INSTITUTO FEDERAL
Brasília

Colunas às quais tive acesso = cor preta

Colunas às quais não tive acesso, mas sei que existem por causa do projeto de pesquisa = cor vermelha

Documentação das dimensões da qualidade dos dados - Sistec IFB

Tipos: Varchar, Strings e Char

valida_texto_nao_nulo (params)

Colunas validadas
3. unidade_ensino
4. dependencia_adm
5. sistema_ensino
6. municipio
13. no_curso
16. tipo_curso
18. ds_tipo_nivel
19. nome_ciclo
nome_aluno
24. modalidade_pagto
25. situacao_matricula
30. modalidade_ensino
status_do_ciclo_de_matricula

- Completude — Início → Não é uma sigla valida?
- Conformidade/ Exatidão — Início → O valor não é uma string valida
- Comformidade/ Exatidão — Início → Inicia com carácter especial (/ *-...)?

valida_estado (params)

Colunas validadas
7. estado

- Completude — Início → O valor é nulo
- Comformidade/ Exatidão — Início → Não é uma sigla valida?

valida_sexo_aluno (params)

Colunas validadas
37. sexo_aluno

- Completude — Início → O valor é nulo
- Comformidade/ Exatidão — Iní cio → O valor é diferente "M" ou "F" ?

valida_texto (params)

Colunas validadas
programa_associado
39. nome_completo_agrupador
40. sigla_agrupador
nacionalidade

- Comformidade/ Exatidão — Início → O valor não é string?
- Comformidade/ Exatidão — Início → O valor inicia com carácter especial (/ *-...)?

valida_numero_cpf (params)

Colunas validadas
numero_cpf

Colunas auxiliares
situacao_cpf

- Completude — Início → A coluna situacao_cpf está vazia ?
- Consistência/ Completude — Início → situacao_cpf é igual a 'COM CPF' ? numero_cpf é nulo?
- Comformidade/ Exatidão — Início → situacao_cpf é igual a 'COM CPF' ? numero_cpf é != 11 dígitos?
- Comformidade/ Exatidão — Início → situacao_cpf é igual a 'COM CPF' ? numero_cpf é == 11 dígitos? numero_cpf não é valido?
- Consistência — Início → situacao_cpf é igual a 'SEM CPF' ? numero_cpf não é nulo?

valida_nu_cnpj (params)

Colunas validadas
nu_cnpj

- Comformidade/ Exatidão — Início → O valor não é um CNPJ válido?

valida_tipo_oferta (valores)

Colunas validadas
28. tipo_oferta

Colunas auxiliares
15. co_tipo_curso

- Completude — Início → O valor é nulo
- Consistência — Início → co_tipo_curso == 3 e valor != no_tipo_oferta_curso
- Consistência — Início → co_tipo_curso != 3 e valor != hifen(-)

valida_diploma_certificado (params) (falta terminar)

Colunas validadas
data_geracao_codigo_diplomacertificado
codigo_diploma_certific ado

Colunas auxiliares
15. co_tipo_curso
25. situacao_matricula

- Conformidade/ Exatidão — Início → data_geracao_codigo_dip lomacertificado é diferente de vazio? data_geracao_codigo_dip loma não é uma data valida?
- Consistência Comformidade Exatidão — Início → data_geracao_codigo _diplomacertificado é vazio? codigo_diplomacertifi cado é diferente de nulo?
- Consistência Comformidade Exatidão — Início → data_geracao_codigo_diplomace rtificado é diferente de vazio? data_geracao_codigo_diploma é uma data valida? codigo_diploma_certificado é nulo ou "-"?
- Consistência — Início → data_geracao_codigo_diplomacert ificado é diferente de vazio? data_geracao_codigo_diploma é uma data valida? codigo_diploma_certificado é diferente nulo ou "-"? co_tipo_curso !=3 e situacao_matricula != concluida?

valida_subdependencia (params)

Colunas validadas
subdependencia

Colunas auxiliares
4. dependencia_adm

- Consistência — Início → O valor de dependencia_admin != "Artigo 240 - Sistema S" ou "Militar" O valor de dependencia_admin == "Pública" ou "Privada" O valor de subdependencia != nulo e != "-" ?
- Consistência/ Completude — Início → O valor de dependencia_admin == "Artigo 240 - Sistema S" ou "Militar" O valor de subdependencia é nulo ou -

