

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. A base contém duas abas, uma chamada “suplemento_cursos_tecnicos” (a qual estou usando) e “microdados_unidade_coleta”. Abaixo, Figura 1, contém um preview da base com suas devidas colunas:

```

<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 1. Colunas da Base.

3. Abstração de Tarefas

Abaixo na Figura 2, temos o review da base em excel com demonstrativo da aba “suplemento_cursos_tecnicos”.

N	Nome da Variável	Descrição da Variável	Tipo	Tam. ⁽¹⁾	Categoria
1	NU_ANO_CENSO	Ano do Censo	Num	4	
2	NO_REGIAO	Nome da Região Geográfica	Char	20	
3	CO_REGIAO	Código da Região Geográfica	Num	1	
4	NO_UF	Nome da Unidade da Federação	Char	50	
5	SG_UF	Sigla da Unidade da Federação	Char	2	
6	CO_UF	Código da Unidade da Federação	Num	2	
7	NO_MUNICIPIO	Nome do Município	Char	150	
8	CO_MUNICIPIO	Código do Município	Num	7	
9	TP_LOCALIZACAO	Localização	Num	1	1 - Urbana 2 - Rural
10	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
11	TP_DEPENDENCIA	Dependência Administrativa	Num	1	1 - Federal 2 - Estadual 3 - Municipal 4 - Privada

Figura 2. Dicionário da base com informações de cada coluna.

Na Figura 3, temos a continuação das outras linhas explicativas das colunas.

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

Figura 3. Continuação dos detalhes sobre a base em excel.

4. Codificação Visual

4.1. Mapeamento tarefas x codificação

Aqui serão apresentados e justificada a escolha das codificações visuais desenvolvidas. É importante observar que quase sempre vou usar a tarefa de número 12 (NO_ENTIDADE), para buscar dados com base no código da escola. Já o valor de número 13 (CO_ENTIDADE) refere a outra aba.

Tabela 1. Lista de codificações visuais

ID	Tarefa(s) associada (id)	Codificação visual	Justificativa
1	1, 2, 12	Gráfico de Barras	Número de Escolas por Região.
2	1,9, 12	Gráfico de Pizza	Número de Escolas por Zona (Urbana/Rural).
3	1, 11, 12	Gráfico de Barras	Número por Dependência Administrativa.
4	1, 2, 12	Gráfico de Linhas	Análise por Ano e Região.
5	1, 4, 12	Gráfico de Barras	Número de Cursos por Estado.
6	16, 19	Gráfico de Barras Horizontal	Matrículas por Curso.
7	4, 19	Gráfico de Barras	Número de Alunos por Estado.
8	16, 19	Gráfico de Barras Horizontal	Número de Alunos por Curso.
9	4, 20, 26	Gráfico de Barras empilhados	Número de Cursos Técnicos por Modalidade.
10	11, 19	Gráfico de Barras	Matrículas por Dependência Administrativa.
11	1, 19	Gráfico de linhas	Evolução de Matrículas ao Longo dos Anos.
12	2, 16, 19	Gráfico de barras empilhados	Cursos com Maior Número de Matrículas por Região.
13	5, 16, 19	Gráfico de barras empilhados	Cursos com Maior Número de Matrículas por Estado.
14	16, 19	Gráfico de barras Horizontal	Cursos com Menor Número de Matrículas.

4.2. Visualizações desenvolvidas e insights obtidos

Para cada visualização, colocar um print e descrever que insight(s) você conseguiu extrair do gráfico.

