



Engenharia de Dados MVP

Micheline Carvalho Barroso Pereira
Abril de 2025

Sumário

1. Introdução.....	3
2. Objetivo.....	3
3. Plataforma.....	4
4. Detalhamento.....	4
4.1 Busca pelos dados.....	4
4.2 Coleta.....	5
4.3 Modelagem.....	5
4.4 Carga.....	8
4.5 Análise.....	9
4.5.1 Qualidade de dados.....	9
4.5.2 Solução do problema.....	11
6. Autoavaliação.....	23
7. Referências.....	23

1. Introdução

Os cursos técnicos são fundamentais para aumentar a empregabilidade, pois oferecem formação prática e especializada em diversas áreas profissionais, permitindo que os alunos adquiram habilidades diretas para o mercado de trabalho. Com uma duração mais curta em relação às graduações universitárias, esses cursos permitem que os formandos entrem no mercado rapidamente, atendendo à alta demanda por profissionais qualificados em setores como tecnologia, saúde, indústria e comércio.

Além disso, os cursos técnicos têm um bom custo-benefício, pois são mais acessíveis e atualizados conforme as mudanças do mercado. Eles também oferecem uma base sólida para o empreendedorismo, permitindo que os profissionais possam abrir seus próprios negócios.

O SISTEC (Sistema Nacional de Informações da Educação Profissional e Tecnológica) tem por finalidade servir como mecanismo de registro e divulgação dos dados da educação profissional e tecnológica e de validação de diplomas de cursos de educação profissional técnica de nível médio.[1]

O SISTEC foi instituído e implantado pelo Ministério da Educação (MEC) em 2009 (Resolução CNE/CEB nº 3/2009), por intermédio da Secretaria de Educação Profissional e Tecnológica do Ministério da Educação (Setec/MEC).

Por meio dele, as instituições de ensino ofertantes de educação profissional e tecnológica inserem as informações sobre os cursos técnicos de nível médio e os cursos de qualificação profissional, incluindo matrícula, frequência, concluintes, entre outros dados.

2. Objetivo

A partir de dados públicos e disponíveis sobre unidades de ensino e cursos registrados no SISTEC, é possível mapear a rede de ensino e a oferta de cursos técnicos a partir de vários recortes. Além disso, sua distribuição por estados e municípios pode servir de base para planejar a formação de novos profissionais alinhada às demandas de cada região.

Sendo assim, o principal objetivo deste trabalho consiste em compreender como estão distribuídas as unidades de ensino no território brasileiro, identificando, ao mesmo tempo, os cursos técnicos ofertados com mais frequência por localidade. Para isso, algumas questões sobre Unidades de Ensino e Cursos deverão ser respondidas.

Sobre **Unidades de Ensino**:

- Quais os 5 estados com maior e menor quantidade de unidades de ensino?
- Quais as 5 capitais com maior e menor quantidade de unidades de ensino?
- Quais os 5 estados com maior e menor quantidade de unidades de ensino em municípios diferentes de capitais?
- Como está a distribuição da quantidade de unidades de ensino por sistema de ensino, tipo de escola, dependência administrativa e sub-dependência administrativa?

- Quais as 5 unidades de ensino com maior e menor quantidade de cursos presenciais? Em quais municípios e estados essas unidades estão?
- Quais as 5 unidades de ensino com a maior e menor quantidade de cursos à distância? Em quais municípios e estados essas unidades estão?
- Existem unidades de ensino sem oferta de cursos? Em que estados elas estão?
- Qual o percentual de unidades de ensino com e sem cursos por UF?
- Qual o percentual de unidades de ensino dentro e fora de capitais por UF?

Sobre **Cursos**:

- Quais os 5 estados com maior e menor quantidade cursos?
- Como está a distribuição da quantidade de cursos por eixo tecnológico?
- Como está a distribuição da quantidade de cursos por modalidade?
- Como está a distribuição da quantidade de cursos por carga horária?
- Como está a distribuição da quantidade de cursos por sistema de ensino?
- Como está a distribuição da quantidade de cursos por dependência administrativa?
- Quais os 5 cursos presenciais com maior e menor oferta?
- Quais os 5 cursos à distância com maior e menor oferta?
- Qual a distribuição dos cursos por eixo e região?

