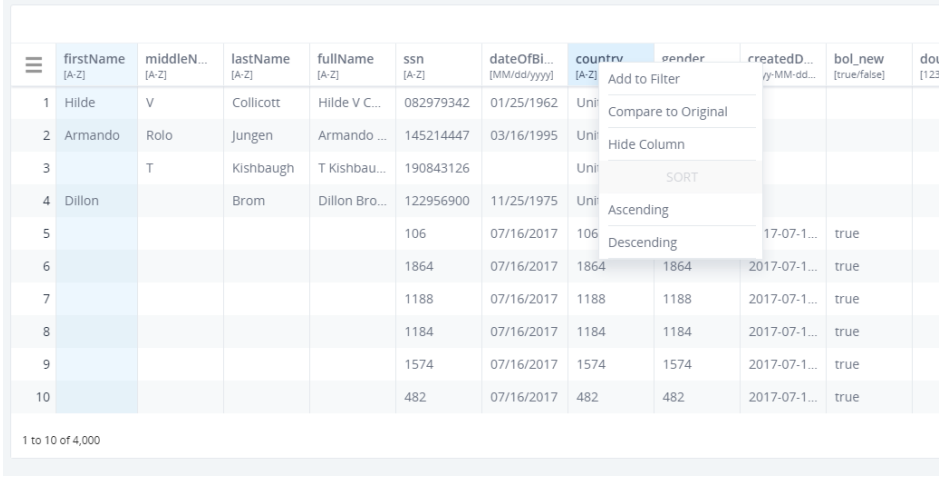
Data Tables

The Data table shows the data for the selected subject in the data tree. It contains columns also named as attributes.

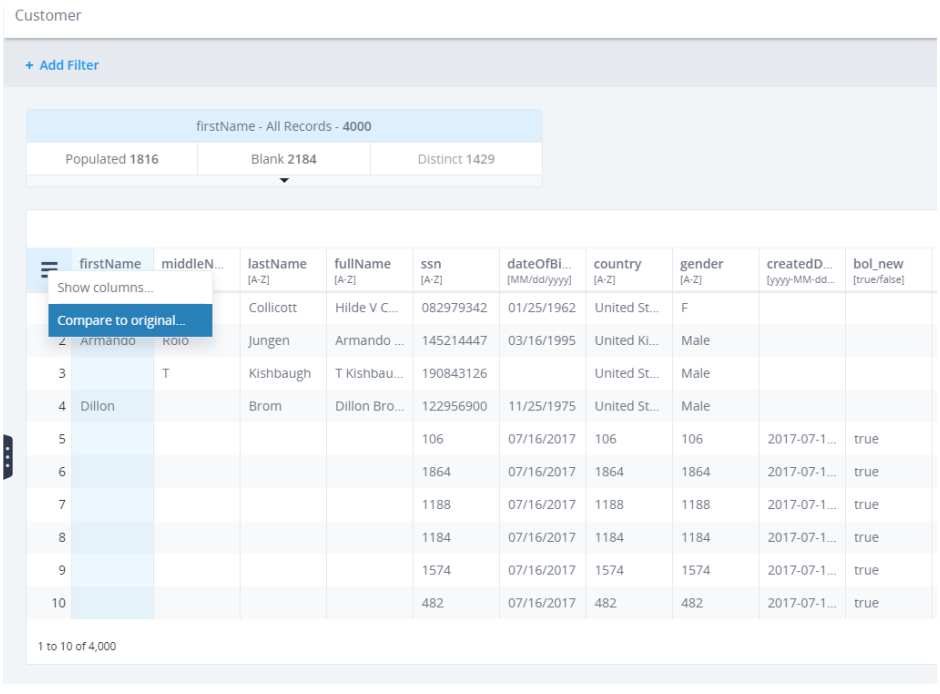
The main operations that can be performed from the Data table include *Add Quick Operation* and *Add to Filter*.



You can sort data in ascending and descending order, and reset sorting (which appears only after some initial sorting was performed). You can also hide some columns or compare them to the original (source).

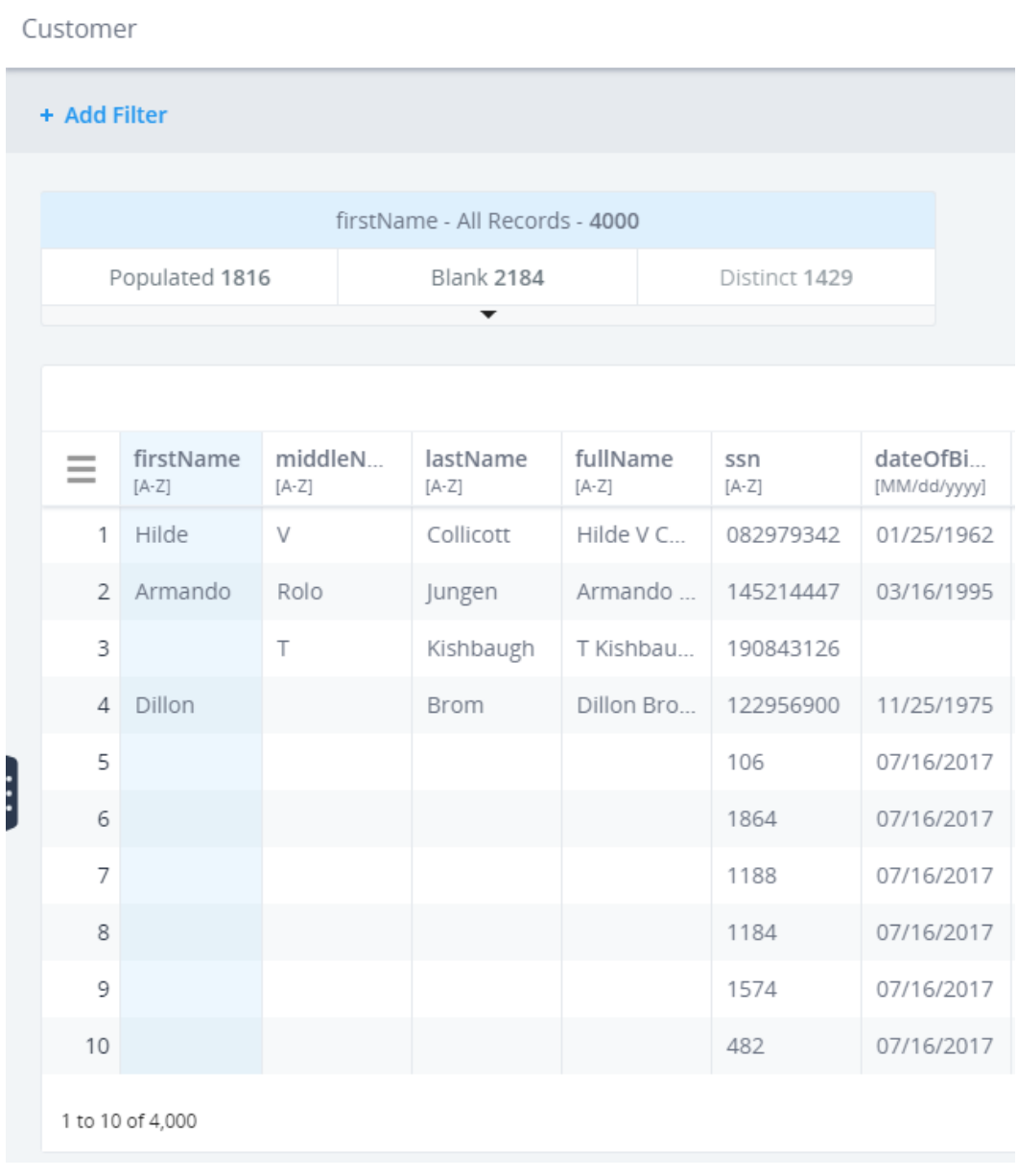


You can also Hide and Show columns, and enable the Compare to Original mode from the menu.



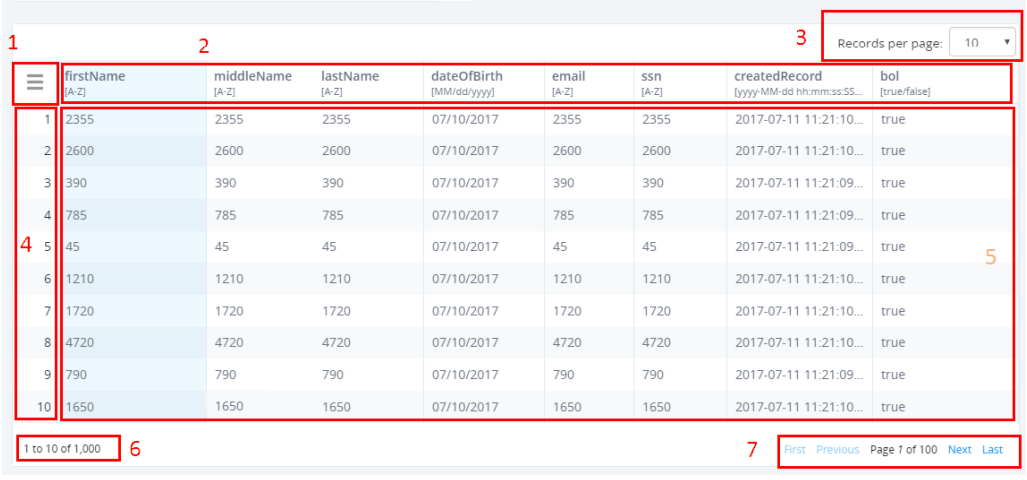
	firstName	middleN...	lastName [A-Z]	fullName [A-Z]	ssn [A-Z]	dateOfBi... [MM/dd/yyyy]	country [A-Z]	gender [A-Z]	createdD... [yyyy-MM-dd...	bol_new [true/false]
			Collicott	Hilde V C...	082979342	01/25/1962	United St...	F		
2	Armando	Rolo	Jungen	Armando ...	145214447	03/16/1995	United Ki...	Male		
3		T	Kishbaugh	T Kishbau...	190843126		United St...	Male		
4	Dillon		Brom	Dillon Bro...	122956900	11/25/1975	United St...	Male		
5					106	07/16/2017	106	106	2017-07-1...	true
6					1864	07/16/2017	1864	1864	2017-07-1...	true
7					1188	07/16/2017	1188	1188	2017-07-1...	true
8					1184	07/16/2017	1184	1184	2017-07-1...	true
9					1574	07/16/2017	1574	1574	2017-07-1...	true
10					482	07/16/2017	482	482	2017-07-1...	true

When you navigate to the Data table page, the first table column is selected by default and the On-Screen filter will be collapsed. You can deselect a column by clicking the column header.



Table

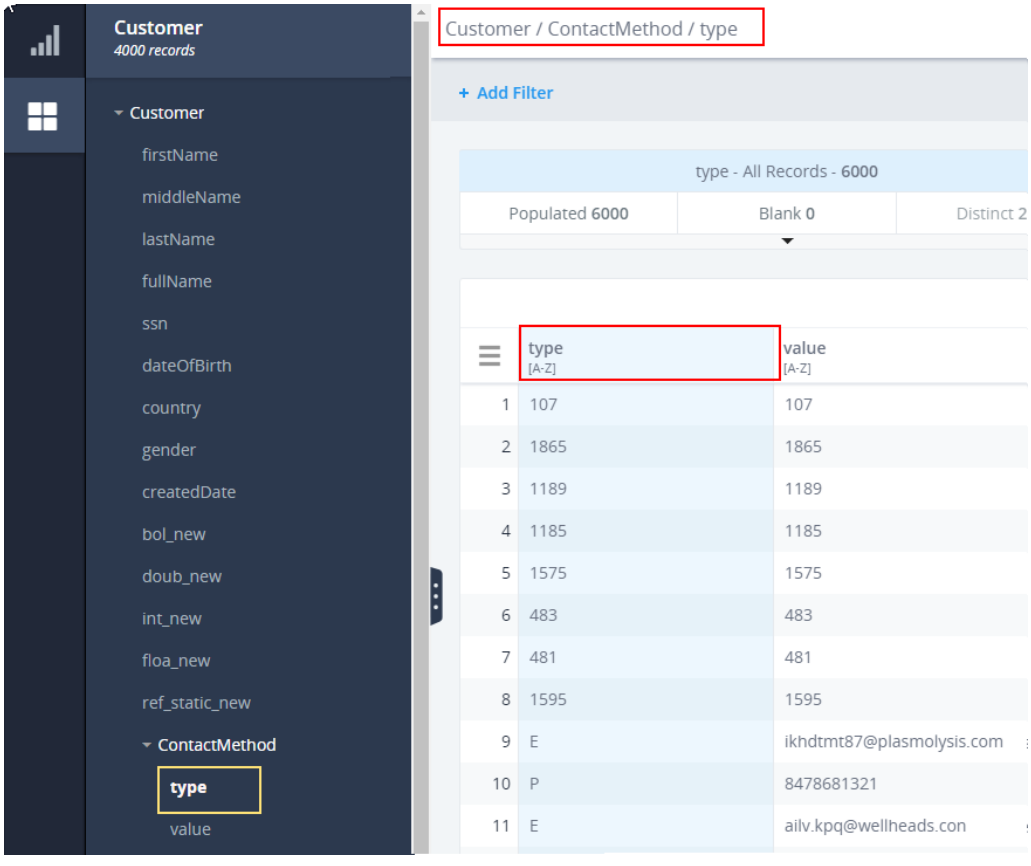
One of the main functions of the Data table is the table where all data is displayed. By default, 100 records are displayed and the first column is selected.



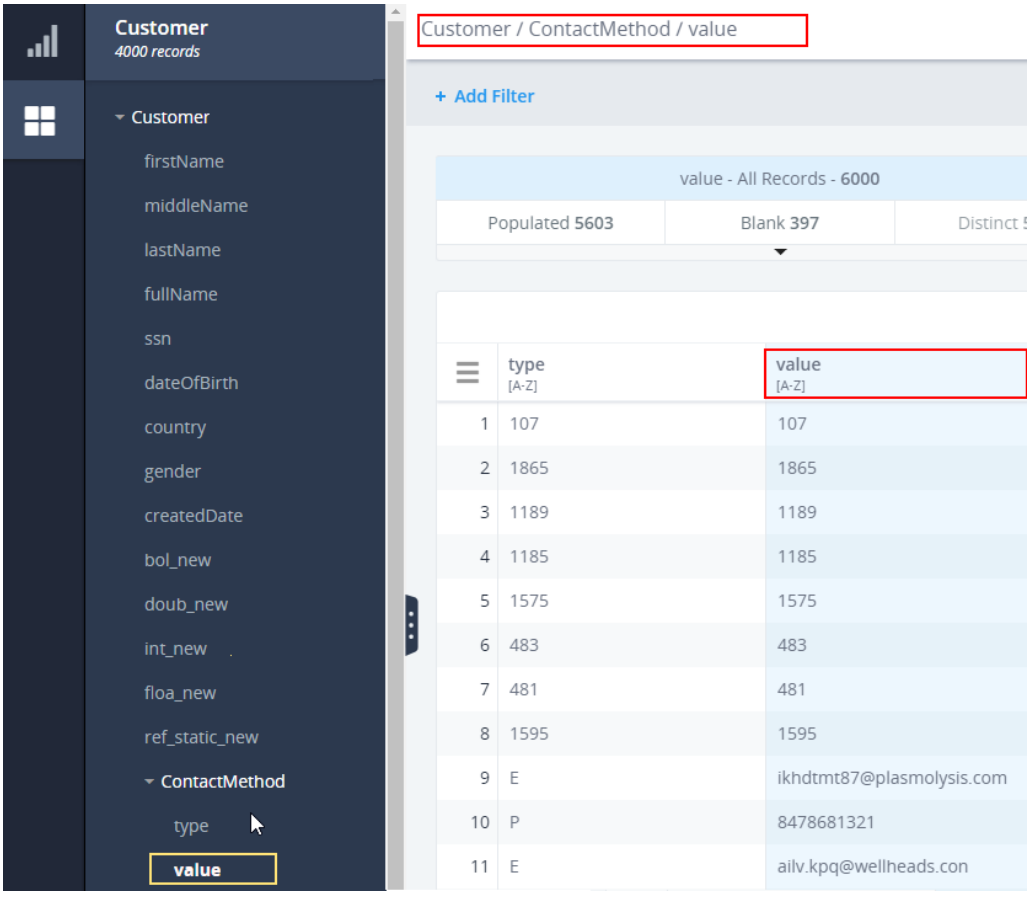
The following list shows the primary components of the table, corresponding to the image above:

1. Menu
2. Table header
3. Number of Records per page
4. Record numbering
5. Data
6. Number of records per page and number of all records
7. Pagination

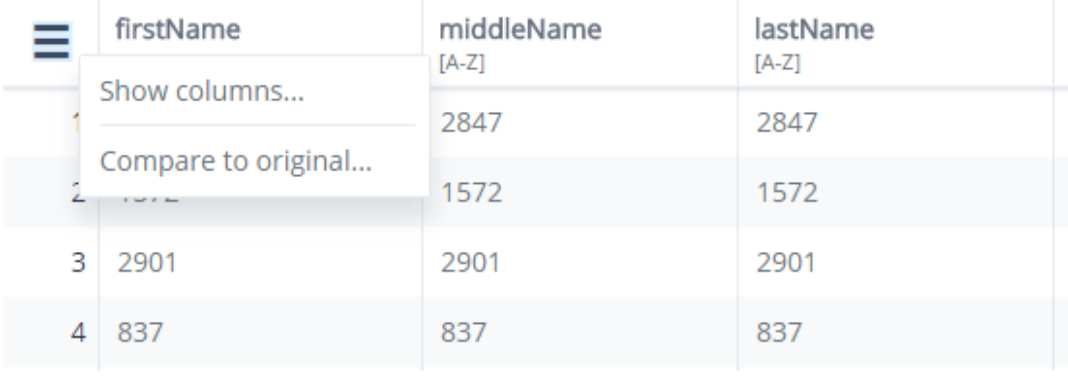
If you select an attribute in the tree, the corresponding breadcrumbs path appears as well as the corresponding table attribute that should be selected:



If you select another table or tree attribute, it will automatically update in the breadcrumbs path, table, and model tree:

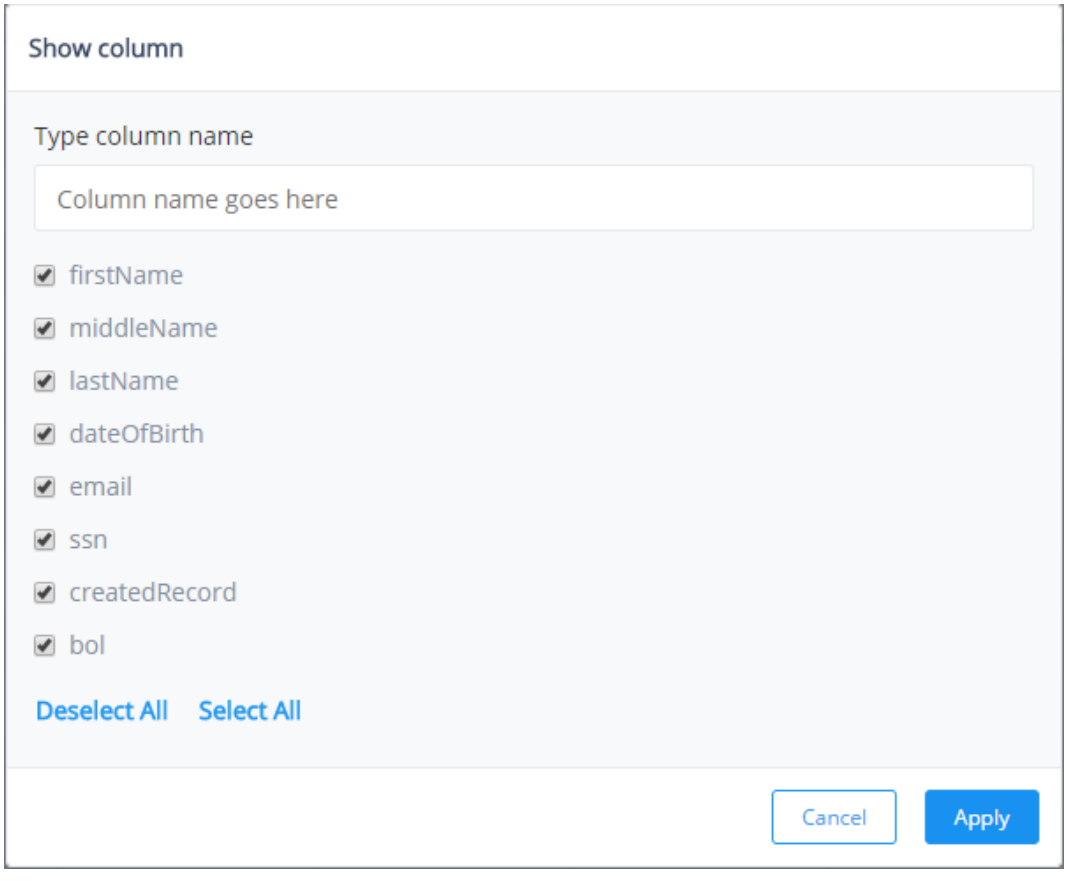


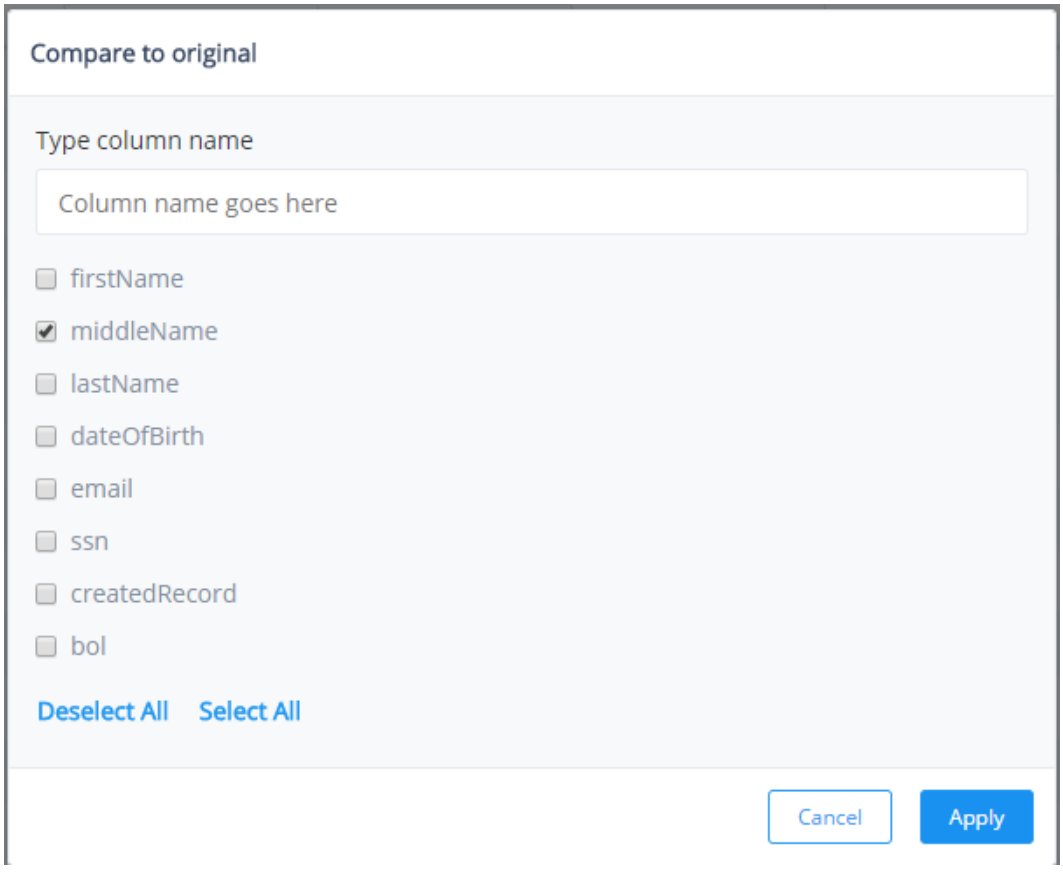
The **menu** allows you to configure the view of the table. Once you click the menu icon, the *Show columns* and *Compare to original* options are available, as shown in the following image.



The *Show column* option opens a dialog consisting of all table columns. The selected columns will appear in the table.

The *Compare to original* option opens a dialog listing all columns which are cleared by default. This dialog turns on Compare to original mode for the selected columns.

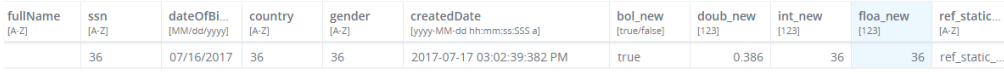




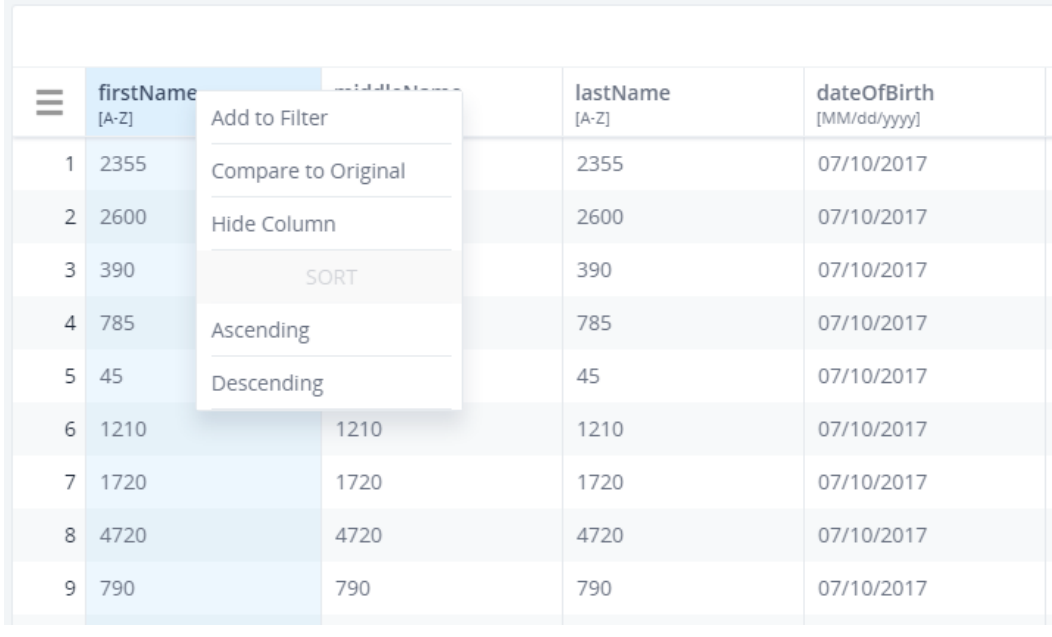
You can right-click on any column header to see the context menu. It contains the following options: Add to Filter, Compare to original, Hide Column, (Sort) Ascending, and Descending.

The Column header consists of a field (attribute) name and type label. Data Types of the attributes (columns values) are displayed in the table header so you can see what data type you are working with:

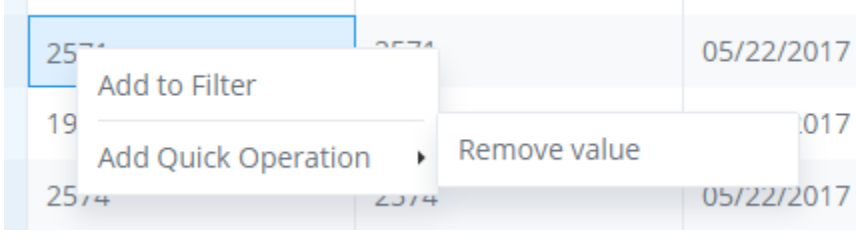
- String data type is represented with [A-Z] characters.
- All numeric data types (Integer, Long, Double, Float) are represented with [123] characters.
- Date data type is represented with [MM-dd-yyyy] labels.
- DateTime data type is represented with the [yyyy-MM-dd hh:mm:ss:SSS a] label.
- Boolean data type is represented with the [true/false] label.



Columns can be resized if needed.



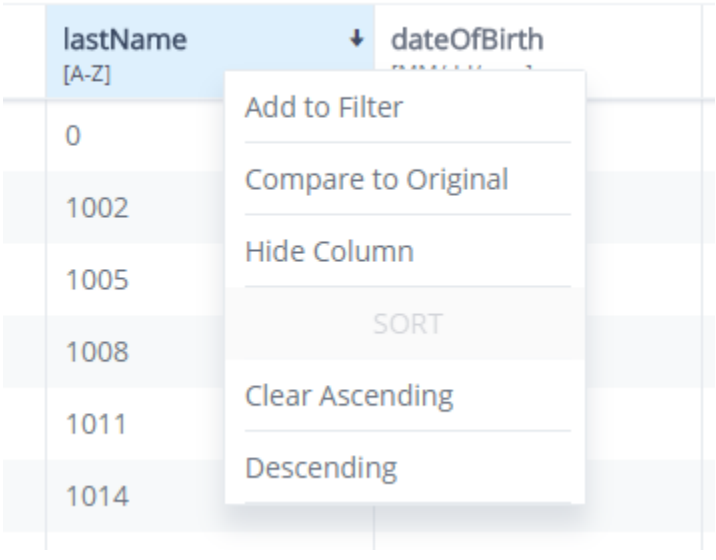
You can open the context menu for every table cell by right-clicking on it.



The following options appear in the context menu: *Add to Filter* and *Add Quick Operation*. The *Add Quick Operation* allows you to select more options containing a list of all available operations (for example, "simple = false") in the corresponding .xml file.

Table Sorting

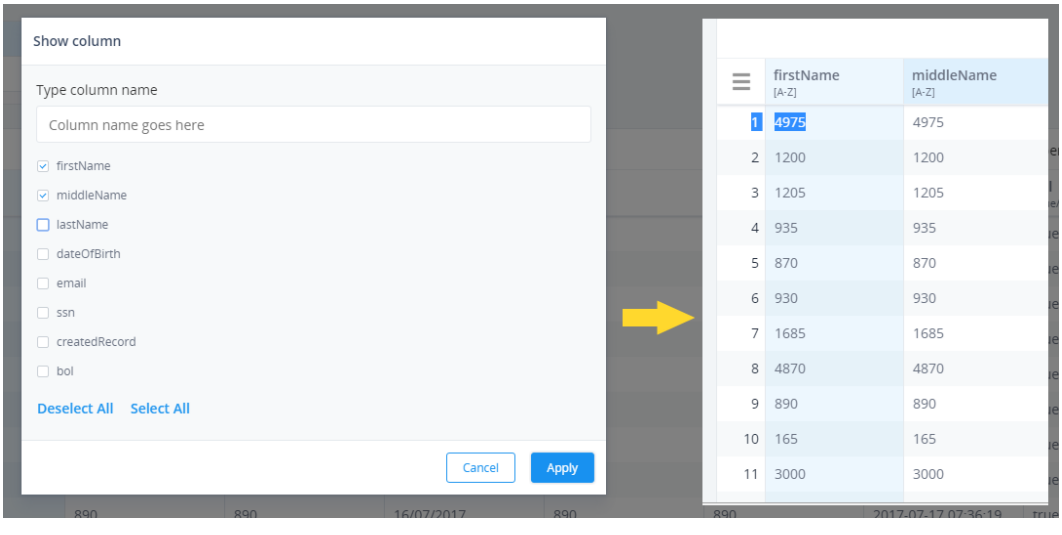
You can sort and reset the table values in ascending or descending order.



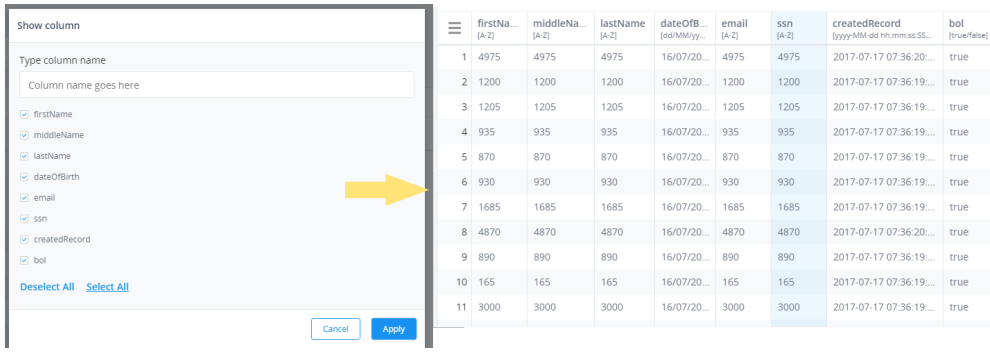
Showing and Hiding Columns

You can show and hide columns in the Data table by right-clicking the column header and selecting *Hide Column* from the context menu.

You selection is saved within a session. The column will be hidden from that moment on.



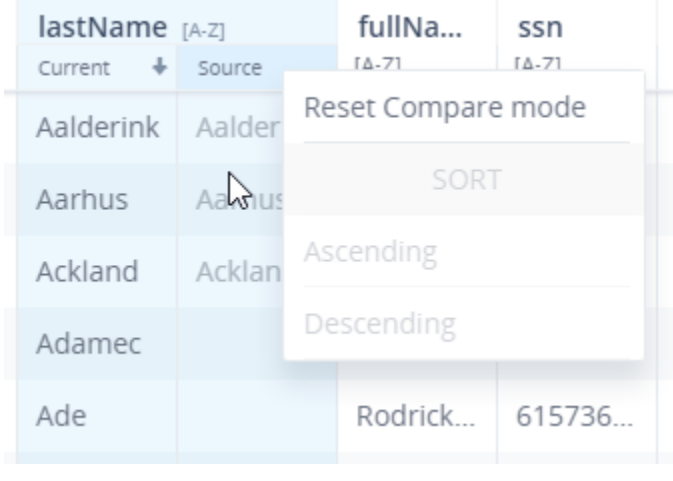
You can make the column visible again by clicking on the menu icon and selecting *Show column*, and then clicking on the columns you wish to view.

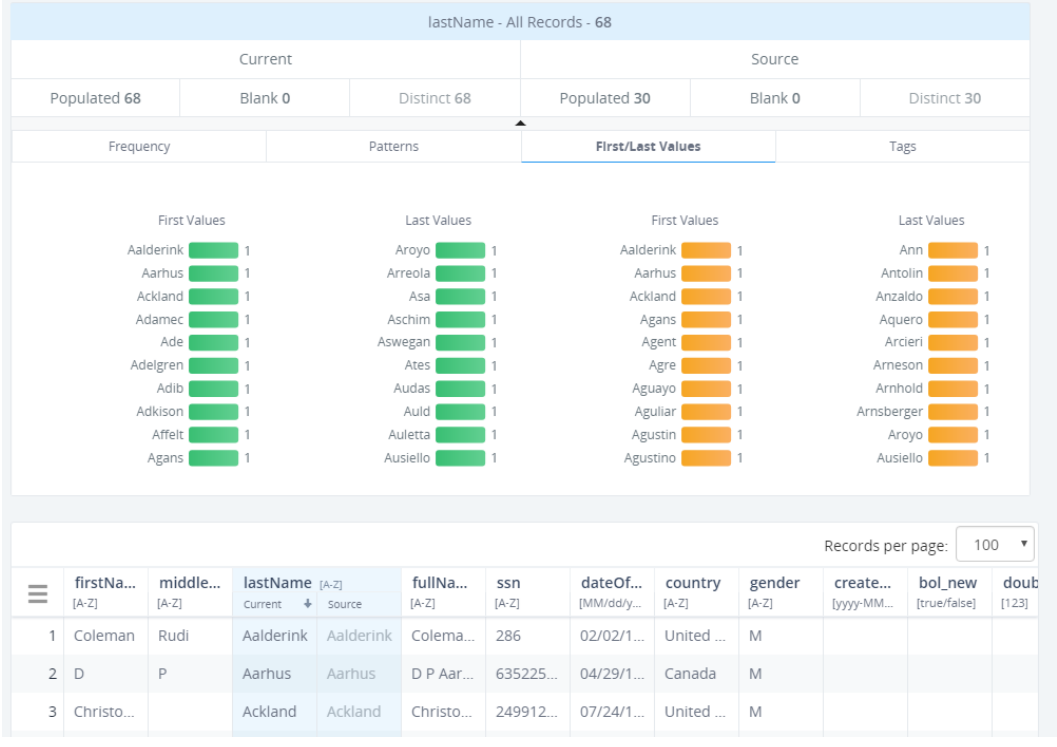


Compare to Original

You can split a column to enable the Source-Current values comparison by right-clicking and selecting *Compare to original*. To revert from the Compare mode, right-click the configured column and select *Reset Compare mode*.

