

## Dicionário da Base Comum de Educação do Ano de 2023

Adjailson Ferreira de Melo

### 1. Caracterização do domínio

Por meio dos base, é possível obter um amplo panorama da educação brasileira e, como importante ferramenta de transparência, é indubitavelmente um rico acervo sobre a Educação Básica do nosso país e uma fonte segura e eficaz de obtenção de dados, acessíveis aos pesquisadores, estudantes, gestores e sociedade em geral.

Esse novo formato dos Microdados, traz dois arquivos de dados que foram dimensionados ao nível de escolas, e contemplam uma série de tabulações que permitem rapidamente obter as principais estatísticas educacionais. Esses dois arquivos consistem em:

- a) Base estatística por escola, denominada “microdados\_ed\_basica\_2023”, contendo, além dos dados cadastrais e de infraestrutura dos estabelecimentos de ensino, dentre outros, o número de matrículas por etapa de ensino, turno, tempo integral, sexo, cor/raça e faixa etária dos alunos. Além disso, é possível consultar também o número de docentes e turmas por etapa de ensino.
- b) Foi incluído como suplemento aos Microdados, uma base denominada “suplemento\_cursos\_tecnicos\_2023” contendo o número de matrículas por curso técnico e entre outras informações. Os dados estão dimensionados por estabelecimento de ensino, permitindo o cruzamento entre as duas bases por meio da variável “CO\_ENTIDADE”, nesse caso não estarei utilizando.

### 2. Abstração de Dados

Aqui contém os resultados da etapa Data Abstraction, onde é feito o mapeamento dos dados originais a tipos abstratos. Abaixo, Figura 1 e Figura 2, contém o que chamamos de dicionário da base:

Nome da Variável	Descrição da Variável	Tipo	Tam. <sup>(1)</sup>	Categoria
NU_ANO_CENSO	Ano do Censo	Num	4	
NO_REGIAO	Nome da Região Geográfica	Char	20	
CO_REGIAO	Código da Região Geográfica	Num	1	
NO_UF	Nome da Unidade da Federação	Char	50	
SG_UF	Sigla da Unidade da Federação	Char	2	
CO_UF	Código da Unidade da Federação	Num	2	
NO_MUNICIPIO	Nome do Município	Char	150	
CO_MUNICIPIO	Código do Município	Num	7	
TP_LOCALIZACAO	Localização	Num	1	1 - Urbana 2 - Rural
TP_LOCALIZACAO_DIFERENCIADA	Localização diferenciada da escola	Num	1	0 - A escola não está em área de localização diferenciada 1 - Área de assentamento 2 - Terra indígena 3 - Área onde se localiza comunidade remanescente de quilombos 8 - Área onde se localizam povos e comunidades tradicionais
TP_DEPENDENCIA	Dependência Administrativa	Num	1	1 - Federal 2 - Estadual 3 - Municipal 4 - Privada

Figura 1. Dicionário da Base.

12	NO_ENTIDADE	Código da Escola	Num	8
13	CO_ENTIDADE	Nome da Escola	Char	100
14	NO_AREA_CURSO_PROFISSIONAL	Nome da Área do Curso Técnico	Char	100
15	ID_AREA_CURSO_PROFISSIONAL	Código da Área do Curso Técnico	Num	3
16	NO_CURSO_EDUC_PROFISSIONAL	Nome do Cursos Técnico	Char	100
17	CO_CURSO_EDUC_PROFISSIONAL	Código do Curso Técnico	Num	5
18	QT_CURSO_TEC	Número de Cursos Técnicos	Num	8
19	QT_MAT_CURSO_TEC	Número de Matrículas em Cursos Técnicos	Num	8
20	QT_CURSO_TEC_CT	Número de Cursos Técnicos - Ensino Médio Integrado à Educação	Num	8
21	QT_MAT_CURSO_TEC_CT	Número de Matrículas em Cursos Técnicos - Ensino Médio Integrado à	Num	8
22	QT_CURSO_TEC_NM	Número de Cursos Técnicos - Ensino Médio - Modalidade	Num	8
23	QT_MAT_CURSO_TEC_NM	Número de Matrículas em Cursos Técnicos - Ensino Médio - Modalidade	Num	8
24	QT_CURSO_TEC_CONC	Número de Cursos Técnicos - Educação Profissional Concomitante	Num	8
25	QT_MAT_CURSO_TEC_CONC	Número de Matrículas em Cursos Técnicos - Educação Profissional	Num	8
26	QT_CURSO_TEC_SUBS	Número de Cursos Técnicos - Educação Profissional Subsequente	Num	8
27	QT_MAT_TEC_SUBS	Número de Matrículas em Cursos Técnicos - Educação Profissional	Num	8
28	QT_CURSO_TEC_EJA	Número de Cursos Técnicos - Curso Técnico Integrado na Modalidade EJA	Num	8
29	QT_MAT_TEC_EJA	Número de Matrículas em Cursos Técnicos - Curso Técnico Integrado na	Num	8

**Figura 2. Dicionário da Base.**

### 3. Abstração de Tarefas

Abaixo na Figura 3, temos o review do dataset processado no Codelab do Google e também um gráfico que melhor representa.

