



ONE – Workshop

DQ Monitoring Project

Prepared for: v15.4

Prepared by: Ataccama

Dated: November 2024

Contents of the Document

Introduction	3
Tasks	3
1. Create a Monitoring Project	3
2. Review the DQ results in the Monitoring Project	8
Conclusion	12

Introduction

In this ONE workshop, you will create a Monitoring Project to monitor the data quality of your catalog items and review the results in the report.

Tasks

1. Creating a Monitoring Project

In the first task of this lab, we will create a **Monitoring Project** using the catalog items from the previous workshops. In the context of ONE Web application, *Projects* consist of selected Catalog items and allow regular monitoring of their data quality. You can assign new rules to the attributes of these Catalog Items. They can also be used for detecting anomalies or structural changes.

Let's start by creating a new **Monitoring Project** and selecting the data items for it:

- › Navigate to the **Data Quality** section - **Monitoring Projects**.
- › Click on the **Create** button in the upper right corner.
- › Fill in the Project's **Name** (e.g. **<prefix>_Training project**) and optionally a **Description**.
- › Confirm your changes by clicking the **Save** button.

Next, we need to choose the input data for it. Let's add several items from the Knowledge Catalog:

- › Within your new Monitoring Project, navigate to the **Configuration & Results** tab
- › Click the **Add Catalog Items** button located at the bottom.

Data Quality

- Rules
 - Detection Rules
 - DQ Evaluation Rules
- DQ Dimensions
- Components
- Transformation Plans
- DQ Firewalls
- Monitoring Projects
- Reconciliation Projects
- Lookup Items

Monitoring projects

xy_Training Project

Overview

Configuration & Results

Report

Export

Notifications

History

Run monitoring

Overview

Latest

Finish configuration first

Tables 0

Attributes 0

Checks 0

Structure

No checks applied

Anomaly Detection

No checks applied

Data Quality

No checks applied

Configuration of Invalid Results Samples

Not configured yet

To get results first add catalog items you want to monitor, then assign required checks to important attributes.

Add catalog items

› Select the following two catalog items from the list:

- **Customers**
- **Products**

Add datasets you want to monitor

Type here to search full-text for Catalog items

Terms

Data Quality

Data Source

Location

Number of Attributes

Number of Records

Profile Date

Anomalies

Suggested terms

Stewardship

Standard view

Hidden columns

	Name	Terms	# Attributes	# Records	Origin	Location
<input checked="" type="checkbox"/>	customers	Country First name Surname +5	15	122	postgres	Sources
<input checked="" type="checkbox"/>	products	xy_Product Line xx_Scale	9	110	postgres	Sources
<input type="checkbox"/>	orders	-	8	326	postgres	Sources
<input type="checkbox"/>	CREDIT_CARDS_ATTRITED	Gender Marital status Personal Data	23	1 303	DO_Snowflake	Sources
<input type="checkbox"/>	US SEC_FAILS TO DELIVER	CUSIP Number Stock Symbol Finance	6	14 364 675	DO_Snowflake	Sources