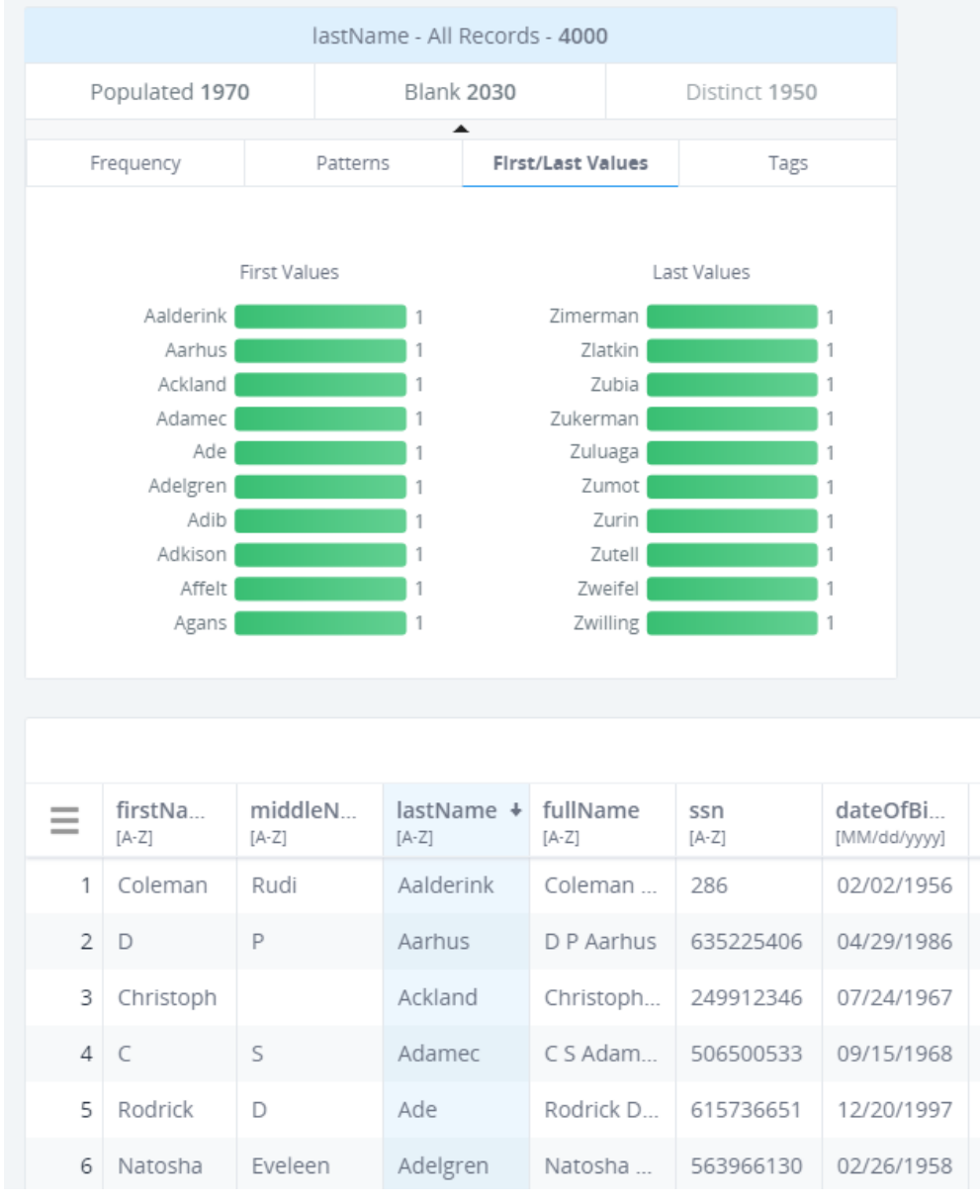
The *Add to Filter* and *Sorting* functionalities are not implemented yet.





Filter Panel

The Filter panel divides data availability (Populated/Blank/Distinct) and specific characteristics (for example, frequency, patters, tags, and first/last values). Each tab has a list of bar charts which helps you understand and visualize filtering.



The filters are represented as a panel with headers named according to the selected column. For example:

- *firstName –All records-6001*

- tree filter tab: Populated <number populated records>
- Blank <number of blank records>
- Distinct <number of distinct records>

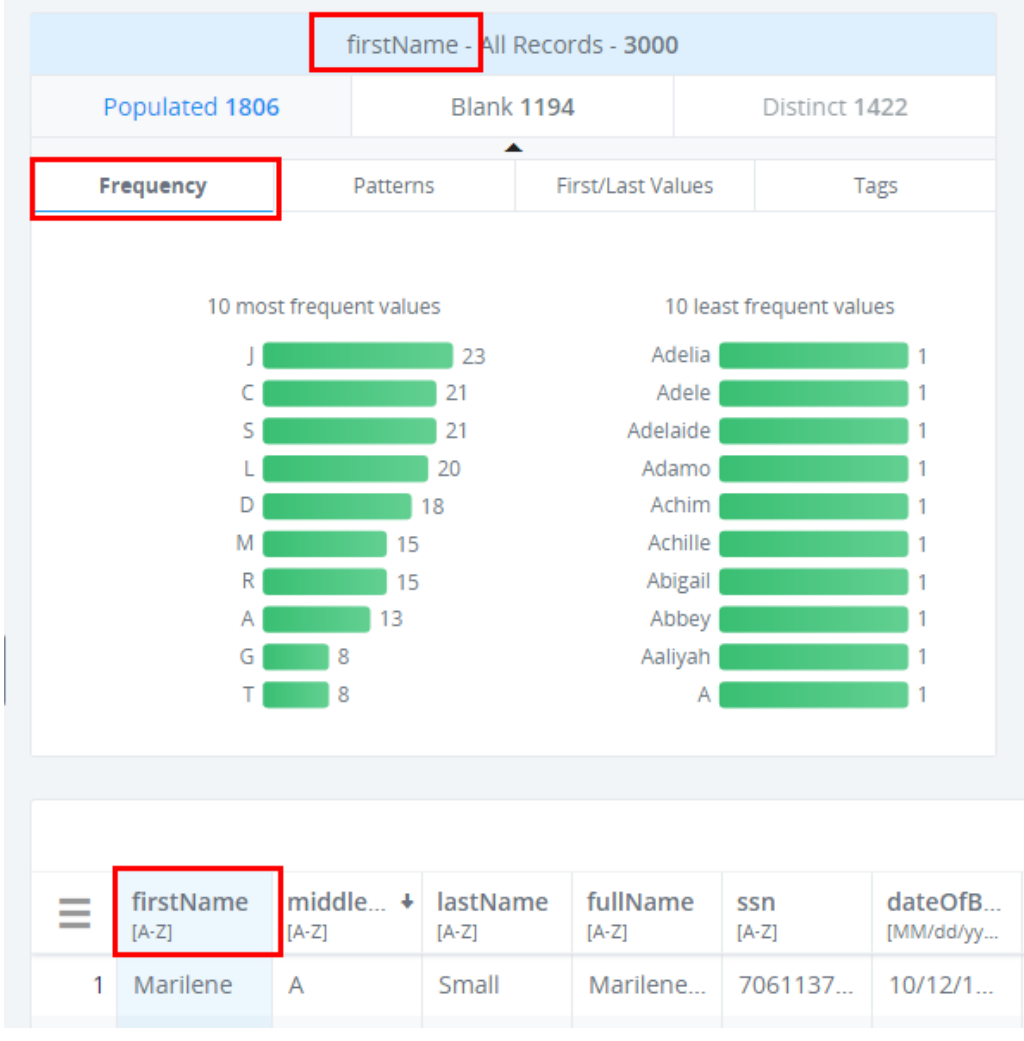
Each filter tab contains several sections:

- Frequency
- Patterns
- First/Last Values
- Tags

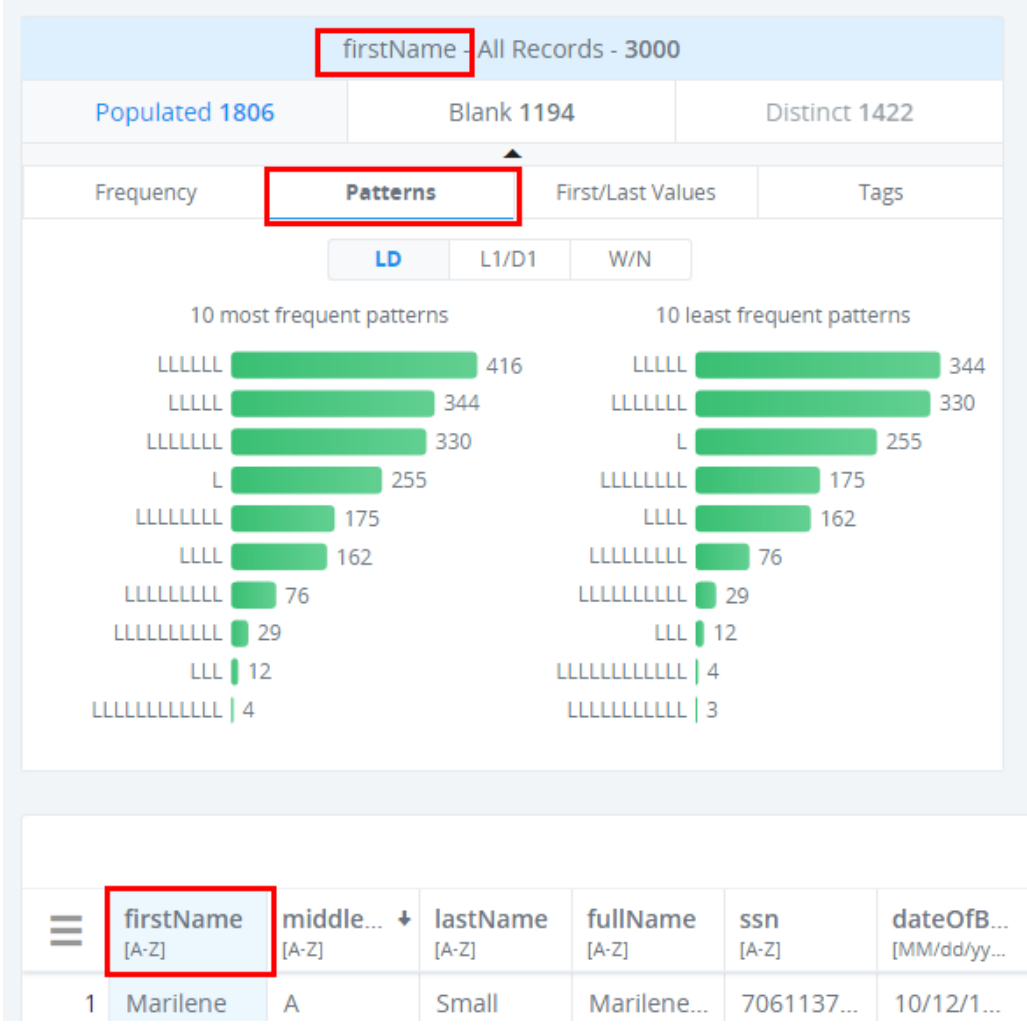
Each tab also contains bar charts with labels and a quantity of data records

The panel appears only for selected columns. If the column is deselected, filter panel will not be displayed.

The Frequency tab displays 10 most frequent values and 10 least frequent values for the selected column.

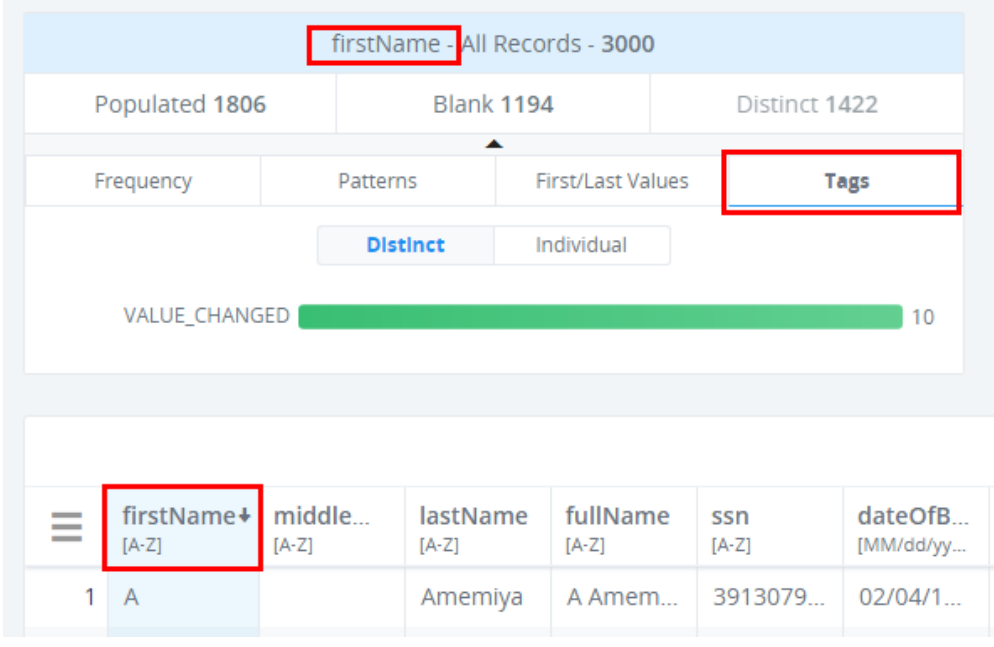


The Pattern tab displays the 10 most frequent patterns and the 10 least frequent patterns for the selected column. Patterns are divided by groups: L/D, L1/D1, and W/N.

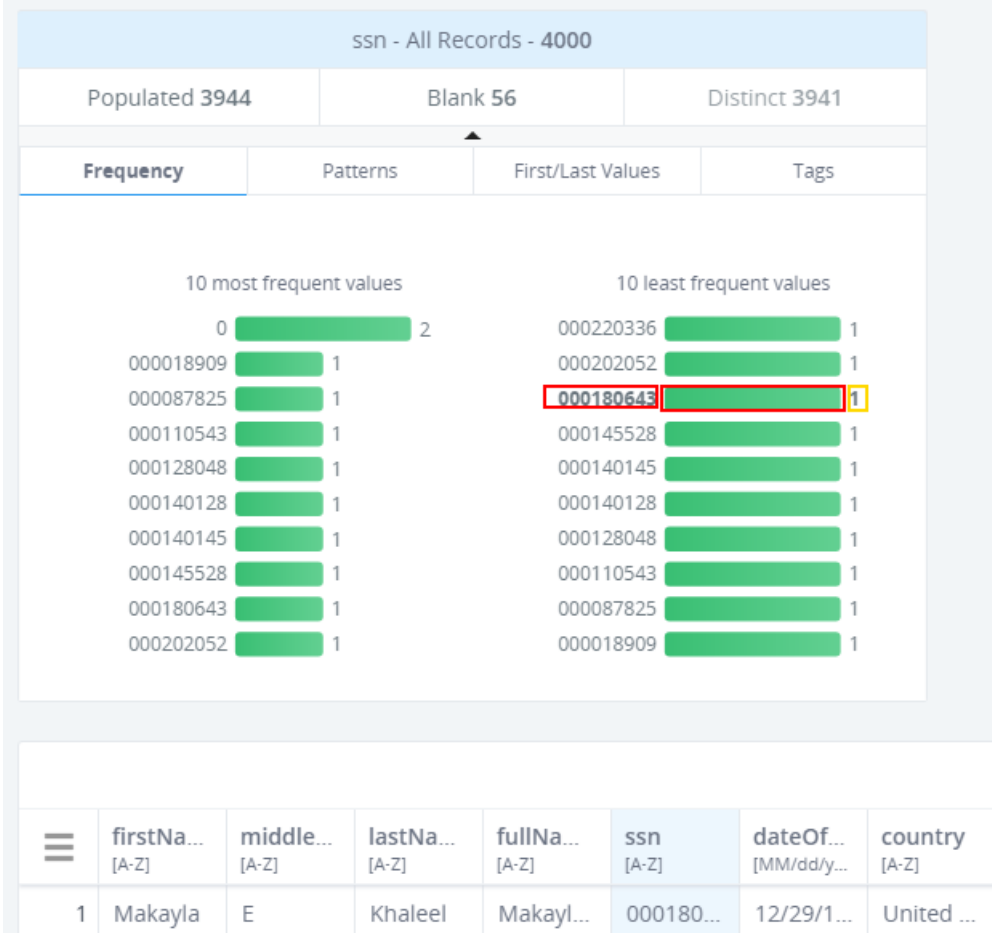


The First/Last Values tab displays the first and last values for a selected column.

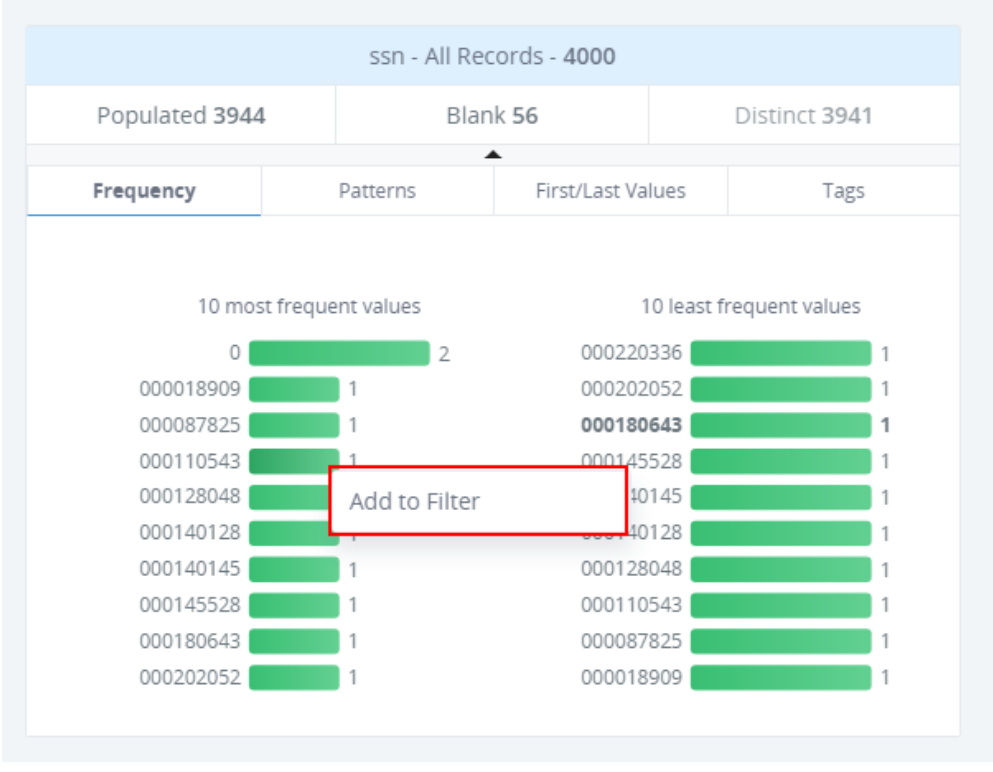
The Tags tab displays tags related to the values of a selected column. Tags are divided into the following groups: Distinct and Individual.



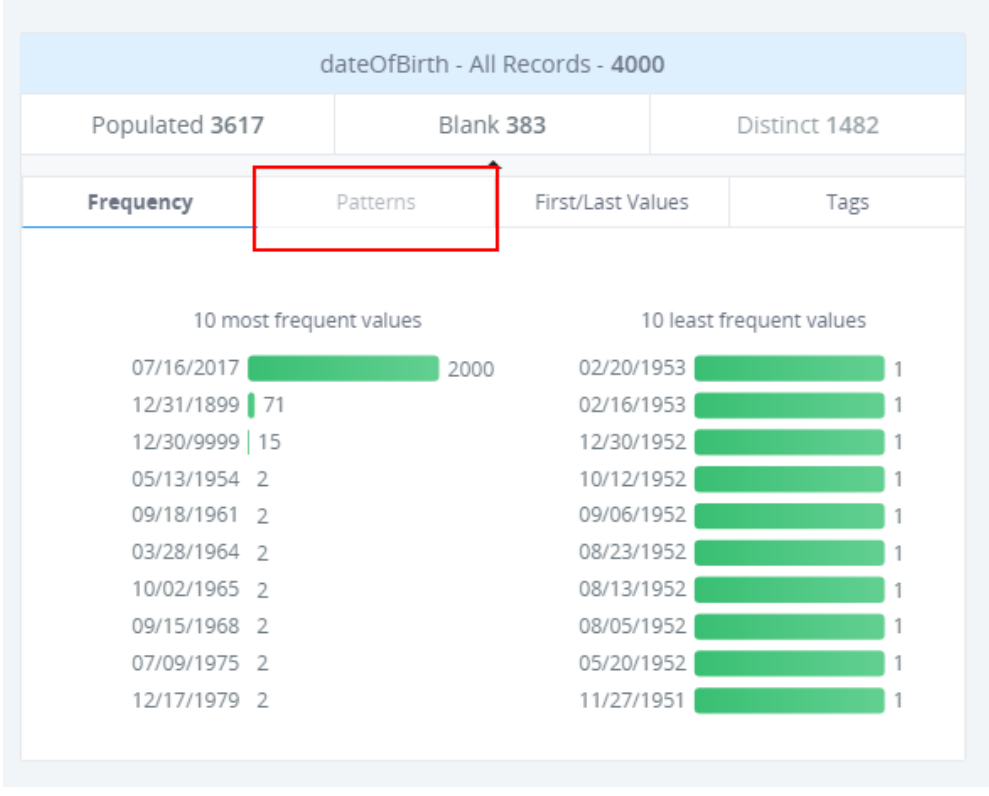
The Quantity, Label, and bar charts are highlighted once they are selected.



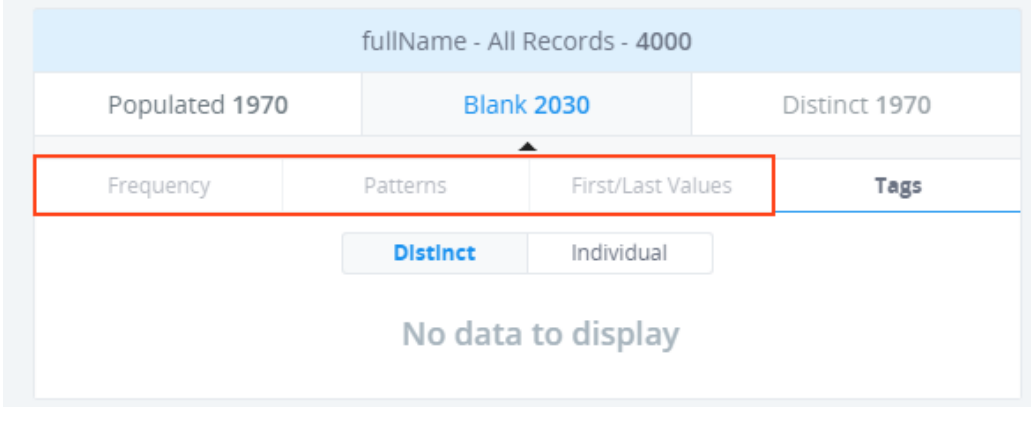
If you right-click on an item, the context menu appears, where you can select the *Add to Filter* option.



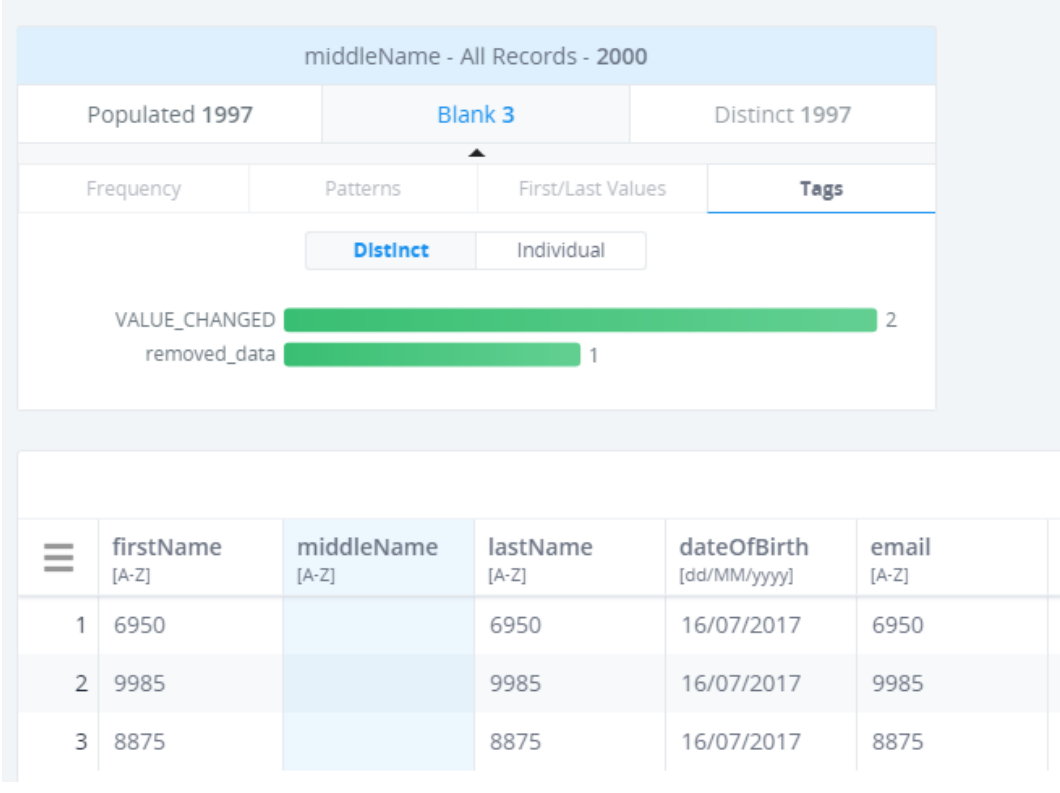
However, not all tabs are available for every attribute type. For example, the patterns tab is not displayed for Date, DateTime, and Boolean data types.



There is also a blank tab available from the filter, but only the Tags tab is enabled. The Frequency, Patterns, and First/Last Values tabs are displayed because there is nothing to show for blank values.



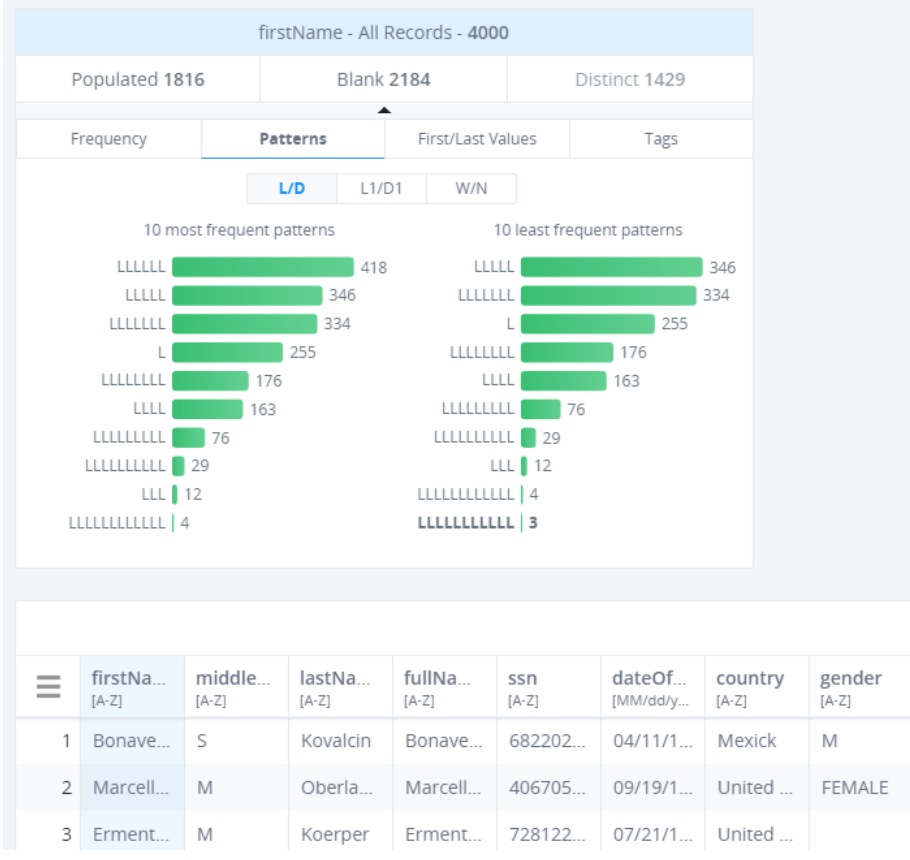
If there were executed rules with operations that were tagged, the following information will appear in the Tag tab.



The Distinct tab in the filter is currently unavailable.

Using Filters

You can right-click the Populated tab to filter the Data table to show only rows with filled-in values for the selected column. Once you select a specific bar chart, the Data table will be filtered again, and show only rows filled with the selected value for the selected column.



You can right-click the blank tab to filter the Data table so only rows with empty cells for the selected column are shown.

Using the Add to Filter Option

All Column profile elements that can be used for filtering can also be right-clicked. Then, select *Add to Filter* from the context menu.

If you add a filter from the Frequency or First/Last Values tabs, then the filter tag will be added with property = <column name of the selected column>, operand *Equals* and value=<value user clicked on>.

If you add a filter from the Pattern tab, then a filter tag is added with the property = <column name of the selected column>, operand *like*, and value=<pattern user clicked on converted in RegEx format>.

If you add a filter from the Tags tab, a filter tag is added with the property = <column name of the selected column>, operand *Tagged with*, and value=<value user clicked on>.

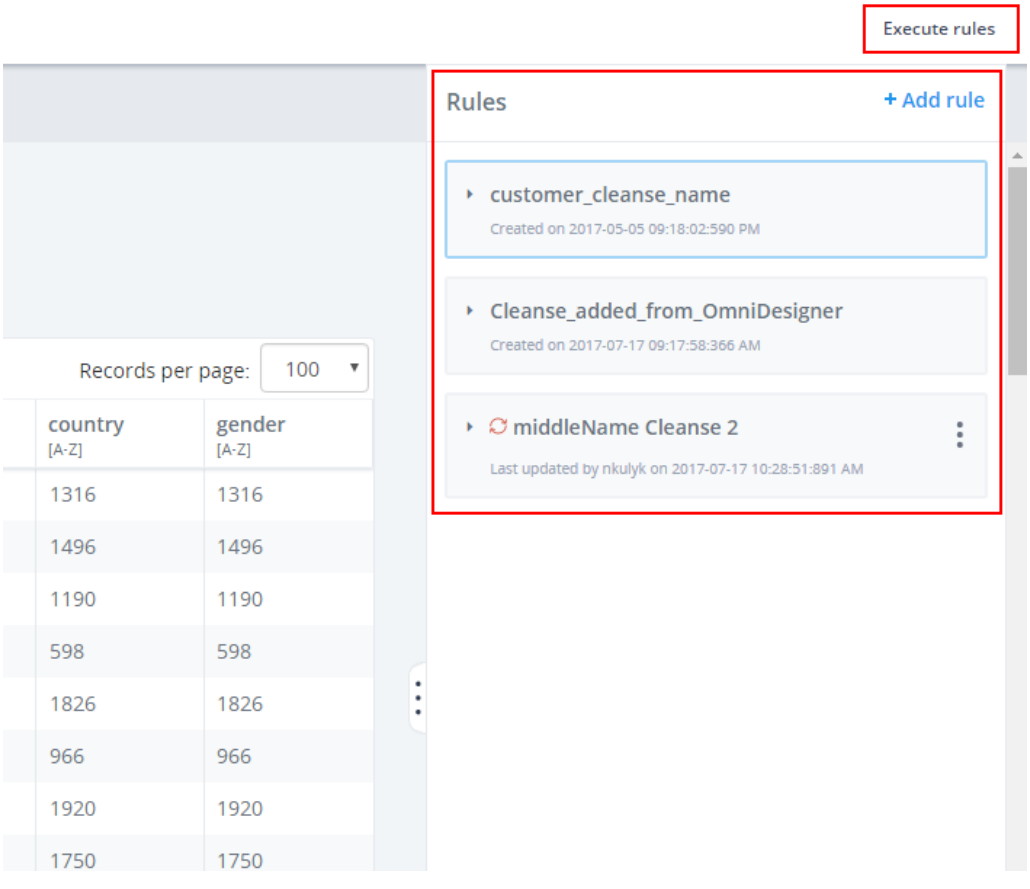
Rules and Operations

If you identify problems or issues with the data and wish to change it, you can use the Rules functionality.

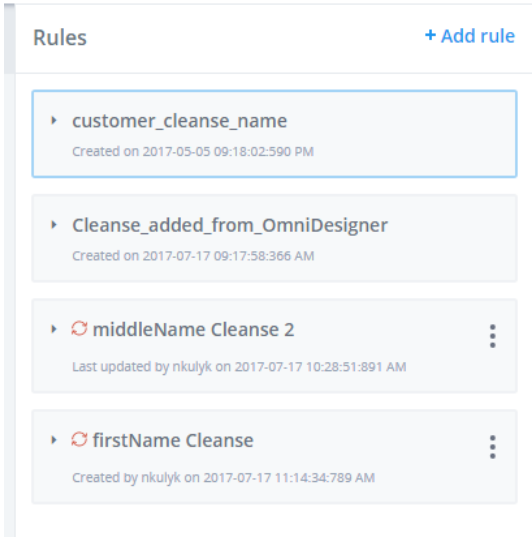
The Rules functionality allows you to edit, mark, and remove data from the interface without using DQS.

It is also used to work with data records (update/remove) by creating data quality rules (Cleansing).

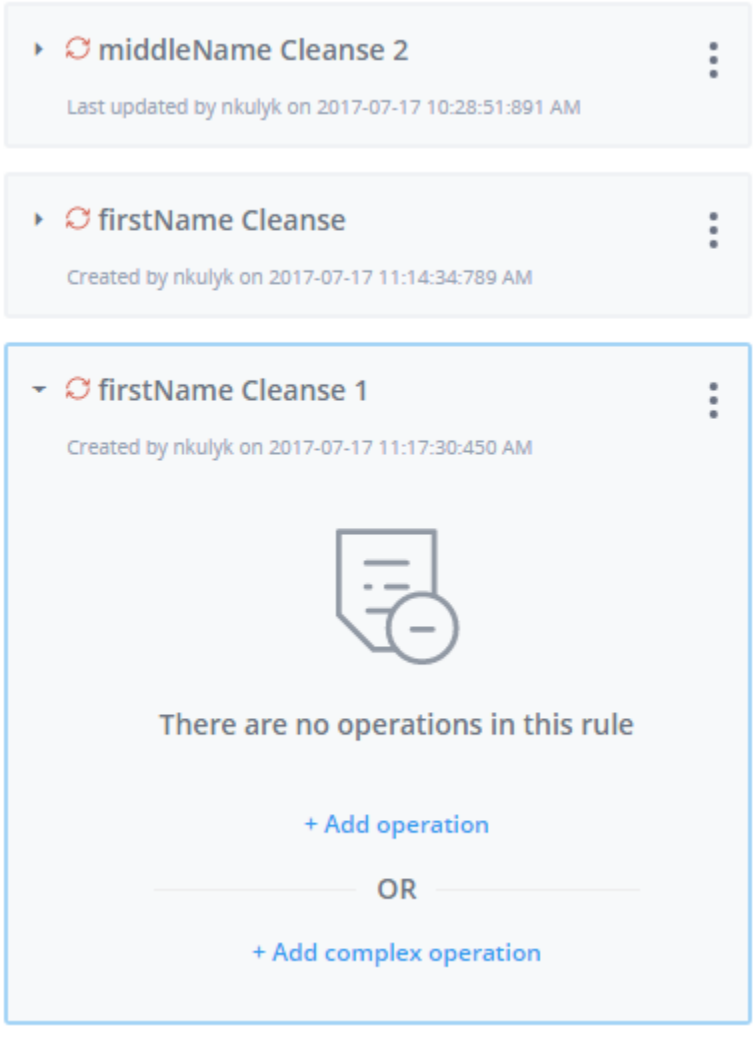
The Rules functionality is represented with the Rules panel and the Execution process.