Tipo: Int

valida_inteiro_nao_nulo (params)

Colunas validadas
co_unidade_ensino
1.Aluno
2. co_ciclo_matricula
12. co_curso
15. co_tipo_curso
17. co_tipo_nivel
21. carga_horaria
co_aluno
co_matricula
co_portfolio

- Completude — Início → O valor é nulo
- Comformidade/ Exatidão — Início → O valor não é número

valida_inteiro (params)

Colunas validadas
cod_municipio
34. vagas_ofertadas
35. total_inscritos
co_inep

- Comformidade/ Exatidão — Início → O valor não é número

valida_ds_eixo_tecnologico (params)

Colunas validadas
11. ds_eixo_tecnologico

Colunas auxiliares
15. co_tipo_curso

- Consistência/ Completude — Início → O co_tipo_curso == (3 ou 4)? O ds_eixo_tecnologico não existe?
- Consistência/ Comformidade/ Exatidão — Início → O co_tipo_curso == (3 ou 4)? ou é diferente dos eixos?

valida_co_endereco_co_ies (params)

Colunas validadas
co_ies
co_endereco_e-mec

Colunas auxiliares
4. dependencia_adm
5. sistema_ensino

- Consistência/ Completude/ — Início → O sistema_ensino é Federal e a dependencia_ad min é Privada? O valor de co_ies. co_endereco_e- mec é nulo?
- Comformidade/ Exatidão — Início → O valor co_ies. co_endereco_e-mec não é um número?

valida_no_responsavel_emissao (params)

Colunas validadas
no_responsavel_emissao

Colunas auxiliares
codigo_diploma_certificado

- Consistência/ Completude — Início → O valor de codigo_diploma_certificado é diferente de nulo? O valor de no_responsavel_emis el_emissao é nulo?

Tipo: Boolean

valida_bool_nao_nulo (params)

Colunas validadas
experimental
contrato_aprendizagem
atestado_baixa_renda
sg_etec
st_ativo

- Completude — Início → O valor é nulo
- Comformidade/ Exatidão — Início → O valor não é boolean (SIM ou NÃO)?

valida_bool (params)

Colunas validadas
carga
sg_proeja
st_deferimento_ue

- Comformidade/ Exatidão — Início → O valor não é boolean (SIM ou NÃO)?

valida_tipo_cota (params)

Colunas validadas
26. tipo_cota

- Completude — Início → O valor é nulo
- Comformidade/ Exatidão — Início → O valor não é boolean sim ou NÃO, com cota ou sem cota e a lista de cotas

valida_sg_brasilpro (params)

Colunas validadas
sg_brasilpro

Colunas auxiliares
48. ano_convenio_brasilpro

- Consistência — Início → O valor na coluna ano_convenio_brasil pro é ! = nulo? O valor da coluna sg_brasilpro é != de sim?
- Consistência — Início → O valor na coluna ano_convenio_brasil pro é ! = nulo? O valor da coluna sg_brasilpro é != de nao?

Tipo: Date

valida_dt_cadastro_ciclo (params)

Colunas validadas
20 dt_cadastro_ciclo

Colunas auxiliares
14. dt_deferimento_curso

- Completude — Início → O valor é nulo?
- Comformidade/ Exatidão — Início → O valor não é uma data valida?
- Consistência — Início → dt_cadastro_ci clo < dt_deferimento _curso

valida_dt_deferimento_curso (params)

Colunas validadas
14. dt_deferimento_curso

Colunas auxiliares
dt_deferimento_ue

- Completude — Início → O valor é nulo?
- Comformidade/ Exatidão — Início → O valor não é uma data valida?
- Consistência — Início → dt_deferimento_cu rso < dt_deferimento_ue?

valida_dt_cadastro_aluno_sistema

Colunas validadas
22. dt_cadastro_aluno_sistema

Colunas auxiliares
20. dt_cadastro_ciclo

- Completude — Início → O valor é nulo?
- Comformidade/ Exatidão — Início → O valor não é uma data valida
- Consistência — Início → O valor de dt_cadastro_aluno_si stema < dt_cadastro_ciclo

valida_dt_data_fim_previsto (params)

