

# 11. Análise da Qualidade dos Dados

Nossa solução se utiliza do InfoSphere® Information Analyzer com o intuito de fornecer recursos para traçar o perfil e analisar dados para fornecer informações confiáveis aos usuários.

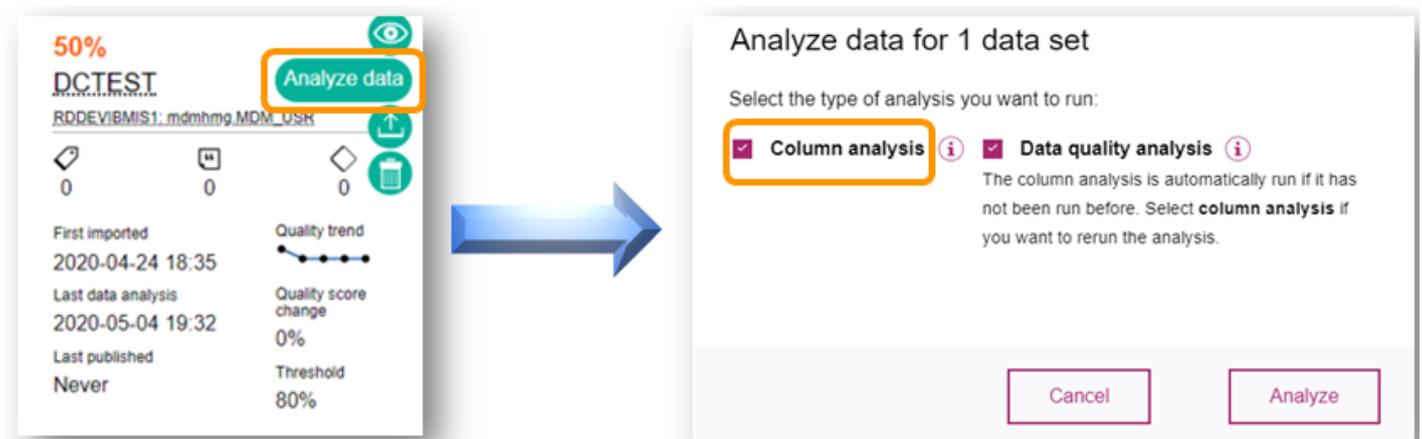
Por intermédio de amostras e volumes completos de dados a qualidade e estrutura dos mesmos é avaliada. A análise das informações geradas permite que sua organização corrija problemas de estrutura ou validade antes que afetem seu projeto de integração de dados.

O Information Analyzer foi desenvolvido para auxiliar o analista de negócios e de dados a entender o conteúdo, a qualidade e a estrutura das fontes de dados através de um processo de análise automática dos dados.

A ferramenta possui interface amigável e diversas formas de análise onde o usuário só necessita apontar as fontes de dados os quais pretende analisar.

A arquitetura da suite IBM InfoSphere Information Server permite que os resultados das análises sejam compartilhados com as outras ferramentas da suite. A partir desse compartilhamento, é possível utilizar o IBM InfoSphere Information Governance Catalog para realizar consultas e relatórios Ad-Hoc de forma prática.

Seguem imagens ilustrando exemplos de utilização da ferramenta no contexto da solução MD2 MDM.



O passo inicial é executar Column Analysis nas fontes de dados que deseja analisar.



Os resultados do Column Analysis geram indicadores e informações relacionadas a qualidade dos dados analisados.

The first screenshot shows the 'Data quality' section for dataset 'DCTEST'. It displays a 'Data quality' score of 50% against a threshold of 80%. The last data quality analysis was performed on 2020-05-22 at 18:33. Below this, there are sections for 'Data Type Violations' (2 findings | 50%) and 'Data Class Violations' (No violations).

The second screenshot shows the 'Data quality' section for dataset 'POC\_RDSL\_LGPD'. It displays a 'Data quality' score of 99% against a threshold of 80%. The last data quality analysis was performed on 2020-05-04 at 20:52. Below this, there is a table showing 'Data Type Violations' and 'Data Class Violations'.

Os indicadores e informações podem ser consultados de forma rápida e interativa na interface web da ferramenta.

This screenshot shows a detailed view of the 'Column properties' for dataset 'POC\_RDSL\_LGPD'. The interface displays a table with the following columns: COLUMN, DATA CLASS, TERM, FORMAT, NULLABLE, UNIQUENESS, MINIMUM, MAXIMUM, and DISTINCT VALUE.