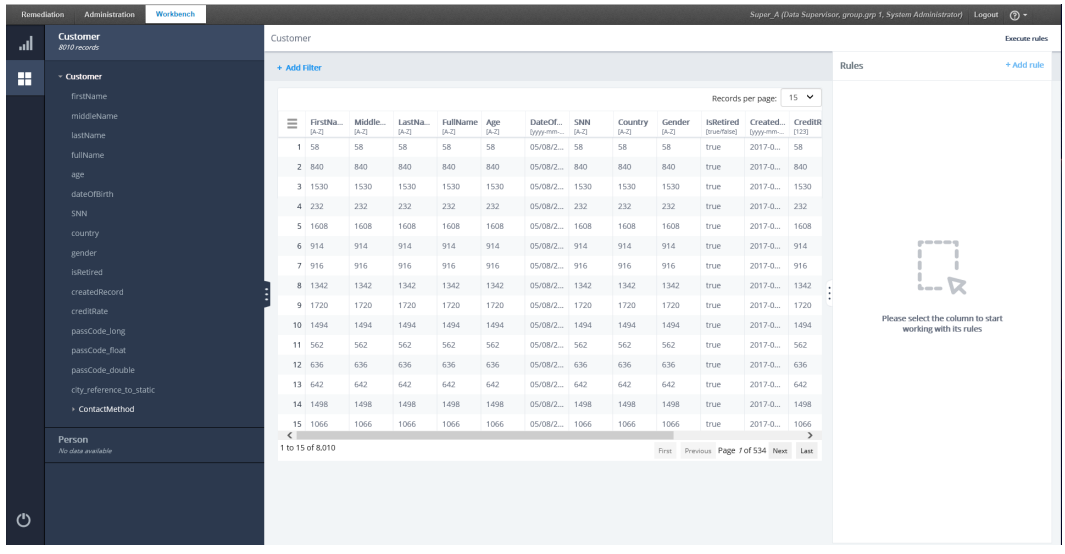
The Rules panel is available for the selected column in the Data table and is collapsed by default. The Rules panel displays a list of rules associated with the selected column. The Rules are collapsed by default when navigating to the Rule panel after working with filters or sorting.



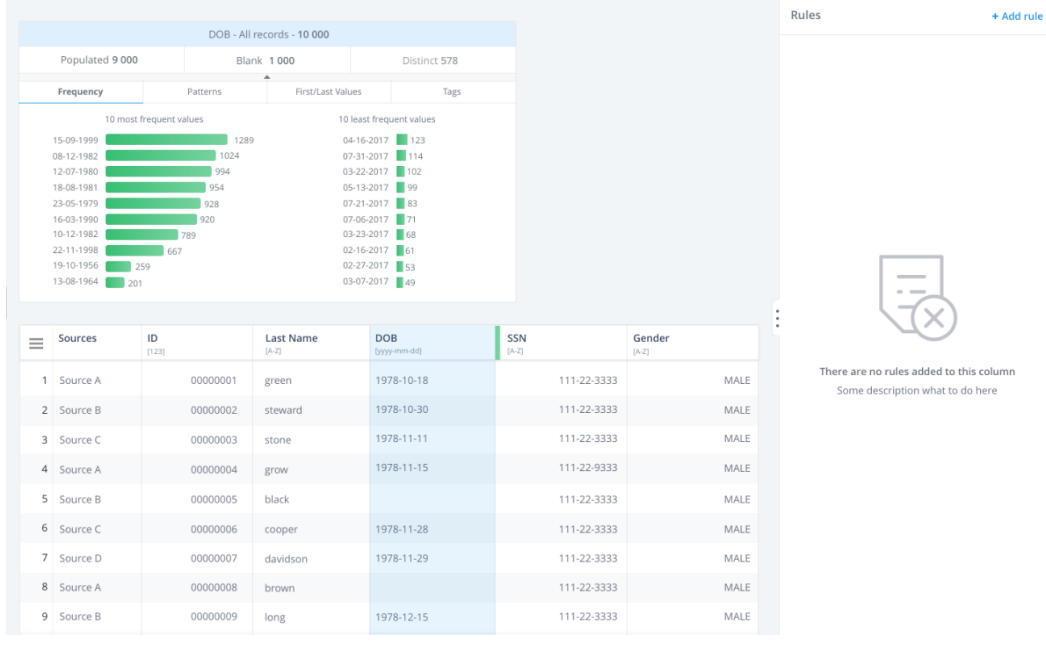
If you are adding a new rule, it is added in the expanded state.



If the column is not selected, the corresponding information is displayed in the Rules panel:



If there are no rules associated with the selected column, the corresponding information will be displayed.

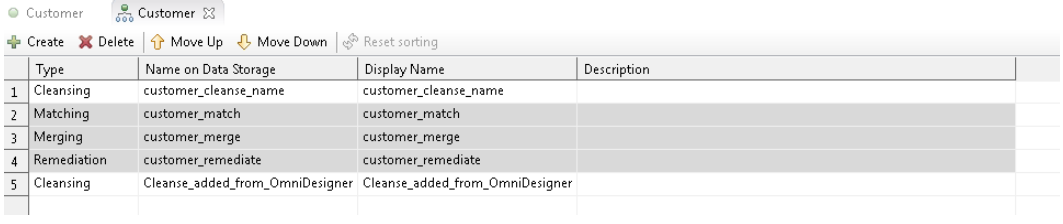


There are two types of data quality rules:

- External rules
- Standard rules (rules that you can create from the UI)

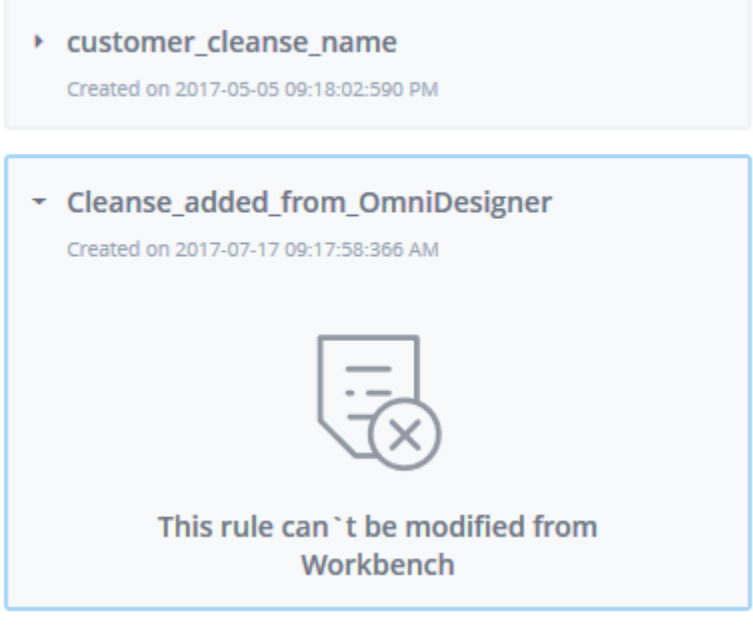
External rules

External rules are created and processed using the Omni Designer tool and deployed to OmniGen.

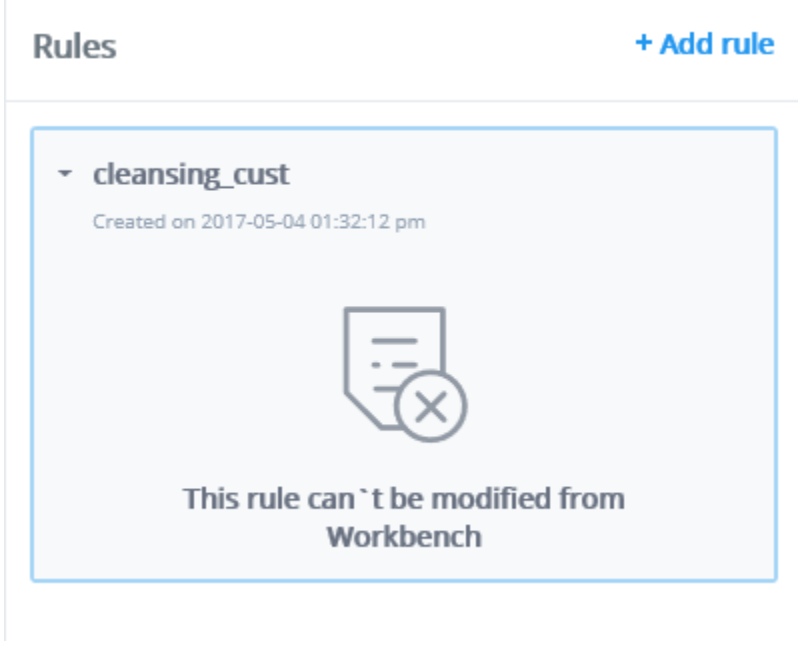


	Type	Name on Data Storage	Display Name	Description
1	Cleansing	customer_cleanse_name	customer_cleanse_name	
2	Matching	customer_match	customer_match	
3	Merging	customer_merge	customer_merge	
4	Remediation	customer_remediate	customer_remediate	
5	Cleansing	Cleanse_added_from_OmniDesigner	Cleanse_added_from_OmniDesigner	

You can view External rules with limited information, but you will not be able to edit them or add operations.

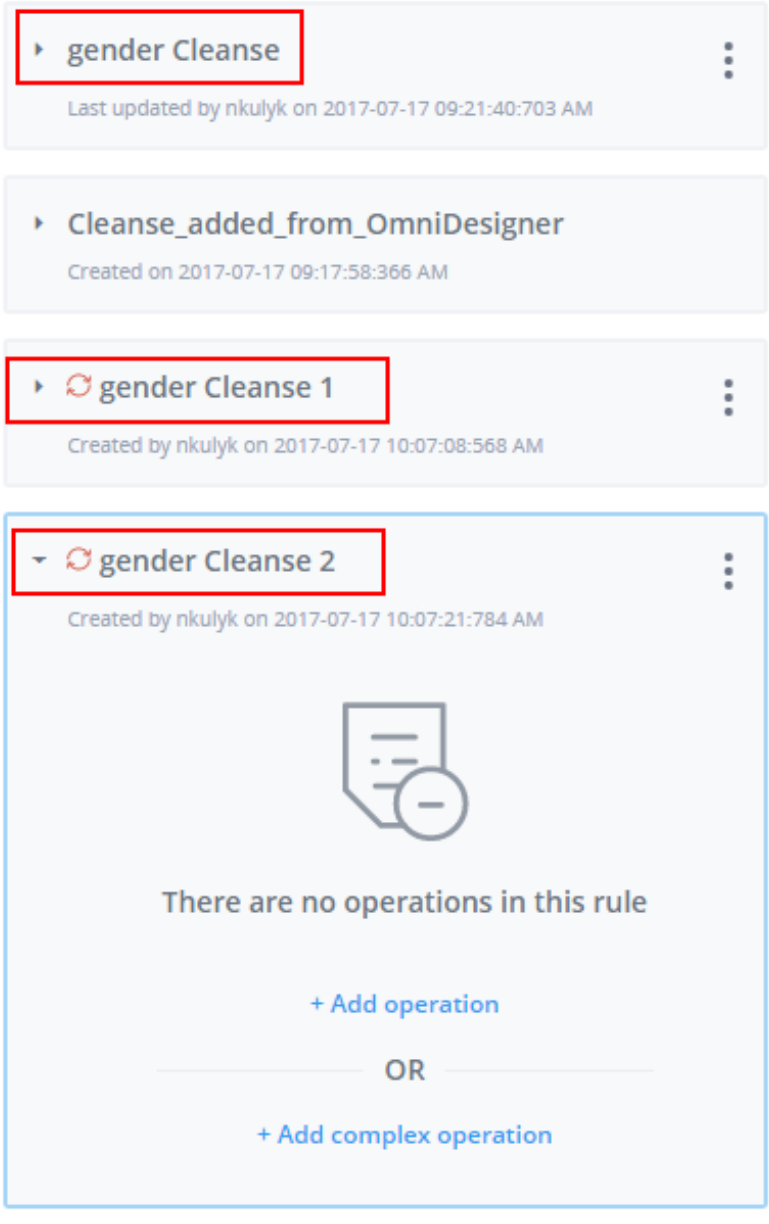


A notification message appears on the rule stating that rules cannot be modified. The Menu button is not available for such rules.

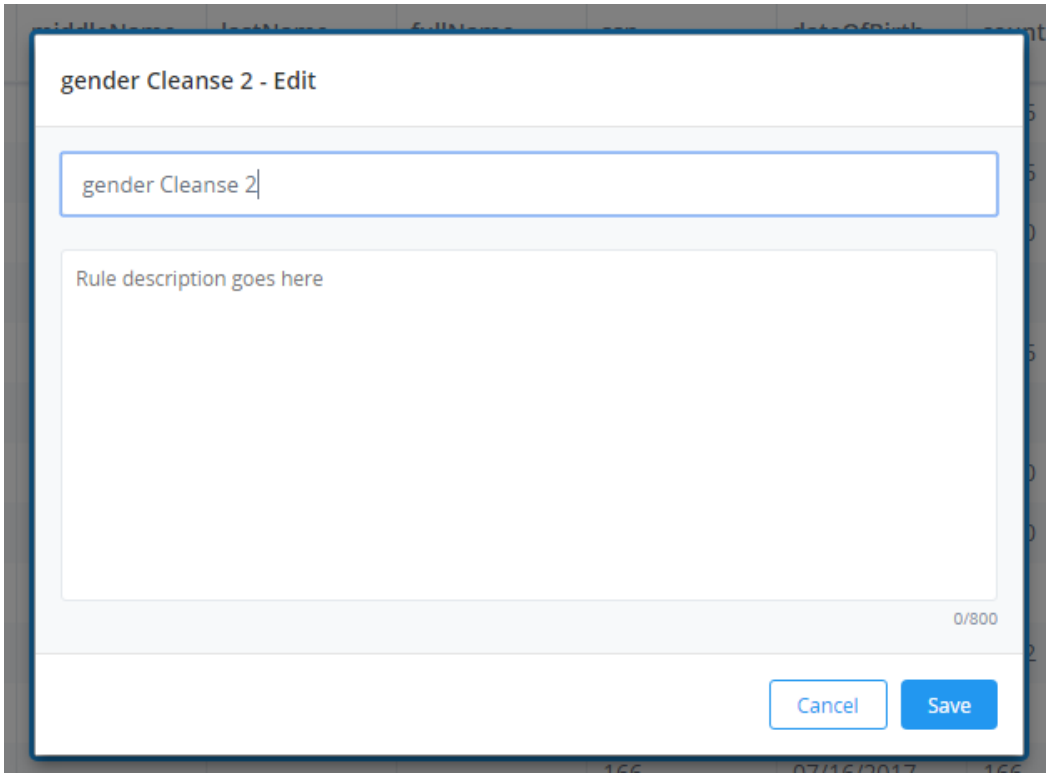


Standard Rules

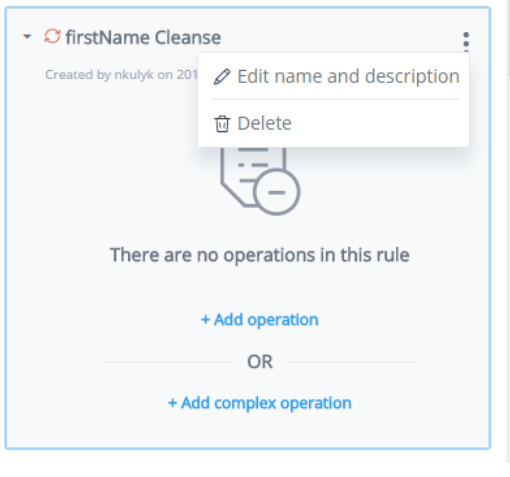
To add a new rule, select the column, expand the Rules panel and click the *Add rule* button. A new rule is created with a predefined name, for example: <*Column name*> *Cleanse* [number representing the last existing rule with this name +1].

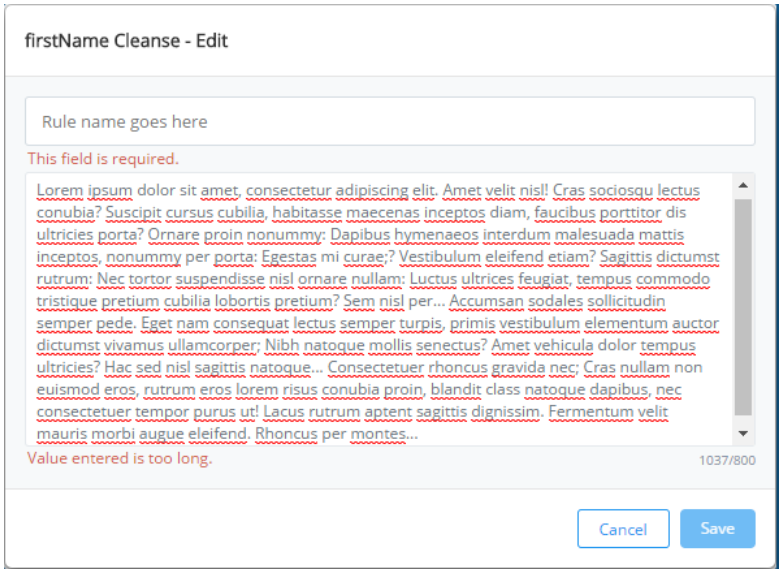


The Rule description is empty by default.

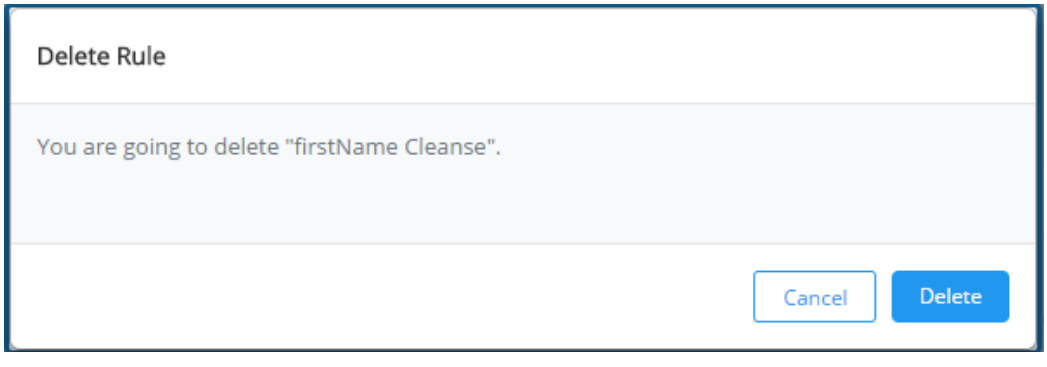


The Rule name and description can be edited. A simple validation is added to the rule edit mode.

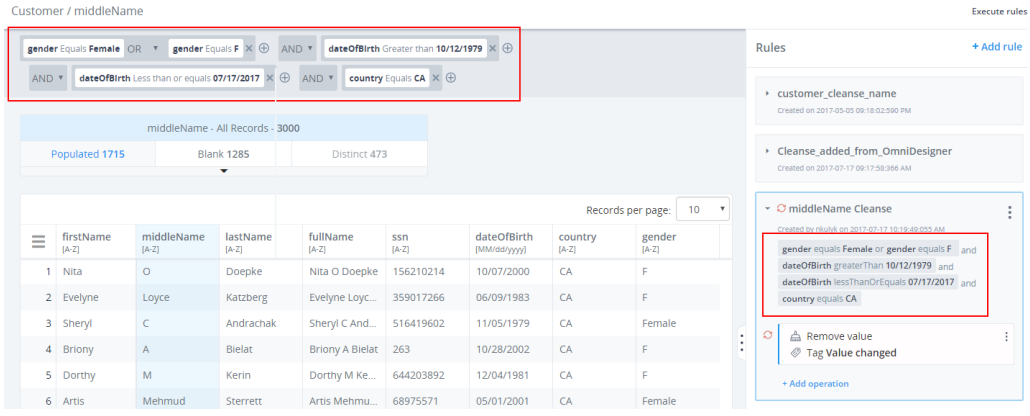




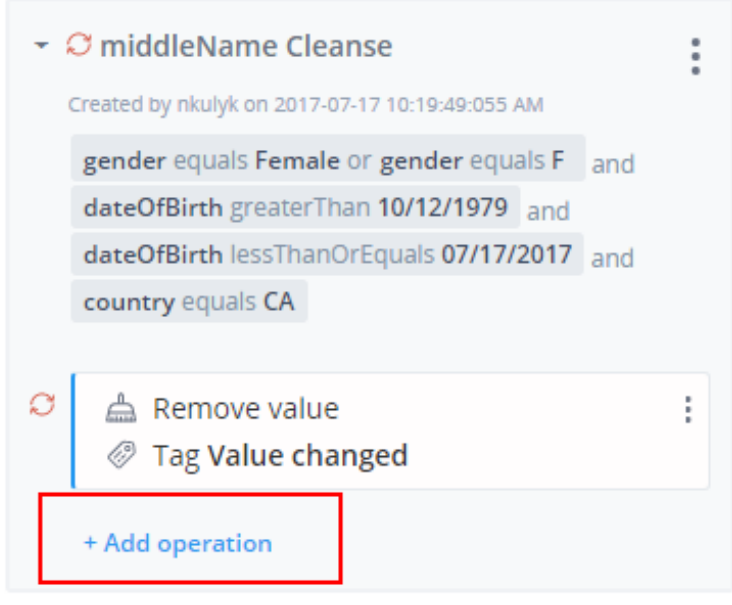
Rules can also be deleted. A confirmation dialog appears when you click *Delete*.

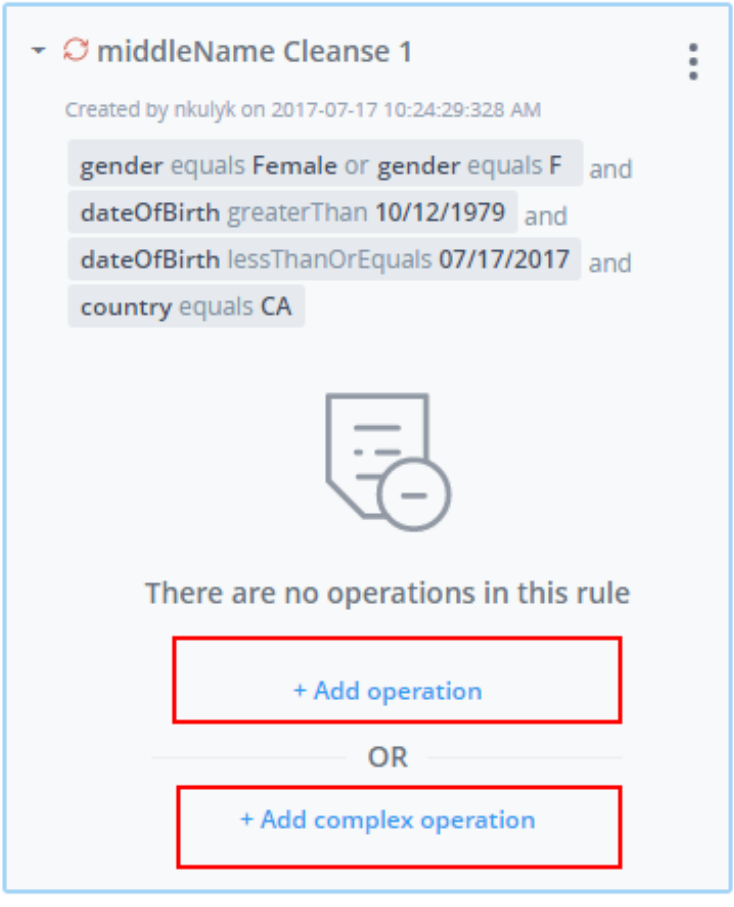


The rule inherits the global filter of the dataset that is added at the moment of adding the rule.

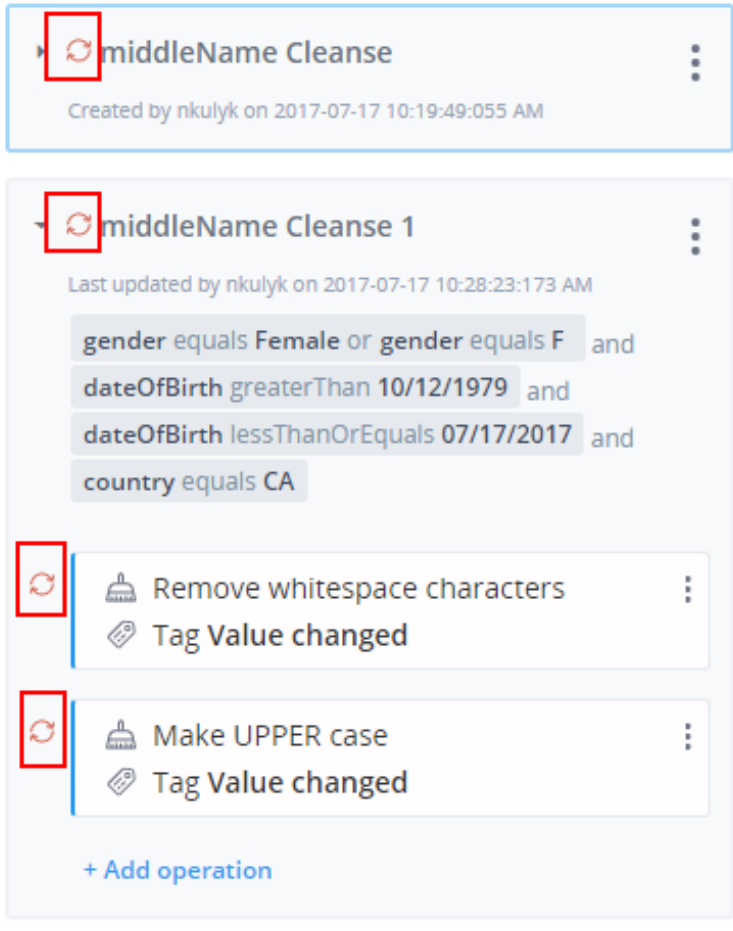


You can click *Add operation* to add a new operation.

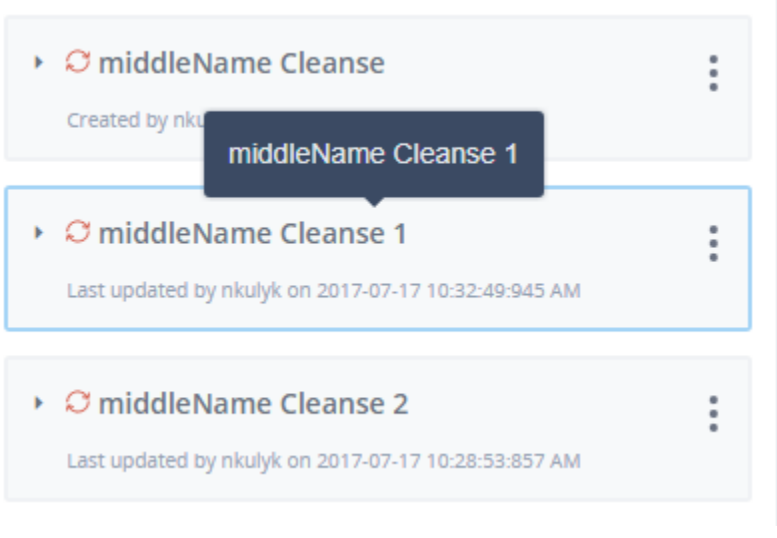


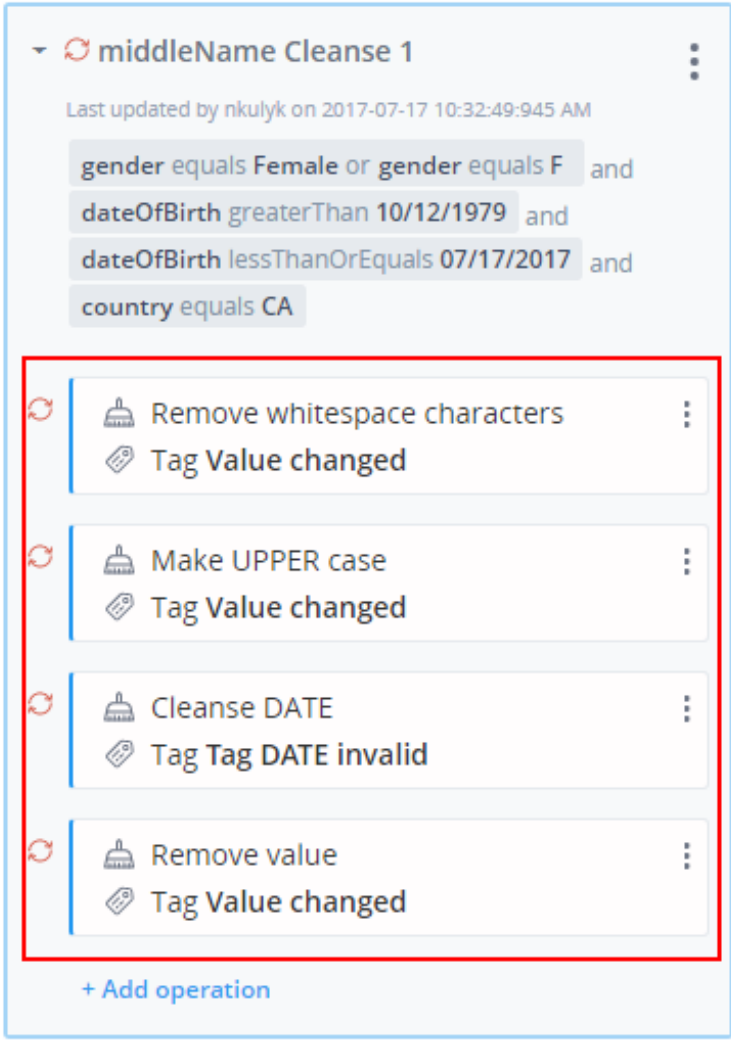


Once you select the rule or operation, it becomes highlighted. The new rule is selected as an uncommitted (orange color) icon as well as a newly added operation.


middleName Cleanse
Created by nkulyk on 2017-07-17 10:19:49:055 AM
middleName Cleanse 1
Last updated by nkulyk on 2017-07-17 10:28:23:173 AM
gender equals Female or gender equals F and
dateOfBirth greaterThan 10/12/1979 and
dateOfBirth lessThanOrEquals 07/17/2017 and
country equals CA
Remove whitespace characters
Tag Value changed
Make UPPER case
Tag Value changed
+ Add operation

Expand the rule to view a list of operations you can choose from to add to the rule.





If there are no operations added to the rule, the following message will appear:

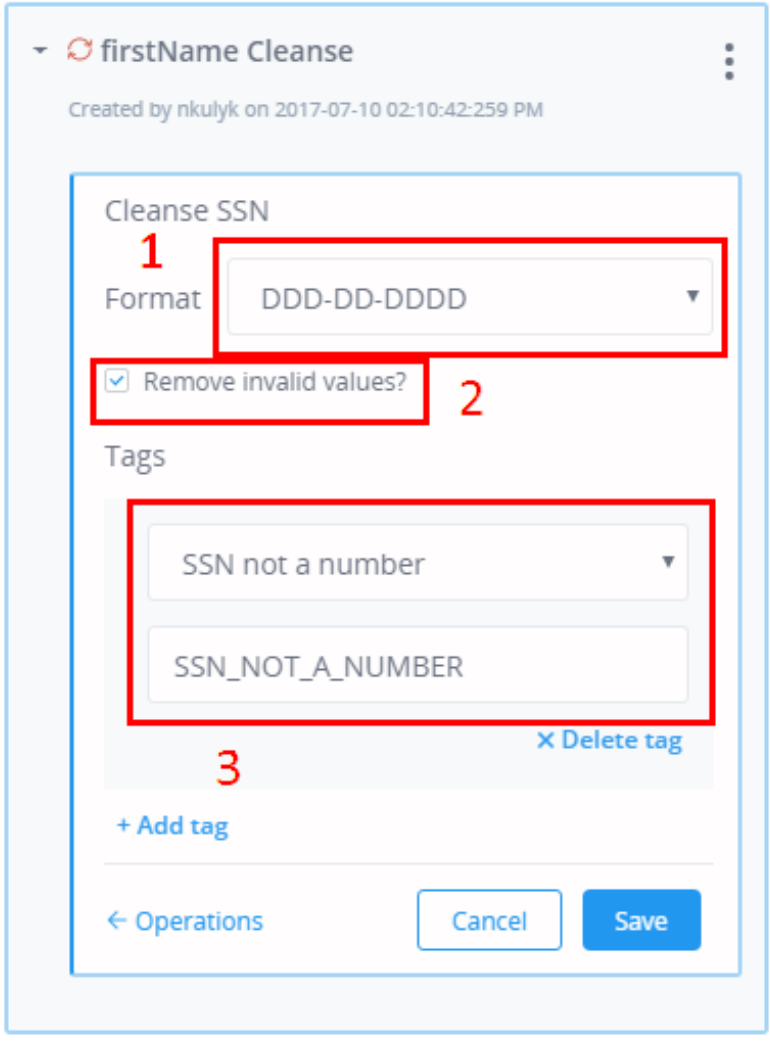
```
There are no operations in the rule.
```

Working With Operations

The Operation is a single action that allows you to customize the way you view data.

For example, the Uppercase operation turns all data in the selected column into Uppercase characters.

Each operation updates data for a selected column in certain way defined by options (for example, inputs, values, radio buttons, check boxes, and so on) which are set in corresponding XML files and help to distinguish and update data.



For example, the *Cleanse SSN* operation contains the following options:

1. **Format.** Indicates what format of data should appear after an operation is applied (nine Digits will be displayed in a specified way).
2. **Remove invalid values?** Select this check box if you want to remove data that does not satisfy a specified format (less or more than nine digits, characters, or special symbols).

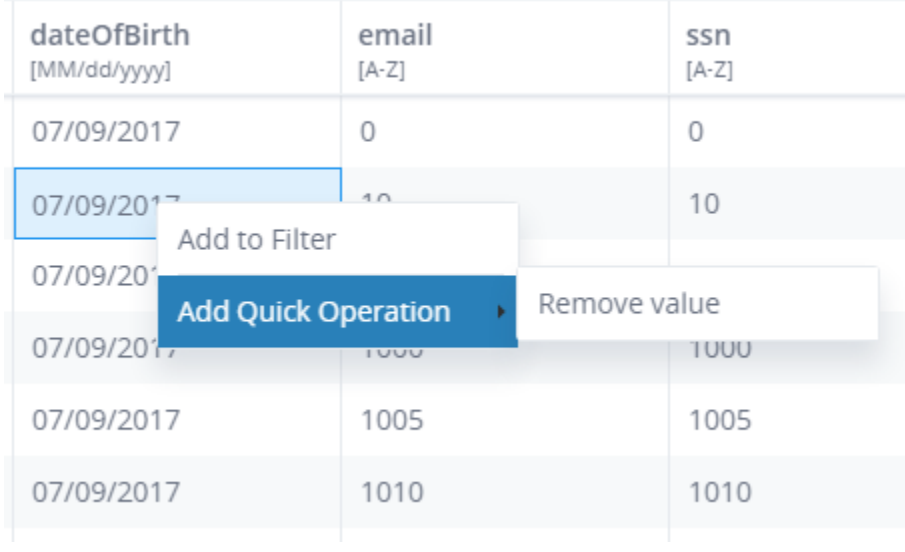
3. **Tags.** Added to data and characterizes it according to the set tag value.

The operation content corresponds to the column data type you select. You can set the parameters for any operation you choose.

You can edit the following parameter types:

1. **Behaviour parameters.** What the operation has to execute.
2. **Tags.** Signs off the execution of an operation.
3. **Output format.** A set of possible output format options for the content type.

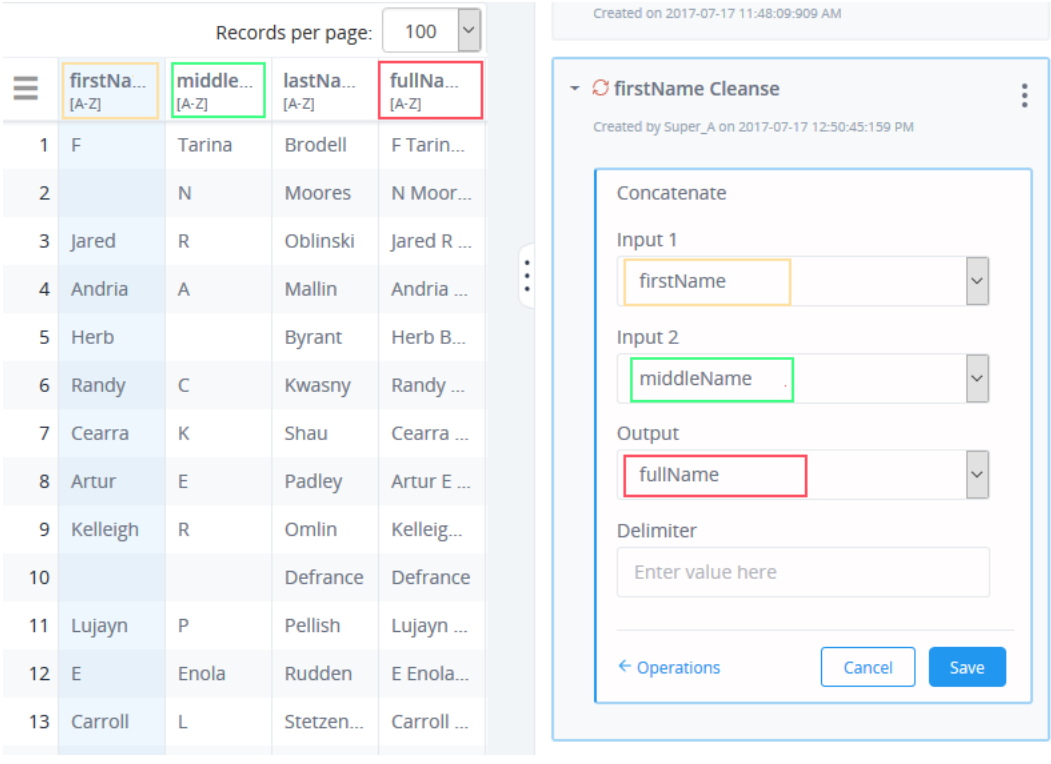
There are two ways to add an operation: add as quick operation from the table context menu and add from the rule.