Add

Cancel




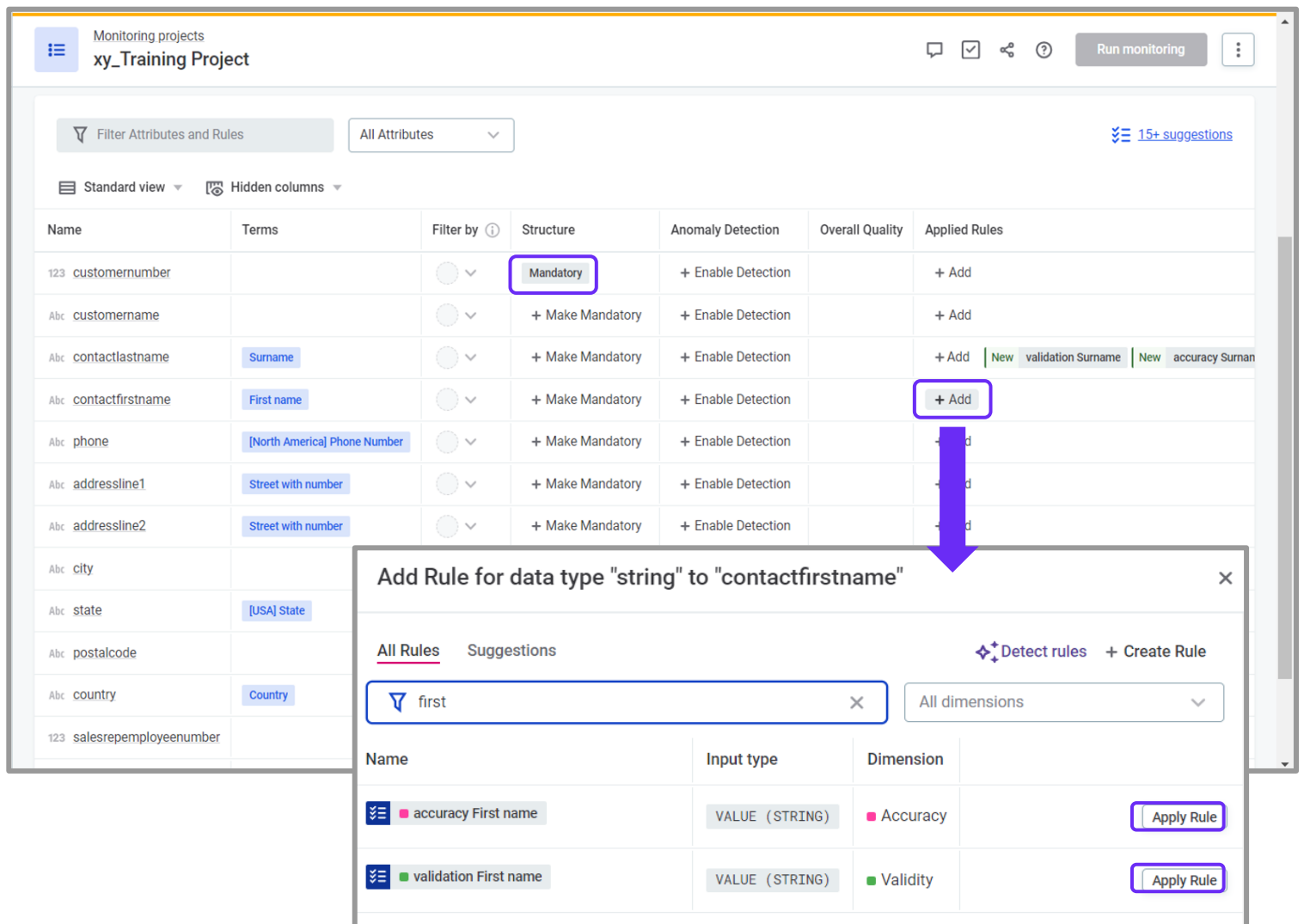
HINT

If you have multiple catalog items with identical names, you can open the details of each item in an additional panel on the right via the **Details** button or button. The panel shows the source's **Location** and other information.

Once the items are added, you can browse and configure its monitoring features. By clicking on each of the Catalog Items you can configure the data quality rules, and turn on the anomaly detection and structure checks for it.

Let's start configuring your data evaluation:

- ▶ In the **Configuration & Results** tab, open the **customers** Catalog Item.
- ▶ Choose the primary key column (e.g. **customernumber**) and turn **Structure checks** on for it. This will make the attribute mandatory for evaluation.
- ▶ You can enable the **Anomaly Detection Check** for this Catalog item – click on the  button for any attribute of your choice (e.g. **creditlimit**).
- ▶ Choose some of the attributes (e.g. **firstname**, **surname**, and **e-mail**) and add some **DQ Checks** – e.g. add **accuracy First name** and **validation First name** rules and optionally validation for the **e-mail** attribute. Use the filter to search through the configured rules:



Monitoring projects
xy_Training Project

Filter Attributes and Rules All Attributes 15+ suggestions

Standard view Hidden columns

Name	Terms	Filter by	Structure	Anomaly Detection	Overall Quality	Applied Rules
123 customernumber			Mandatory	+ Enable Detection		+ Add
Abc customername			+ Make Mandatory	+ Enable Detection		+ Add
Abc contactlastname	Surname		+ Make Mandatory	+ Enable Detection		+ Add New validation Surname New accuracy Suman
Abc contactfirstname	First name		+ Make Mandatory	+ Enable Detection		+ Add
Abc phone	[North America] Phone Number		+ Make Mandatory	+ Enable Detection		+ Add
Abc addressline1	Street with number		+ Make Mandatory	+ Enable Detection		+ Add
Abc addressline2	Street with number		+ Make Mandatory	+ Enable Detection		+ Add
Abc city						
Abc state	[USA] State					
Abc postcode						
Abc country	Country					
123 salesremployeenumber						

Add Rule for data type "string" to "contactfirstname"

All Rules Suggestions Detect rules + Create Rule

first All dimensions

Name	Input type	Dimension	
accuracy First name	VALUE (STRING)	Accuracy	Apply Rule
validation First name	VALUE (STRING)	Validity	Apply Rule

- Once you have finished configuring the item, **Publish** the project to save the changes.