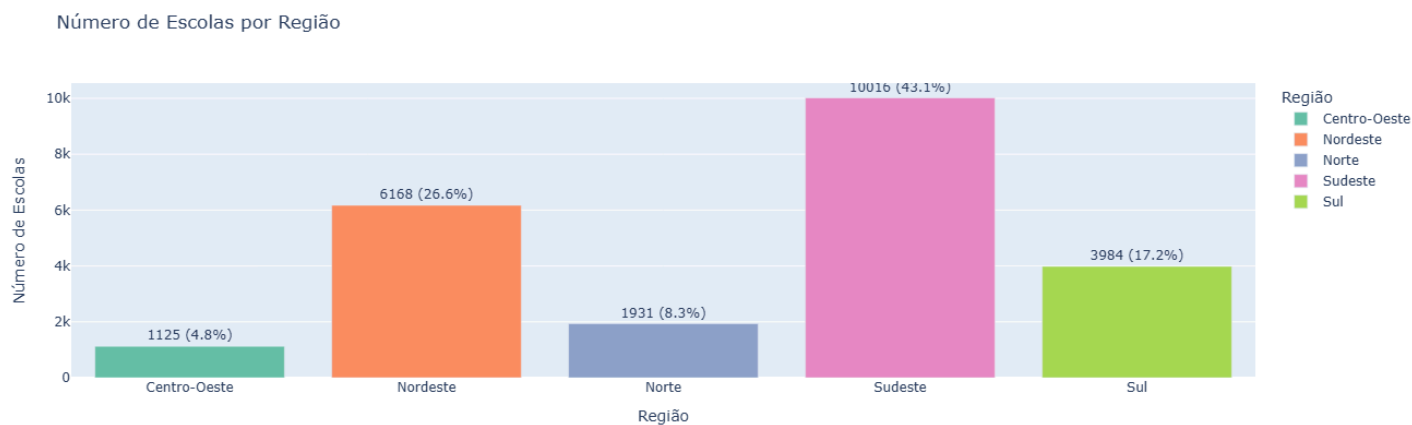


Figura 4. Número de escolas por região.

Distribuição de Escolas por Zona Urbana/Rural

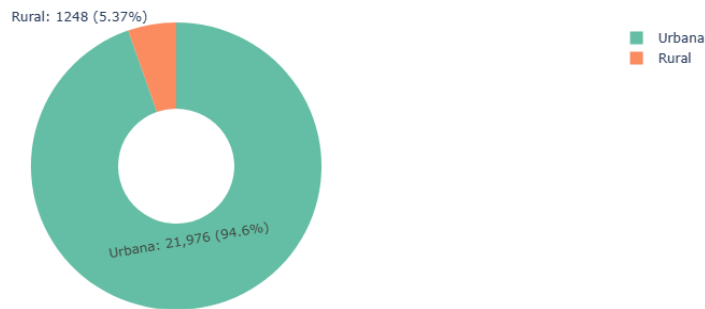


Figura 5. Distribuição de escolas por zona rural e urbana.

Comparação por Dependência Administrativa

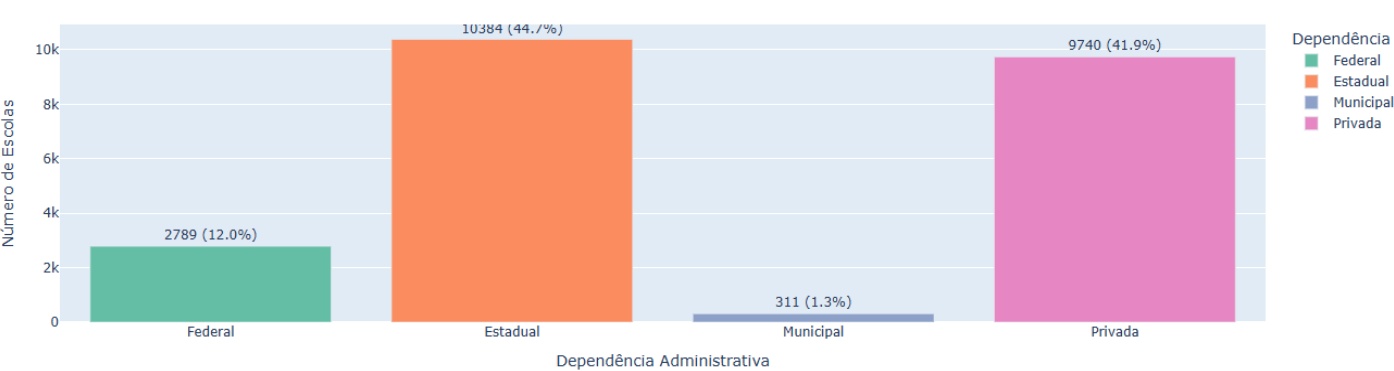


Figura 6. Número de dependências administrativas.