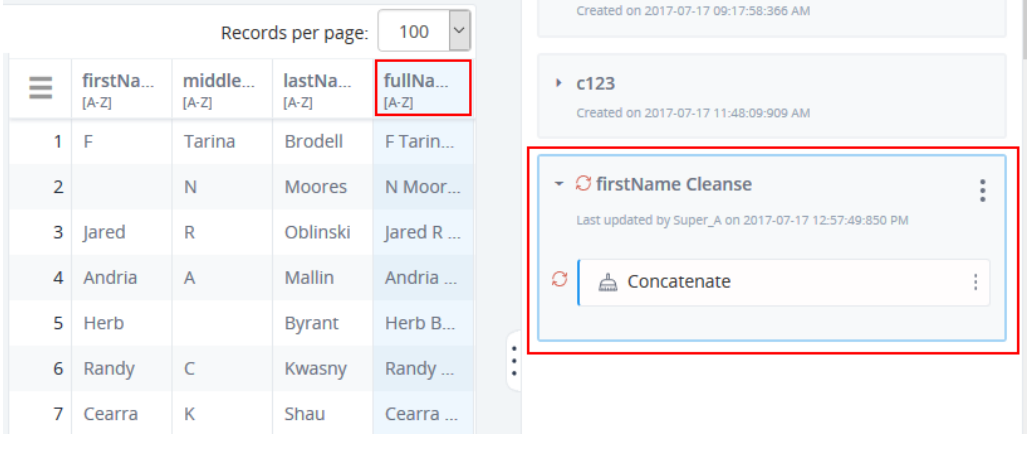
There are also two types of operations depending on the affected areas: Simple and Complex.

Simple operations are related to one Attribute and after execution of the rule will affect data only for the selected attribute.

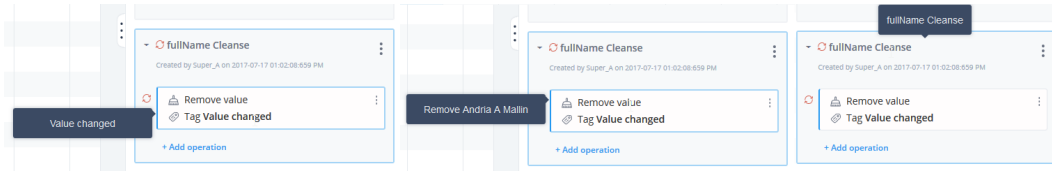
Complex operations involve several attributes and works with data from those attributes that are listed in the inputs and outputs of the operations.



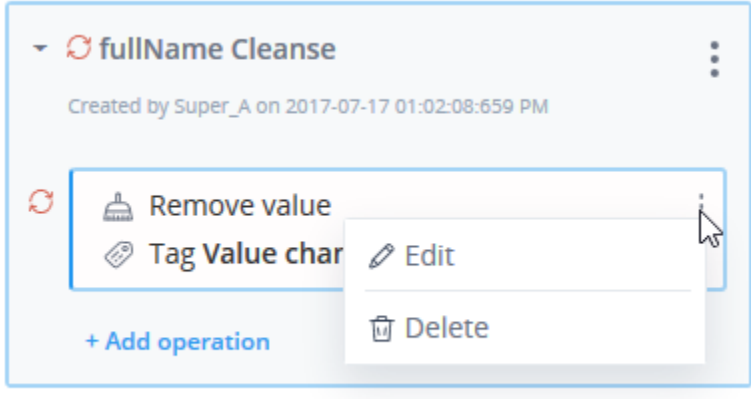
The rule that contains a complex operation will be created for each involved attribute.



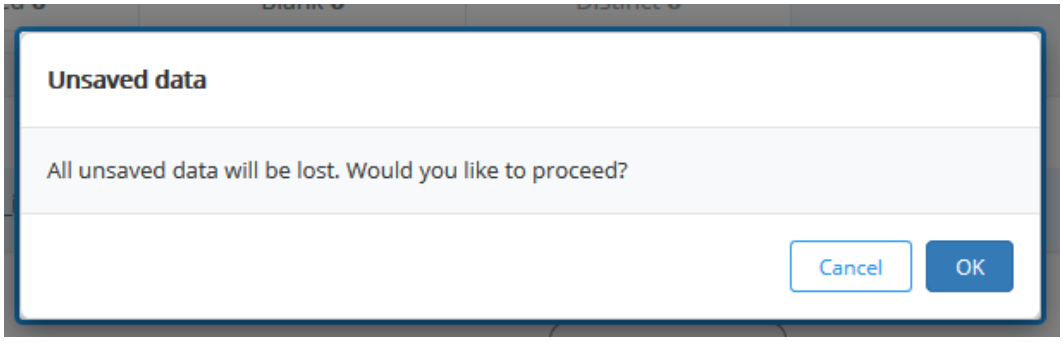
You can hover your cursor over the operation and see the name, parameters, and tags inside the corresponding tooltip.



Any operation can be edited and deleted.



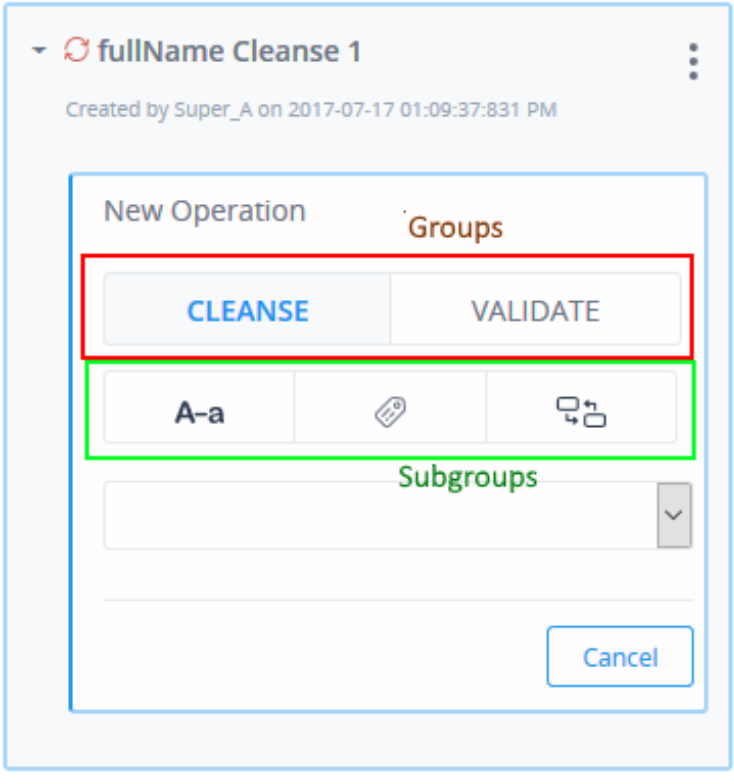
Only one operation per time can be opened in the Edit mode. If you try to open another operation or do another action while the operation is opened, you will be asked to save changes before that.



Adding Operations from the Rules panel

In order to add a rule from Rules panel, select either the *Add operation* or *Add complex operation* buttons.

The Operations are displayed by groups and sub groups, as shown in the following image.

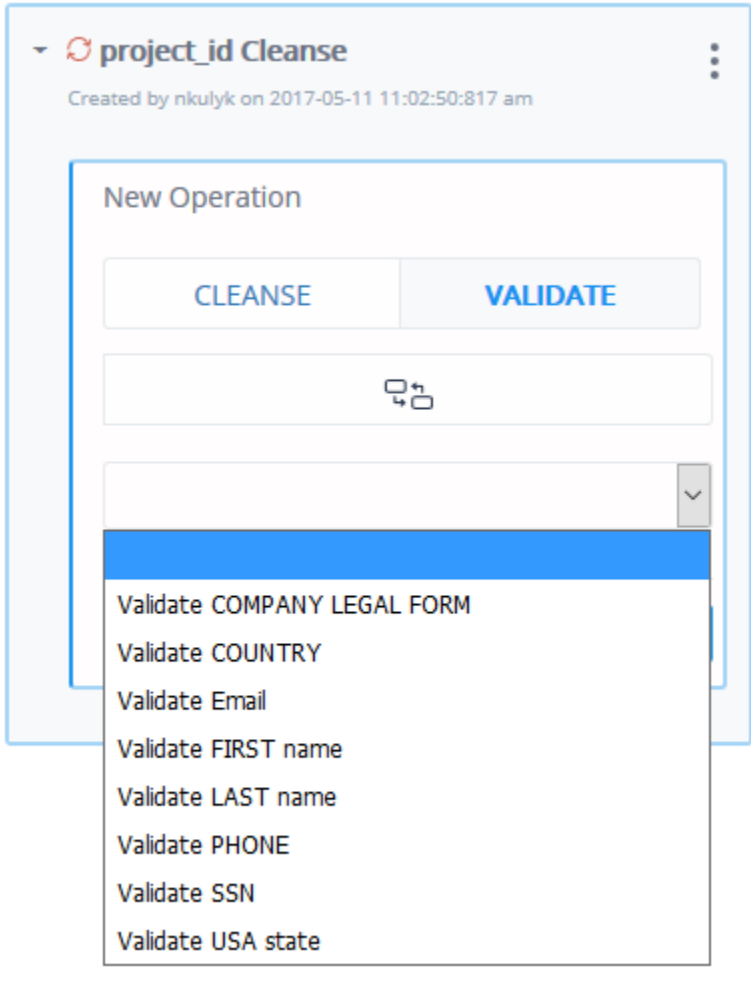


Currently, there are two groups:

- Cleanse
- Validation

The number of subgroups can be different. Groups and sub groups are configured in the corresponding XML file of the operation in a specific folder where the OmniGen system is installed.

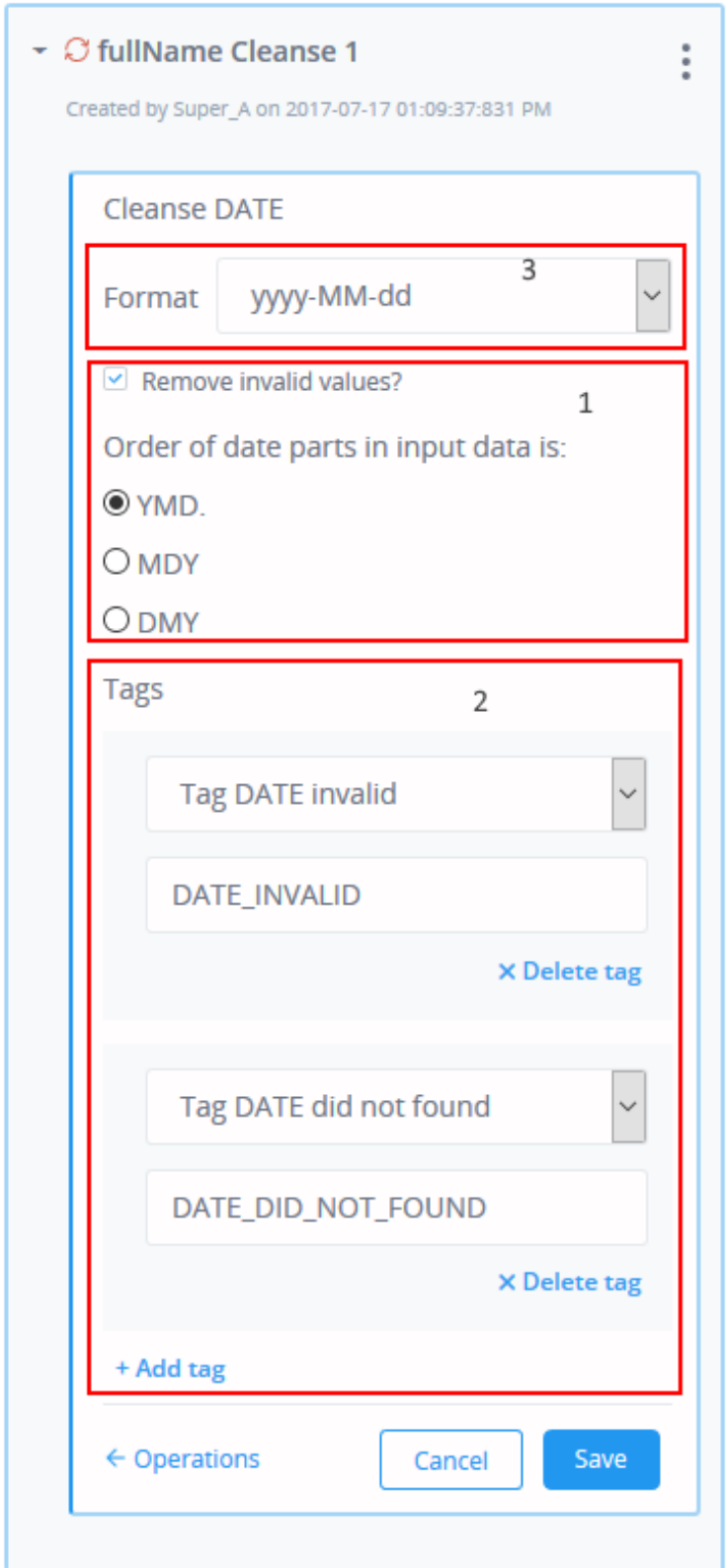
If number of operations is more than five, then the operations will be displayed in a drop-down list.



The operation content corresponds to the column data type you select. You can set the parameters for any operation you choose.

You can edit the following parameter types:

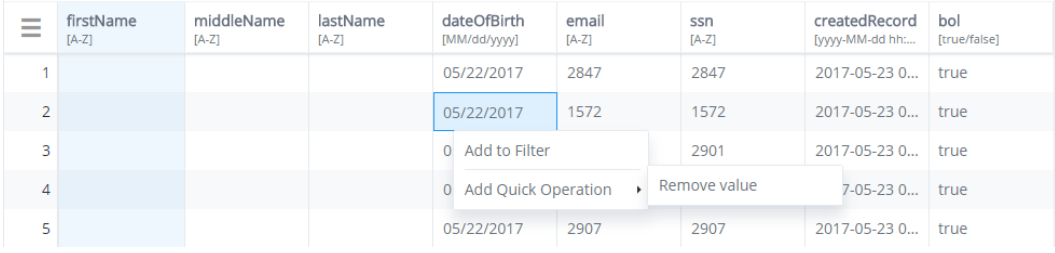
1. **Behaviour parameters.** What the operation has to execute.
2. **Tags.** Signs off the execution of an operation.
3. **Output format.** A set of possible output format options for the content type.


fullName Cleanse 1
Created by Super_A on 2017-07-17 01:09:37:831 PM
Cleanse DATE
Format
yyyy-MM-dd
3
Remove invalid values?
1
Order of date parts in input data is:
YMD.
MDY
DMY
Tags
2
Tag DATE invalid
DATE_INVALID
Delete tag
Tag DATE did not found
DATE_DID_NOT_FOUND
Delete tag
+ Add tag
Operations
Cancel
Save

For changes that were made with the operation/rule, they will be marked with an uncommitted icon.

Adding Quick Operations From the Data Table

You can add basic operations from the data table value through the context menu:



The mechanism of adding operation is the same as for the Rules panel. When you select the *Add Quick Operation* option, the Rules panel will be expanded and the operation will be added.

The following list describes the ways how quick operations will be added to a new or existing rule:

- The rule is NOT present in the rules panel or there are just external rules present.

  A new rule and the operation will be added to the new rule. The system automatically generates the name of the rule, for example <Column name>, Cleanse [number representing the last existing rule with this name +1], and empty description. The rule will inherit the Global Filter of the dataset that you used at the moment of adding the operation and the rule is automatically selected and highlighted (it means that user works with this rule at the moment).

- The rule is present in the rules panel.

  When the rule is selected:

  - If the filter of the rule matches the global filter, then the operation will be added to the SELECTED rule.
  - If filter of the rule does not match the global filter, then the latest rule in the rules panel will be verified. If the latest filter of the ruler matches the Global filter, then the operation will be added to the LATEST rule, if not, the operation will be added to the NEW rule.

  When the rule is NOT selected, then the latest rule in the rules panel will be verified:

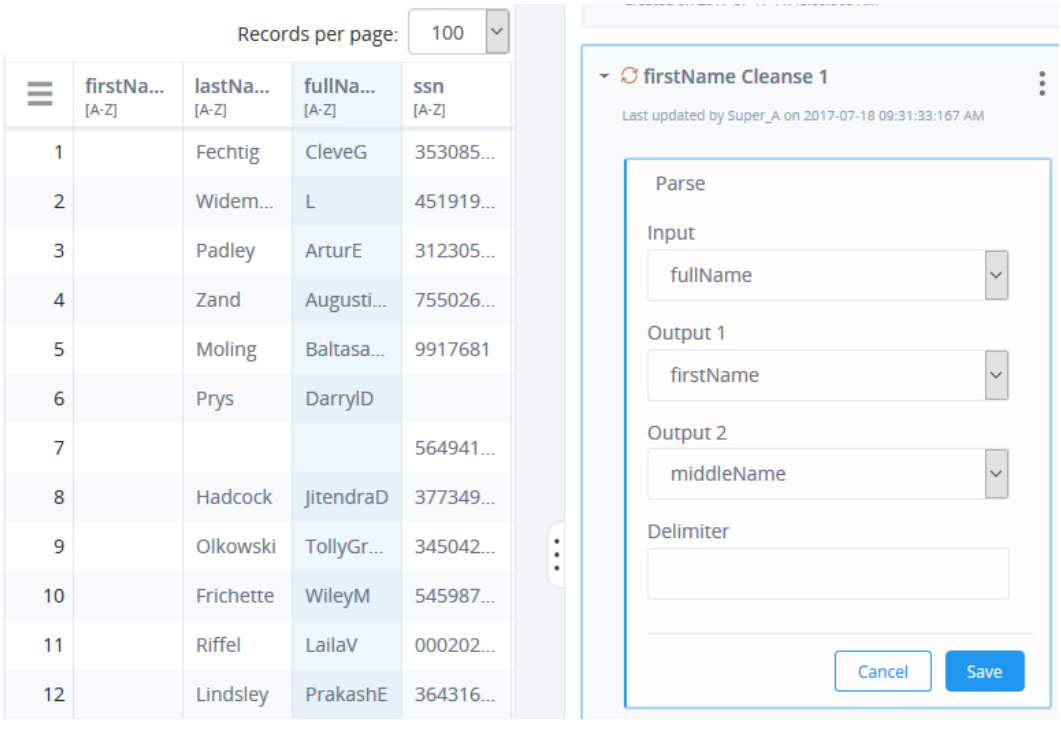
  - If the latest filter of the rule matches the global filter, then the operation will be added to the LATEST rule.

- If the latest filter of the rule does not match the global filter, then the operation will be added to the NEW rule.

You can add as many quick operations as you wish. The added operation and the rule will be marked as uncommitted.

Adding Complex Operations

Complex operations affect more than one input attribute (for example, First name, Middle name, and Last name). This type of operation can be added to every involved attribute as well as attributes that have been created.



Operations can be edited from every involved attribute.

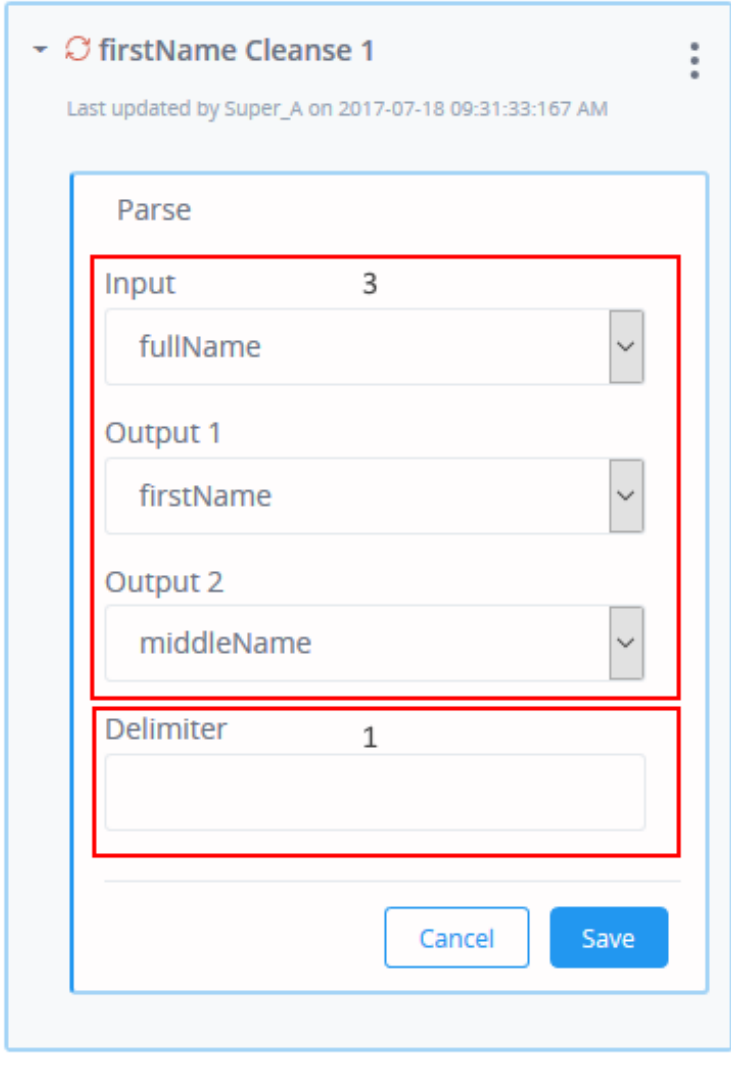
Only one complex operation can be added to one rule.

Complex operations cannot be added to the rules with standard operations.

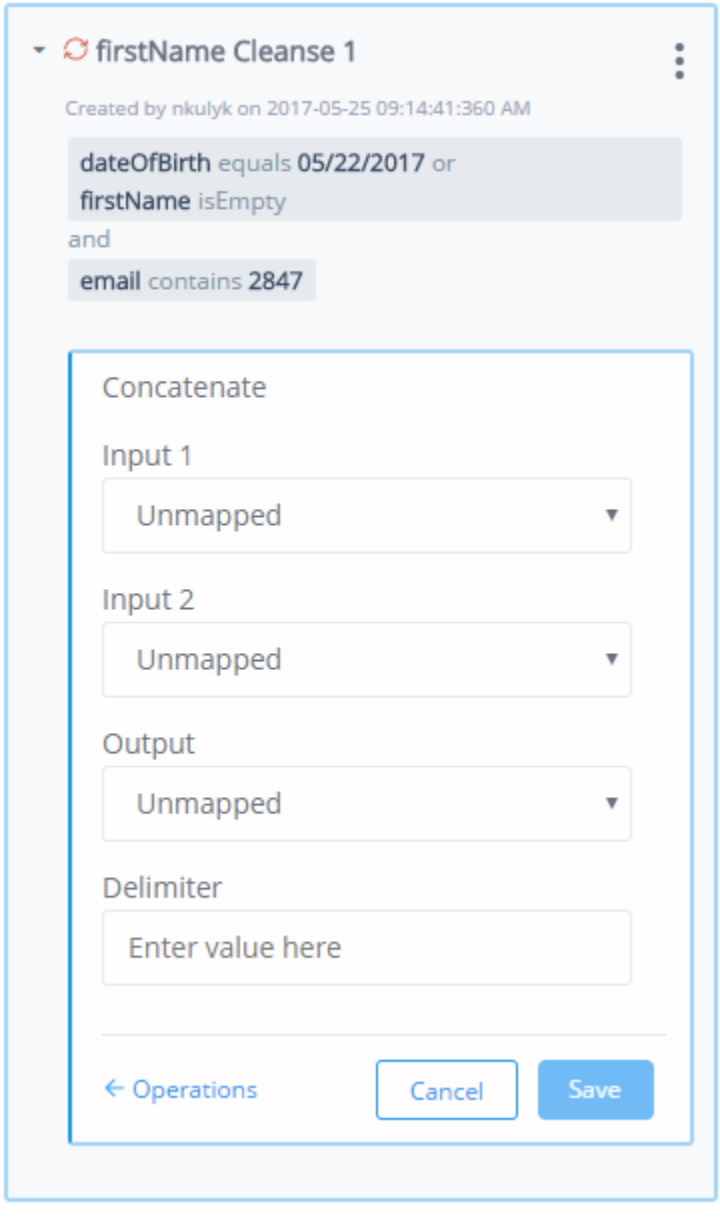
Operations can contain up to three sections:

1. **Behaviour parameters.** What the operation has to execute.
2. **Tags.** Sign-offs on executing the operation.

3. **Mapping**



By default, all input fields are unmapped, as shown in the following image.

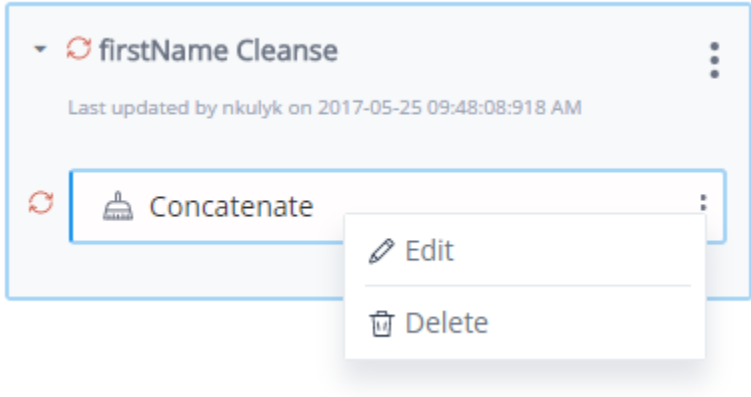


You have to map standard labels for the operation input attributes with particular examples for the data model (for example: First name= Name1, Middle name = Name2, Last name = Name3).

You can add only fields with the same data type as selected.

Any field can stay unmapped.

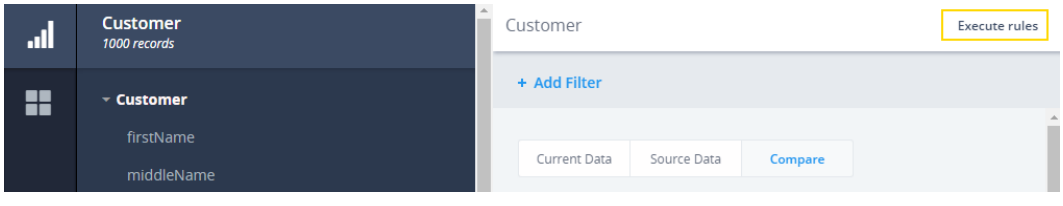
You can add as many Simple operations as needed, but only one Complex operation can be added.



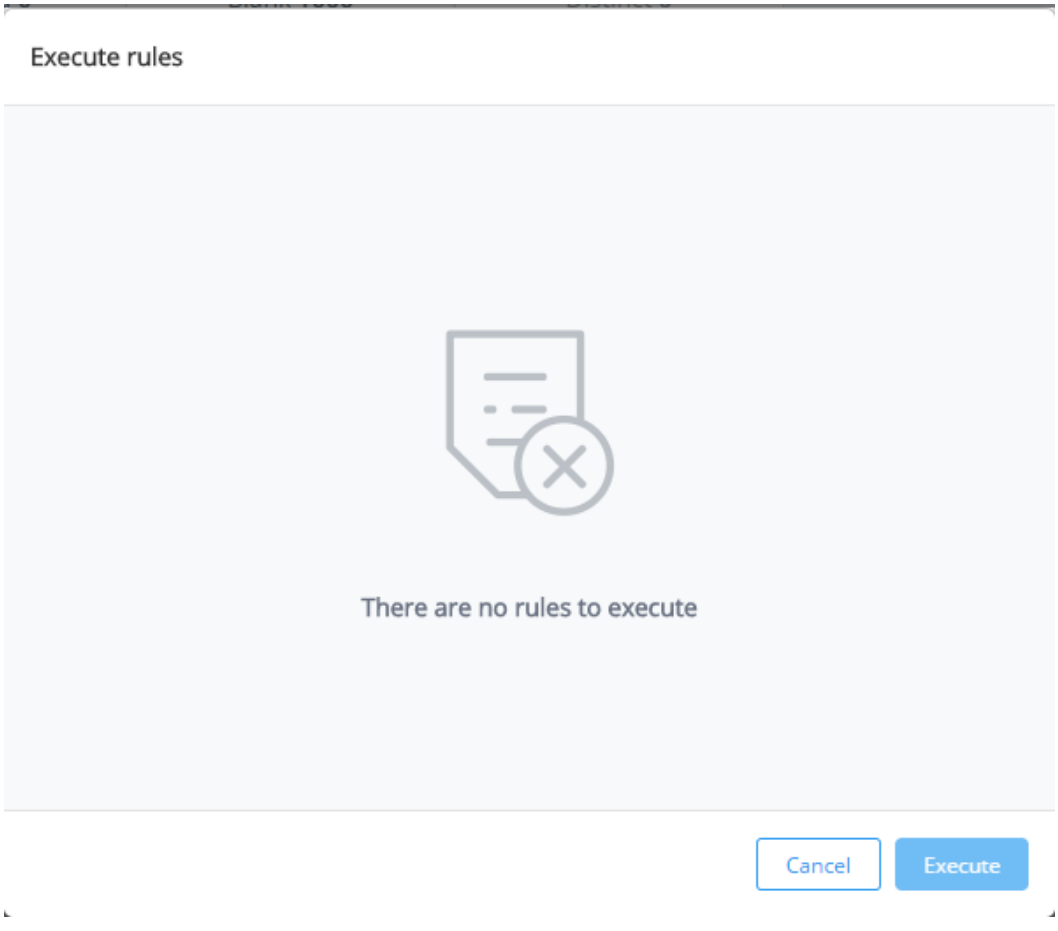
Processing and Executing Rules

If added rules are not executed, then the data will not be changed. To apply all changes, click *Execute rules*.

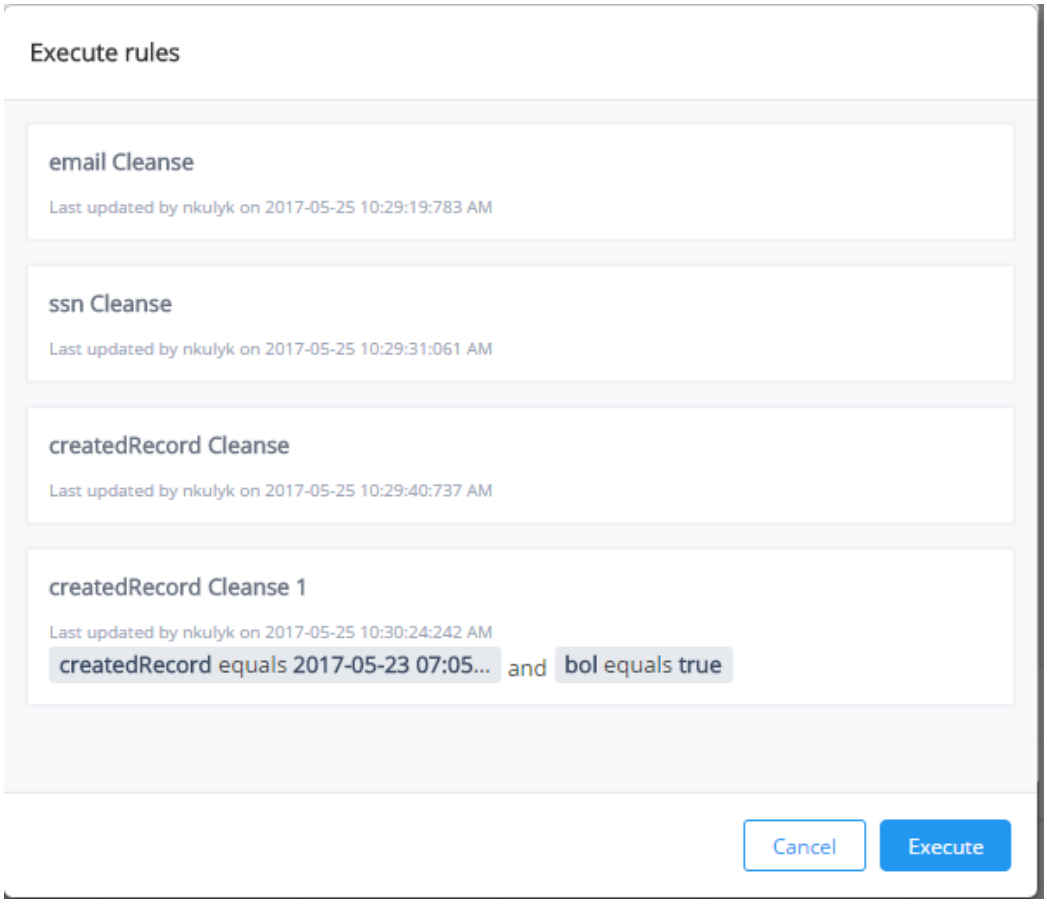
The *Execute rules* button is available for every Workbench page, as shown in the following image.



The Execute rules dialog appears. If there are no rules to execute, the corresponding message will be shown.



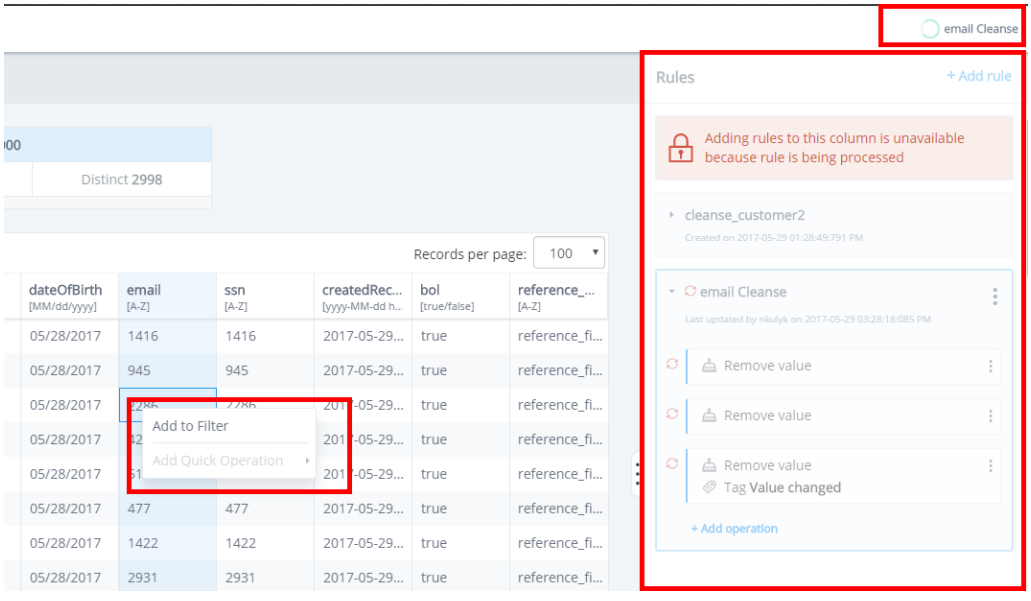
Every time rules are executed, all created rules will be processed but only those which have modifications will be displayed in the Execute rules dialog, as shown in the following image.



If there were any prior rules created and processed, and you have removed some rules but did not create or modify a new one, then all added rules will be displayed in the Execute rules dialog. Note that removing executed rules will affect the data.