3. Plataforma

A plataforma escolhida para implementação desse MVP foi **Databricks Community Edition**, disponível em <https://community.cloud.databricks.com/>

4. Detalhamento

4.1 Busca pelos dados

A fim de atingir os objetivos propostos anteriormente, foram escolhidos dados abertos do SISTEC, publicados no Portal de Dados Abertos do governo brasileiro [2].

Dados abertos são aqueles livremente disponíveis para todos utilizarem e redistribuírem como desejarem, sem restrição de licenças, patentes ou mecanismos de controle. Os aspectos de linhagem e acesso devem ser preservados.

Conjunto de dados: Educação profissional e tecnológica (SISTEC) - Regulação Escolas

Apresenta informações sobre as instituições de ensino que ofertam educação profissional e tecnológica no país, cadastradas no Sistema Nacional de Informações da Educação Profissional e Tecnológica.

Formato:CSV. Catalogação: 06/09/2024

Disponível em: <https://dados.gov.br/dados/conjuntos-dados/educacao-profissional-e-tecnologica--sistec---regulacao-escolas>

Conjunto de dados: Educação profissional e tecnológica (SISTEC) - Regulação Cursos

Apresenta informações sobre os cursos de educação profissional e tecnológica no país cadastrados no Sistema Nacional de Informações da Educação Profissional e Tecnológica.

Formato:CSV. Catalogação: 06/09/2024

Conjunto de dados disponível em: <https://dados.gov.br/dados/conjuntos-dados/educacao-profissional-e-tecnologica--sistec---regulacao-cursos>

4.2 Coleta

Os dados foram coletados a partir de downloads realizados no Portal Brasileiro de Dados Abertos [2], conforme endereços (URLs) listados na seção anterior para os 2 (dois) conjuntos de dados considerados no estudo.

4.3 Modelagem

Os dados desse estudo foram organizados segundo a arquitetura “medalhão”, que descreve um conjunto de camadas, cujo propósito é denotar a qualidade dos dados armazenados no lakehouse.

Essa arquitetura garante atomicidade, consistência, isolamento e durabilidade à medida que os dados passam por várias camadas de validações e transformações antes de serem armazenados em uma disposição otimizada para uma análise eficiente. Os termos bronze (bruto), prata (validado) e ouro (enriquecido) descrevem a qualidade dos dados em cada uma dessas camadas [3].

Para o estudo em questão, as seguintes camadas e tabelas desnormalizadas (planas/flats) foram criadas. Tabela plana/flat é uma estrutura de dados onde todos os dados são armazenados em uma única tabela, sem a necessidade de relacionamentos complexos entre as informações.

Camada BRONZE/BRONZE

- **ESCOLAS_SISTEC:** tabela criada com estrutura e conteúdo extraídos diretamente do arquivo original “Regulacao_Escolas_2024.csv”, disponível para download no Portal Brasileiro de Dados Abertos [2];
- **CURSOS_SISTEC:** tabela criada com estrutura e conteúdo extraídos diretamente do arquivo original “Regulacao_Cursos_2024.csv”, disponível para download no Portal Brasileiro de Dados Abertos [2].

Camada PRATA/SILVER

- **ESCOLAS:** tabela criada através de processo de ETL e de qualificação dos dados da tabela ESCOLAS_SISTEC (camada BRONZE);
- **CURSOS:** tabela criada através de processo de ETL e de qualificação dos dados da tabela CURSOS_SISTEC (camada BRONZE).

Camada OURO/GOLD

- **ESCOLAS_FINAL:** tabela flat criada para o conceito de Unidades de Ensino (escolas) através da junção das tabelas ESCOLAS (camada SILVER) e ESCOLAS_QTD_CURSOS (tabela temporária da camada GOLD) pelo campo CODIGO_UNIDADE_DE_ENSINO;
- **CURSOS_FINAL:** tabela flat criada para o conceito de Cursos através da junção das tabelas CURSOS e ESCOLAS (ambas da camada SILVER) pelo campo CODIGO_UNIDADE_DE_ENSINO;
- **ESCOLAS_QTD_CURSOS:** tabela temporária contendo informações consolidadas de cursos por unidades de ensino e criada através da junção das tabelas ESCOLAS e CURSOS (ambas da camada SILVER) pelo campo CODIGO_UNIDADE_DE_ENSINO.