Finish it on your own. Go through the other catalog items and add some DQ checks to it based on your demands or interests. Don't forget to publish the changes afterwards!

Name	Terms	Filter by ⓘ	Structure	Anomaly Detection	Overall Quality	Applied Rules
customernumber		⊙ ▼	Mandatory	+ Enable Detection		+ Add
customername		⊙ ▼	+ Make Mandatory	+ Enable Detection		+ Add
contactlastname	Surname	⊙ ▼	+ Make Mandatory	+ Enable Detection		+ Add New validation Surname New accuracy Surname
contactfirstname	First name	⊙ ▼	+ Make Mandatory	+ Enable Detection		+ Add New validation First name New accuracy First name
phone	[North America] Phone Number	⊙ ▼	+ Make Mandatory	+ Enable Detection		+ Add
addressline1	Street with number	⊙ ▼	+ Make Mandatory	+ Enable Detection		+ Add
addressline2	Street with number	⊙ ▼	+ Make Mandatory	+ Enable Detection		+ Add
city		⊙ ▼	+ Make Mandatory	+ Enable Detection		+ Add
state	[USA] State	⊙ ▼	+ Make Mandatory	+ Enable Detection		+ Add
postalcode		⊙ ▼	+ Make Mandatory	+ Enable Detection		+ Add
country	Country	⊙ ▼	+ Make Mandatory	+ Enable Detection		+ Add
salesremployeenumber		⊙ ▼	+ Make Mandatory	+ Enable Detection		+ Add
creditlimit		⊙ ▼	+ Make Mandatory	Enabled		+ Add
src_sys_id		⊙ ▼	+ Make Mandatory	+ Enable Detection		+ Add
email	E-mail	⊙ ▼	+ Make Mandatory	+ Enable Detection		+ Add New validation E-mail

If you are happy with the project configuration and published the changes, you can click on **Run monitoring** to see the DQ results.

Monitoring projects

xy_Training Project

ⓘ

✓

🔗

?

Run monitoring

⋮

Overview

Configuration & Results

Report

Export

Notifications

History

xy_Training Project > customers

Latest

⌚

Run the project to see results

Attributes 15

Checks 8

Structure ⓘ

⌚ Run monitoring

1 check

Anomaly Detection ⓘ

⌚ Run monitoring

1 attribute and 1 catalog item check

Data Quality ⓘ

⌚ Run monitoring

5 checks

Once triggered, you can see the current progress of the monitoring in the **Processing center**. Progress can also be monitored in the project itself:

Q

Data Quality

▼

Rules

▼

Detection Rules

▼

DQ Evaluation Rules

▶

DQ Dimensions

▶

Components

▶

Transformation Plans

▶

DQ Firewalls

▶

Monitoring Projects

▶

Reconciliation Projects

▶

Lookup Items

Monitoring projects

xy_Training Project

Overview

Configuration & Results

Report

Export

Notifications

History

xy_Training Project

>

customers

≡

November 21, 2024, 1:28:23 AM (Latest)

⋮

Monitoring project is running

Attributes 15

Checks 8

Structure

OK

1 check

Anomaly Detection

Running

1 attribute and 1 catalog item check

Data Quality

Running

5 checks

Filter Attributes and Rules

All Attributes

Check for Rule suggestions

Standard view

Hidden columns

Name	Terms	Filter by	Structure	Anomaly Detection	Overall Quality	Applied Rules
123 customernumber		⋮	Mandatory	+ Enable Detection		+ Add
Abc customername		⋮	+ Make Mandatory	+ Enable Detection		+ Add
Abc contactlastname	Surname	⋮	+ Make Mandatory	+ Enable Detection		+ Add New validation Surname
Abc contactfirstname	First name	⋮	+ Make Mandatory	+ Enable Detection		+ Add New validation First name
Abc phone	[North America] Phone Number	⋮	+ Make Mandatory	+ Enable Detection		+ Add

When you are on the Project’s **Configuration & Results** page, you can see the progress for all Catalog Items and DQ Checks. The results are being updated as the Monitoring continues.