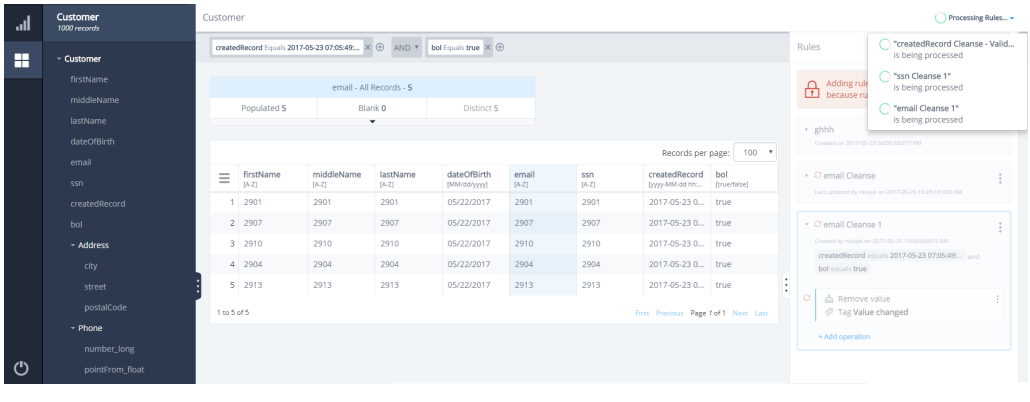
Click *Execute* to begin executing the rules.

Once Rule processing is started, the Rule panel for the selected column or several involved columns (for Complex operations) are blocked until the rule is processed. This also includes adding Quick operations for a selected column.

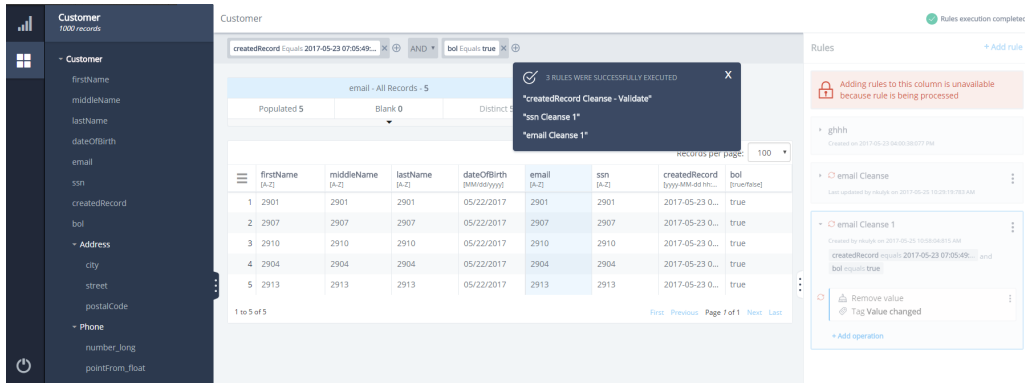


You can receive notifications to the reason behind why the panel was blocked. You can work with other columns and rules for background processing and create rules for other columns not involved while processing. However, the new rules cannot be executed until processing of the previous rules are complete.

While processing, a *Rule processing* indicator (loader) appears in the bread crumbs area. If only one rule is processed, its name will be displayed instead of the Processing rule. There are no limits for the number of rules that can run in the background simultaneously. If many rules are processed simultaneously, the *Rules processing* indicator will contain all rules and will be collapsed by default. You can expand it to see the detailed information about the rules which were processed:



When the rule processing is completed, a notification appears about the completed changes, and the Rules panel is unblocked, as shown in the following image.



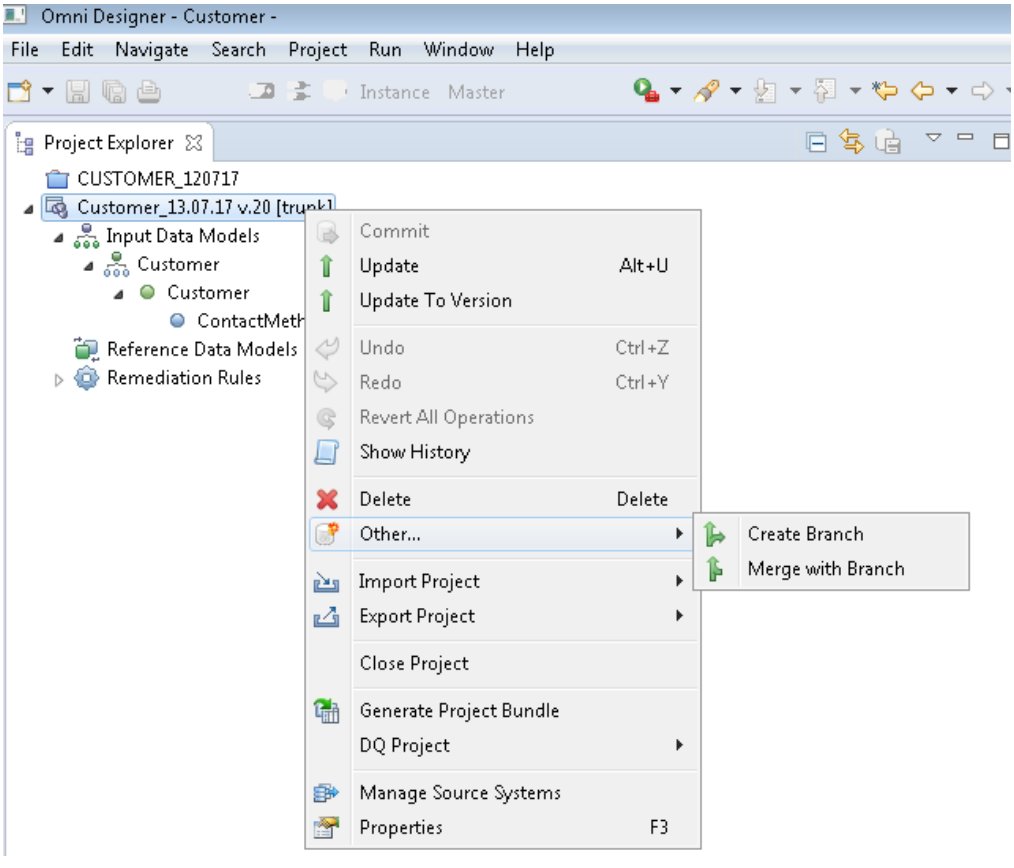
You cannot cancel or revert the rules process.

When rules are added from the Workbench interface, it will be visible to you only. Data affected by the executed rules will be also visible only to you. To have those rules and applied data visible to everyone, the project must be re-deployed and installed to the Omni Console.

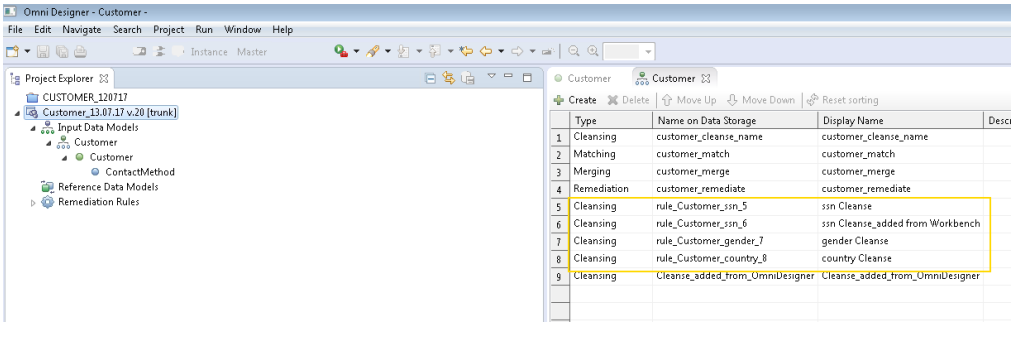
Procedure: How to Re-deploy the Project

You can add rules with the required operations from the Workbench and execute them. To re-deploy the project:

1. Navigate to the workstation where OmniGen is installed using the OmniDesigner tool (eclipse) and select the project from the Navigation panel (the project that was uploaded to the OmniConsole).

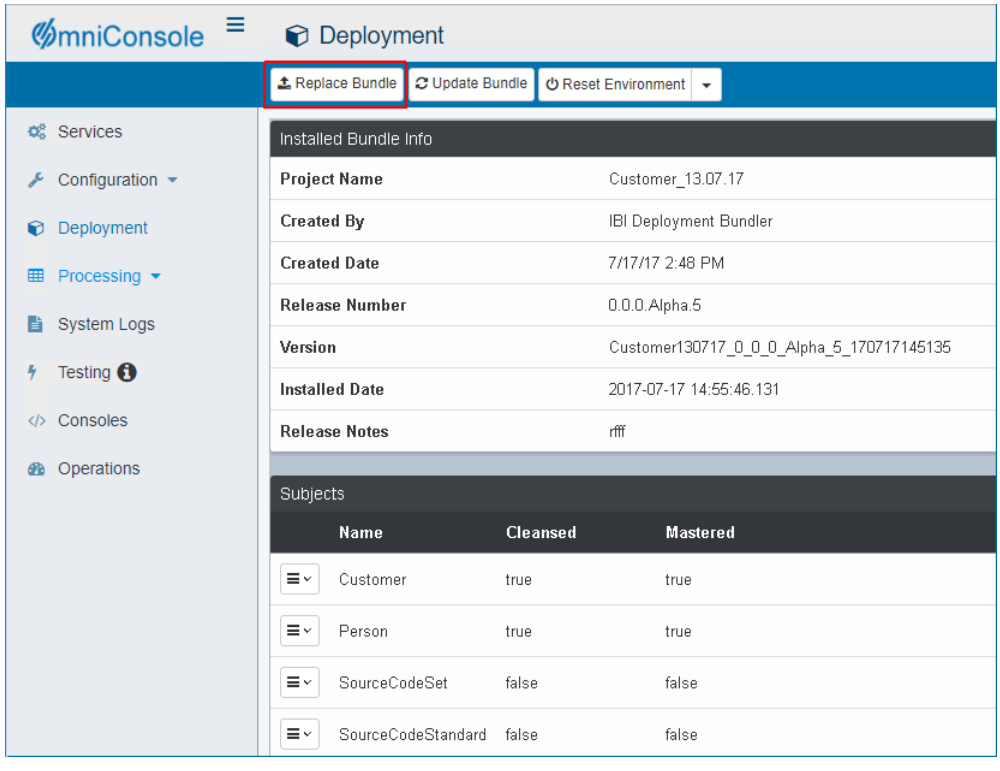


2. Right-click on the project, select *Other*, and then click *Merge with Branch* and find the branch you wish to merge with, from the list that appears.

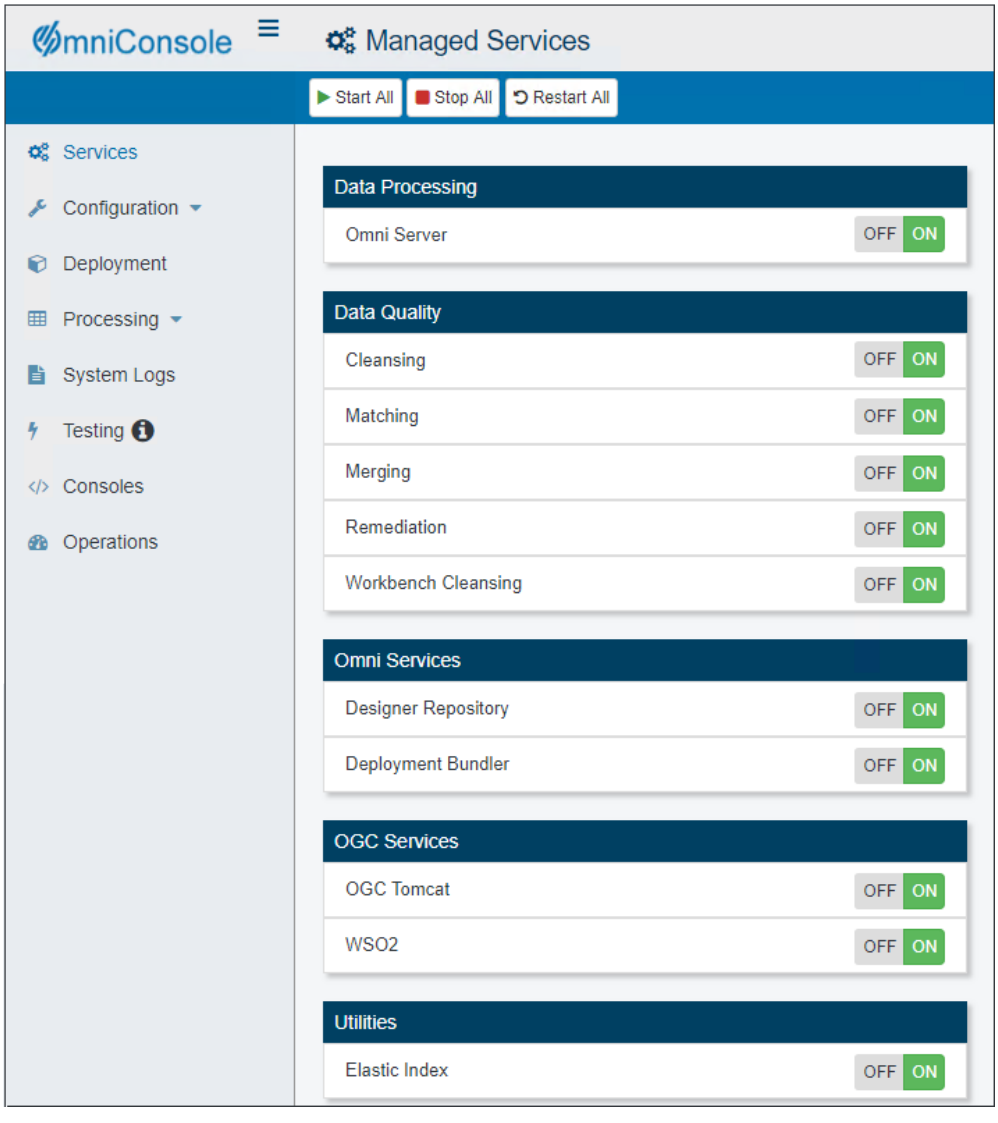


3. Right-click on the project and click *Commit*.
4. Right-click on the project again and click *Generate Project Bundle*.

5. Navigate to the Deployment Generator tool in your browser, for example: *https://<host>:9554/*.
6. Refresh the page and select just the generated Release bundle or the Generate Deployment bundle from the project (the green triangle button).
7. Navigate to the OmniConsole, select *Deployment*, and then click *Replace Bundle* (you can find it through the deployment bundle destination mentioned in og_configuration.properties, for example, *deploytool.dest.dir=C:/OGS/omnigen*).

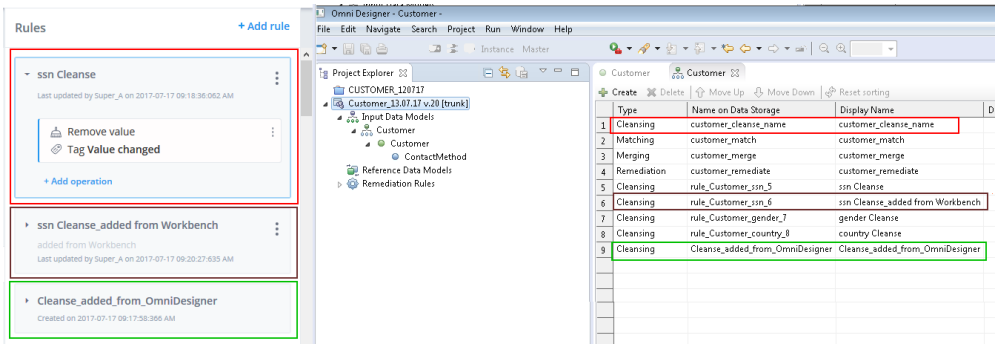


8. Start all OmniGen processes (*https://<host>:9600/Services*) and generate data using the DQS tool.

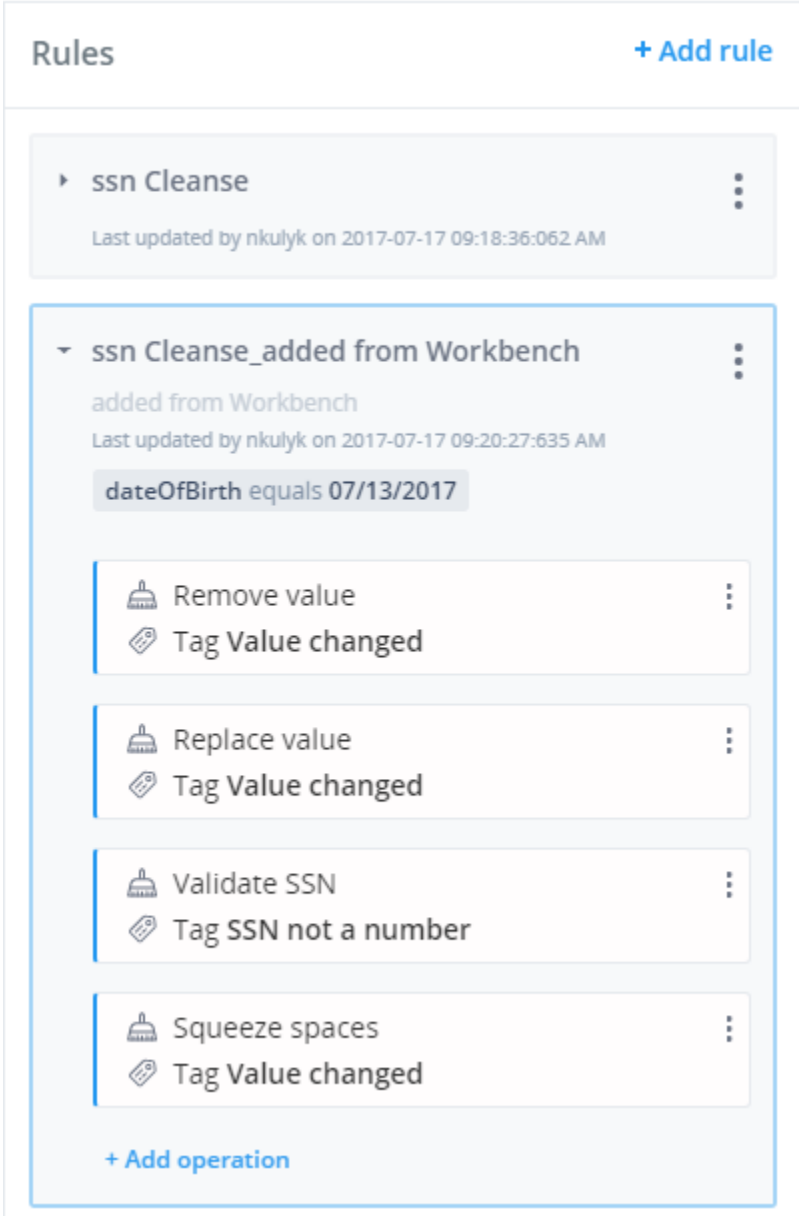


9. Restart the Workbench service and the OGC Tomcats.
10. Navigate to the Workbench and verify that the added Cleanse rules are available..

    The deployed rules should not have uncommitted icons.

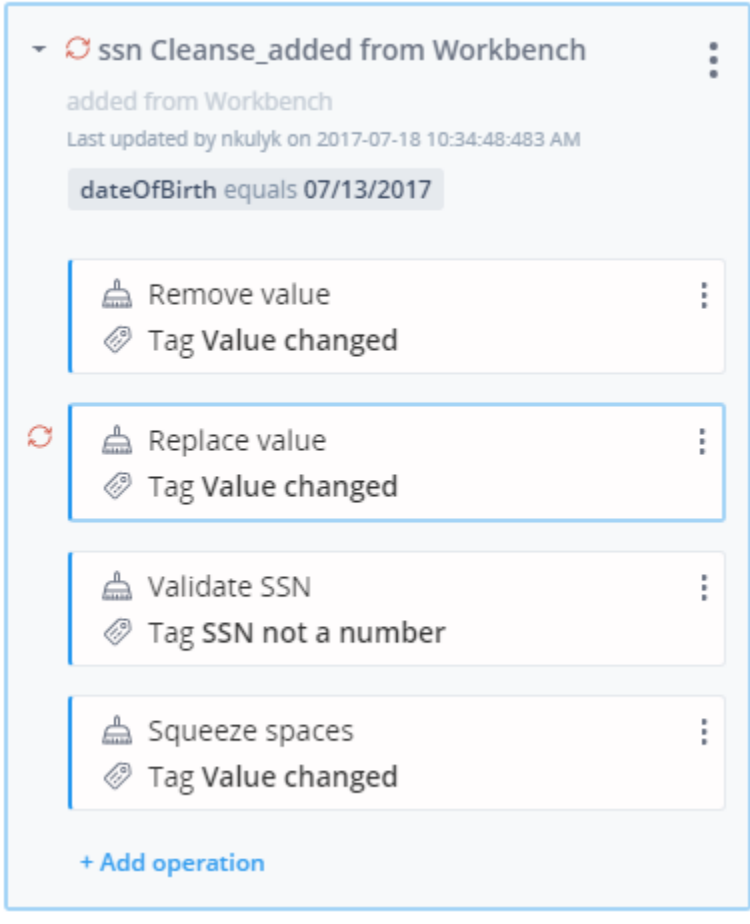


The operations for the deployed rules should also not have uncommitted icons

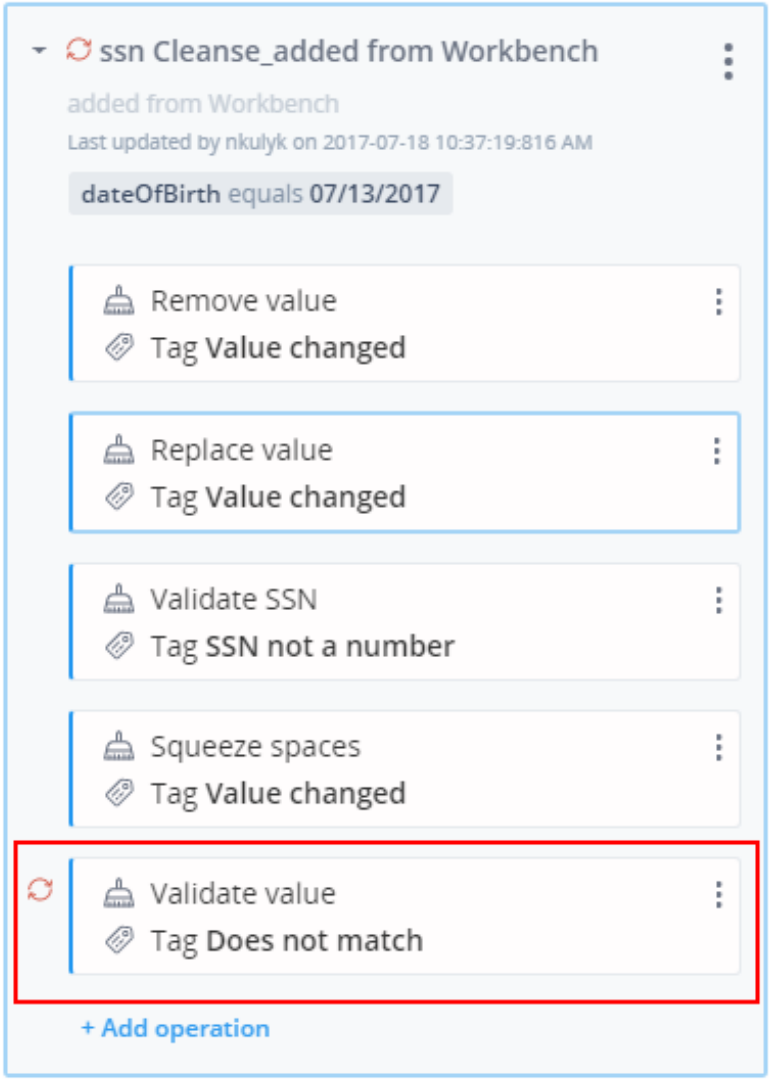


The deployed rules can be edited and re-executed in the same way as the new rules that were created from the Workbench.

After the operations or rules are updated, the uncommitted icon will appear:



If a new operation was added to the deployed rule, it will be marked with an uncommitted icon as well as the deployed rule, as shown in the following image.

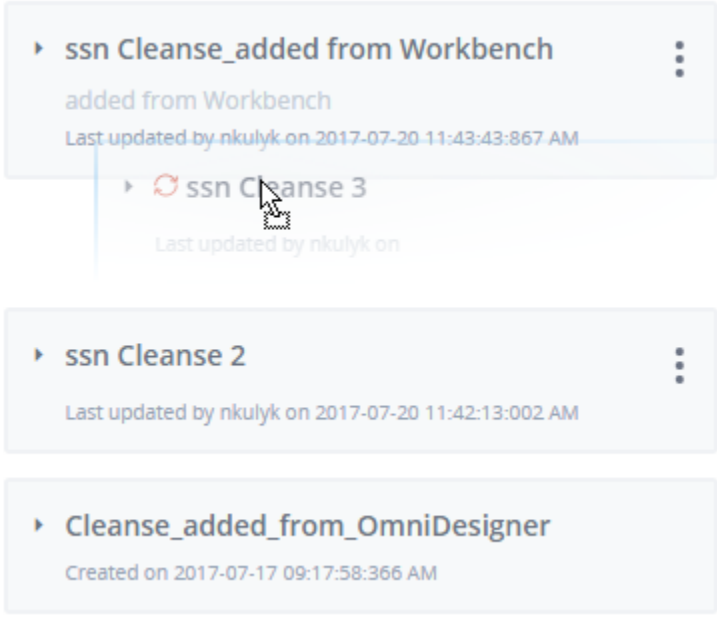


Reordering Rules and Operations

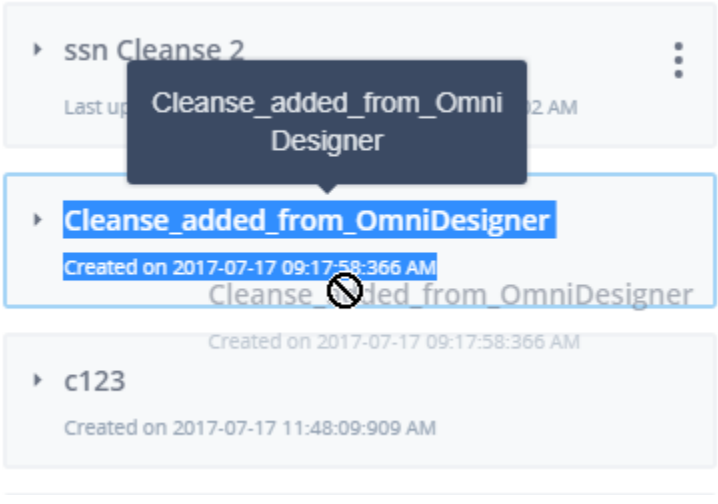
Reordering rules and operations allows you to change the position of Rules and Operations in order to execute them in another sequence.

Dragging and Dropping Rules

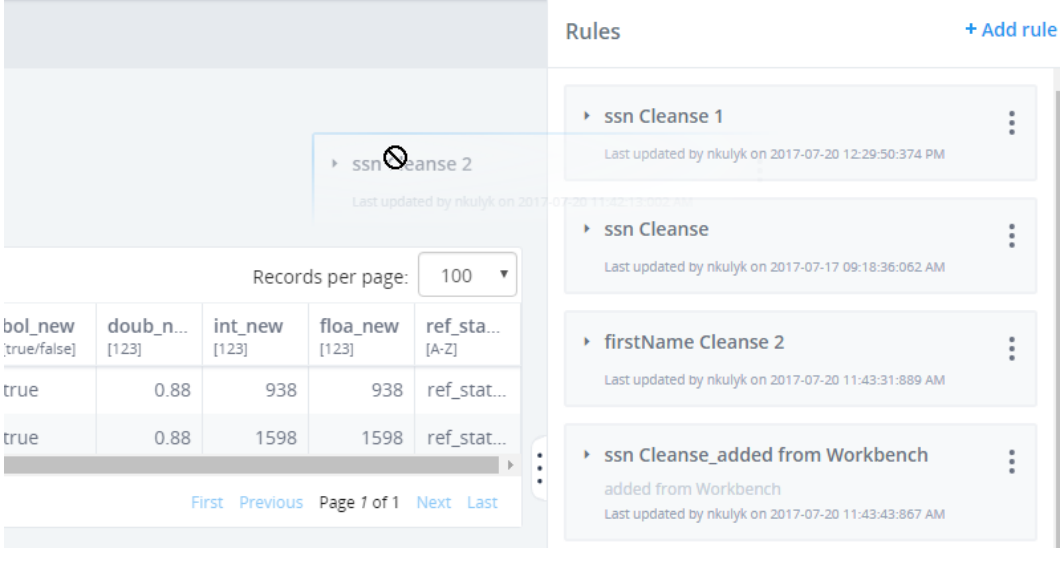
To move the rule to another area, select the rule, and then drag and drop it to any area you wish. A special icon appears for the Rule that is being dragged.



Only new and deployed rules can be moved. However, you will not be able to move External rules (created from Omni Designer tool).

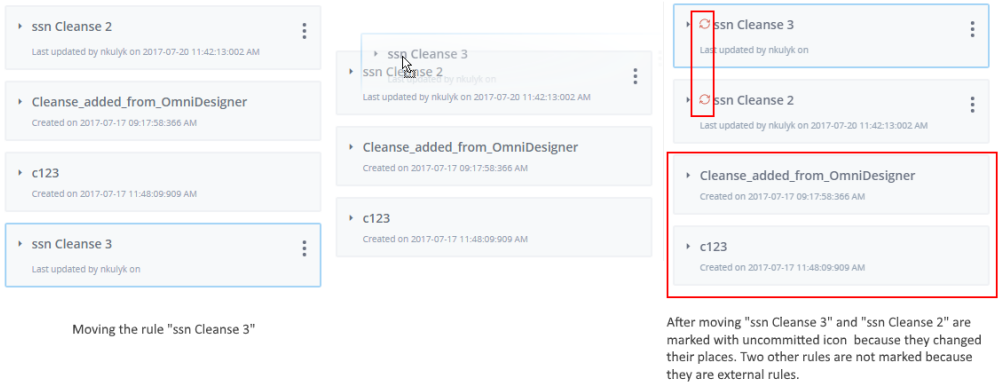


Rules can be moved to any location in the Rules panel, but you cannot drop the rule outside of it.



Rules can be moved even if they are either expanded or collapsed.

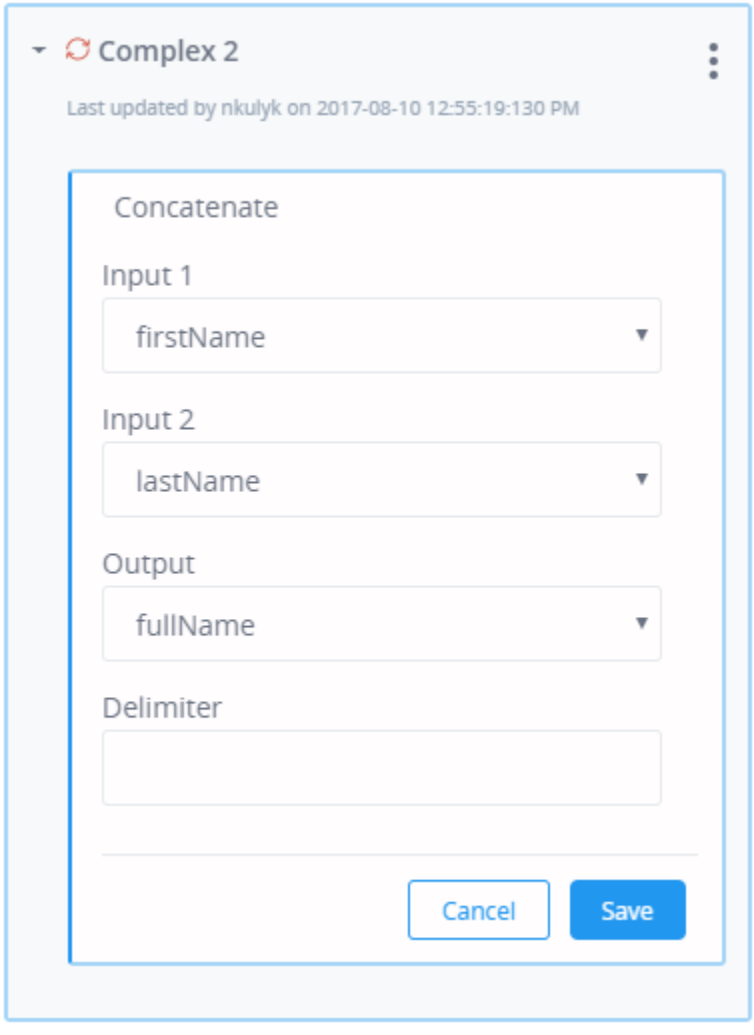
When the rule is moved, all other remaining rules in the rules panel will be marked with an uncommitted icon. External rules will not be marked.



Moving the rule "ssn Cleanse 3"

After moving "ssn Cleanse 3" and "ssn Cleanse 2" are marked with uncommitted icon because they changed their places. Two other rules are not marked because they are external rules.

Reordering rules with simple operations will not affect the list of rules for other columns since simple operations are related to single attributes, except in a few situations when rules contain complex operations that are related to several attributes.



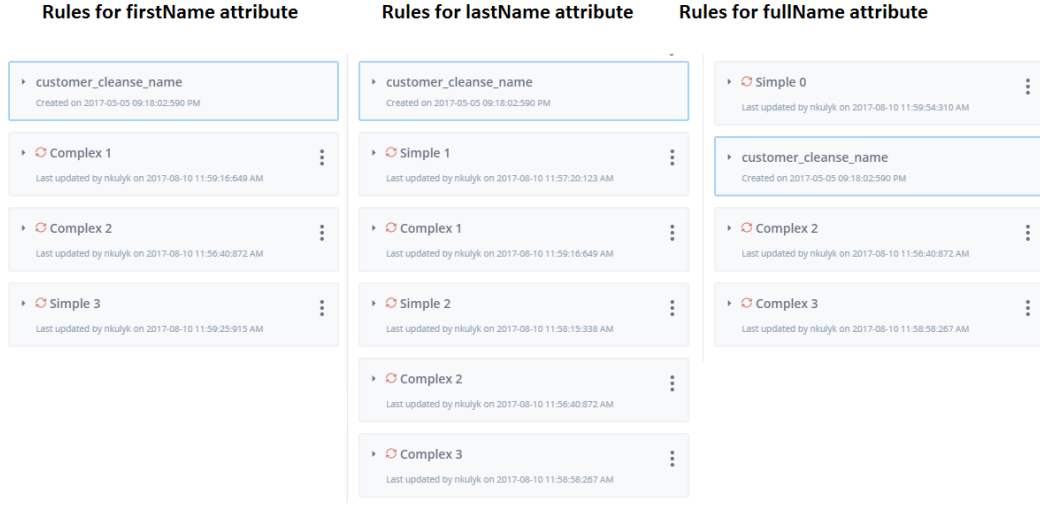
For example, if *multicolumn rule* was a complex operation with multiple columns and rules, reordering them will impact the order of the rules for other attributes.

Reordering Multicolumn Rules

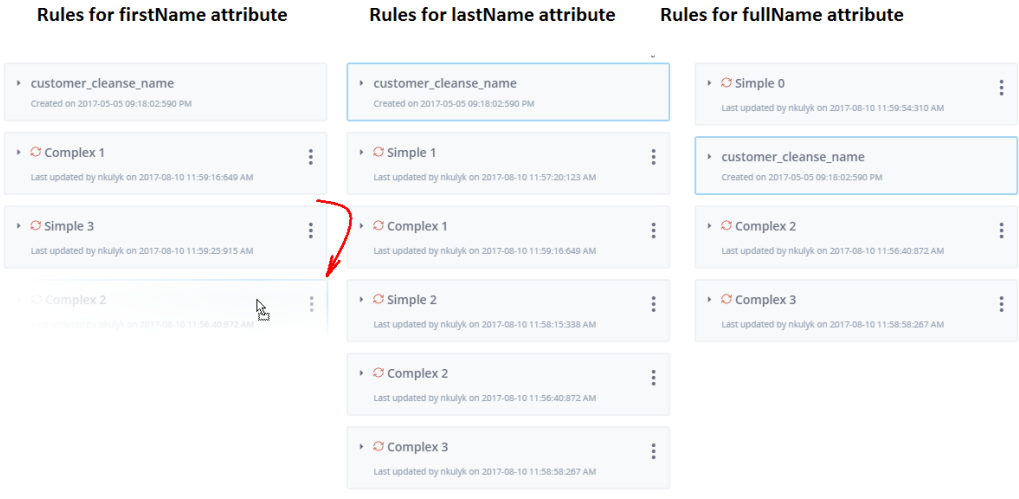
Rules containing a Complex operation affects several columns mentioned within the complex operation. There are several main principles:

- The rule containing the Complex operation is moved below the rule with simple operations and will not affect the position of the rule in other columns, as described below in Example 1.
- The rule containing the Complex operation is moved below the rule with complex operations and affects the position of the rules in other columns, as described below in Example 2.
- The Rule containing the Complex operation is moved below the external rule, and the rule with complex operations will affect the position of the rules in other columns, as described below in Example 3.