COLUMN	DATA CLASS	TERM	FORMAT	NULLABLE	UNIQUENESS	MINIMUM	MAXIMUM	DISTINCT VALUE
<input type="checkbox"/> NR_CEP		-	99999999	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0	99600000	8445
<input type="checkbox"/> CD_DDI		-	NA	<input checked="" type="checkbox"/>	<input type="checkbox"/>	null	null	0
<input type="checkbox"/> CD_DDD		-	NA	<input checked="" type="checkbox"/>	<input type="checkbox"/>	null	null	0
<input type="checkbox"/> NR_TEL		-	9999999999	<input checked="" type="checkbox"/>	<input type="checkbox"/>	(2) 2533-5321	999995723	7792
<input type="checkbox"/> TX_CTT_ELET	EMAIL-Regex	-	NA	<input checked="" type="checkbox"/>	<input type="checkbox"/>	--	zonasuladm07...	7164

É possível que o usuário escolha a granulometria da avaliação (Tabela, coluna).

This screenshot shows the search and filter options for a data asset in the IBM InfoSphere Information Governance Catalog. The interface includes a search bar, a list of search criteria, and a table of search results.

**TIPO DE RECURSO:** Coluna do Banco de Dados

**PROPRIEDADES DISPONÍVEIS:**

- Nome
- Descrição Curta
- Descrição Detalhada
- Tabela ou Visualização de Banco d
- Pontuação da qualidade
- Rótulos

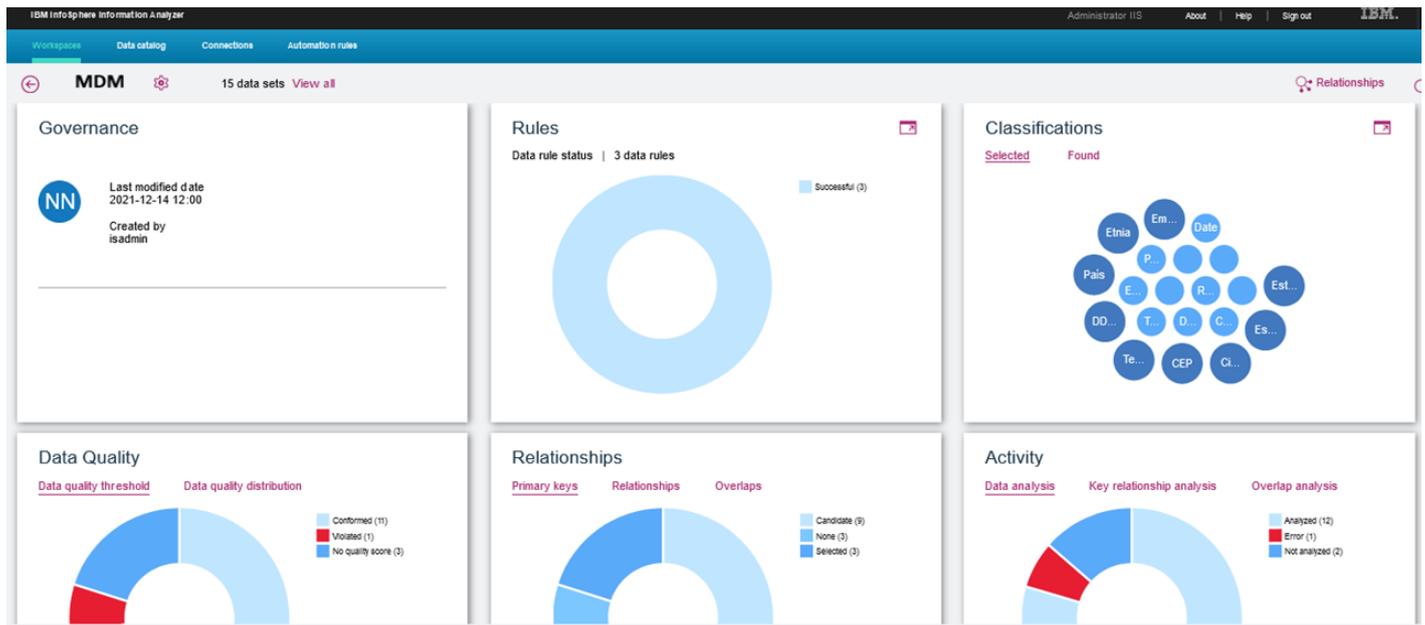
**EXIBIR CRITÉRIOS**

**PROPRIEDADES EXIBIDAS:**

- Nome
- Tabela ou Visualização de Banco de Dados - Nome
- Pontuação da qualidade
- Número de Formatos Distintos
- Número de Valores Distintos
- Comprimento Inferior
- Formato Inferior

Exibir apenas os valores de propriedade que correspondem aos critérios de consulta

A partir da publicação dos resultados é possível utilizar o módulo de Consultas do Information Governance Catalog para gerar relatórios Ad Hoc acerca das análises de qualidade realizadas pelo IBM Information Analyzer.



A ferramenta também possibilita a gestão das análises de profiling e qualidade de dados realizadas por intermédio de um dashboard presente na interface Web do Information Analyzer e também pela possibilidade de gerar reports pelo cliente da aplicação.

Revision #8

Created 7 November 2022 00:09:04 by Gustavo Santos

Updated 13 December 2022 11:01:47