Colunas validadas
10. dt_data_fim_previsto

Colunas auxiliares
9. dt_data_inicio
15. co_tipo_curso

- Conformidade/ Exatidão — Início → dt_data_fim_previsto é diferente de nulo? dt_data_fim_previsto é uma data não válida?
- Consistência — Início → dt_data_fim_previsto é diferente de nulo? dt_data_fim_previsto é uma data válida? Se co_tipo_curso == 2 e dt_data_fim_previsto < dt_data_inicio
- Consistência — Início → dt_data_fim_previsto é diferente de nulo? dt_data_fim_previsto é uma data válida? Se co_tipo_curso != 2 e dt_data_fim_previsto > dt_data_inicio Se co_tipo_curso != 2 e dt_data_fim_previsto <= dt_data_inicio

valida_data_nao_nulo (params)

Colunas validadas
9. dt_data_inicio
dt_deferimento_ue
20.dt_cadastro_ciclo
38. dt_data_nascimento
31.dt_ocorrencia_ciclo
32. dt_ocorrencia_matricula
33. data_ultima_alteracao

- Completude — Início → O valor é nulo?
- Comformidade/ Exatidão — Início → O valor não é uma data valda?

valida_ano_convenio_brasil_pro (params)

Colunas validadas
ano_convenio_brasilpro

- Comformidade/ Exatidão — Início → O valor do ano não é valido
- Consistência — Início → O ano é maior que o atual?

valida_dt_validacao_conclusao

Colunas validadas
dt_validacao_conclusao

Colunas auxiliares
data_geracao_codigo_diploma_certificado

- Comformidade/ Exatidão — Início → dt_validacao_co nclusao não é data válida?
- Consistência Completude — Início → O valor data_geracao_codigo_dipl o_certificado existe e a dt_validacao_conclusao não existe?

Fonte: Equipe do projeto de pesquisa do IFB, com base no TED 10718/2021, adaptada pela autora.

O que cada dimensão representa

Accuracy/ Exatidão — A exatidão dos dados é verificada comparando-os a um valor de referência conhecido. Em alguns casos, como no gênero, é fácil medir a precisão, mas em outros, sem referência, isso se torna difícil. Com a proximidade com que os dados são representados de um evento da vida real para o qual um valor de dado de atributo é atribuído.

Completeness / Completude — A completude é definida como a razão entre os valores não ausentes.

Consistency / Consistência — Refere-se à correção e completude dos relacionamentos entre dados correlacionados. Em bancos de dados, significa que os dados em diferentes áreas de armazenamento devem ser equivalentes. A sincronização de dados é o processo de garantir essa equivalência.

Conformity/ Conformidade — Um problema típico de conformidade de dados que pode ser enfrentado são os campos de data e hora, que podem ter formatos diferentes dependendo da fonte de dados. Além disso, como os dados são não estruturados, campos que exigem um formato específico, como e-mail, número de telefone e código postal, podem não estar necessariamente preenchidos corretamente. Assim, definimos conformidade como o grau em que os dados respeitam as regras e as restrições do seu ambiente. A expressão regular de cada campo deve ser especificada, pois ela representa o padrão do campo com uma sintaxe específica.

Uniqueness/ Unicidade — Unicidade mede a ausência de duplicidades em um conjunto de dados.

Métrica de cada dimensão

Accuracy/Exatidão = Número de valores corretos/Número de valores totais

Completeness/Completude = Número de valores não vazios/ Número de Valores totais

Consistency/Consistência = Número de valores consistentes/ Número de valores totais

Conformity/Conformidade = Número de valores com formato correto/ Número de valores totais

Uniqueness/ Unicidade = Número de valores únicos/ Número de valores totais

Colunas: ano, periodo, cadastro_matricula não foram validadas, pois não foram consideradas tão relevantes.

Outras dimensões, talvez mais subjetivas, a serem analisadas — https://docs.google.com/document/d/1EgcGPqUoyiXxkMlLDeHyr-JERLhhCmUOL9wYIpeYuR8/edit?usp=sharing

Descrição das Inconsistências — https://docs.google.com/document/d/1SJnLqt24pMfcs1piE5QAGquwQIOAnHseVn7CB1E2ZrA/edit?usp=sharing