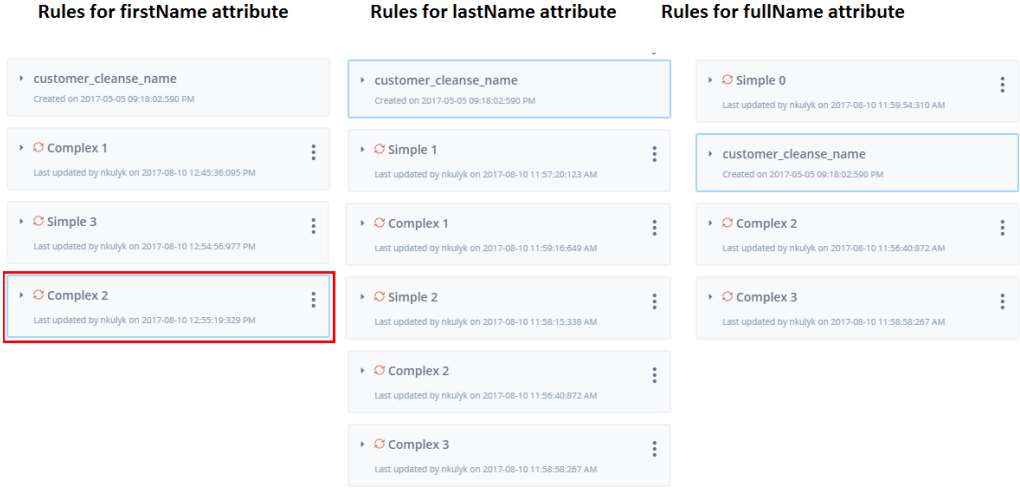
The following lists shows how the lists of rules for the three attributes appear before reordering:



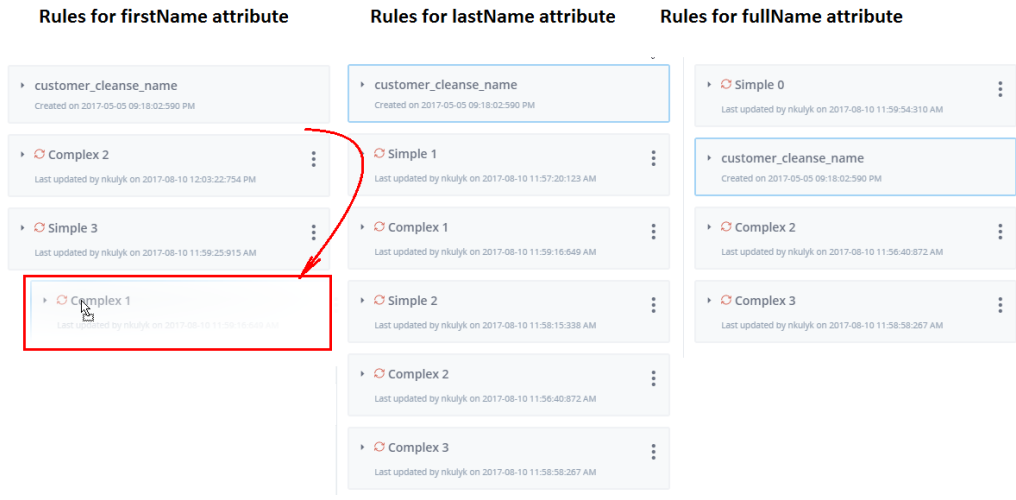
Example 1. The following image shows the *firstName* attribute where a complex operation rule called *Complex_2* was moved below a rule with a simple operation called *Simple_3*.



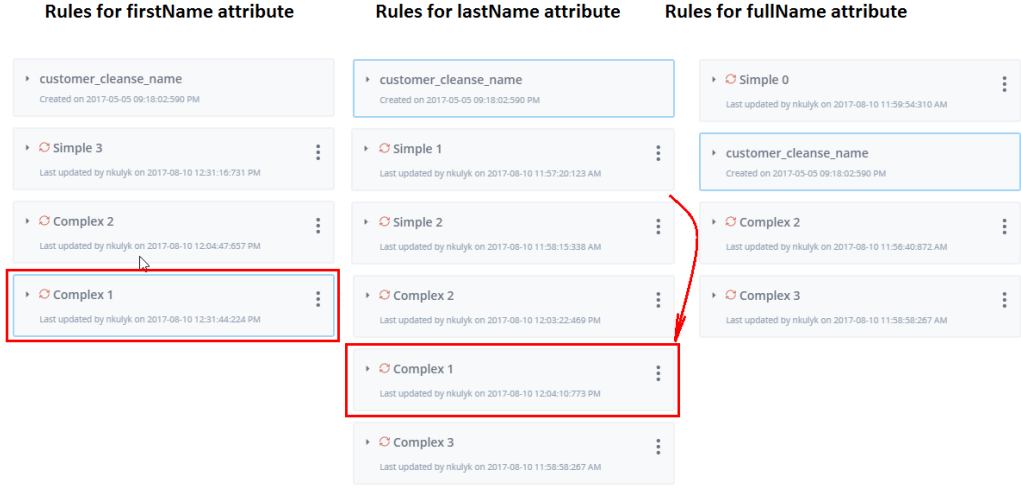
Only the firstName column is affected:



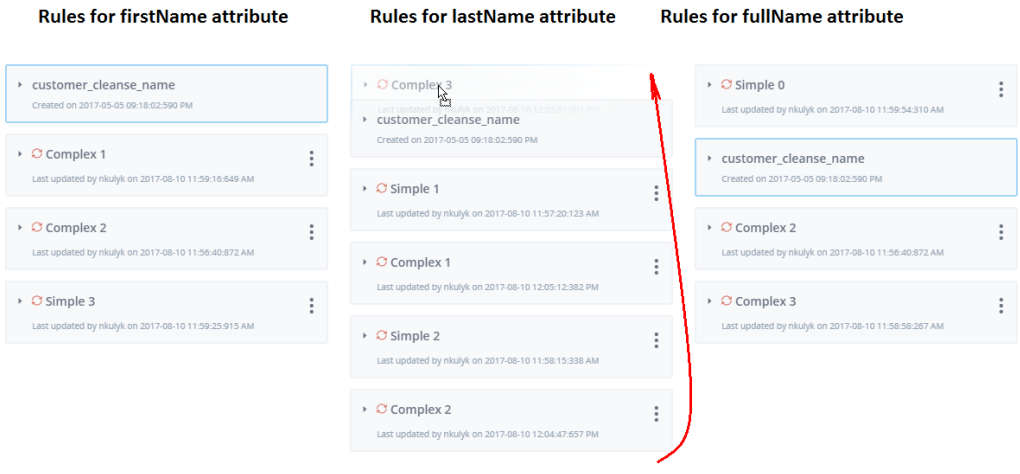
Example 2. The following image shows the *firstName* attribute where the rule with a complex operation called *Complex_1* was moved below a rule with simple operations called *Simple_2* and another rule with a complex operation called *Complex_2*.



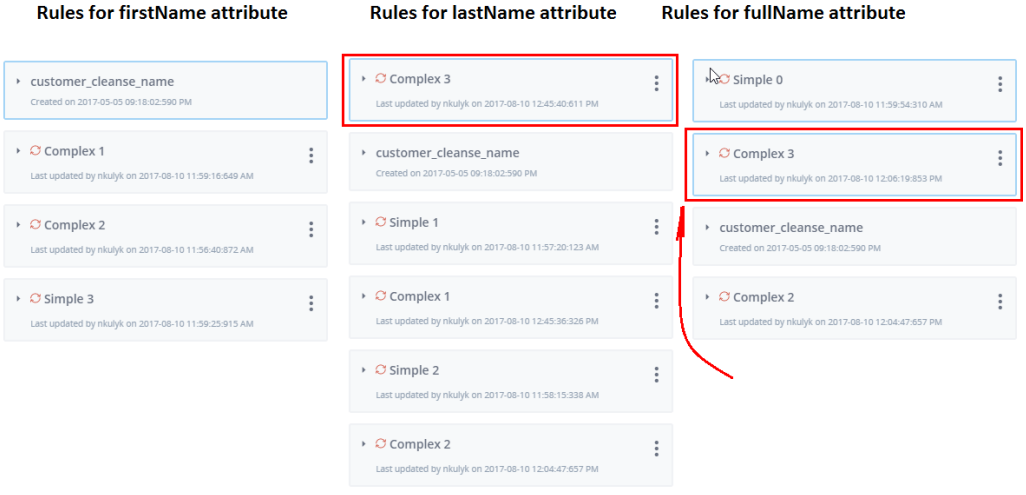
The firstName and lastName attributes are affected, as shown in the following image.



Example 3. The following image shows the *lastName* attribute where a rule with a complex operation called *Complex_3* was moved to the top, below the External rule and other rules with simple and complex operations.

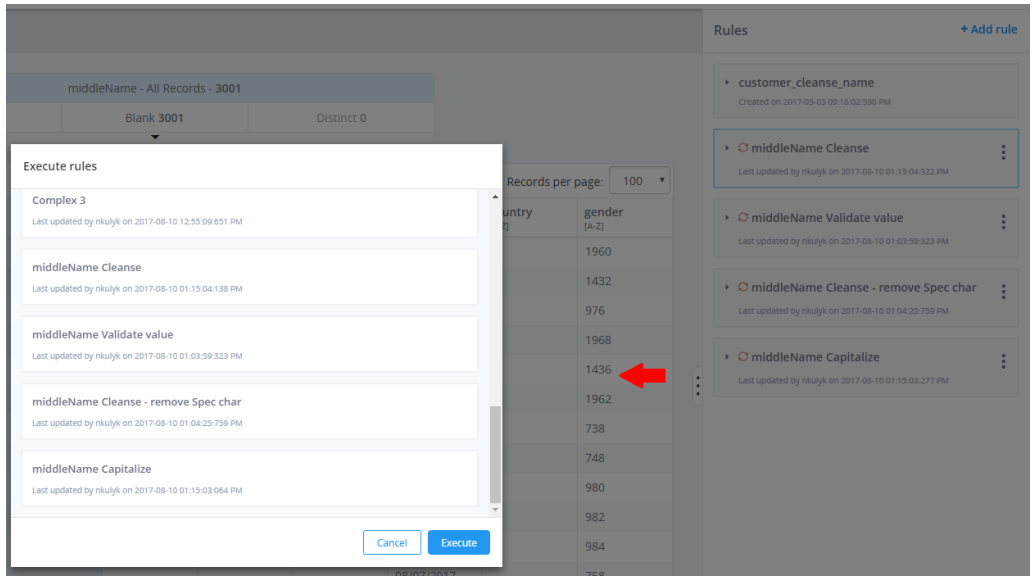


The lastName and fullName columns are affected, as shown in the following image.

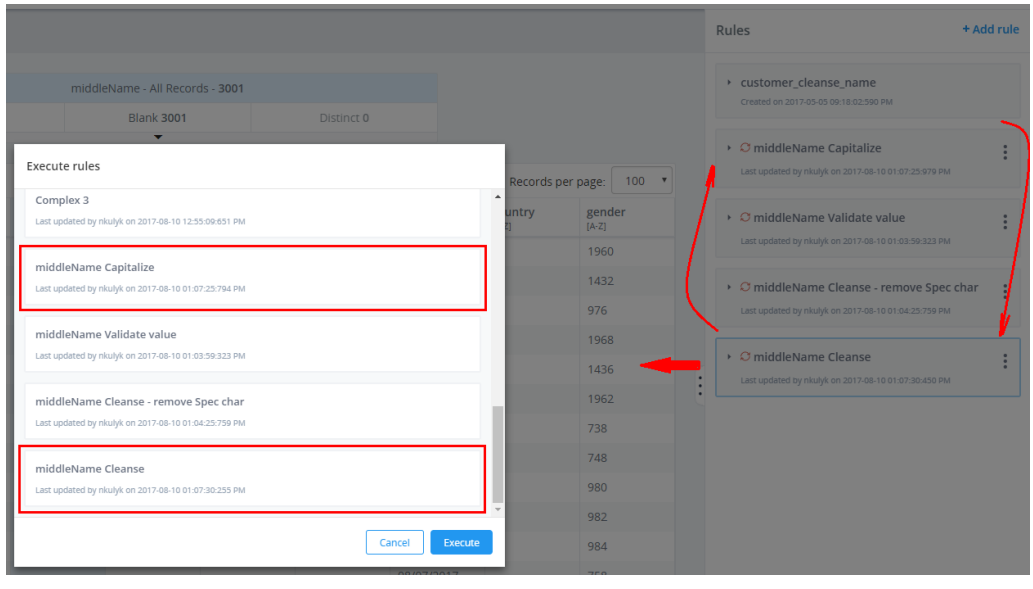


Reordering and Execution

Once Rules are reordered in the Rules panel, they will change their place in the Execute rules dialog as shown in the following image. As a result, Rules will be executed in a new order and operations will be applied in different way.

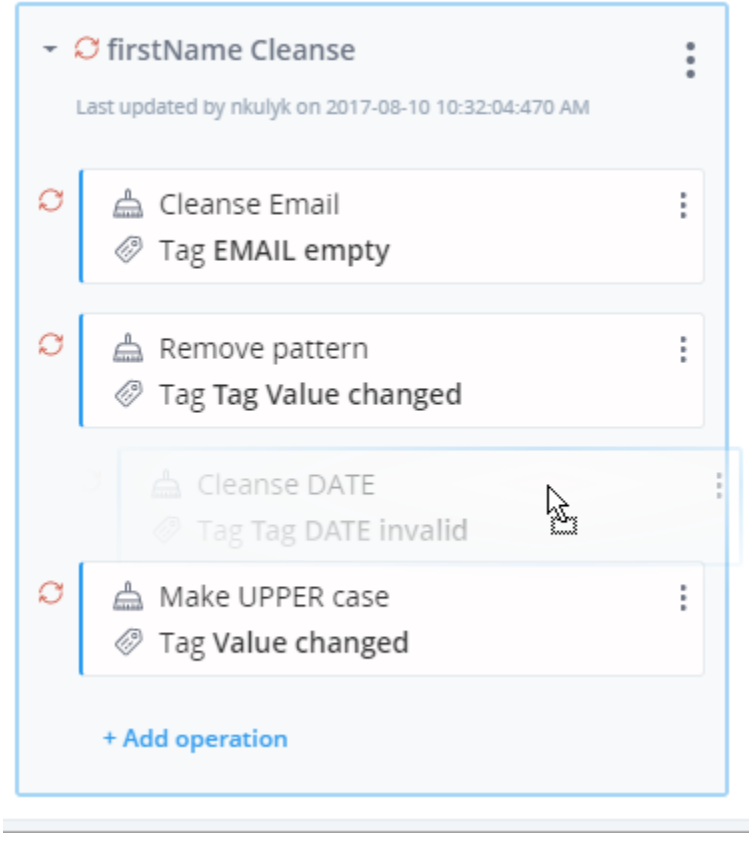


After rules are reordered in the Rules panel, they will change their place in the Execute rules dialog according to the order of their new position or location.



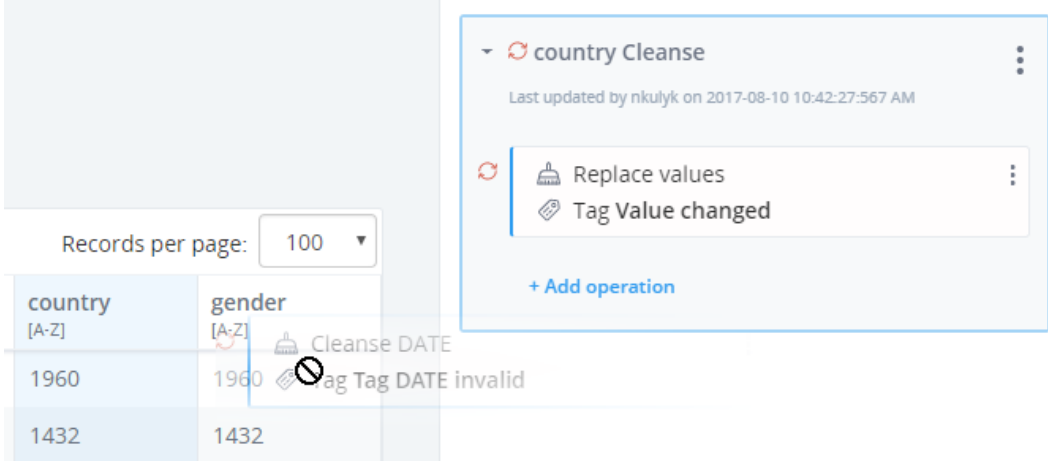
Dragging and Dropping Operations

To move an operation to another location or position, select the operation and then drag and drop it in the order or location you wish. A special icon appears for the Rule that is being dragged.



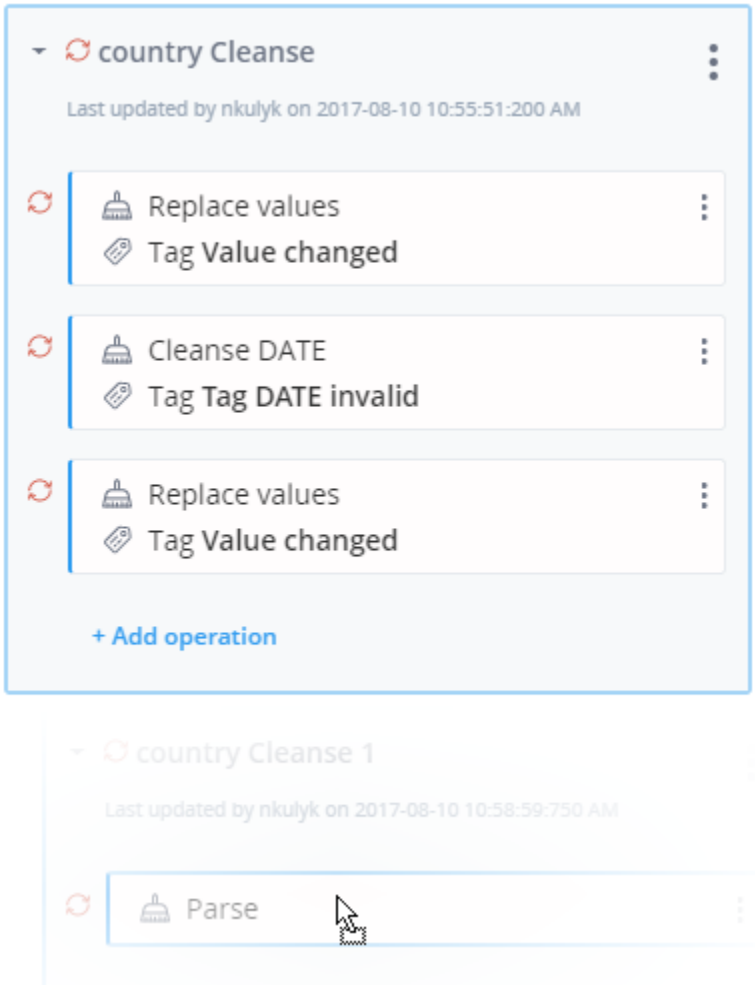
Only collapsed operations can be moved. Operations cannot be moved if other operations inside the rule is opened in configuration mode.

Operations can be moved to any place within the Rules. You will not be able to drop the operation outside the Rules panel.



Operations can be moved within the new and deployed rules. If an operation is moved inside the deployed rule, the Rule operation and other operations that changed places, will marked with an uncommitted icon.

Complex operations cannot be moved. If you try to drag and drop a complex operation, all rules will be dropped.



Content Types

Functionality Content Types is aimed to create alternative types of data related to real data attributes. This is some custom data type that you can relate to any attribute or operation that works better with specific information and avoid problems with inconsistent data types.

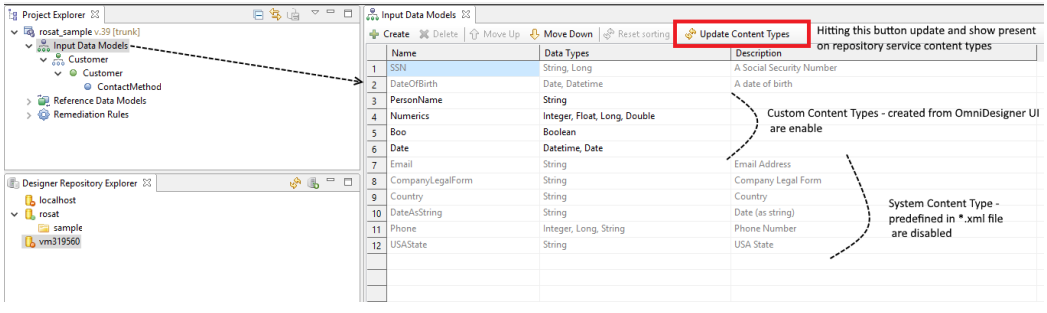
The difference between data type and content type is that data type is a predefined column in data base, while content type is specific type of data that can be mapped to real data, or data that comes in a system-source data.

Sometimes data can have strange data types or data type that differentiates from data type sets in a database, and it is better to map content types with data so the system can work with data without errors.

For example, the attribute *DateOfBirth* can have either date or the datetime data type, but there is information that it has a date data type in the database. When real data appears, containing a datetime data type, it can cause problems. To avoid this situation, the *dateOfBirth* content type with both date or datetime data type is created and mapped to the attribute.

Content Types in Omni Designer

Content types are taken from the repository service and can be created in the corresponding *.xml file or directly in Omni Designer.



There are two available kinds of Content Types:

- **System Content Type.** A content type taken from contentTypes.xml which is located by following path:

  \omnigen\OmniDesignerRepository\webapps \com.iwaysoftware.omni.designer.repositoryservice\WEB-INF\classes\contentTypes.

  The contentTypes.xml is extracted from the com.iwaysoftware.omni.designer.repositoryservice.war file and contains the following content:

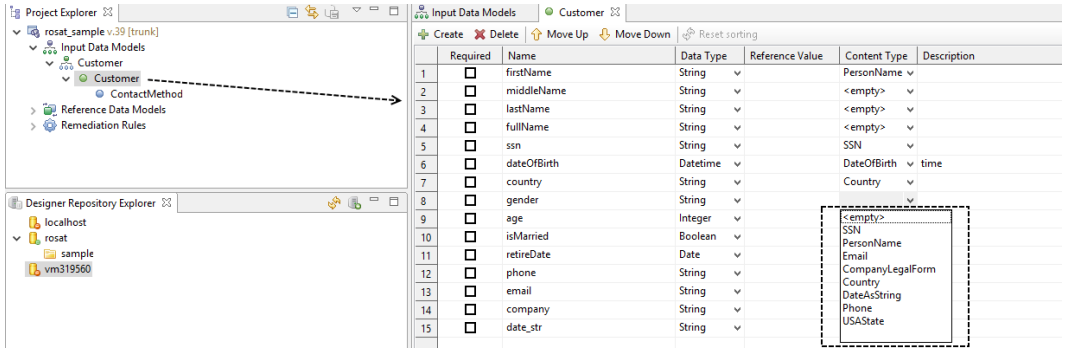
```
<contentTypes>
    <contentType>
        <name>SSN</name>
        <description>Social Security Number</description>
        <dataTypes>
            <dataType>String</dataType>
            <dataType>Long</dataType>
            <dataType>Integer</dataType>
        </dataTypes>
    </contentType>
    <contentType>
        <name>DateOfBirth</name>
        <description>Date of Birth</description>
        <dataTypes>
            <dataType>Date</dataType>
            <dataType>Datetime</dataType>
        </dataTypes>
    </contentType>
    <contentType>
        <name>Email</name>
        <description>Email Address</description>
        <dataTypes>
            <dataType>String</dataType>
        </dataTypes>
    </contentType>
    <contentType>
        <name>CompanyLegalForm</name>
        <description>Company Legal Form</description>
        <dataTypes>
            <dataType>String</dataType>
        </dataTypes>
    </contentType>
```

You can add as many Content Types as is needed into the XML file and edit or remove existing types once. After updating the OmniDesigner/InputDataModels editor, the Update Content Types button will be available on the OmniDesigner user interface.

- **Current Content Type.** a content type created from the OmniDesigner/InputDataModels editor page.

  The Data type(s) must be obviously specified for both types of content types.

Available Content Types can be mapped to Subject fields from the subject editor page, as shown in the following image.



Only content types having the same data type as the field could be mapped to this field.

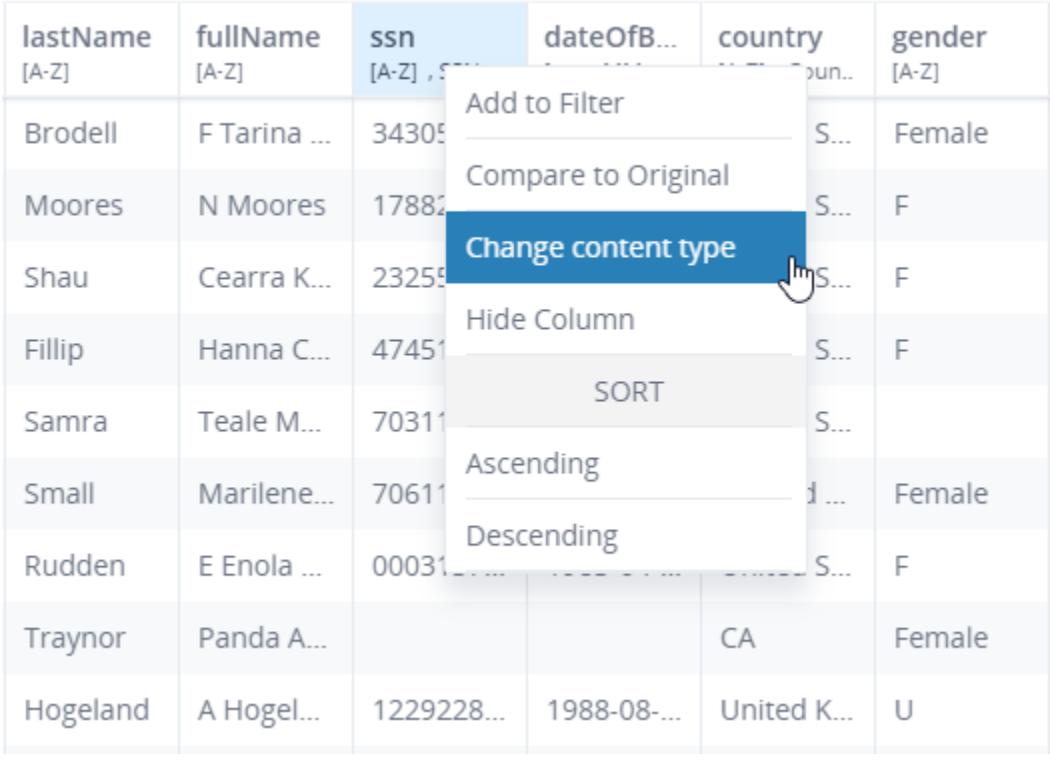
Note: Fields with data types *reference* and *identified* do not have mapped content types.

To have Content Types added or mapped from OmniDesigner and appear in the Workbench, you must generate project and deployment bundles and deploy bundles to OmniGen.

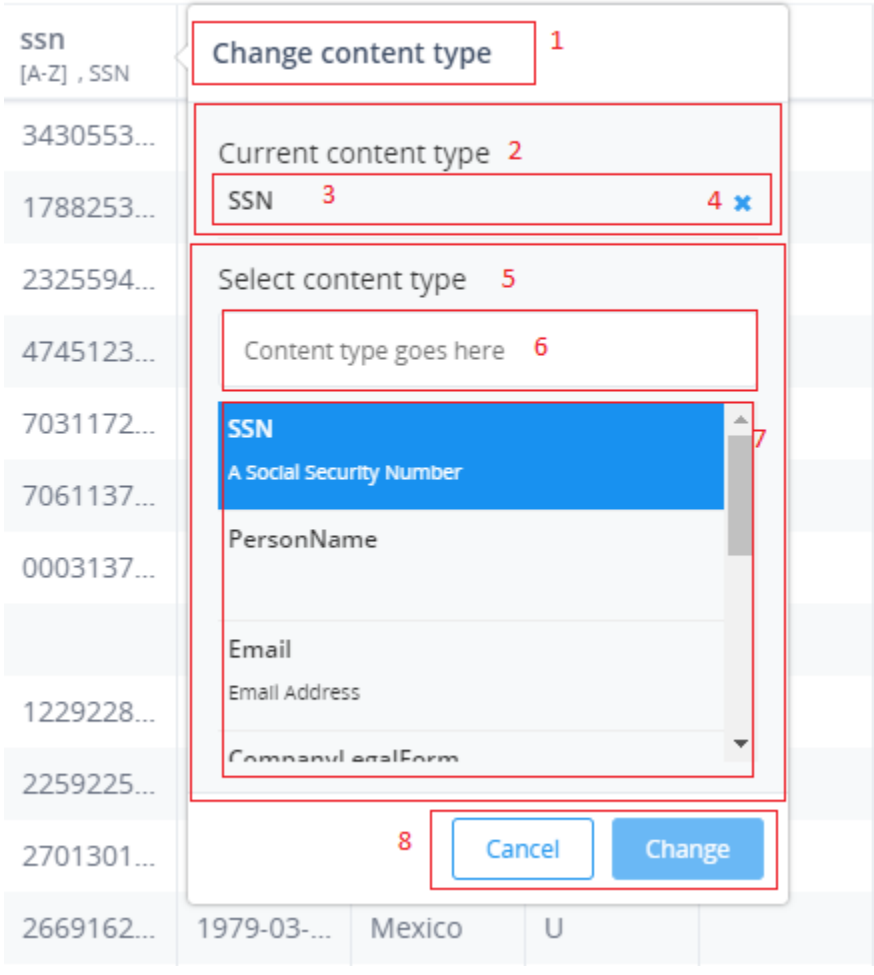
Content Types in the Workbench

You can change the content type in the Workbench by clicking *Change content types* in the context menu from the repository service.

Content types can be added from table header content menu, as shown in the following image.



When you select the *Change content type* option in context menu, a dialog opens where you can select and change the content type.

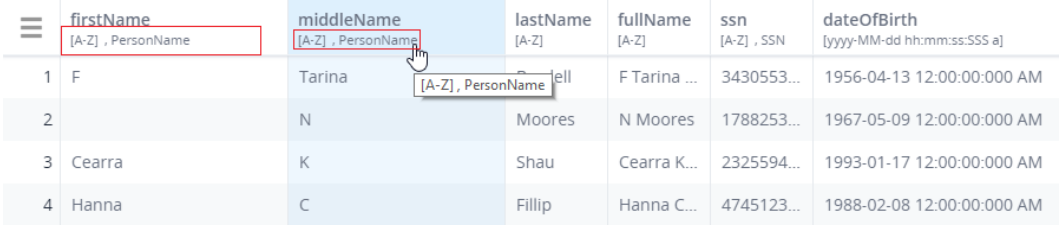


The following list shows the corresponding sections to the image above.

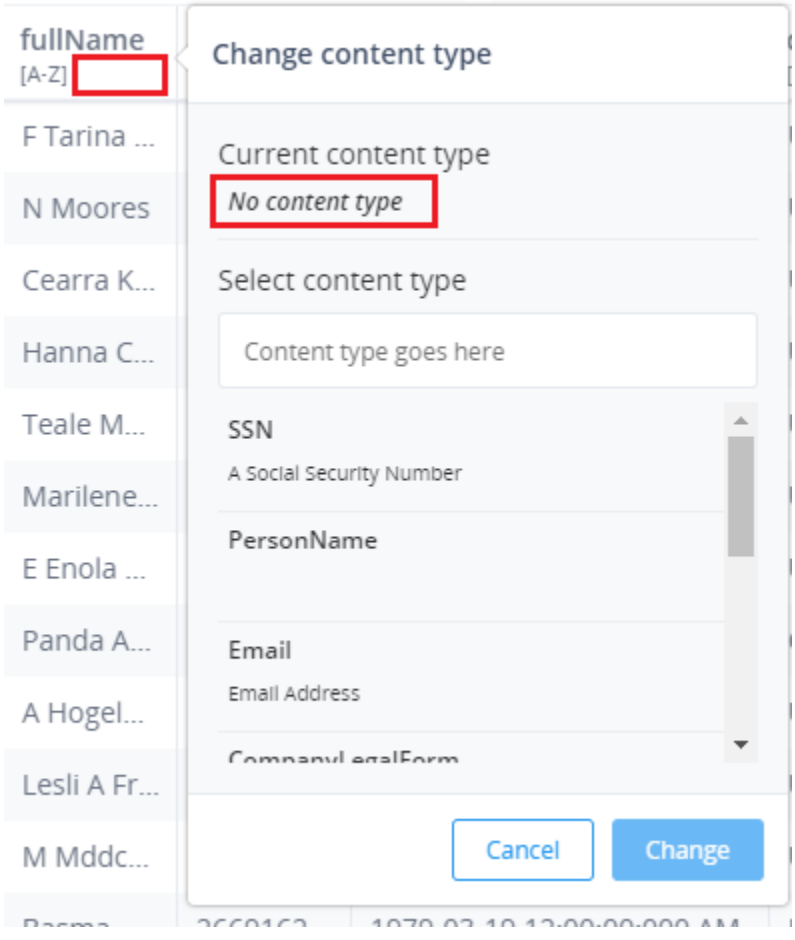
1. Change content type header
2. Current content type section
3. Selected current content type
4. Delete current content type section
5. Selected current content type
6. Search content type field

7. Content type list
8. Cancel /Change buttons

The Current content type section displays information regarding the current content type mapped to the attribute. If there is a content type already added, it will appear in the table header, as shown in the following image.

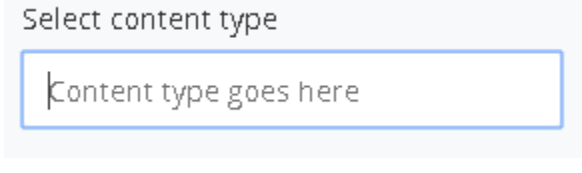


If there is no content type added, there will be no content type in the table header, and the corresponding information will be displayed in the Change content type dialog.

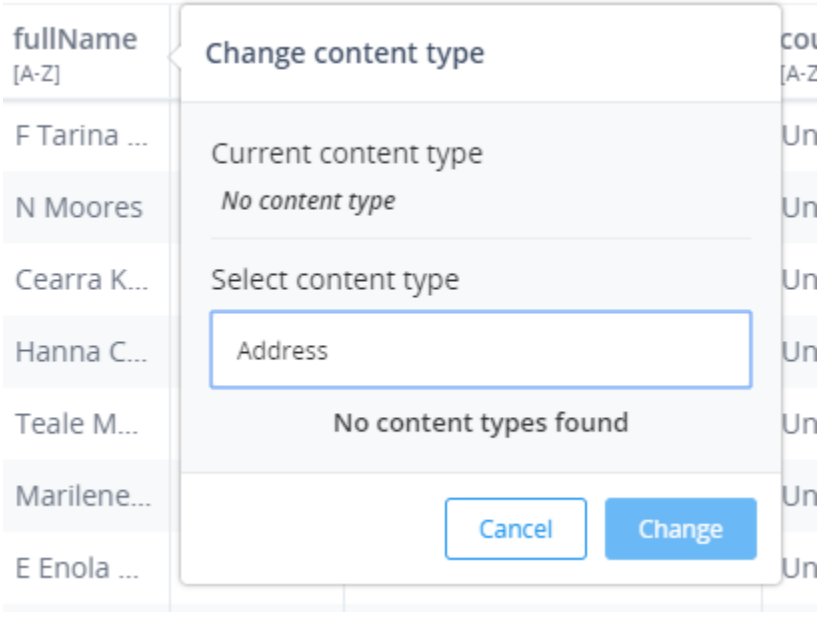


The Select content type section contains information regarding all available content types.