Número de Escolas por Ano e Região

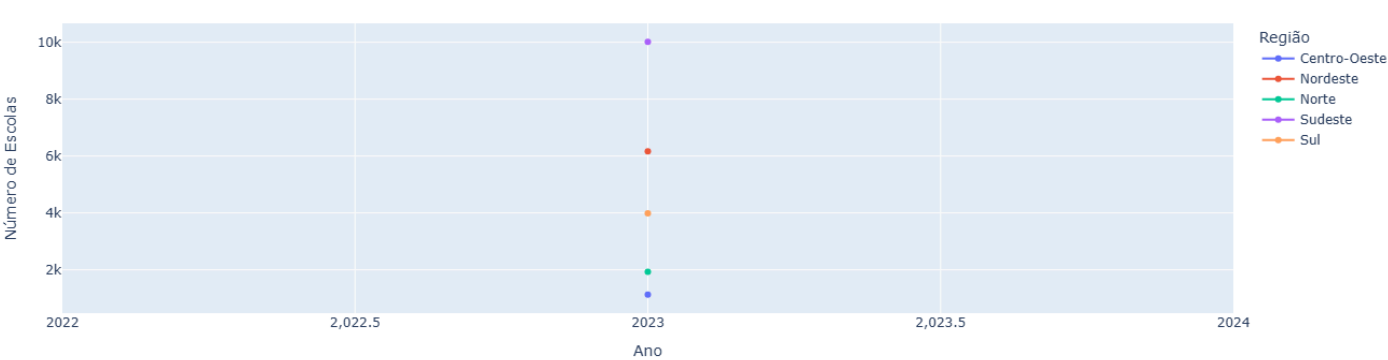


Figura 7. Número de escolas por ano e região.

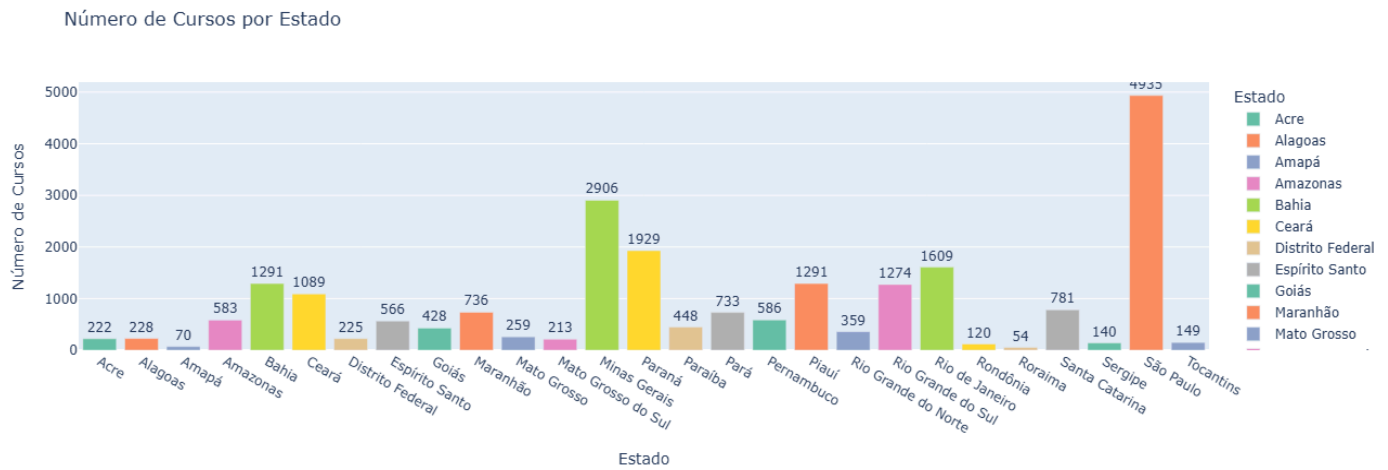


Figura 8. Número de cursos por estado.

