

## Module 3

# Analyzing Profiling Results

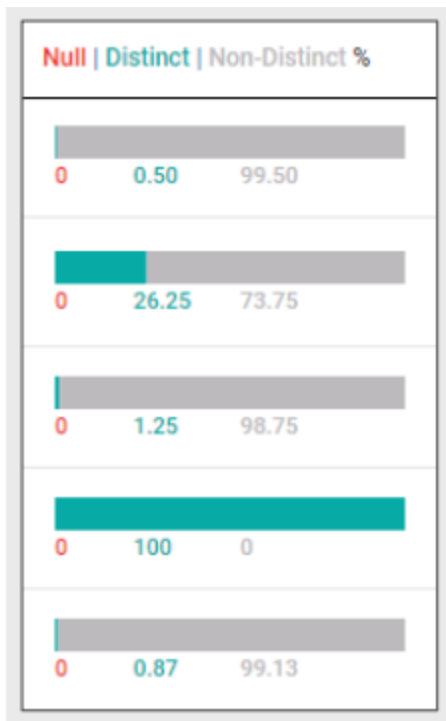
## Profiling Results :

### Why do we need Data Profiling?

- Data Profiling enables you to analyze the data within the asset to determine what it looks like and what problems exist within it

تحليل البيانات الموجوده في الداتا عندي ومعرفة المظهر العام و تحديد المشاكل

- Find data issues before they become major data problems
- Uses profiling to check the quality of data for the asset
- Displays percentage of null values, distinct values, and non-distinct values in each field
- View number of rows associated with each percentage of values



اللون الاحمر يرمز اللي ال Nulls بمعنى كم من خلية فارغة

القيم الاخضر الفاتح ترمز الي القيم الفريدة

القيم الرصاصي ترمز الي القيم المتكررة

## What Profiling Results can you view for an Asset?

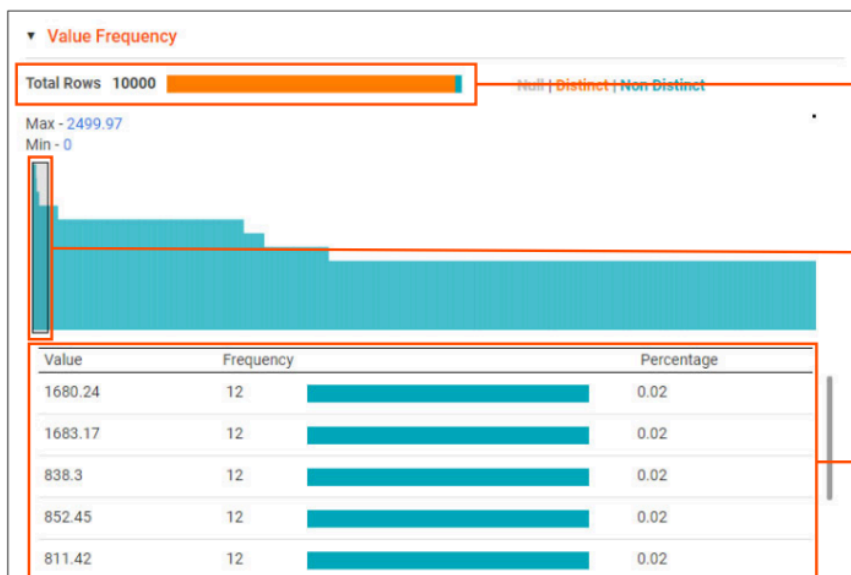
- Value Frequency • Similar Columns • Patterns • Unique Keys • Inferred Data Types

### 1- Value Frequency :

View following information:

- distribution of values across data asset
- maximum and minimum values in the data asset
- distribution of distinct values in the data asset
- list of values and their frequency in a table

#### Value Frequency (continued)



Value distribution chart

Slider

Value frequency table

9

© Informatica. Proprietary and Confidential.