2. Review of the DQ results in the Monitoring Project

In this task, we will go through the monitoring results of our Monitoring Project and have a look at some potential DQ issues.

After the monitoring run is finished, if any issue is found, you will see a warning message on the **Configuration & Results** tab.

You can notice that in our case the issues were found for only one table as expected (The **products** catalog item doesn't have any rules assigned to it yet).

- ▶ Click on the **configuration & Results > customers** to open the details.

You can now observe the issues identified from various data quality perspectives, including structure, anomalies, and adherence to data quality rules. Specifically, regarding data quality rules, you can find the pass/fail rate for each rule applied to individual attributes. Furthermore, you have the capability to dive deeply into the analysis of invalid samples.

- ▶ Click on **show invalid samples**.

The screenshot shows the Data Quality Monitoring Project interface. The left sidebar contains navigation options like Rules, Detection Rules, DQ Evaluation Rules, DQ Dimensions, Components, Transformation Plans, DQ Firewalls, Monitoring Projects, Reconciliation Projects, and Lookup Items. The main panel displays the 'Configuration & Results' tab for the 'xy_Training Project'. The 'Overview' section shows a green checkmark and the message 'Great! Your data is in good shape'. It also displays statistics: Tables 2, Attributes 24, and Checks 8. The 'Data Quality' section shows a green checkmark and the message 'OK' with '5 checks on 1 table'. A 'Show invalid samples' button is highlighted in the Data Quality section. The 'Items to monitor' section shows a table with columns: Name, Structure, Anomalies, DQ Rules, Overall, VAL, and ACC. The table lists two items: 'customers' and 'products'. The 'customers' item has 1 check, 2 checks, 5 DQ Checks, 11+ suggestions, 74% Overall, 74% VAL, and 84% ACC. The 'products' item has 9 structure checks, 1+ suggestion, and is not calculated yet for Overall, VAL, and ACC.

Name	Structure	Anomalies	DQ Rules	Overall	VAL	ACC
customers xy_training > postgres > public Personal Data	1 check 15 122	2 checks	5 DQ Checks 11+ suggestions	74%	74%	84%
products xy_training > postgres > public	9		1+ suggestion	Not calculated yet	Not calculated yet	Not calcul

In the new window that opens, you can go through the invalid data. The individual values will be highlighted in red color.

The **Filter** can also be applied to narrow down the list of displayed records. It can be configured with the following filtering options:

- By default, you see only **Attributes with failed rules**. This can be changed to display **Attributes with applied rules** or **All Attributes**.
- You can choose from the list of **Attributes** to show/hide the individual attributes you are interested in.
- You can specify which **Rules** you are interested in. That is useful especially when you have many DQ checks applied to the Catalog item.

You can also quickly access all the Catalog Items in your Monitoring Project – simply click on the Catalog item name in the left panel.

Invalid Samples

customers 74%

Sources > xy_training > postgres > public

Apply a filter

customername	contactlastname 84%	contactfirstname 91%	phone	addressline1
	Surname	First name	+1	Street with number
Signal Gift Stores	King	Jean IS_INVALID: validation... +1	7025551838	8489 Strong St
Baane Mini Imports	Bergu			
Havel & Zbyszek Co	Piest			
Mini Wheels Co.	Murph	IS_INV		
Porto Imports Co.	de Ca	IS_INV		
Daedalus Designs Im...	Ranc	IS_INV		
La Corne D'abondanc...	Bertr	IS_INV		

Apply a filter

DQ Rules (4/5)

Narrow down options by name

☐ Select All ☐ Select Only Failed

☐ Validity ☐ Completeness ☐ Uniqueness ☒ Accuracy ☐ Timeliness

☒ accuracy First name

☒ accuracy Surname

☐ validation E-mail

☒ validation First name

☒ validation Surname

Attributes (2/15)

Narrow down options by name

☐ Select All

☐ addressline1

☐ addressline2

☐ city

☒ contactfirstname

☒ contactlastname

☐ country

☐ creditlimit

☐ customernumber

☐ customernumber

Apply

When you are done with exploring the issues and their details, let's get back to the project.

- Close the **Invalid Samples** window.

Now that you are back to your monitoring project, let's modify it to monitor the **products'** Data Quality too. In this regard, we need to add rules to its attributes. Therefore, we decided to consider the application's suggestions.