	NU_ANO_CENSO	NO_REGIAO	CO_REGIAO	NO_UF	SG_UF	CO_UF	NO_MUNICIPIO	CO_MUNICIPIO	TP_LOCALIZACAO	TP_LOCALIZACAO_DIFERENCIADA
0	2023	Norte	1	Rondônia	RO	11	Ariquemes	1100023	1	0
1	2023	Norte	1	Rondônia	RO	11	Ariquemes	1100023	1	0
2	2023	Norte	1	Rondônia	RO	11	Ariquemes	1100023	1	0
3	2023	Norte	1	Rondônia	RO	11	Ariquemes	1100023	1	0
4	2023	Norte	1	Rondônia	RO	11	Ariquemes	1100023	1	0
...	...	...	...	...	...	...	...	...	...	...
23219	2023	Centro-Oeste	5	Distrito Federal	DF	53	Brasília	5300108	1	0
23220	2023	Centro-Oeste	5	Distrito Federal	DF	53	Brasília	5300108	1	0

**Figura 3. View da base de dados.**

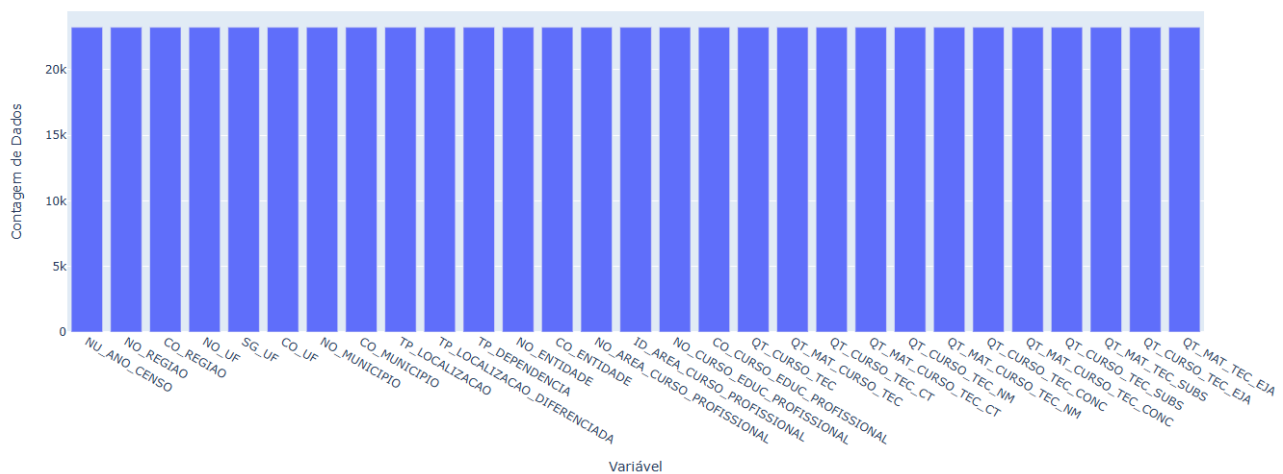
```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23224 entries, 0 to 23223
Data columns (total 29 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   NU_ANO_CENSO                             23224 non-null  int64
1   NO_REGIAO                                23224 non-null  object
2   CO_REGIAO                                23224 non-null  int64
3   NO_UF                                     23224 non-null  object
4   SG_UF                                     23224 non-null  object
5   CO_UF                                     23224 non-null  int64
6   NO_MUNICIPIO                             23224 non-null  object
7   CO_MUNICIPIO                             23224 non-null  int64
8   TP_LOCALIZACAO                           23224 non-null  int64
9   TP_LOCALIZACAO_DIFERENCIADA              23224 non-null  int64
10  TP_DEPENDENCIA                           23224 non-null  int64
11  NO_ENTIDADE                              23224 non-null  object
12  CO_ENTIDADE                              23224 non-null  int64
13  NO_AREA_CURSO_PROFISSIONAL               23224 non-null  object
14  ID_AREA_CURSO_PROFISSIONAL               23224 non-null  int64
15  NO_CURSO_EDUC_PROFISSIONAL               23224 non-null  object
16  CO_CURSO_EDUC_PROFISSIONAL               23224 non-null  int64
17  QT_CURSO_TEC                             23224 non-null  int64
18  QT_MAT_CURSO_TEC                         23224 non-null  int64
19  QT_CURSO_TEC_CT                          23224 non-null  int64
20  QT_MAT_CURSO_TEC_CT                     23224 non-null  int64
21  QT_CURSO_TEC_NM                          23224 non-null  int64
22  QT_MAT_CURSO_TEC_NM                     23224 non-null  int64
23  QT_CURSO_TEC_CONC                       23224 non-null  int64
24  QT_MAT_CURSO_TEC_CONC                   23224 non-null  int64
25  QT_CURSO_TEC_SUBS                       23224 non-null  int64
26  QT_MAT_TEC_SUBS                         23224 non-null  int64
27  QT_CURSO_TEC_EJA                        23224 non-null  int64
28  QT_MAT_TEC_EJA                          23224 non-null  int64
dtypes: int64(22), object(7)
memory usage: 5.1+ MB

```

**Figura 4. Gráfico representativo da Base.**

Resumo da Base: Suplemento Cursos Técnicos 2023



**Figura 5. Gráfico representativo da Base.**

4. Codificação Visual

4.1. Mapeamento tarefas x codificação

Aqui serão apresentados e justificada a escolha das codificações visuais desenvolvidas.

Tabela 3. Lista de codificações visuais

ID	Tarefa(s) associada (id)	Codificação visual	Justificativa
1	1,2	Gráfico de Barras	Adequado para comparar uma métrica para várias categorias...

4.2. Visualizações desenvolvidas e insights obtidos

Referências