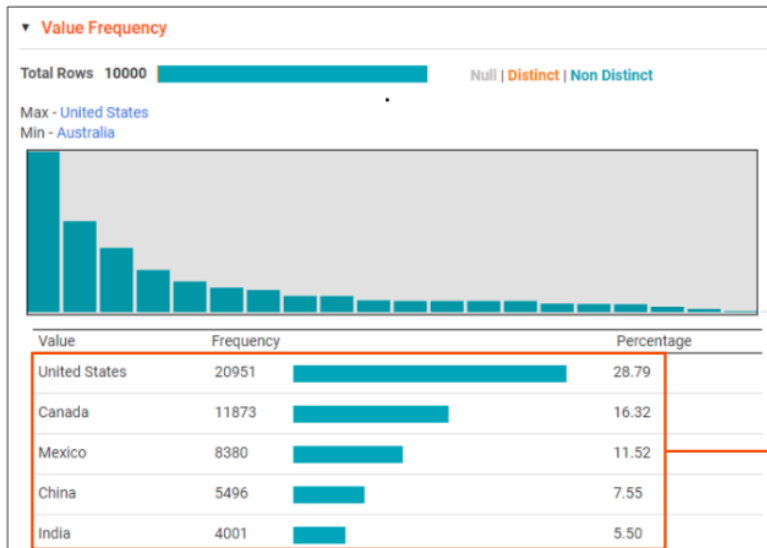
### Value Frequency Chart for Columns with Unique Values

"Pasted image 20250610155501.png" could not be found.

هنا ده البروفایلنج بتاع عمود ال ID فال Frequency كلها 1 عشان ده عمود primary key

### Value Frequency Chart for Columns with Non-Distinct Values

# Value Frequency Chart for Columns with Non-Distinct Values



Frequency of each country value (in values and in percentages)

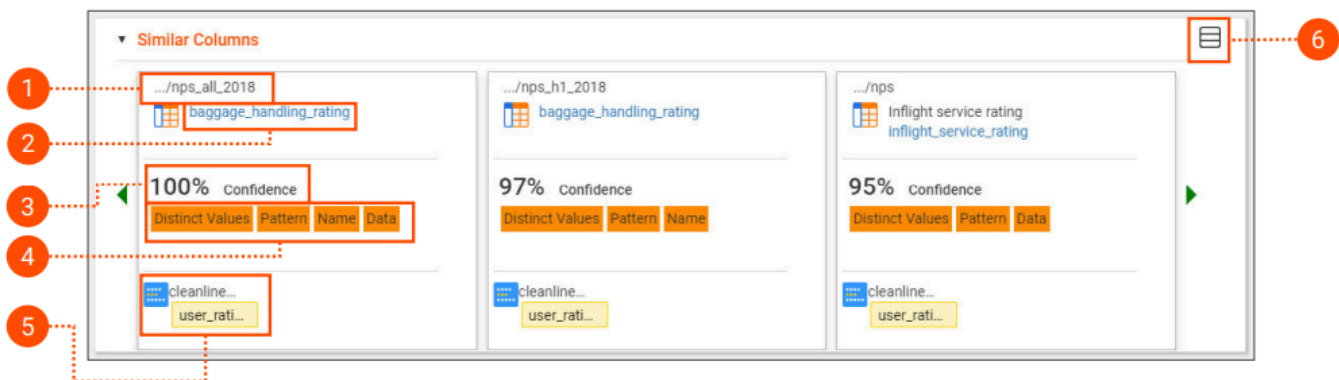
11

© Informatica. Proprietary and Confidential.



## Similar Columns :

- Displays columns that are similar to the current column
- Similarity is based on column names, column patterns, unique values, and value frequencies



1. Name of the resource
2. Name of the column
3. Percentage of Confidence
4. Parameters based on which the percentage is calculated
5. Data domains inferred or assigned to column
6. Toggle switch to view the similar columns as a list or as blocks.


7. Note: Current view is as blocks.

View the following information about columns in the Similar Columns section:

- Column Name: Displays the column name.
- Business Term: Displays the associated business term.
- Data Domain: Displays the inferred data domains for the column and the percentage of rows that conform to the data domain.
- Confidence %: Displays the score and the parameters based on which the score is calculated.