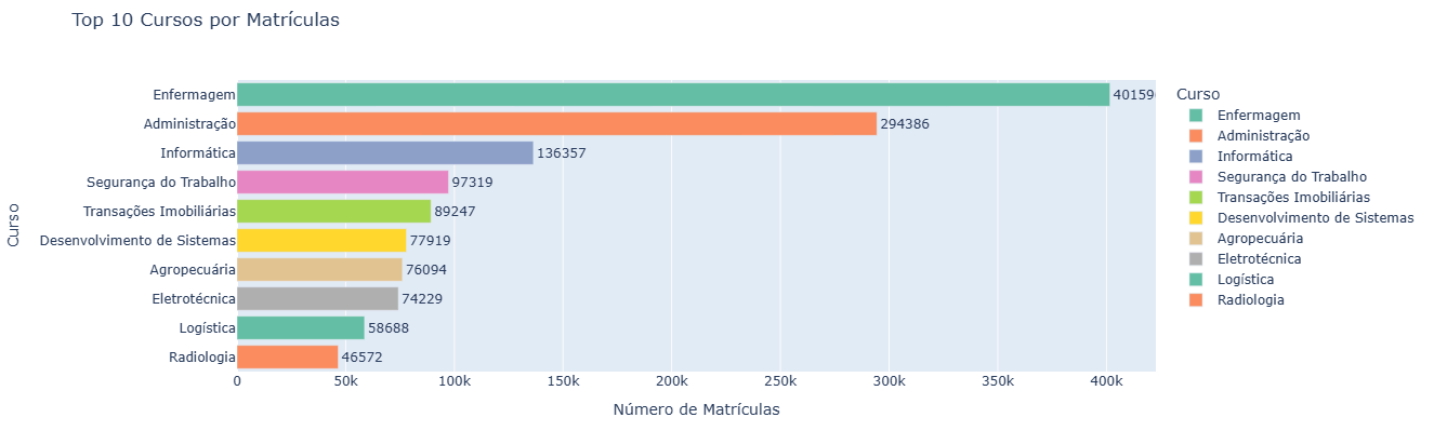


Figura 9. Os 10 cursos mais matriculados.

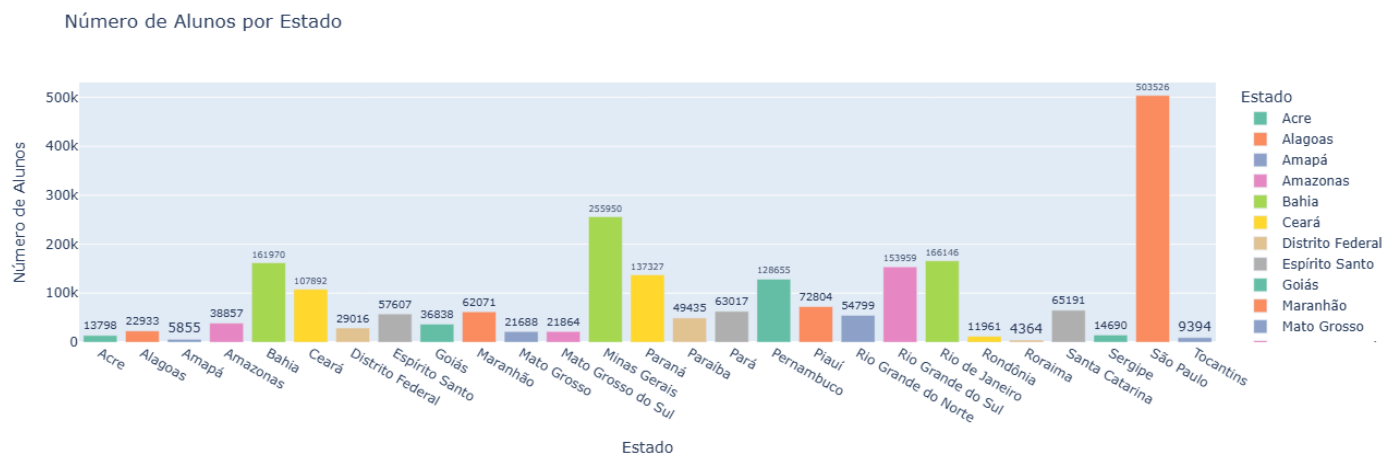


Figura 10. Número de alunos matriculados no estado.