Catálogo de Dados: Descrição das tabelas ESCOLAS_FINAL e CURSOS_FINAL, ambas da camada GOLD, preparadas para realização de análises.

Tabela CURSOS_FINAL (camada GOLD)

Campo	Tipo	Descrição
NOME_SUBTIPO_DE_CURSOS	VARCHAR(255)	Nome do Subtipo de curso técnico ofertado cadastrado no Sistec. Domínio: Técnico.
CODIGO_CURSO	INT	Número do código do curso técnico cadastrado no Sistec conforme CNCT – Catálogo Nacional de Cursos Técnicos.
CURSO	VARCHAR(255)	Nome do curso técnico ofertado cadastrado no Sistec conforme CNCT – Catálogo Nacional de Cursos Técnicos.
EIXO_TECNOLOGICO	VARCHAR(255)	Nome do eixo tecnológico do curso técnico ofertado. Domínio: Produção Cultural e Design, Desenvolvimento Educacional e Social, Produção Industrial, Infraestrutura, Controle e Processos Industriais, Turismo, Hospitalidade e Lazer, Recursos Naturais, Ambiente e Saúde, Informação e Comunicação, Gestão e Negócios, Militar, Produção Alimentícia, Segurança.
MODALIDADE	VARCHAR(100)	Nome da modalidade de ensino do curso técnico ofertado. Domínio: Todos, Educação Presencial, Educação a Distância.
CARGA_HORARIA_CURSO	INT	Carga horária curso técnico ofertado. Domínio: 800, 1000, 1200.
SITUACAO_ATIVO	BOOLEAN	Nome da situação do curso técnico (Ativo). Domínio: True, False.

NOME_SISTEMA_DE_ENSINO	VARCHAR(255)	Sistema de Ensino da Unidade de Ensino. Domínio: ESTADUAL/DISTRITAL, MUNICIPAL, FEDERAL.
NOME_TIPO_ESCOLA	VARCHAR(255)	Tipo da escola da Unidade de Ensino. Domínio: Escolas de Oferta Exclusiva FIC, Escolas de Oferta Outros Cursos, IPES, Escolas Técnicas.
CODIGO_UNIDADE_DE_ENSINO	INT	Código da Unidade de Ensino ofertante no Sistec.
UNIDADE_DE_ENSINO	VARCHAR(255)	Nome da Unidade de Ensino ofertante no Sistec.
NOME_DEPENDENCIA_ADMINISTRATIVA	VARCHAR(255)	Nome da dependência administrativa. Domínio: PÚBLICA, PRIVADA, MILITAR, ARTIGO 240, SISTEMA S.
NOME_SUBDEPENDENCIA_ADMINISTRATIVA	VARCHAR(255)	Nome da sudependência administrativa. Domínio: SENAR, Marinha, SESI, SENAT, SESCOOP, SENAI, Aeronáutica, SENAC, SESC.
UF	CHAR(2)	Sigla da UF da unidade de ensino.
CODIGO_MUNICIPIO	INT	Código do Município da unidade de ensino.
MUNICIPIO	VARCHAR(255)	Nome do Município da unidade de ensino.

Tabela ESCOLAS_FINAL (camada GOLD)