## Unique Keys :

- You can generate unique key from the columns in the source data object
- View unique key interference results to understand data quality issues
- You can accept or reject the inferred unique interference results
- You can also reset the unique key interference results to restore the inferred status


 Add Business Title FREQUENT\_FLIERS 5 stars (5) Follow Refresh Close

>> Frequent\_Fliers > infaorcl > SRC\_OMNIBUS

OverviewColumnsKeysLineage and ImpactRelationshipsReviewsQuestions

Unique Keys

Show DocumentedFind

Key Name	Columns	Description	Confidence Score	Type
1 FRONT_FLIER_PK	FREQUENT_FLIER_ID		100%	 Documented

1-1 of 1< 1 of 1 > 15

## Curating Unique Keys :

You can accept or reject inferred unique keys

- Orange – inferred, Green – accepted, Red – rejected
- You can undo curation and bring it back to the inferred state
- You can filter the list to view:
  - All keys
  - Documented keys
  - Inferred keys
  - Accepted keys
  - Rejected keys

Add Business Title  
CUSTOMERS  
★★★★★ (1)  
Loyalty\_Store > Infaorci > TDWSRC

Follow | | |

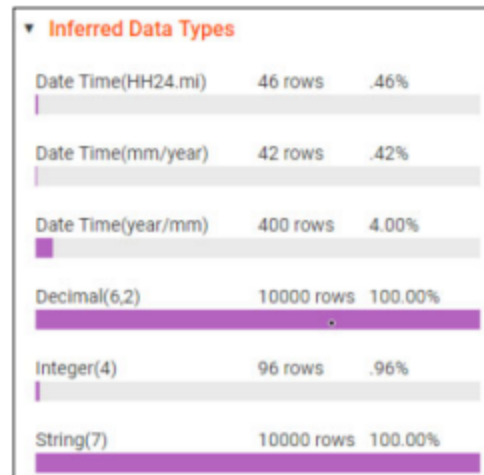
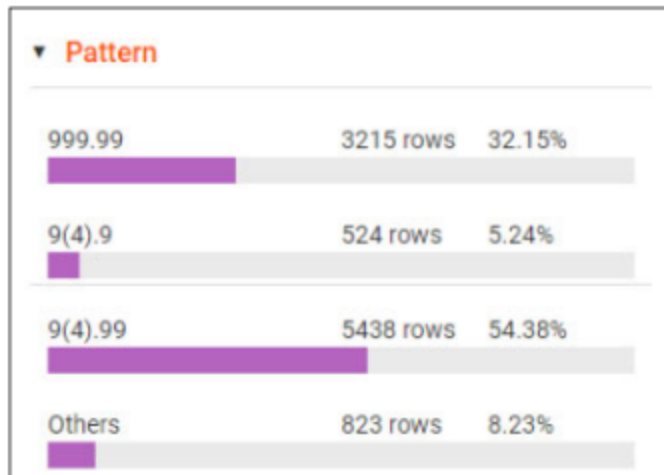
Overview Columns **Keys** Lineage and Impact Relationships Reviews Questions

**Unique Keys** Show All Find

Key Name	Columns	Description	Confidence Score	Type	
1 SYS_C00393903			100%	Documented	
2 UK_CUSTOMERS_CREDITLIMIT	CREDITLIMIT		99.2%	Inferred	
3 UK_CUSTOMERS_CUSTOMER_ID	CUSTOMER_ID		100.0%	Accepted	Undo Accept
4 UK_CUSTOMERS_PHONE_NUMB...	PHONE_NUMBER		99.6%	Accepted	Undo Accept
5 UK_CUSTOMERS_ADDRESS_LIN...	ADDRESS_LINE1		99.4%	Rejected	Undo Reject
6 UK_CUSTOMERS_FIRST_NAME	FIRST_NAME		99.8%	Rejected	Undo Reject

1 - 6 of 6 < 1 of 1 > 15

## Patterns and Inferred Data Types :



هنا يوجد رمزين للانماط:

X : وتعني حرف

9: تعني رقم

امثلة :

X(5) : Ahmed

X(4) : Loka

X(4)@X(5).com : [loka@gmail.com](mailto:loka@gmail.com)

X(2)9(3): XX126

يوجد قاعدة ان ما بعد رقم 4 يكتب كرقم  
توضيح

تم كتابة الحرف 3 مرات XXX  
تم كتابة الحرف و الرقم Ahmed : X(5)

امثلة علي الارقام

9(4).99: meaning

ان 4 ارقام قبل العلامة و رقمين بعد العلامة 2421.45