Top 10 Cursos por Matrículas

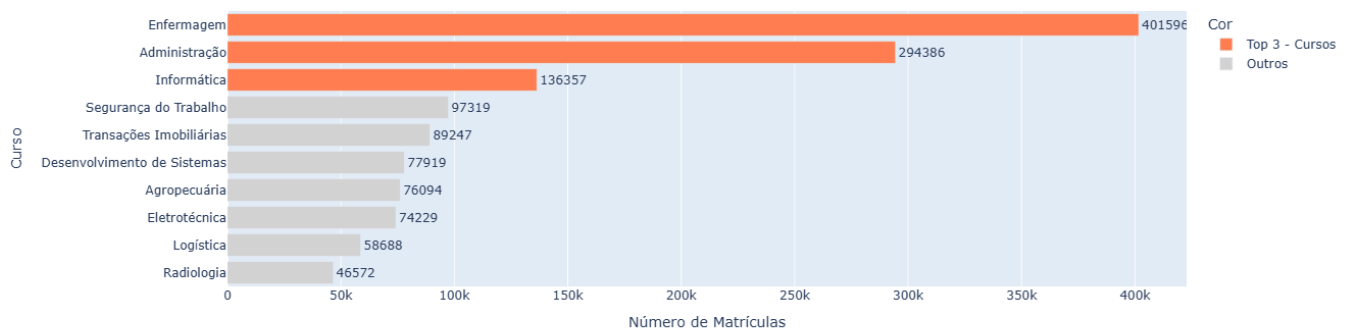


Figura 11. Dez cursos mais matriculados.

Número de Cursos Técnicos por Modalidade e Estado

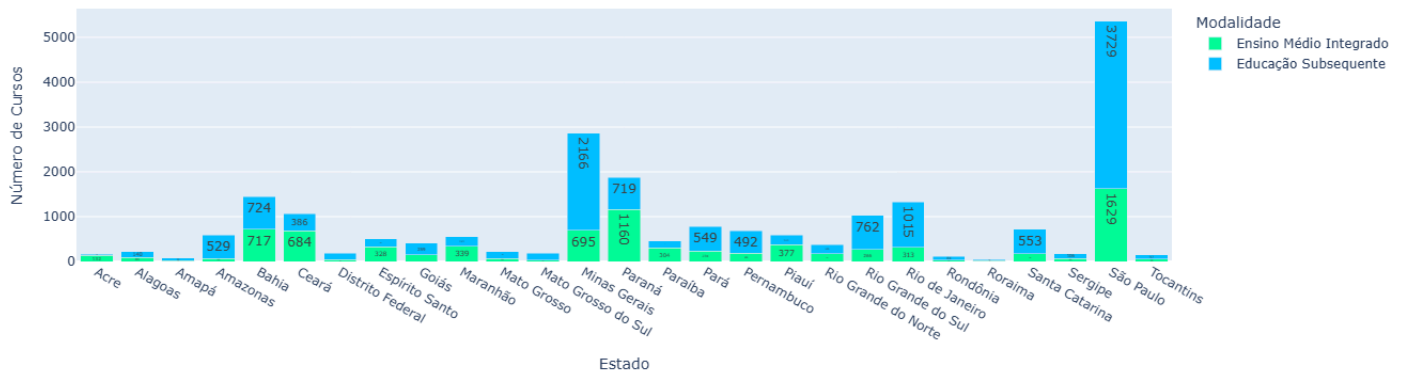


Figura 12. Número de alunos do subsequente e médio integrado.

Número de Matrículas por Dependência Administrativa

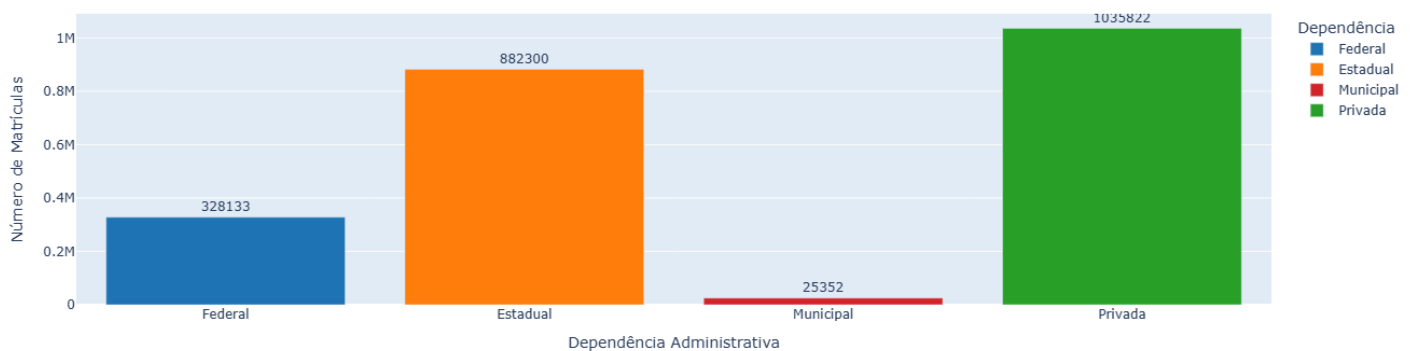


Figura 13. Número de alunos matriculados por rede ou dependência.