Campo	Tipo	Descrição
NOME_SISTEMA_DE_ENSINO	VARCHAR(255)	Sistema de Ensino da Unidade de Ensino. Domínio: ESTADUAL/DISTRITAL, MUNICIPAL, FEDERAL
NOME_TIPO_ESCOLA	VARCHAR(255)	Tipo da escola da Unidade de Ensino. Domínio: Escolas de Oferta Exclusiva FIC, Escolas de Oferta Outros Cursos, IPES, Escolas Técnicas.
CODIGO_UNIDADE_DE_ENSINO	INT	Código da Unidade de Ensino ofertante no Sistec.
UNIDADE_DE_ENSINO	VARCHAR(255)	Nome da Unidade de Ensino ofertante no Sistec.
NOME_DEPENDENCIA_ADMINISTRATIVA	VARCHAR(255)	Nome da dependência administrativa. Domínio: PÚBLICA, PRIVADA, MILITAR, ARTIGO 240, SISTEMA S.
NOME_SUBDEPENDENCIA_ADMINISTRATIVA	VARCHAR(255)	Nome da sudependência administrativa. Domínio: SENAR, Marinha, SESI, SENAT, SESCOOP, SENAI, Aeronáutica, SENAC, SESC.
UF	CHAR(2)	Sigla da UF da unidade de ensino.
CODIGO_MUNICIPIO	INT	Código do Município da unidade de ensino.
MUNICIPIO	VARCHAR(255)	Nome do Município da unidade de ensino.
QTD_CURSOS_PRESENCIAIS	INT	Quantidade de cursos presenciais ofertados pela unidade de ensino.
QTD_CURSOS_DISTANCIA	INT	Quantidade de cursos à distância ofertados pela unidade de ensino.
QTD_CURSOS_TOTAIS	INT	Quantidade total de cursos ofertados pela unidade de ensino.

4.4 Carga

O processo de ETL seguiu os princípios da arquitetura medalhão.

Camada BRONZE: Dados brutos, sem transformação.

Código: 1-Engenharia-de-Dados-MVP-Criação-Carga-Qualidade-Bronze

- Passo 1: Realizado download dos arquivos “Regulacao_Escolas_2024.csv” e “Regulacao_Cursos_2024.csv” a partir do Portal Brasileiro de Dados Abertos, utilizando os links citados na seção 4.1.
- Passo 2: Realizado upload dos arquivos “Regulacao_Escolas_2024.csv” e “Regulacao_Cursos_2024.csv” na plataforma Databricks, através da opção "+ New" -> "Add or upload".
- Passo 3: Criação da camada de banco de dados BRONZE.
- Passo 4: Leitura dos arquivos e gravação dos dados nas tabelas correspondentes, ESCOLAS_SISTEC e CURSOS_SISTEC, da camada BRONZE.
- Passo 5: Realizada análise de qualidade de dados (será descrita em detalhes na seção 4.5.1).

Camada SILVER: Dados limpos e parcialmente transformados.

Código: 2-Engenharia-de-Dados-MVP-Criação-Carga-Qualidade-Silver

- Passo 1: Criação da camada de banco de dados SILVER.
- Passo 2: Criação das tabelas ESCOLAS e CURSOS a partir das tabelas ESCOLAS_SISTEC e CURSOS_SISTEC da camada BRONZE. O resultado da análise de qualidade dos dados da camada BRONZE, motivou a seleção de campos, o ajuste dos tipos de dados e a transformação dos valores de algumas colunas, incluindo tratamento de nulos.
- Passo 3: Realizada análise de qualidade de dados (será descrita em detalhes na seção 4.5.1).

Camada GOLD: Dados refinados, prontos para consumo de negócio.

- Passo 1: Criação da camada de banco de dados GOLD.

- Passo 2: Criação da tabela CURSOS_FINAL a partir dos dados das tabelas CURSOS e ESCOLAS da camada SILVER. Os dados de Unidades de Ensino presentes na tabela CURSOS foram substituídos pelos dados e informações adicionais sobre Unidades de Ensino, presentes na tabela ESCOLAS. A junção entre as tabelas foi realizada como inner join/junção interna, através do campo CODIGO_UNIDADES_DE_ENSINO. Neste caso, não foram carregados os Cursos, cujas Unidades de Ensino não existiam na tabela ESCOLAS.
- Passo 3: Criação da tabela temporária ESCOLAS_QTD_CURSOS com dados consolidados de Cursos por Unidade de Ensino. A junção entre ESCOLAS e CURSOS se deu como left join/junção à esquerda, através do campo CODIGO_UNIDADES_DE_ENSINO.
- Passo 4: Criação da tabela ESCOLAS_FINAL a partir dos dados das tabelas ESCOLAS e ESCOLAS_QTD_CURSOS da camada SILVER e GOLD respectivamente. A junção entre as tabelas foi realizada como left join/junção à esquerda, através do campo CODIGO_UNIDADES_DE_ENSINO.
- Passo 5: Remoção da tabela ESCOLAS_QTD_CURSOS.