- › Navigate to the **configuration & Results > products** and click on **suggestions**
- › Accept rule suggestions (**xy_Product Line, String Completeness**) for the **products** catalog item in your monitoring project.

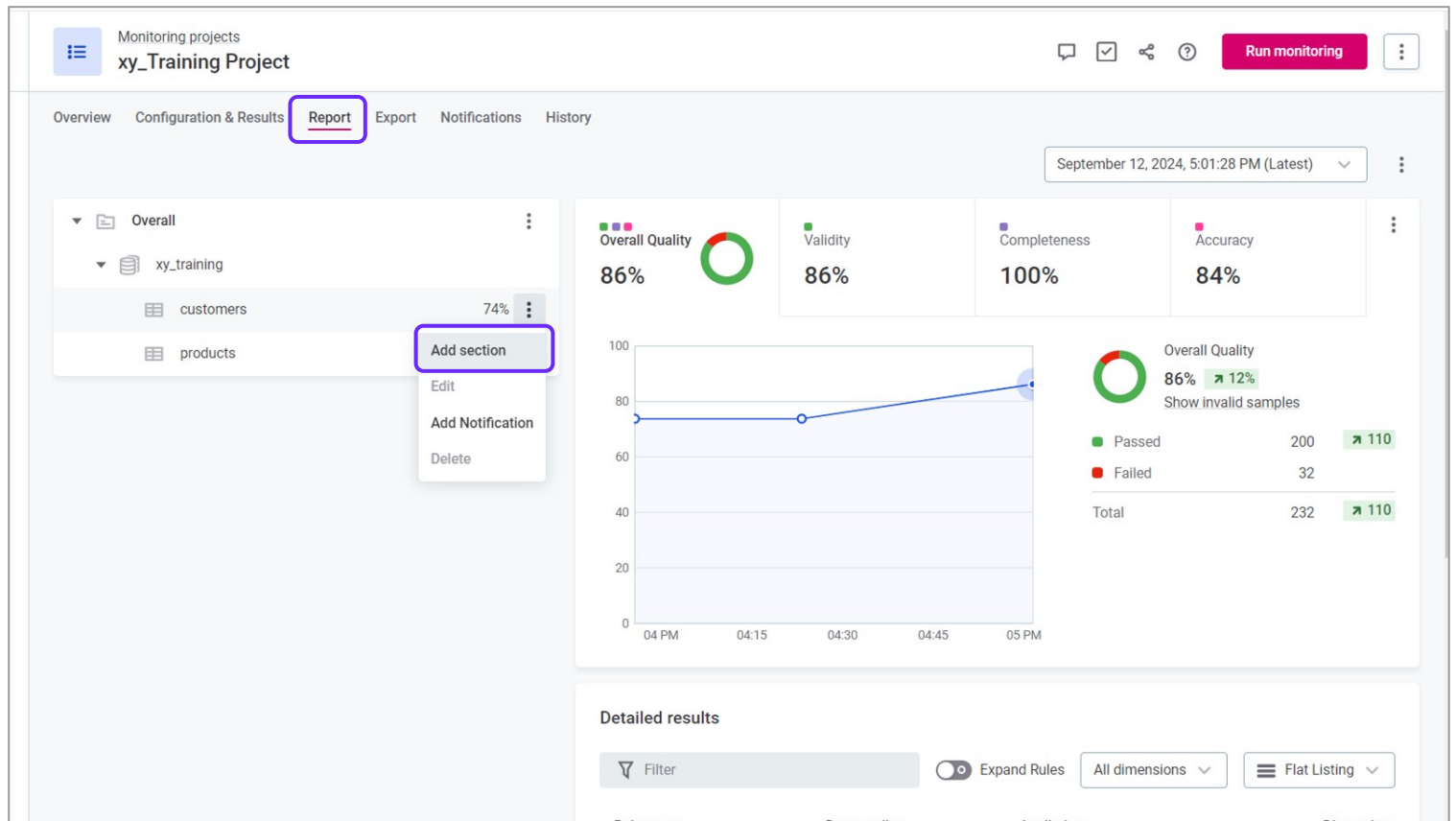
Rule suggestions			
<div> <div>products</div> <div>Sources > xy_training > postgres > public</div> </div>			
Name	Applied DQ Checks	Terms	Suggested Rules
▼ products			
<input type="checkbox"/> productcode			<div>Length not longer than ✓ X</div> <div>Pattern match ✓ X</div> <div>Not empty (String) ✓ X</div> <div>Mask match ✓ X</div> <div>Length not shorter than ✓ X</div>
<input type="checkbox"/> productname			<div>Length not longer than ✓ X</div> <div>Not empty (String) ✓ X</div> <div>Mask not match ✓ X</div> <div>Length not shorter than ✓ X</div> <div>Unique (String) ✓ X</div>
<input type="checkbox"/> productline	xy_Product Line		<div>Length not longer than ✓ X</div> <div>✓ xy_Product Line ↗</div> <div>Not empty (String) ✓ X</div> <div>Mask match ✓ X</div> <div>Length not shorter than ✓ X</div>
<input type="checkbox"/> productscale	xy_Scale		<div>Length not longer than ✓ X</div> <div>Pattern match ✓ X</div> <div>Not empty (String) ✓ X</div> <div>✓ String Completeness ↗</div> <div>Mask match ✓ X</div>
<input type="checkbox"/> productvendor			<div>Length not longer than ✓ X</div> <div>Not empty (String) ✓ X</div> <div>Length not shorter than ✓ X</div>
<input type="checkbox"/> productdescription			<div>Length not longer than ✓ X</div> <div>Not empty (String) ✓ X</div> <div>Mask not match ✓ X</div> <div>Length not shorter than ✓ X</div>
<input type="checkbox"/> quantityinstock			<div>Not empty (Integer) ✓ X</div> <div>Unique (Integer) ✓ X</div> <div>Less than or equal to (Integer) ✓ X</div> <div>Less than (Integer) ✓ X</div>
<input type="checkbox"/> buyprice			<div>Greater than or equal to (Float) ✓ X</div> <div>Greater than (Float) ✓ X</div> <div>Less than or equal to (Float) ✓ X</div> <div>Less than (Float) ✓ X</div>
<input type="checkbox"/> msrp			<div>Greater than or equal to (Float) ✓ X</div> <div>Greater than (Float) ✓ X</div> <div>Less than or equal to (Float) ✓ X</div> <div>Less than (Float) ✓ X</div>