Cursos com Maior Número de Matrículas por Região

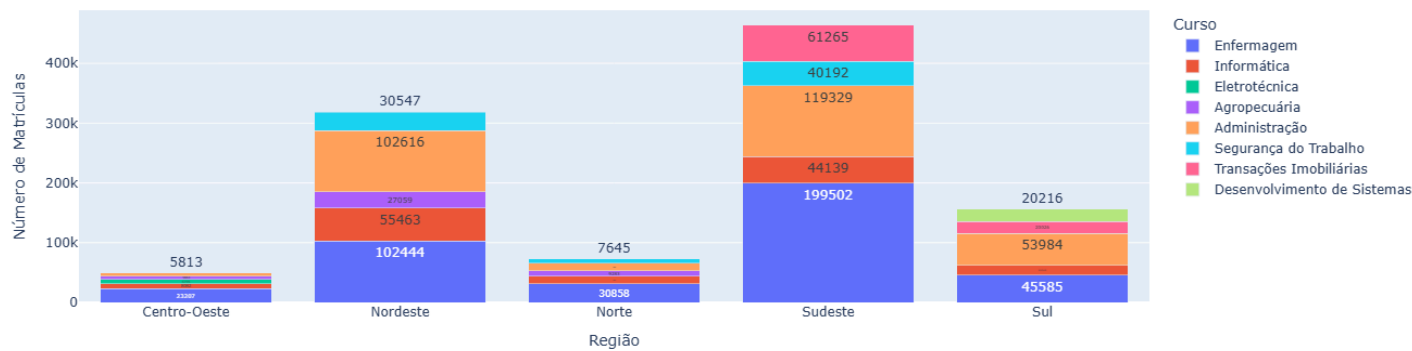


Figura 14. Cinco maiores cursos com matrícula.

Cursos com Maior Número de Matrículas por Estado

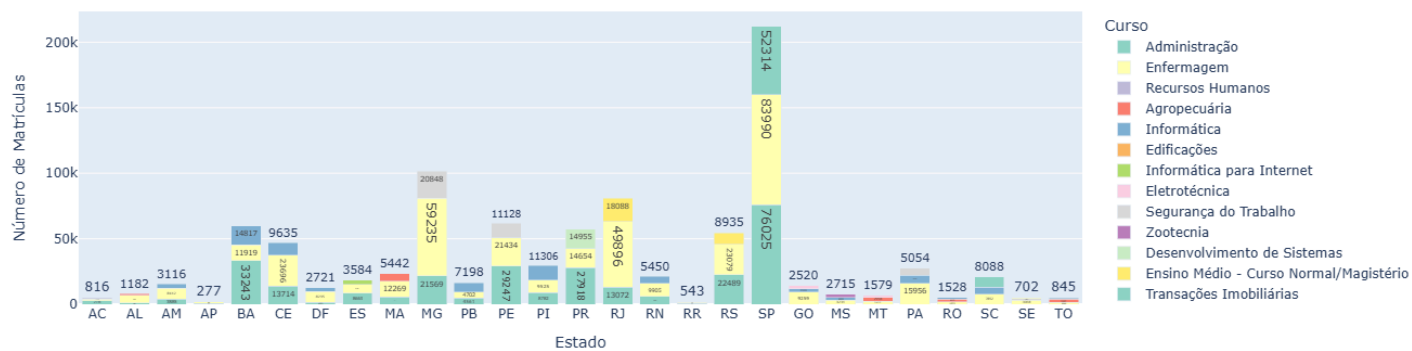


Figura 15. Três cursos por estado com maior número de matrículas.