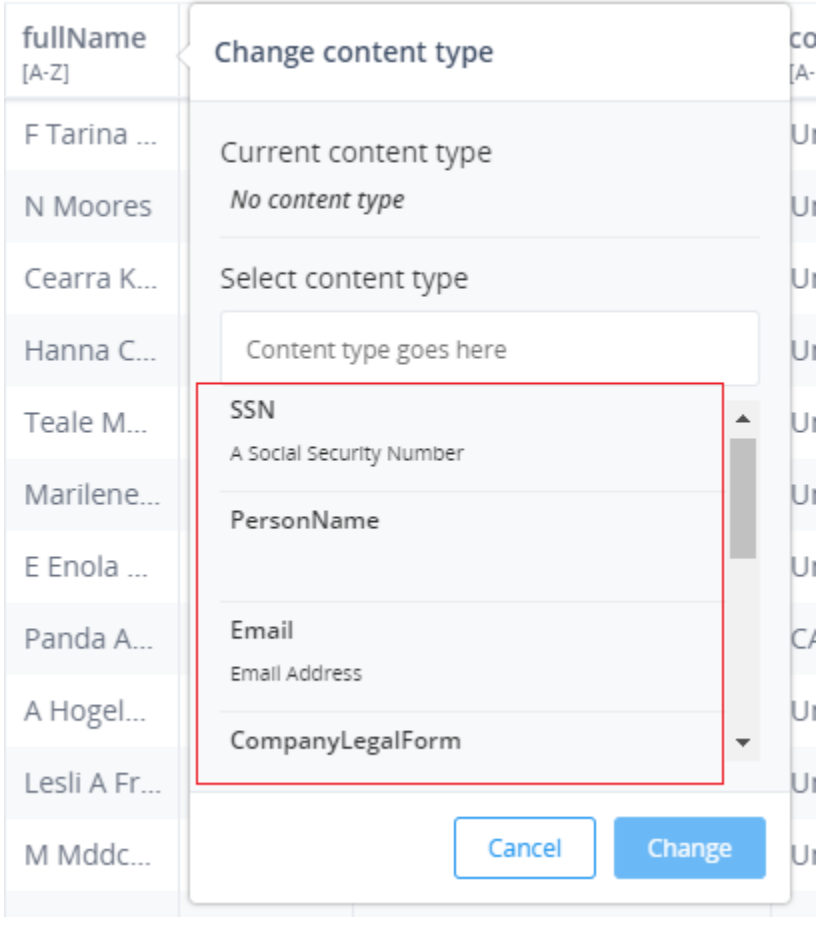
There is also a search field that allows you to search for content types.



If there is no content type found, then the appropriate information appears.



The following image shows a list of available content types (created or uploaded on the repository service). The Content type name and description are displayed for each content\ item.



The Name and Description of content types are predefined in the contentTypes.xml file and in OmniDesigner.

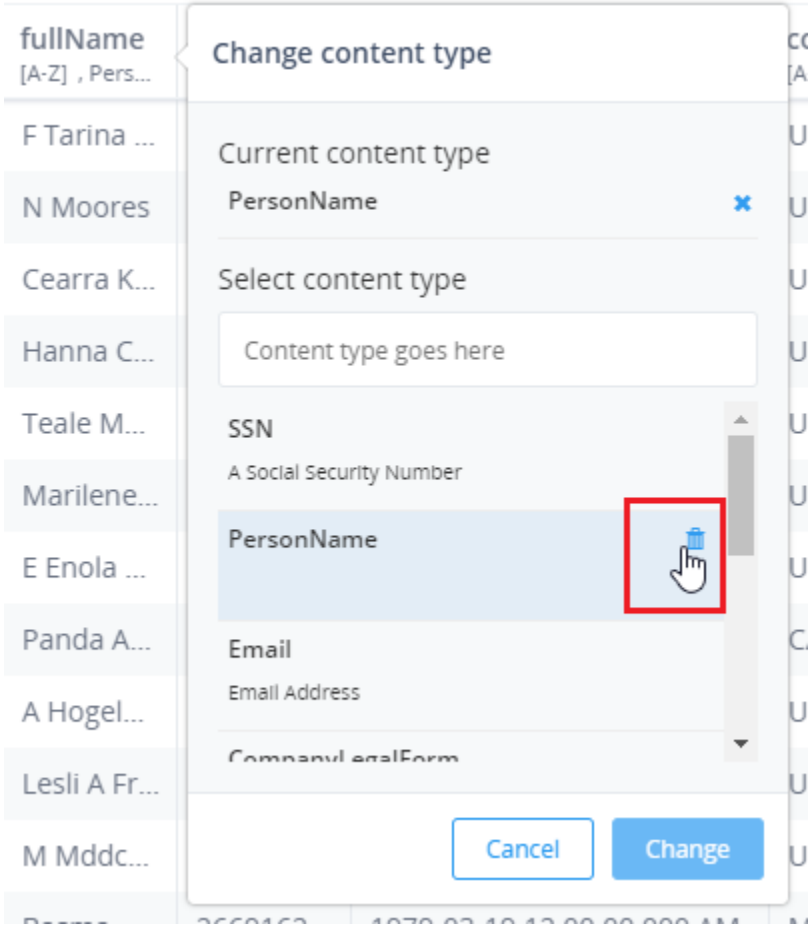
If the FieldType was specified inside the contentTypes.xml file on the OmniDesigner side, then the content types displayed in the list of the dialog corresponds to the data type of the selected attribute. If FieldType was not specified in the context_type.xml file, then this content type will be displayed for all attribute data types.

Once you select a content type in the list, it will appear in the Current content type section. Clicking *Change* will make it appear in the Column header.

firstName [A-Z] , PersonName	middleName [A-Z] , PersonName	lastName [A-Z]	fullName [A-Z] , Pers...	ssn [A-Z] , SSN
F	Tarina	Brodell	F Tarina ...	3430553...
	N	Moores	N Moores	1788253...
Cearra	K	Shau	Cearra K...	2325594...
Hanna	C	Fillip	Hanna C...	4745123...

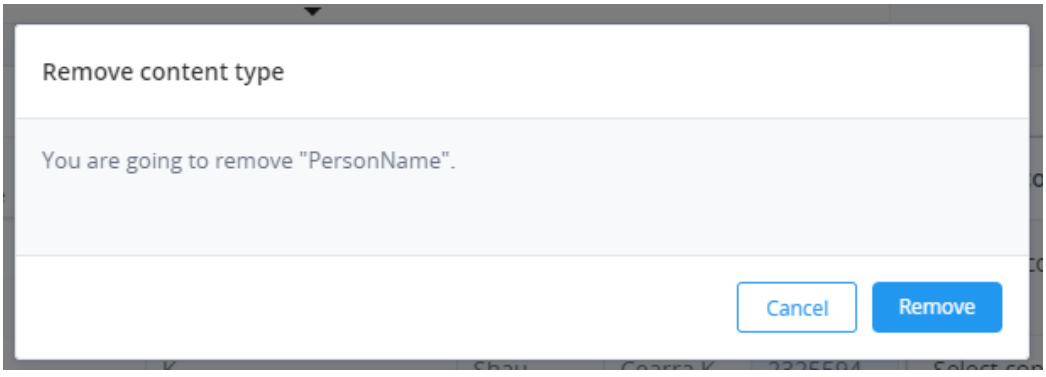
To delete content type from the selected column, delete it from Current content type section and click *Change*.

To delete content type from all over the attributes, Enter the content type in the *Select content type* field from the list and click *Delete*. The Delete button appears when you hover the cursor over the content type.



Only Current Content Types (created from OmniDesigner) can be removed. System Content Types cannot be removed from the list, because the delete button does not appear for them.

The following confirmation message appears when you delete a Content Type from the list.

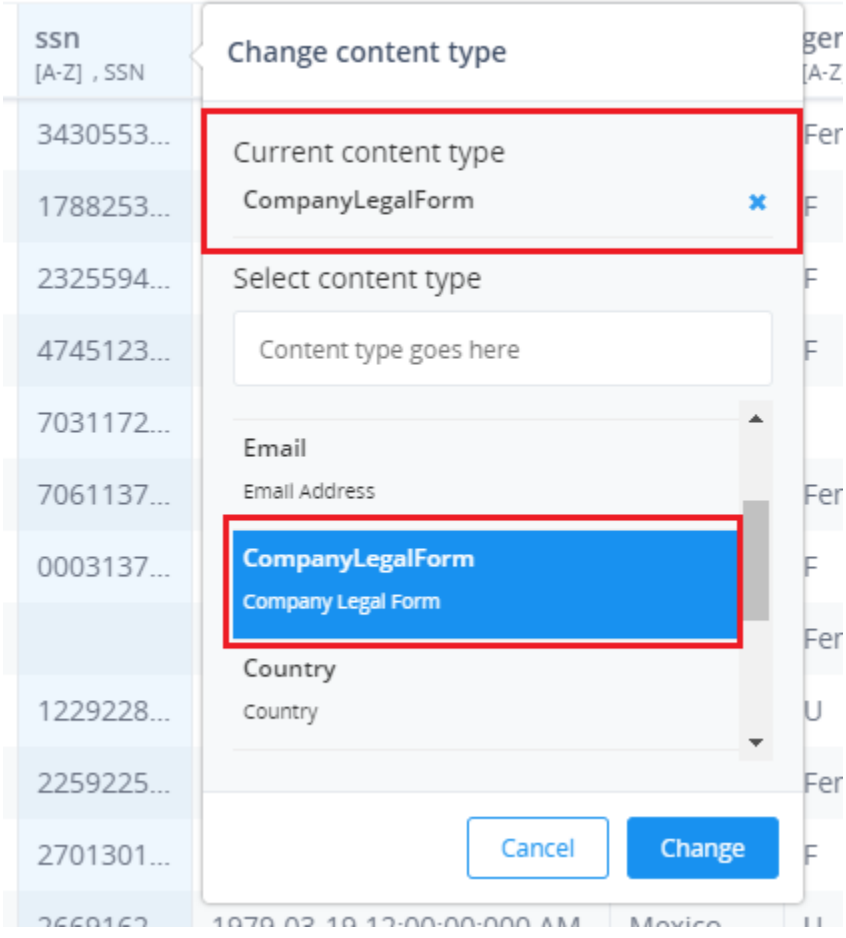


After clicking *Remove*, the rules panel will appear and the content type will be deleted from all the attributes. All operations added for the removed content type will also be deleted.

You can remove content types from all over the application during any user session.

To revert deleting from the application, click *Reset Session*. The application returns to the default state.

To map content types to the selected column, you must first select it from the list and click *Change*. The Content Type will appear in the Current content type section.



One attribute can have only one Content Type assigned, but one content type can be related to several attributes.

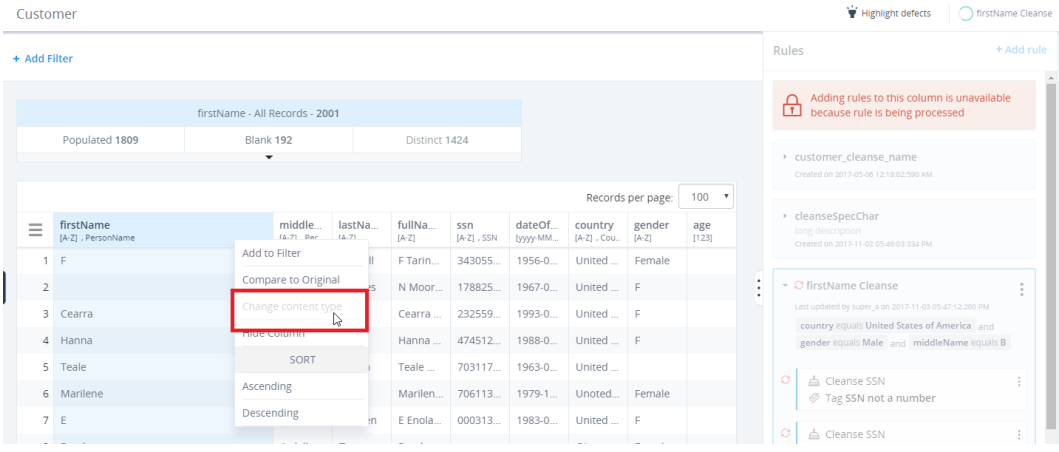
Content Types and Operations

Both Data Type operations and Content Type operations are available for cleansing.

Attributes that have added Content Types can be cleansed using Operations configured for that Content Type, but it cannot work with operations configured for another Content Type. The configuration of the Content Type inside the operation is made in the following property:

```
<contentType> </contentType>
```

Changing the content type of an attribute that is being executing is disabled. However, you can change the Content Type or remove it from other attributes.



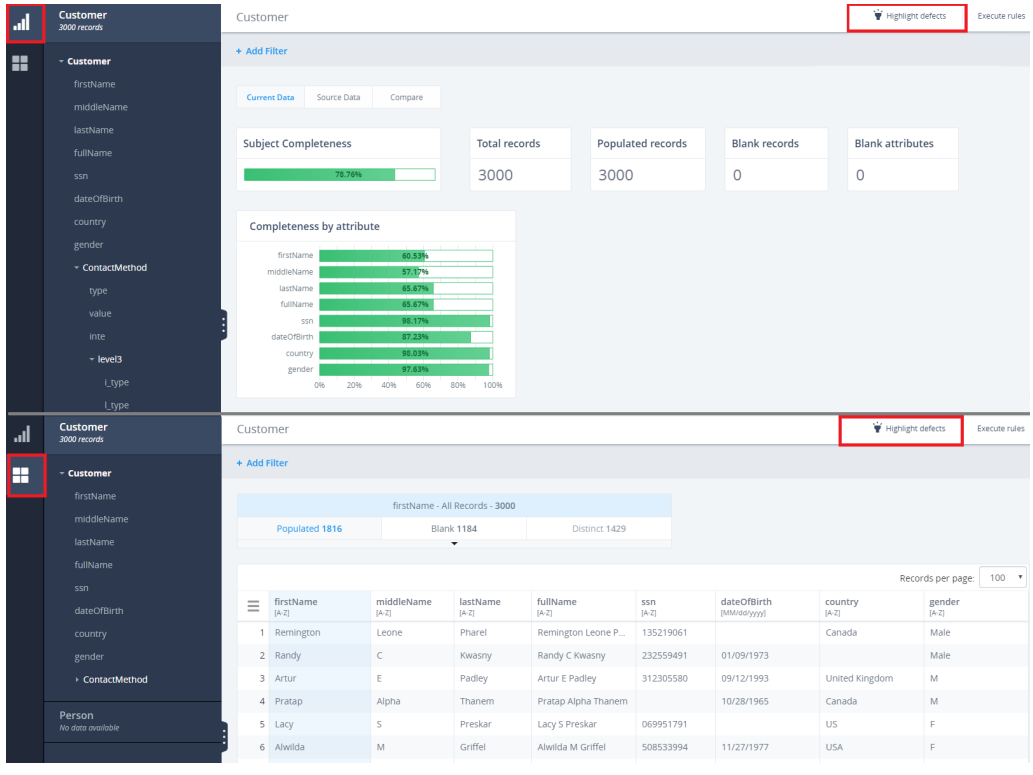
If you remove a Content Type from the list of dialogs, all operations added for that content type will also be removed. Once the rules have been executed, the rules data will be updated accordingly.

Detecting Errors (Highlight Defects)

Highlight Defects is a feature that helps you find inconsistencies and potential errors in the data.

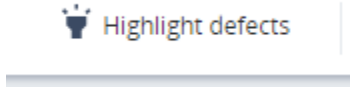
This feature can be accessed using the Highlight defect button in upper-right corner.

The Highlight defect button is available on the Profile page and on the Data table page, as shown in the following image.

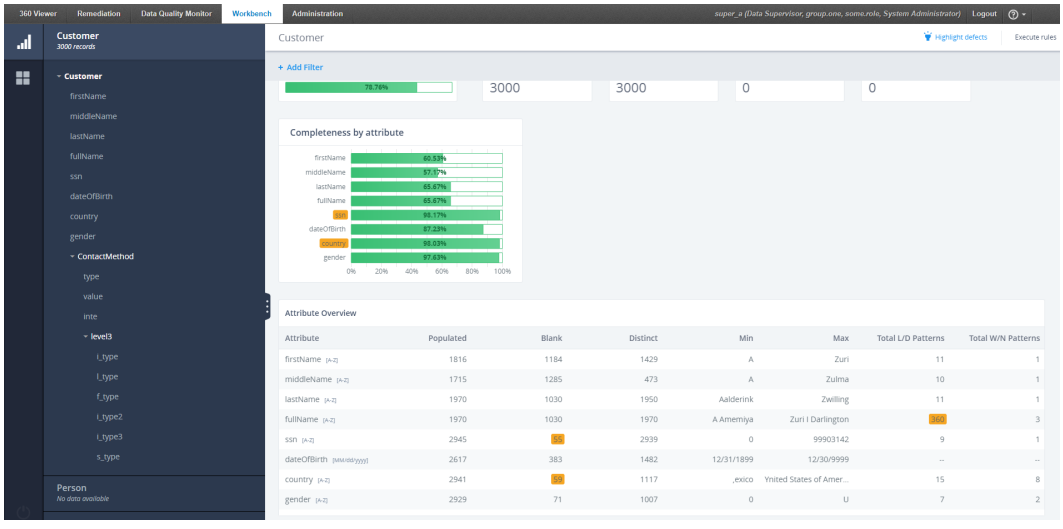


Using the Highlight Defects Feature

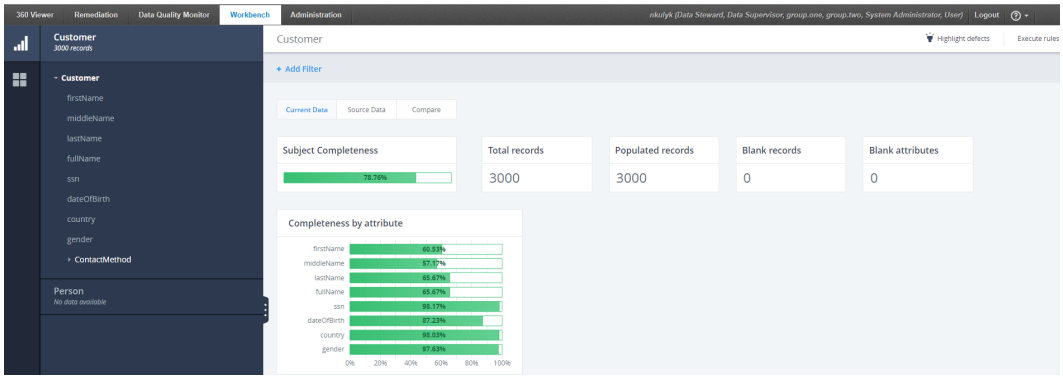
To activate this feature, click the *Highlight defects* button.



When the Highlight defects functionality is enabled, the color of the button is changed to blue and data areas with issues are highlighted in orange, as shown in the following image.



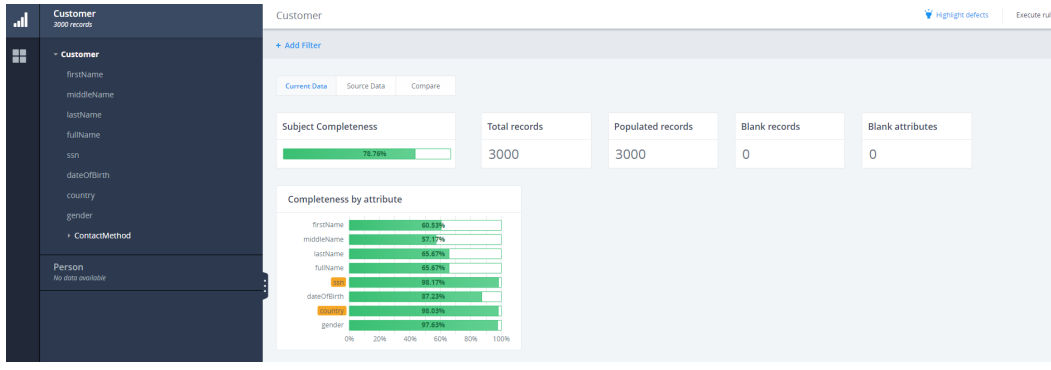
To disable the Highlight defects functionality, re-click the *Highlight defects* button.



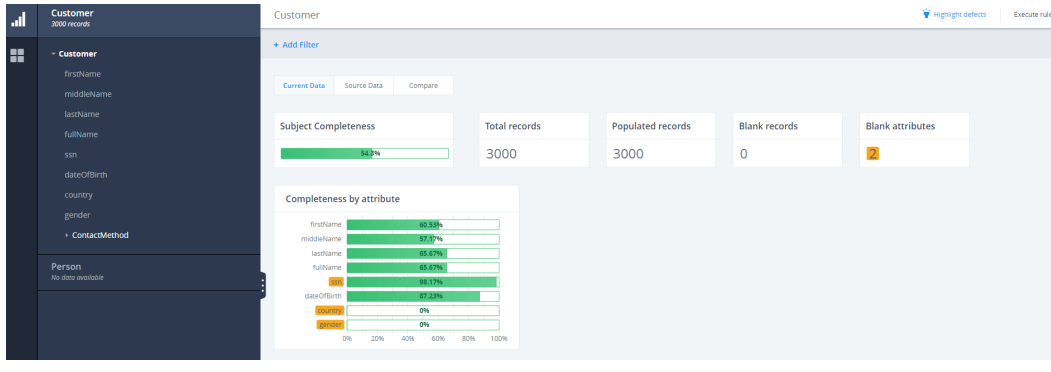
The Highlight defects functionality is saved per session. For example, if you activate Highlight defects and refresh the page, the highlighting defects functionality will be turned off. The same will happen when you decide to reset the session.

The Highlight defects functionality works with the latest data. If you execute some rules, the changes that were applied after processing those rules will also affect highlighting defects. This means that if processing the rules succeeded while the Highlight defects functionality was turned on, the defects highlighting will be re-calculated and the defect will be highlighted based on the new updated data.

The following image shows how a sample highlighting defect for the Customer Subject appeared before Rules were executed:

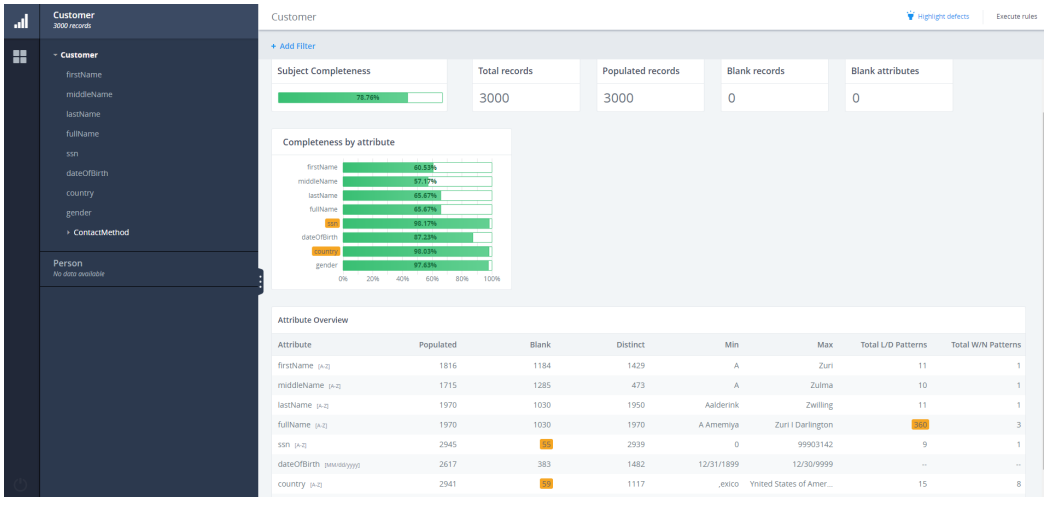


The following image shows how the Highlighting defect for Customer Subject appeared after the Cleanse_SSN operation was applied to the *Country* and *gender* attributes:

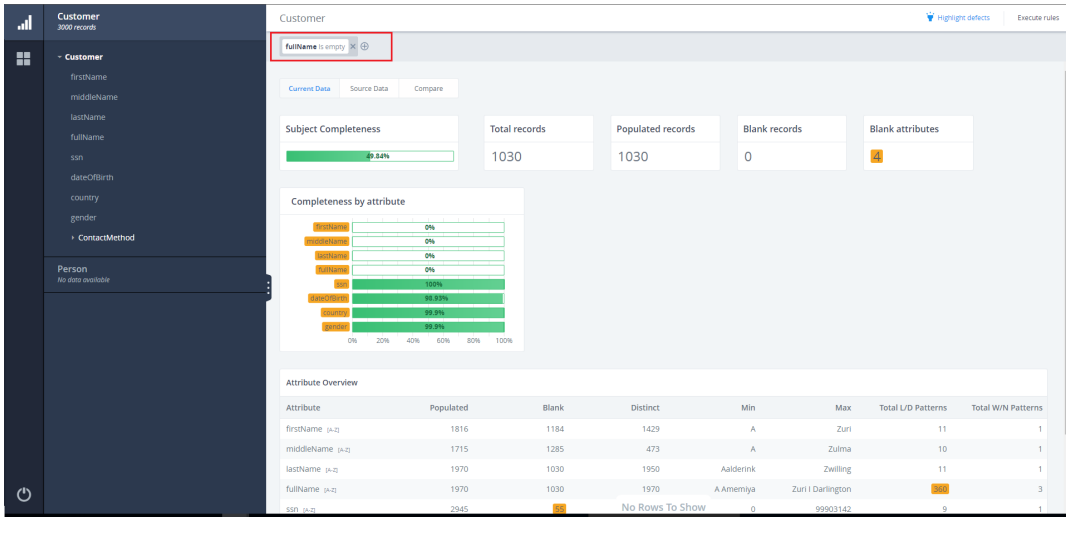


The same will happen if you apply a Filter while the Highlight defects functionality is activated. In this example, the defects highlighting will be re-calculated and the defect will be highlighted based on filtered data.

The following image shows how Highlighting defects for the Customer Subject appeared without an added filter:



The following image shows how the Highlighting defect for Customer Subject appeared after a Filter was added to the *fullName* bar chart:



Defining Rules to Determine Whether a Value Should be Highlighted

The Highlight Defects functionality works by placing a *highlight_defects.json* file in the following directory where the environment is configured:

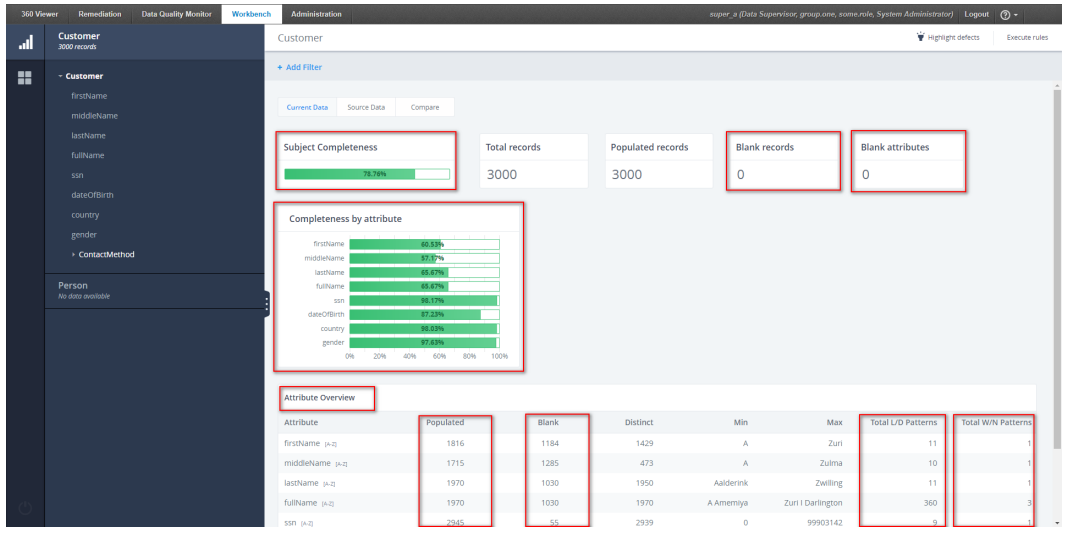
`C:\iWay_OGC\ogc\data\workbench`

The highlight_defects.json file contains a set of expression-rules based on which Highlight functionality made the calculation to determine whether a value is to be highlighted.

The expressions can be edited (if needed) inside the JSON file. After OGC Tomcat is restarted, the calculation of the highlight defects will be made according to the updated expressions.

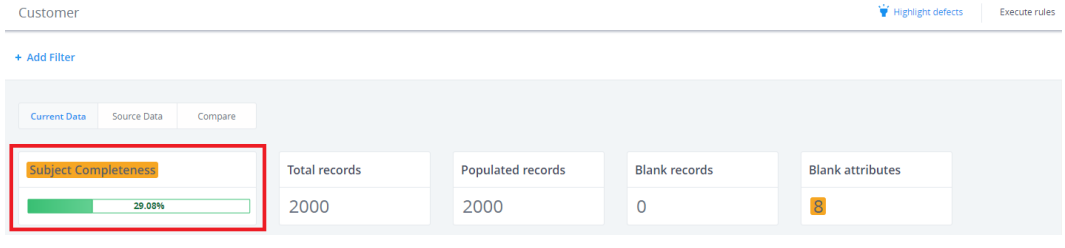
Profiling (Subject)

Highlight defects can be applied to the following bar charts:



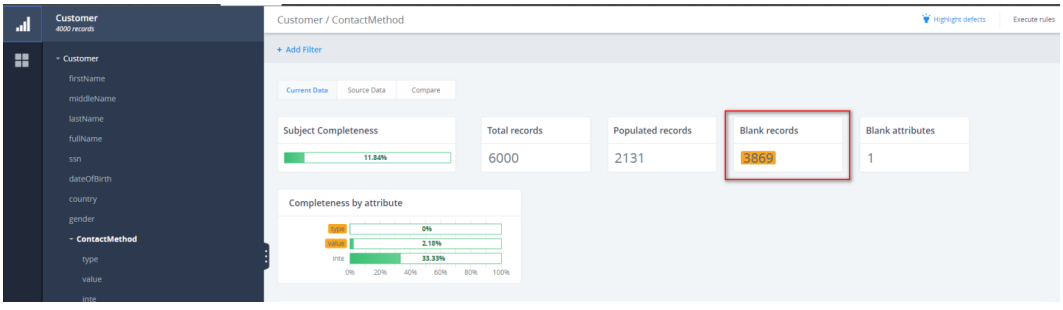
- **Subject Completeness.** The Subject completeness bar chart is highlighted when the value drops below the threshold [30%].

  If the value does not drop above the threshold [30%], it will not be highlighted.



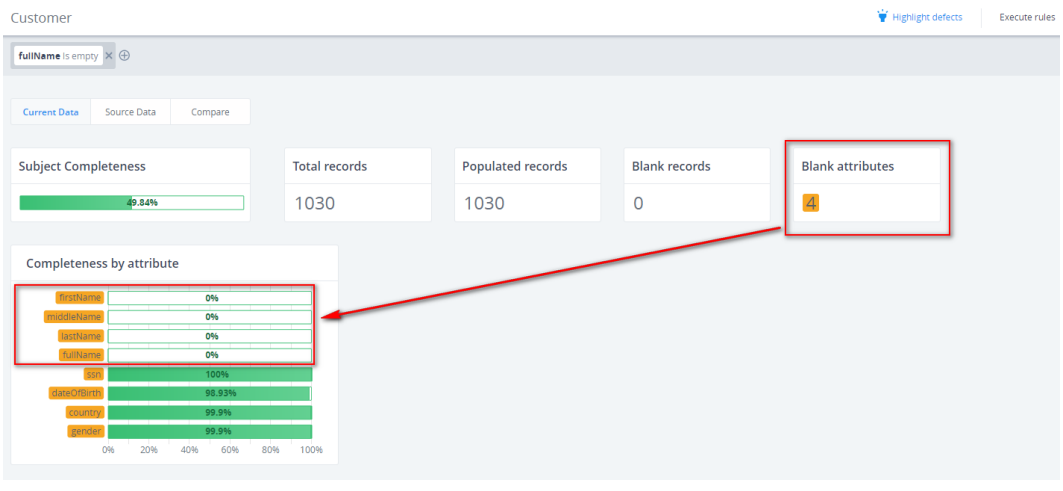
- **Blank records.** The blank records number is highlighted if either of the following occurs:
  - As a percentage of Total Records above the threshold [10%].
  - In total above the threshold [100%].

If the value does not fulfill the criteria above, it will not be highlighted.

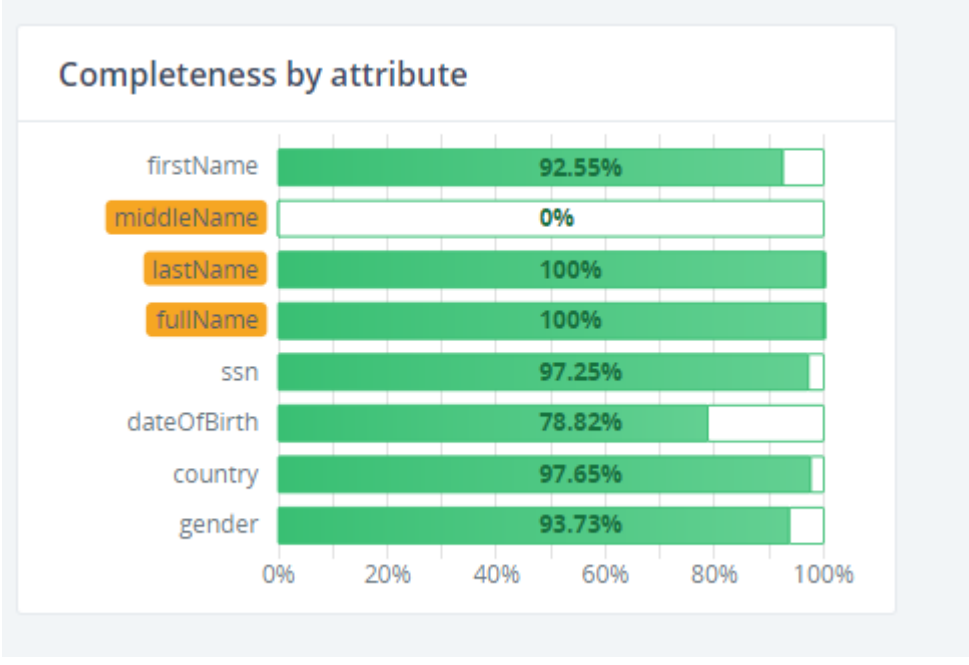


- **Blank attributes.** The Blank attributes number is highlighted if the value is above the threshold [1].

  If the value does not fulfill the criteria above, it will not be highlighted.



- **Completeness by attribute.** These bar charts will be highlighted if the Attribute completeness is either:
  - Below the threshold [10%].
  - Between thresholds [above 98% but below 100%].



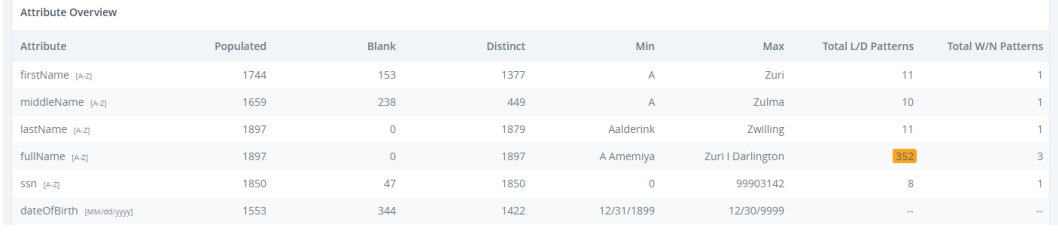
- **Attribute Overview.** This Blank number is highlighted if the corresponding values are either:
  - As a percentage of the Populated/Blank above a threshold [90%]
  - As a percentage of the Populated/Blank between thresholds [below 2% but above 0%]

Attribute Overview

Attribute	Populated	Blank
firstName [A-Z]	1816	1184
middleName [A-Z]	1715	1285
lastName [A-Z]	1970	1030
fullName [A-Z]	1970	1030
ssn [A-Z]	2945	55
dateOfBirth [MM/dd/yyyy]	2617	383
country [A-Z]	2941	59
gender [A-Z]	2929	71

The Total number for patterns is highlighted for the following instances:

- Total W/N Patterns is above the threshold [20].
- Total L/D Patterns is above the threshold [20].