- › **Publish and run** the monitoring project again.

Now you can recheck the results as well as the invalid samples on the configuration & Results tab.

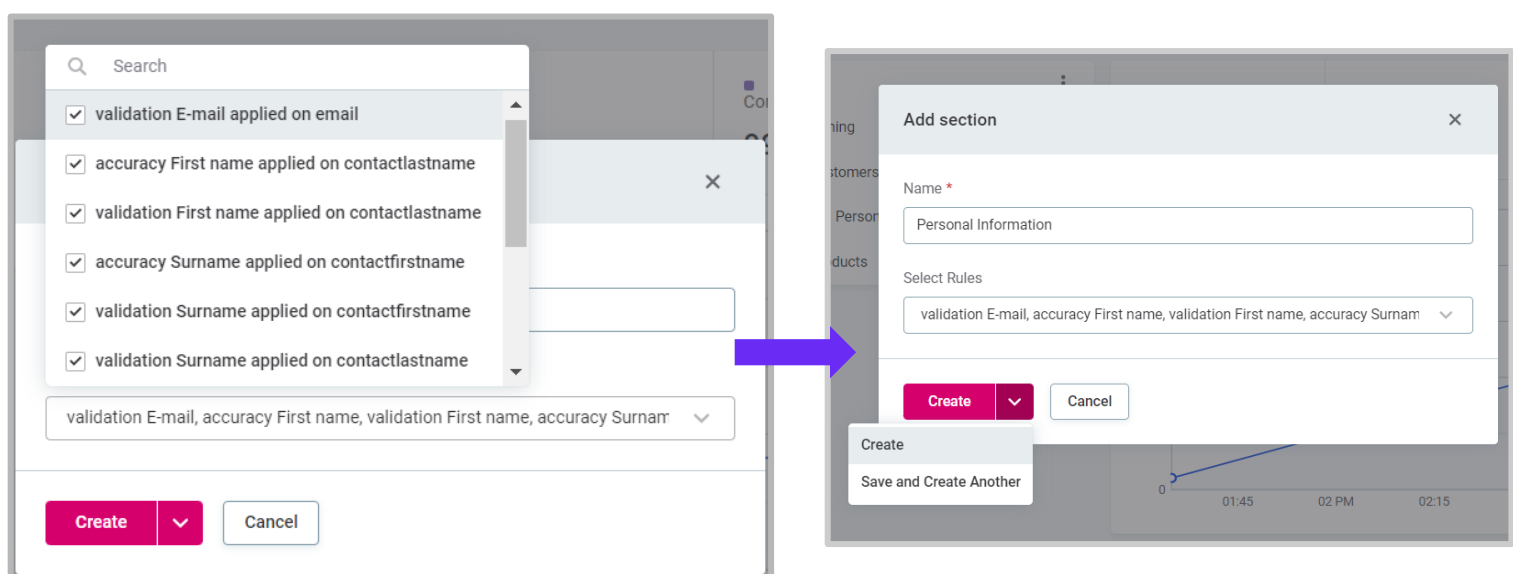
To find out more details about the evaluation, let's have a look at the **Report** that is generated during the monitoring and provides an overview of the data quality over time. There, you can also aggregate the rules into sections to generate multiple reports:

- › Go to the **Report** tab.
- › Expand the **Overall** node to all levels to reveal the full data structure.
- › Press the three dots next to the **customers** Catalog Item and choose the **Add Section:**



In the new popup window, the new Section of yours needs to have some properties defined:

- › Name the section, e.g., '**Personal information**'.
- › Use the **Add Rule** to choose from the available list of DQ rules and assign them to the personal information (First name, Surname, etc.)
- › Press the **Create** button to complete the creation of the **Section** of rules.





Anytime during the Group creation, you can click on the arrow and choose the '**Save and Create Another**' option. This will save the group and let you create a new one right away.

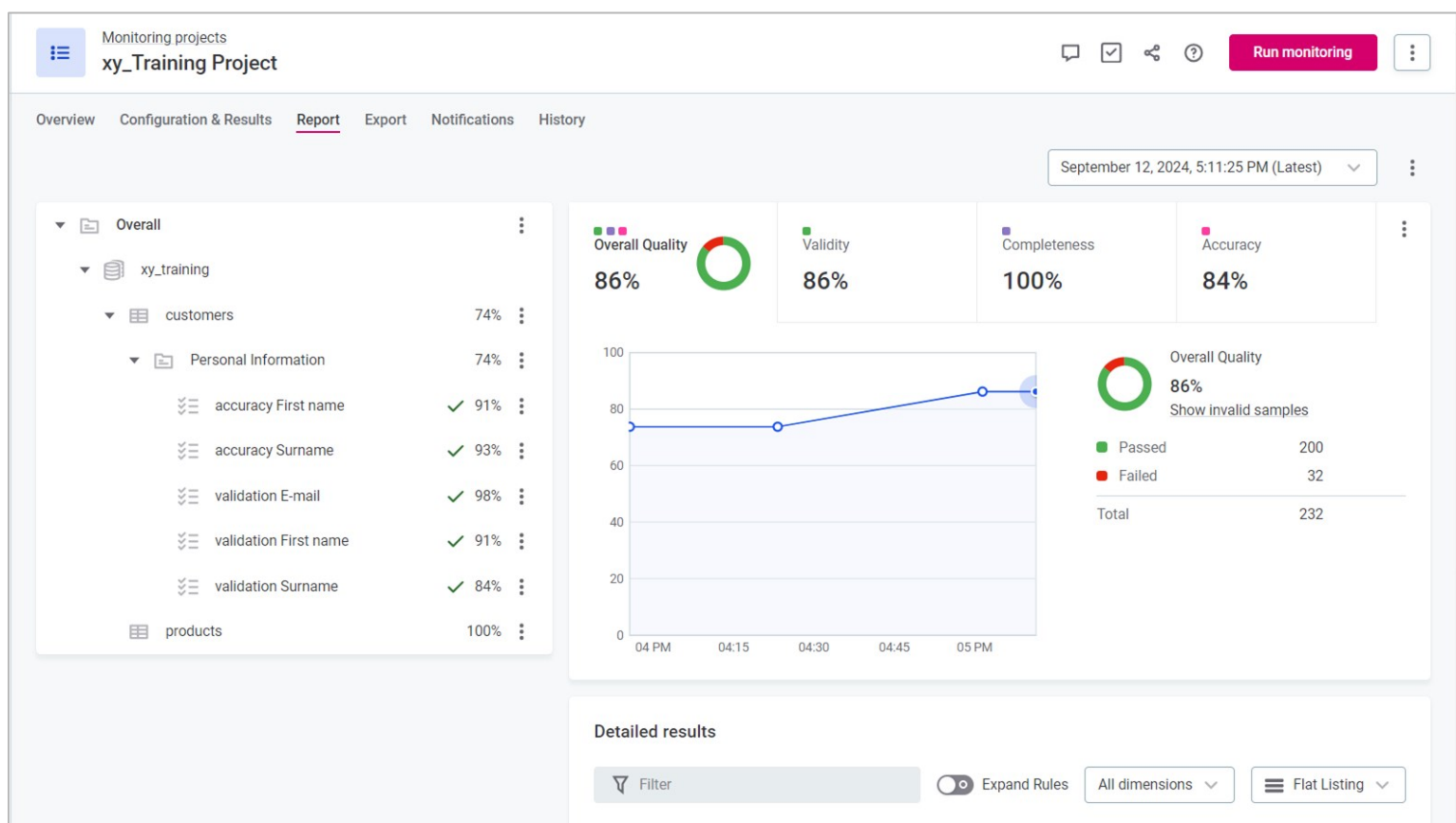
- › **Publish** the changes to your Monitoring Project.