Top 10 Cursos com Menor Número de Matrículas

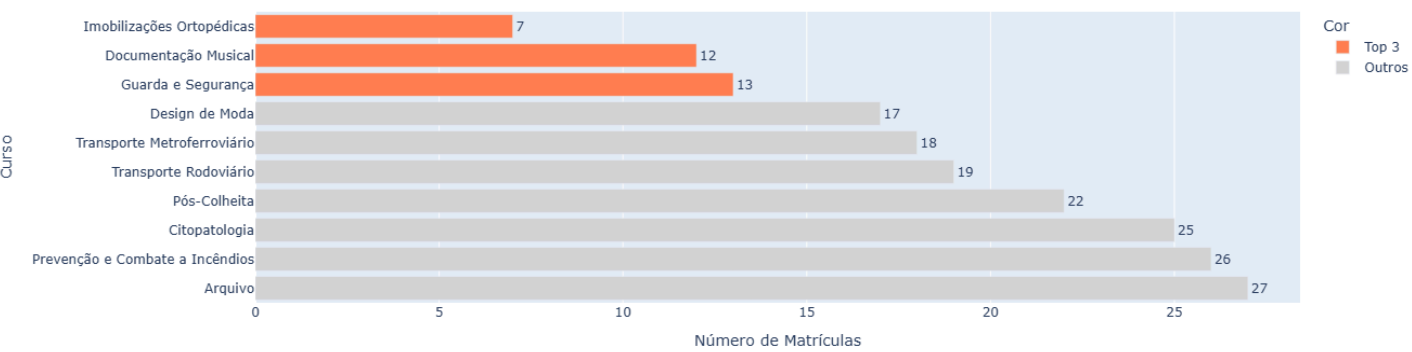


Figura 16. Dez cursos menos procurados no Brasil.

4. Design de Interação

Descrever tecnologia utilizada para implementar interatividade e preencher a tabela abaixo com as técnicas de interatividade suportadas para as visualizações implementadas.

Tabela 2. Lista de recursos interativos implementados

Codificação visual	Tipo de interação	Justificativa
Gráfico de Barras	Comparar valores absolutos entre categorias através de seleções.	Número de Escolas por Região - Permite comparar diretamente a quantidade de escolas em diferentes regiões. Matrículas por Dependência Administrativa - Mostra claramente como as matrículas estão distribuídas entre Federal, Estadual, Municipal e Privada.
Gráfico de Pizza	Mostrar a proporção de cada categoria em relação ao todo. Melhor visualização para valores em porcentagem.	Matrículas por Dependência Administrativa - Apresenta a contribuição percentual de cada tipo de dependência (Federal, Estadual, Municipal, Privada). Distribuição de Matrículas por Curso - Mostra o percentual de alunos nos cursos.
Gráfico de Linhas	Tendências ao longo do tempo e identificação de padrões.	Evolução de Matrículas ao Longo dos Anos - Mostra como o número de matrículas mudou ao longo do tempo, ajudando a identificar tendências de crescimento ou queda.
Gráfico de Barras Horizontal	Quando os nomes das categorias são longos e poderiam dificultar a leitura. Destaque de menores valores e são ideais para classificar valores.	Top 10 Cursos com Menor Número de Matrículas - As barras horizontais destacam os cursos menos procurados. Cursos com Maior Número de Matrículas por Região - Representa as categorias ordenadas em um formato mais legível quando há muitos dados.
Gráfico de Barras Empilhados	Comparação, Clareza visual e Destaque das partes ou do todo.	Mostra o número de matrículas divididas entre Ensino Médio Integrado e Educação Subsequente em cada estado. Ajuda a entender como cada modalidade contribui para o total em cada estado. Permite visualizar os cursos mais relevantes (subcategorias) empilhados em cada região (categoria principal). Ajuda a identificar quais cursos predominam em cada região.

Referências

- PLOTLY. Dash for Python Documentation. Disponível em: <https://dash.plotly.com/>. Acesso em: 16 dez. 2024.
- STASKO, J. T.; CARR, D. B.; CHANG, R.; HANSEN, C. D. Measuring the Effectiveness of Visualization: A Case Study on Visualizing Network Traffic. Disponível em: <https://faculty.cc.gatech.edu/~stasko/papers/infovis05.pdf>. Acesso em: 16 dez. 2024.
- ISENBERG, P.; ISENBERG, T.; CHEN, J.; SEDLMAIR, M.; MOERE, A. V. A Systematic Review on the Practice of Evaluating Visualization. Disponível em: <https://ieeexplore.ieee.org/document/5290695>. Acesso em: 16 dez. 2024.