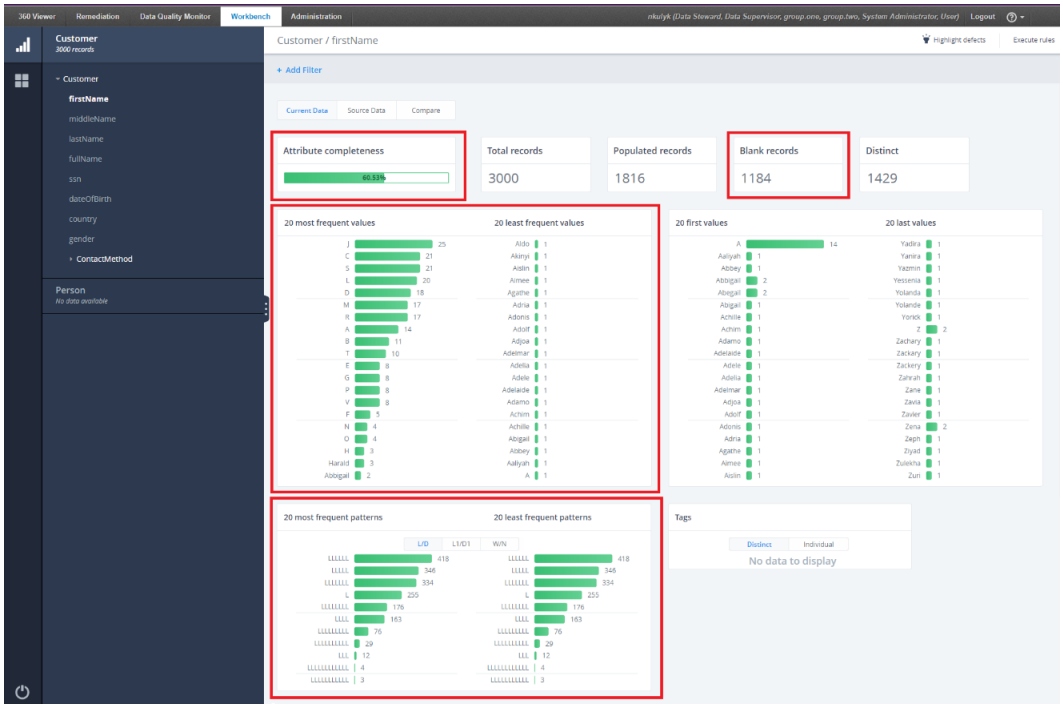


Attribute Overview

Attribute	Populated	Blank	Distinct	Min	Max	Total L/D Patterns	Total W/N Patterns
firstName [A-Z]	1744	153	1377	A	Zuri	11	1
middleName [A-Z]	1659	238	449	A	Zulma	10	1
lastName [A-Z]	1897	0	1879	Aalderink	Zwilling	11	1
fullName [A-Z]	1897	0	1897	A Amemiya	Zuri I Darlington	352	3
ssn [A-Z]	1850	47	1850	0	99903142	8	1
dateOfBirth [MM/dd/yyyy]	1553	344	1422	12/31/1899	12/30/9999	--	--

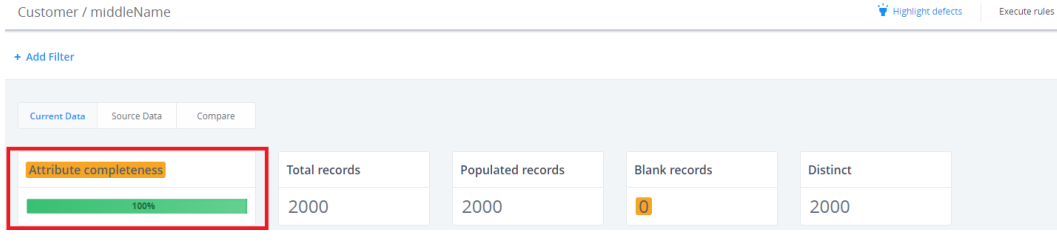
Profiling (Attribute)

The following image shows the profiling bar charts and sections.

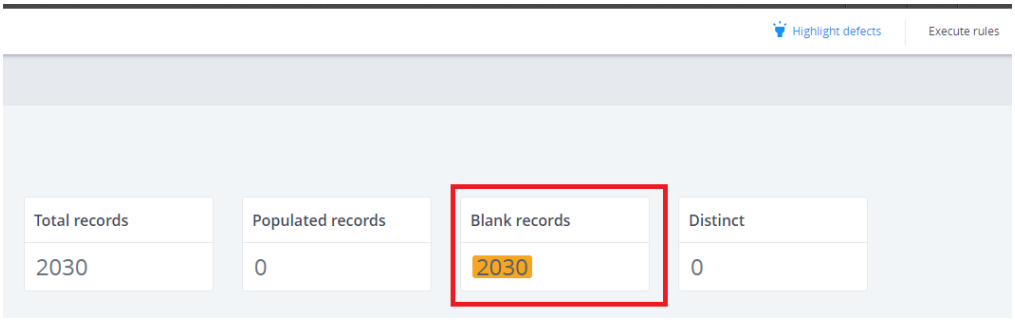


Highlight defects can be applied to the following bar charts:

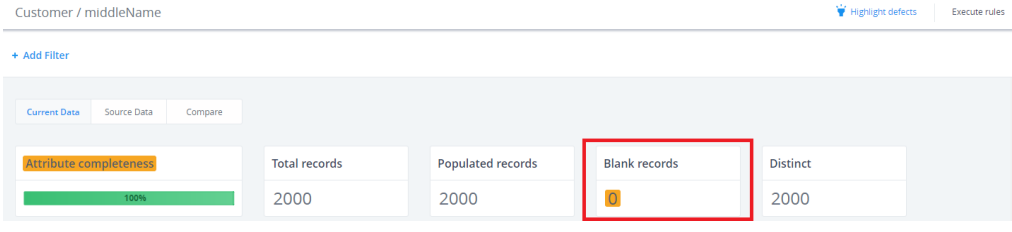
- **Attribute Completeness.** This bar chart is highlighted when the value for Attribute completeness is either:
  - Below a threshold [10%].
  - Between thresholds [above 98% but below 100%].



- **Blank records.** This number is highlighted in either one of the following conditions:
  - If the percentage of the Populated/Blank is above a threshold [90%].



  - If the percentage of the Populated/Blank is between thresholds [below 2% but above 0%] .



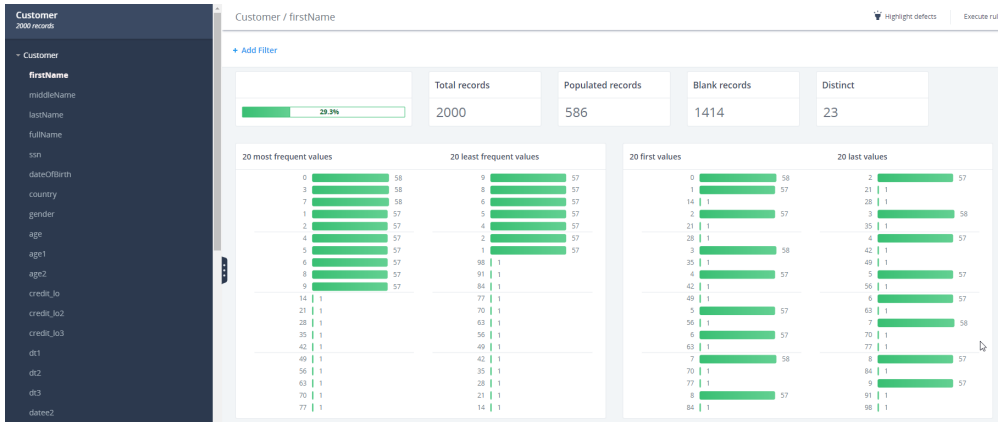
- **The 20 most/least frequent values charts.**

  To define which of the 20 most and least frequent values charts should be highlighted, the following complex computation should be solved.

```
TMP_TOTAL := $total;
TMP_DISTINCT := $distinct;
20_THRESHOLD := TOTAL/5;
for each row from top 5 most frequent
do
if $records > 20_THRESHOLD
do
TMP_TOTAL := TMP_TOTAL - $records;
TMP_DISTINCT := TMP_DISTINCT - 1;
done
done
LOW_THRESHOLD := TMP_TOTAL/TMP_DISTINCT;
for each row in least frequent
do
if $records < LOW_THRESHOLD/10
do
highlight
done
done Let's consider the following example:
```

- **20 most/least frequent patterns charts.**

  The following image shows the highlighting for 20 most/least frequent patterns charts for the *firstName* attribute.

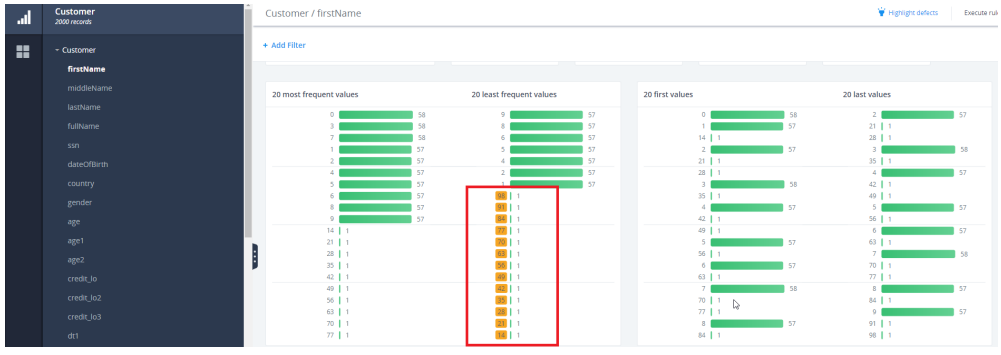


  The following sample calculations identify which values should be highlighted:

```
TMP_TOTAL :=2000
TMP_DISTINCT := 23

20_THRESHOLD := TOTAL/5 = 2000/5 = 400(static and doesn't change)
```

```
for each row from top 5 most frequent
do
if $records > 20_THRESHOLD

if 58 > 400 (false) done
LOW_THRESHOLD := TMP_TOTAL/TMP_DISTINCT;

LOW_THRESHOLD := 2000/23;for each row in least frequent
do
if 1<8.69 (true)

do highlight

if 57<8.69 (false)

done
done
```

The following image shows the results of the calculation containing the highlighted charts with the value of *1*.



Highlighting for 20 of the most and least frequent patterns and charts fulfills the following criteria:

- Patterns with frequency below a threshold [5%].
- If the sum of the 5 most frequent patterns exceeds 60% in total, and they do not contain a certain set of characters (digits/alpha/specials), then highlight all patterns containing those sets of characters.

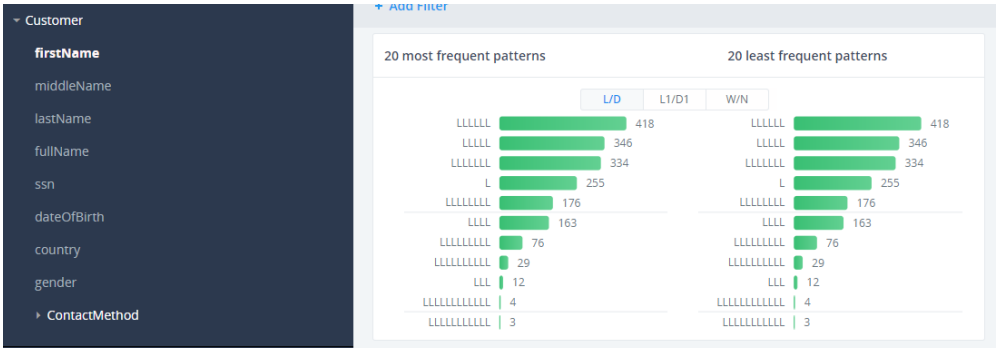
  For example, if the 5 most frequent patterns exceed 60% of the total cases and do not contain digits, then all patterns that do contain digits should be highlighted.
- Patterns with the frequency above the threshold if the number of distinct values is also above the threshold (if the number of distinct values is greater than 100 and the frequency is greater than 25%).

It is easier to define which of the 20 most and least frequent patterns charts should be highlighted based on the following complex computation:

```
TMP_TOTAL := pattern_total = sum(#records for all most frequent patterns)
TMP_DISTINCT := pattern_distinct = # of rows in most frequent patterns
```

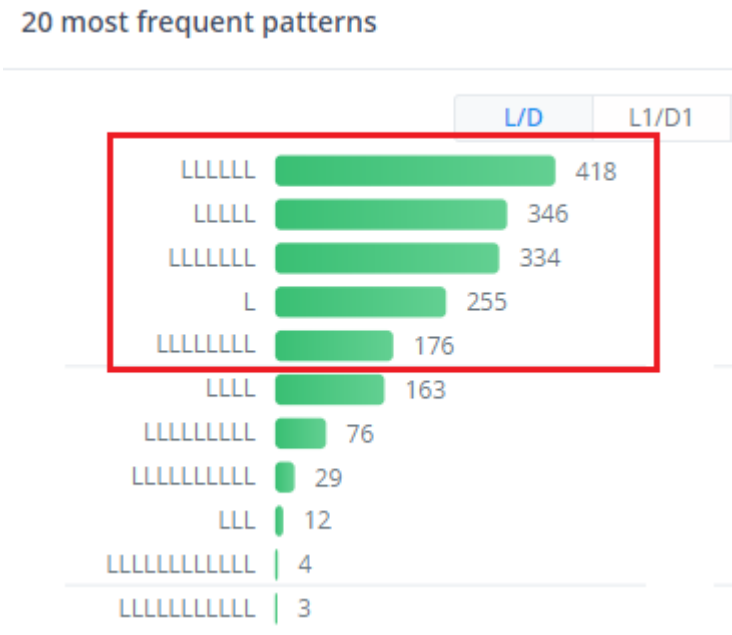
```
20_THRESHOLD := TOTAL/5;
for each row from top 5 most frequent
do
if $records > 20_THRESHOLD
do
TMP_TOTAL := TMP_TOTAL - $records;
TMP_DISTINCT := TMP_DISTINCT - 1;
done
done
LOW_THRESHOLD := TMP_TOTAL/TMP_DISTINCT;
for each row in least frequent
do
if $records < LOW_THRESHOLD/10
do
highlight
done
done Let's consider the following example:
```

The following example shows highlighting for the 20 most/least frequent pattern charts for the *firstName* attribute.

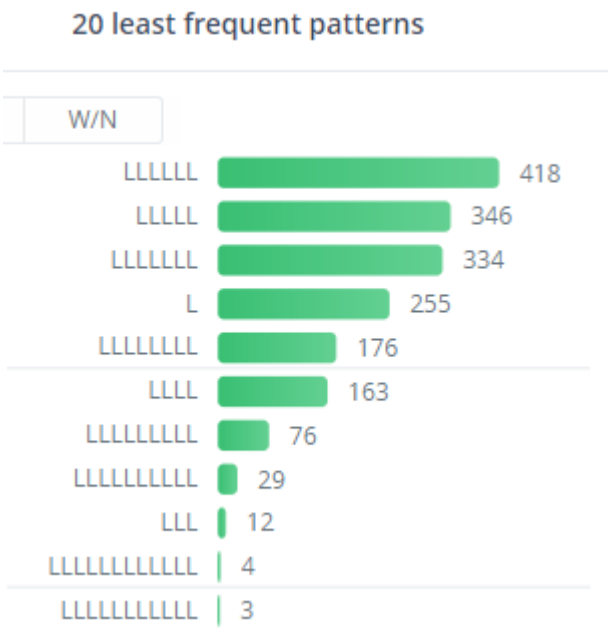


The following calculations are used to identify which values are highlighted:

```
TMP_TOTAL := pattern_total = sum(#records for all most frequent patterns)

TMP_TOTAL :=418+346+334+255+175+163+76+29+12+4+3=1815
TMP_DISTINCT := pattern_distinct = # of rows in most frequent patterns

TMP_DISTINCT := 11

20_THRESHOLD := TOTAL/5 = 1815/5 = 363 (static and doesn't change)
```

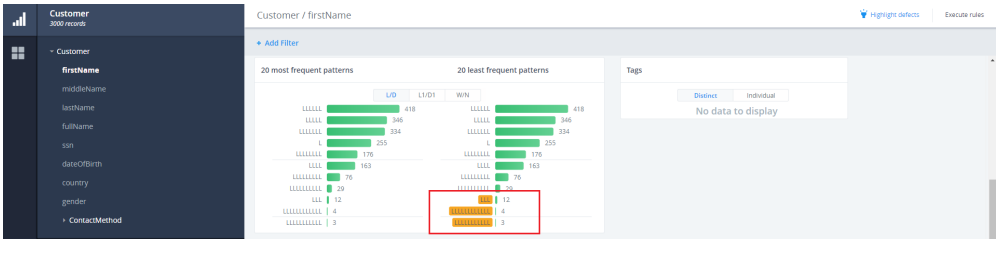



```
for each row from top 5 most frequent
do
if $records > 20_THRESHOLD

if 418 > 363 (true) do
TMP_TOTAL := 1815 – 418 = 1397;

TMP_DISTINCT := 11-1 = 10;

if 346 > 363 (false)
done
done

LOW_THRESHOLD := TMP_TOTAL/TMP_DISTINCT;

LOW_THRESHOLD := 1397/10 = 139.7;
```



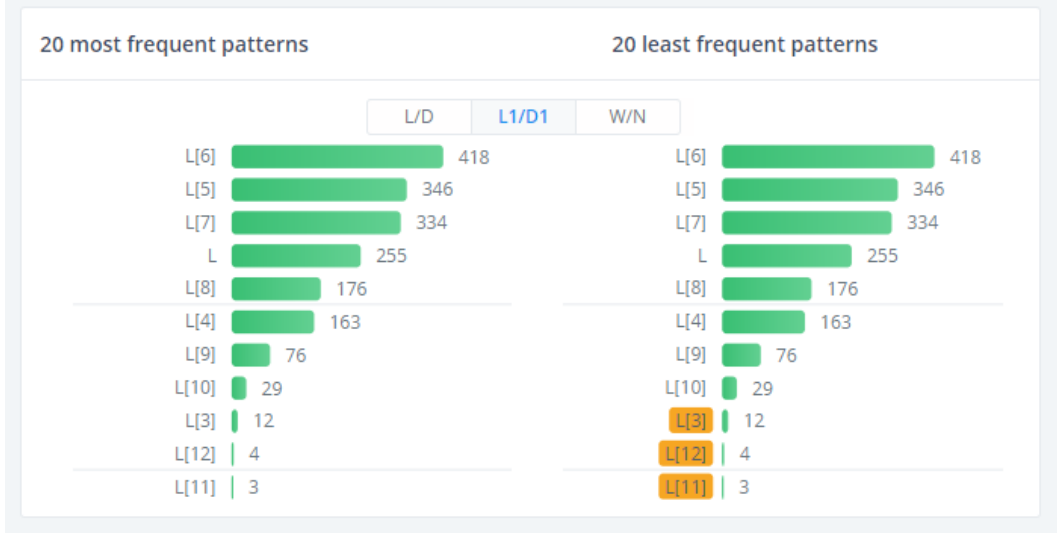

```
for each row in least frequent
do
if $records < LOW_THRESHOLD/10

if 3 < 13.97 (true) do highlight done

if 4 < 13.97 (true) do highlight done

if 12 < 13.97 (true) do highlight done

if 29 < 13.97 (false)

done
done
```

The following image shows the highlighted charts, containing values *3*, *4*, and *12*.

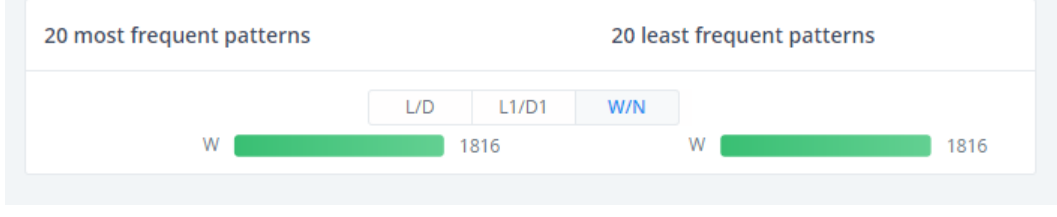


The same calculation can be performed to define highlighting for L1/D1 and W/N patterns.

Since the bar chart numbers of the L1/D1 patterns match bar charts numbers for L/D patterns, the result after highlighting defects will be the same.



Since the most and least frequent values for W/N patterns look different when defining which bar charts should be highlighted, new calculations should be performed.

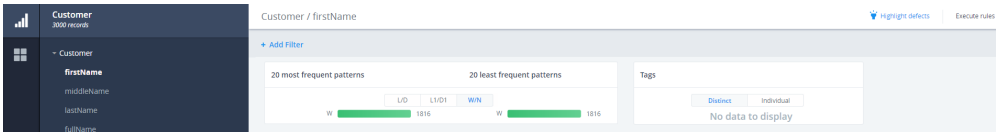


The following calculations are used to identify which values are highlighted:

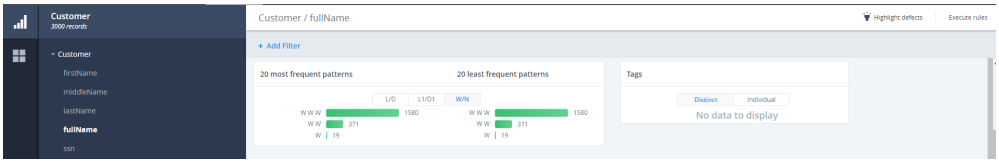
```
TMP_TOTAL := pattern_total = 1816
TMP_DISTINCT := 1


20_THRESHOLD := 1816/5 = 363.2


for each row from top 5 most frequent
do
if 1816 > 363.2 (true)
TMP_TOTAL := 1816 - 1816 = 0;
TMP_DISTINCT := 1- 1 = 0;
```

```
done
done
```

The resulting output is that no bar charts are highlighted, as shown in the following image.



The following image shows a sample bar chart for the most and least frequent W/N patterns to view how highlighting will appear for the *fullName* attribute.

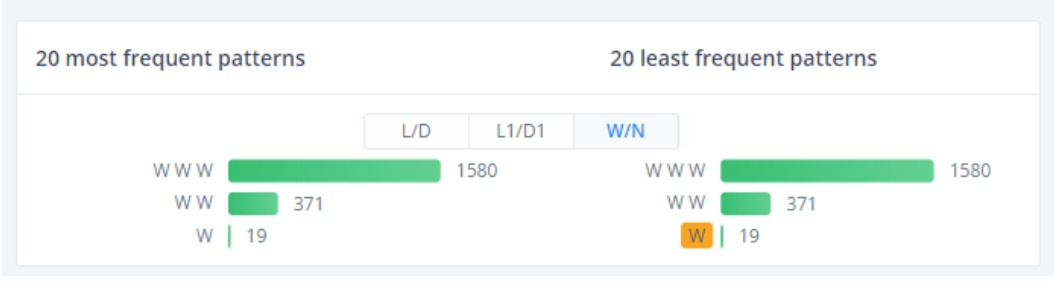


The following calculations are used to identify which values are highlighted:

```
TMP_TOTAL := pattern_total = 1580+371+19 = 1970

TMP_DISTINCT := 3

20_THRESHOLD := 1970/5 = 394

for each row from top 5 most frequent
do
if 1580 > 394 (true)

do
TMP_TOTAL := 1970-1580=390;
TMP_DISTINCT := 2
if 371>390 (false)
do
LOW_THRESHOLD := 390/2=195
for each row in least frequent
do
if 19<19.5 (true)

do
highlight
done

if 371<19.5 (false)

done
```

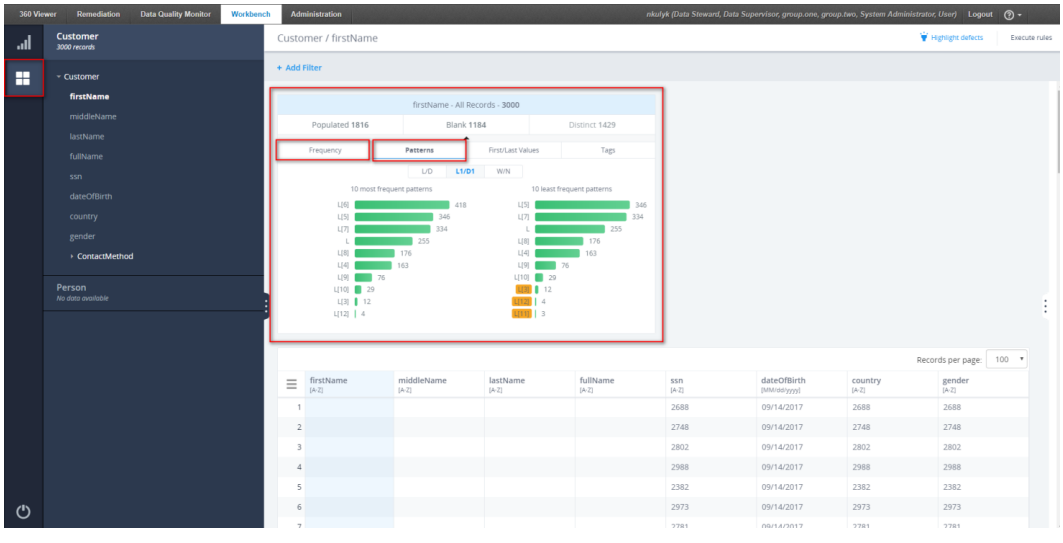
done

The resulting value for the highlighted bar chart is *19*.



Highlight Defects in On-Screen Filters

The Highlight functionality can be also applied to On-screen filters on the Frequency and Patterns tabs.



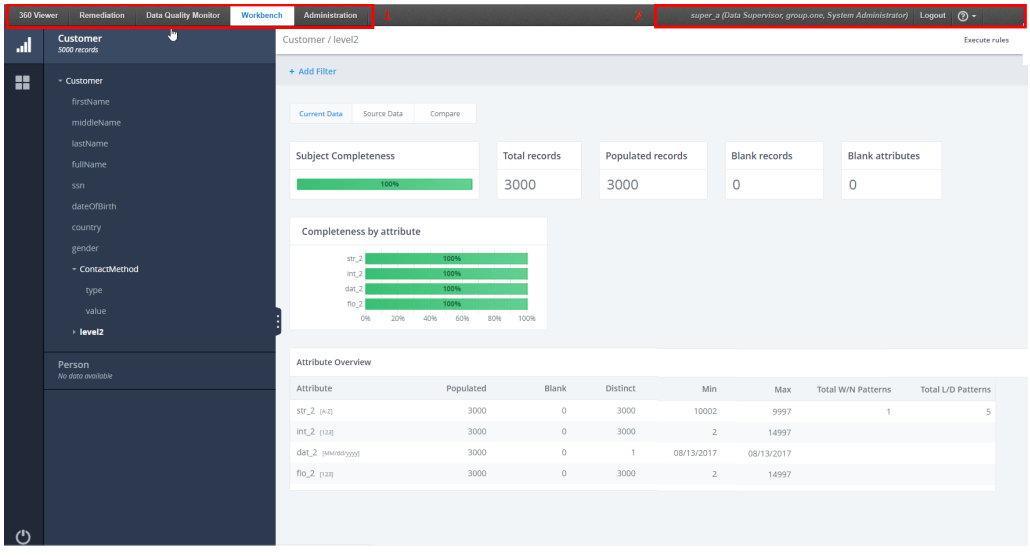
The Rules that define which bar chart will be highlighted are the same as for the Profile page.

Using Data Quality Workbench

This section describes the general options and functions of the Data Quality Workbench User's Guide.

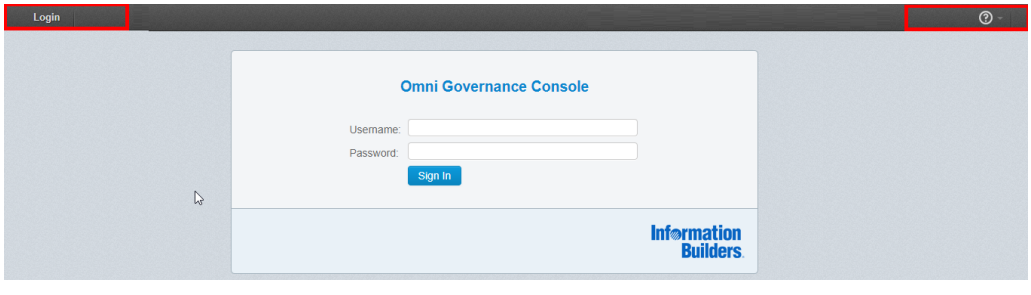
Menu Items

The menu bar is represented with dark grey line in upper part of the page. The available tabs are displayed on the left part of the menu, while your assigned roles, the Logout button, and the About button are shown on the right.

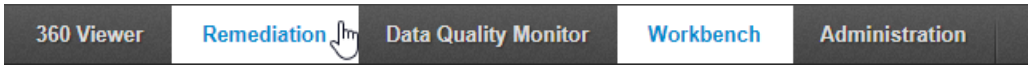


All pages available on the left pane are displayed according to your permissions that were configured in the WSO2 service.

If you are not logged in, then only the About button will be accessible.



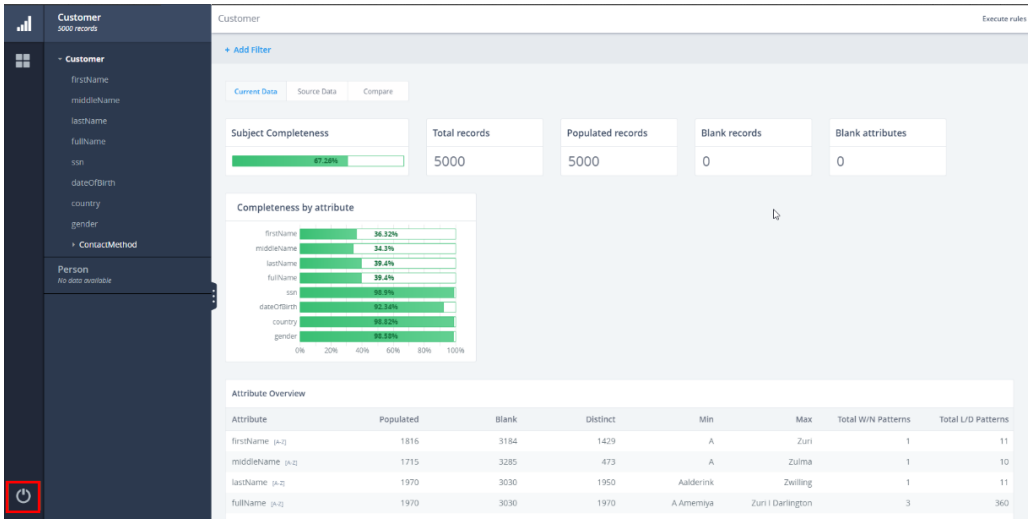
When you hover the mouse cursor over a different tab, it will appear in the same way as an active or current tab that is highlights, showing a white background.



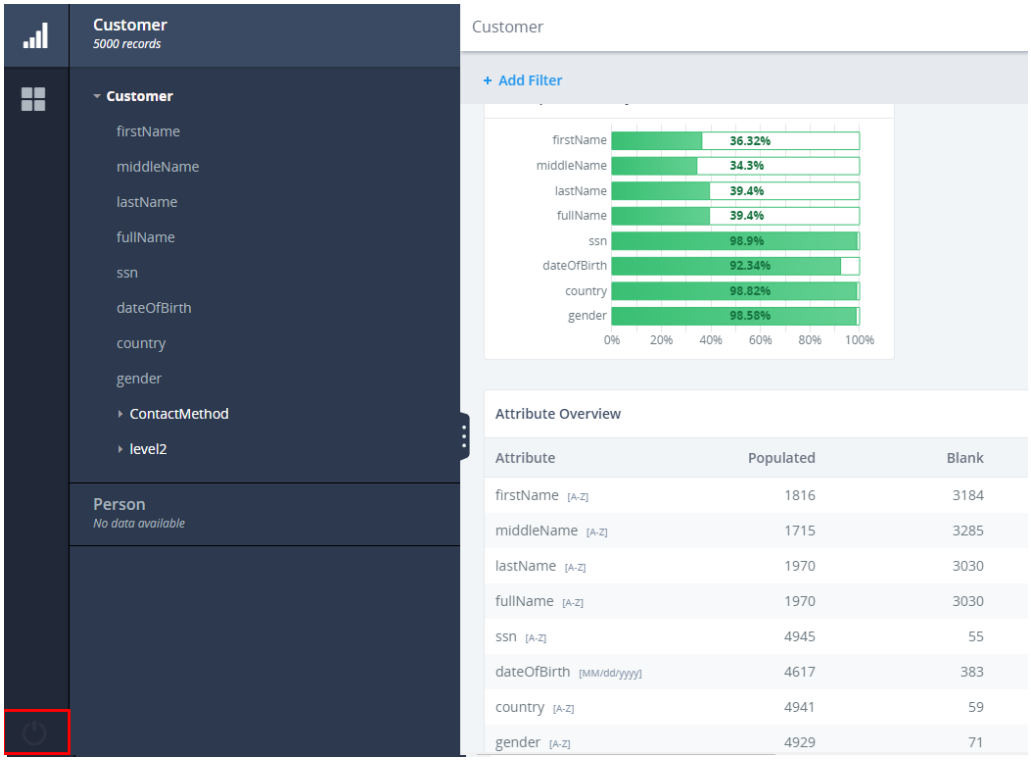
Reset Session

The Reset session button is located in the bottom left corner of the page. The Reset session button is used to return the Workbench to the default state without removing all added rules individually.

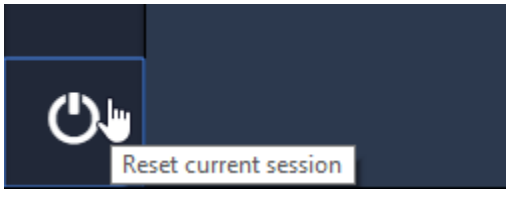
The Reset session button is also available from the Profile and Data Table pages.



When you initially navigate to the Workbench, and no changes were made, the Reset session button will be disabled, indicating that the functionality has already been performed or that there is no data available in the Workbench.



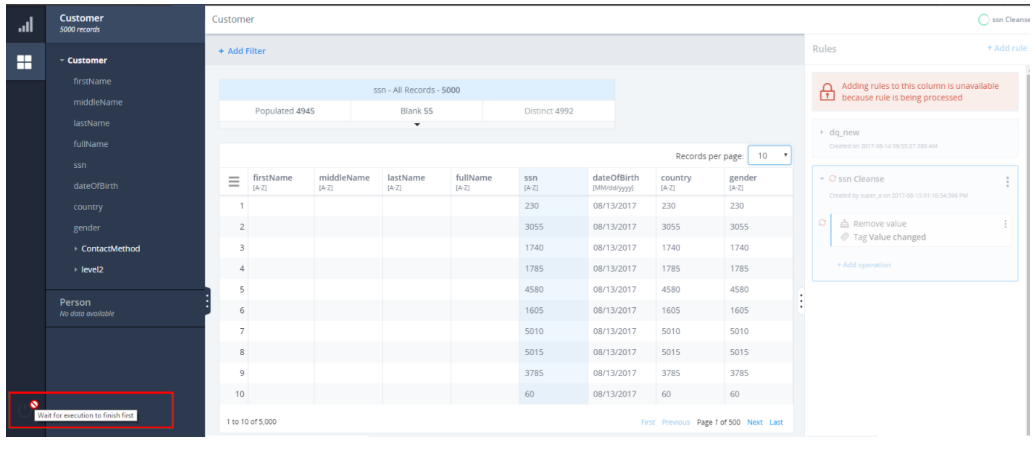
When you hover the mouse cursor on the Reset session button, the *Reset current session* tooltip appears, as shown in the following image.



The Reset session button is enables only after adding a Rule or a Filter. When you click the Reset session button, the following confirmation dialog appears:

After clicking *Reset*, the workbench will be reloaded and all added filters and underplayed rules, either newly created or executed, will be removed and the state of the Workbench application will return to the default setting. The Reset session button then becomes disabled.

The Reset session button is not always available. For example, you cannot reset the session while Rules are processing. If you try to reset it during the rules runtime, a tooltip appears informing you to *Wait for the execution to finish first.*



Uncommitted Icon

The Uncommitted icon appears if a change is made for the rule and operation. It will appear on the CRUD rule/operation.

There is no uncommitted icon for external or deployed rules (operations).

Saving State

The state of the Workbench is automatically saved until the session times out or when you log out. Additionally, if you refresh the page, the current filters and sorting preferences that were set will be removed.

Loader

Whenever a request is being sent (create/delete rule, operation, or another action which takes some time), the loader will appear for the component that has not yet been loaded.

Feedback

Customer success is our top priority. Connect with us today!

Information Builders Technical Content Management team is comprised of many talented individuals who work together to design and deliver quality technical documentation products. Your feedback supports our ongoing efforts!

You can also preview new innovations to get an early look at new content products and services. Your participation helps us create great experiences for every customer.

To send us feedback or make a connection, contact Sarah Buccellato, Technical Editor, Technical Content Management at *Sarah_Buccellato@ibi.com.*

To request permission to repurpose copyrighted material, please contact Frances Gambino, Vice President, Technical Content Management at *Frances_Gambino@ibi.com.*

Information Builders

iWay

Data Quality Workbench User's Guide

Version 3.2 and Higher

DN3502311.0119

Information Builders, Inc.
Two Penn Plaza
New York, NY 10121-2898