If you now check the sections of your report, you can see that the results are not available until you run the monitoring again.



The monitoring will start for the whole project; thus, all Catalog items will be analyzed again.

- › Press the **Run Monitoring** button to re-evaluate the data in your project and navigate through different sections to see different reports you have generated.



Detailed results

Filter

☐ Expand Rules

All dimensions ▾

Flat Listing ▾

Rule name	Data quality	Applied on	Dimension
<div> <div>▼</div> <div>☰</div> accuracy Surname </div>	<div> <div>✓</div> <div> <div></div> <div>93%</div> </div> </div>	<div> <div>☰</div> customers <div>☰</div> <div> <div>></div> <div>contactlastname</div> <div>☰</div> </div> </div>	<div> <div>■</div> Accuracy </div>
<div> <div>93%</div> <div>113/122</div> <div>9/</div> </div> <div> <div></div> </div> <div> <div> <div>■</div> Accurate 113/122 93% </div> <div> <div>■</div> Not Accurate 0/122 0% </div> </div> <div> <div> <div>IS_ACCURATE</div> <div>113/122 93%</div> </div> <div> <div>OTHER</div> <div>0/122 0%</div> </div> </div>			
<div> <div>▼</div> <div>☰</div> validation Surname </div>	<div> <div>✓</div> <div> <div></div> <div>84%</div> </div> </div>	<div> <div>☰</div> customers <div>☰</div> <div> <div>></div> <div>contactlastname</div> <div>☰</div> </div> </div>	<div> <div>■</div> Validity </div>
<div> <div>84%</div> <div>103/122</div> <div>19/</div> </div> <div> <div></div> </div> <div> <div> <div>■</div> Valid 103/122 84% </div> <div> <div>■</div> Invalid 19/122 16% </div> </div> <div> <div> <div>OTHER</div> <div>103/122 84%</div> </div> <div> <div>IS_INVALID</div> <div>19/122 16%</div> </div> </div>			
<div> <div>▶</div> <div>☰</div> accuracy First name </div>	<div> <div>✓</div> <div> <div></div> <div>91%</div> </div> </div>	<div> <div>☰</div> customers <div>☰</div> <div> <div>></div> <div>contactfirstname</div> <div>☰</div> </div> </div>	<div> <div>■</div> Accuracy </div>
<div> <div>▶</div> <div>☰</div> validation First name </div>	<div> <div>✓</div> <div> <div></div> <div>91%</div> </div> </div>	<div> <div>☰</div> customers <div>☰</div> <div> <div>></div> <div>contactfirstname</div> <div>☰</div> </div> </div>	<div> <div>■</div> Validity </div>
<div> <div>▶</div> <div>☰</div> validation E-mail </div>	<div> <div>✓</div> <div> <div></div> <div>98%</div> </div> </div>	<div> <div>☰</div> customers <div>☰</div> <div> <div>></div> <div>email</div> <div>☰</div> </div> </div>	<div> <div>■</div> Validity </div>

Conclusion

We have come to the end of this Ataccama ONE workshop focused on Projects.

We went through the process of creating monitoring projects, learned how to configure them, how to interpret the results, and configured the exports of the DQ results.

Congratulations, you are now ready to start developing your own Ataccama ONE project!