4.5 Análise

4.5.1 Qualidade de dados

A análise de qualidade de dados foi realizada para os dados das camadas BRONZE e SILVER.

Análise de Qualidade da camada BRONZE, tabela ESCOLAS_SISTEC:

- Detalhamento do esquema com atributos e tipos correspondentes. Nesta camada, todos os campos apresentam tipo “string” (tipo original);
- Contagem de registros completa e distinta para verificar duplicidades. Foram encontrados 22.508 registros em ambas as contagens, ou seja, não foram encontradas duplicidades;
- Verificação de nulos para todos os atributos. As seguintes colunas apresentaram valores nulos (quantidade entre parêntesis) = CEP (4), LOGRADOURO (11), NUMERO (18), E_MAIL (23), TELEFONE_1 (20), TELEFONE_2 (12.364);
- Verificação de domínios dos campos categóricos: NOME_SISTEMA_DE_ENSINO, NOME_TIPO_ESCOLA, NOME_DEPENDENCIA_ADMINISTRATIVA, NOME_SUBDEPENDENCIA_ADMINISTRATIVA. Valores de NOME_SUBDEPENDENCIA_ADMINISTRATIVA são válidos apenas para NOME_DEPENDENCIA_ADMINISTRATIVA igual a “ARTIGO 240 - SISTEMA S”;

- Verificação de campos com tipos incompatíveis: campos numéricos com caracteres diferentes de números. Não foram encontrados valores inválidos nos campos numéricos CODIGO_UNIDADE_DE_ENSINO, CODIGO_MUNICIPIO.

Análise de Qualidade da camada BRONZE, tabela CURSOS_SISTEC:

- Detalhamento do esquema com atributos e tipos correspondentes. Nesta camada, todos os campos apresentam tipo “string” (tipo original);
- Contagem de registros completa e distinta para verificar duplicidades. Foram encontrados 83.150 registros na contagem completa e 260 registros na contagem distinta. Como existem duplicidades, é possível que cursos sejam oferecidos mais de uma vez;
- Verificação de nulos para todos os atributos. Nenhuma coluna apresentou valores nulos;
- Verificação de domínios dos campos categóricos:
NOME_SUBTIPO_DE_CURSOS, EIXO_TECNOLOGICO, MODALIDADE, CARGA_HORARIA_CURSO, SITUACAO_ATIVO;
- Verificação de campos com tipos incompatíveis: campos numéricos com caracteres diferentes de números. Não foram encontrados valores inválidos nos campos numéricos CODIGO_CURSO, CODIGO_UNIDADE_DE_ENSINO, CODIGO_MUNICIPIO, CARGA_HORARIA_CURSO.

Análise de Qualidade da camada SILVER, tabelas ESCOLAS e CURSOS:

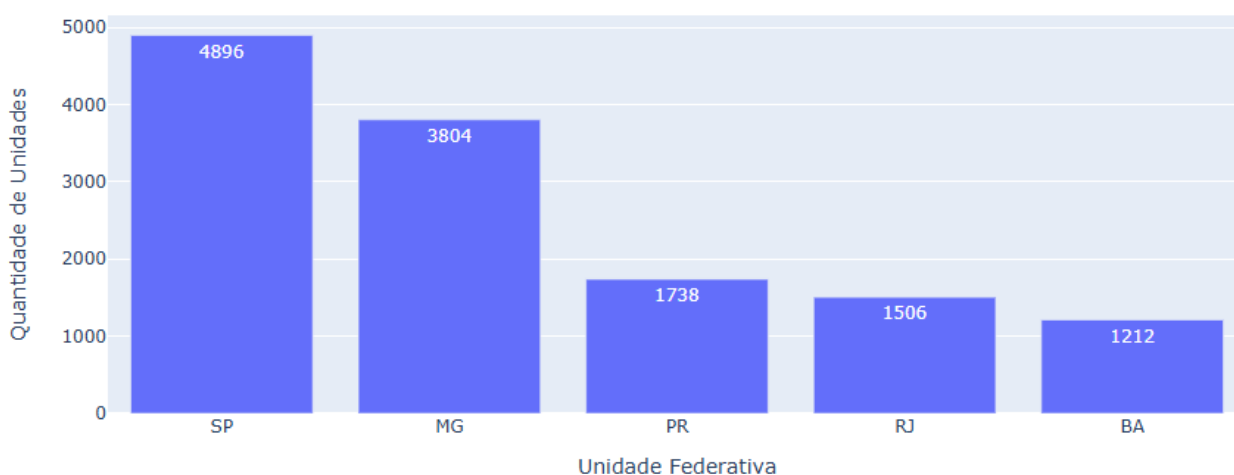
- Verificação de campos em comum entre as tabelas ESCOLAS e CURSOS: não foram encontradas diferenças de conteúdo em campos de mesmo nome que estão presentes em ambas as tabelas. Na verdade, as informações de UNIDADE_DE_ENSINO, UF, MUNICIPIO e CODIGO_MUNICIPIO são provenientes das Unidades de Ensino;
- Verificação de correspondência de CODIGO_UNIDADE_DE_ENSINO da tabela CURSOS na tabela ESCOLAS: 13 Unidades de Ensino da tabela CURSOS não foram encontradas na tabela ESCOLAS e, portanto, não possuem informações complementares para Unidades de Ensino;
- Verificação de correspondência de CODIGO_UNIDADE_DE_ENSINO da tabela ESCOLAS na tabela CURSOS: 8.093 Unidades de Ensino da tabela ESCOLAS não foram encontradas na tabela CURSOS e, portanto, não possuem cursos ofertados.

4.5.2 Solução do problema

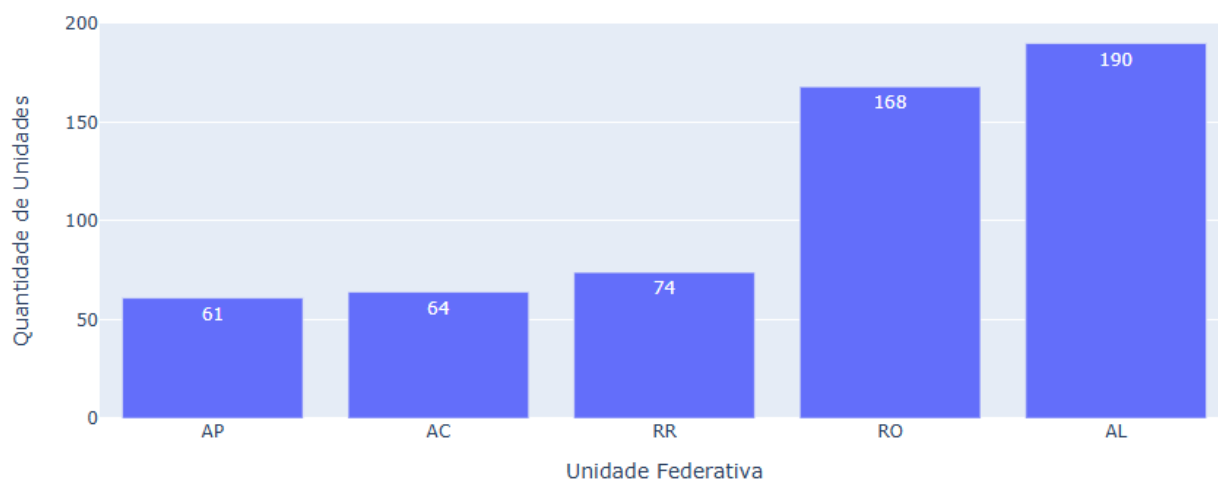
A) Sobre Unidades de Ensino

1) *Quais os 5 estados com maior e menor quantidade de unidades de ensino?*

Ranking de Quantidade de Unidades de Ensino por UF (5 maiores)



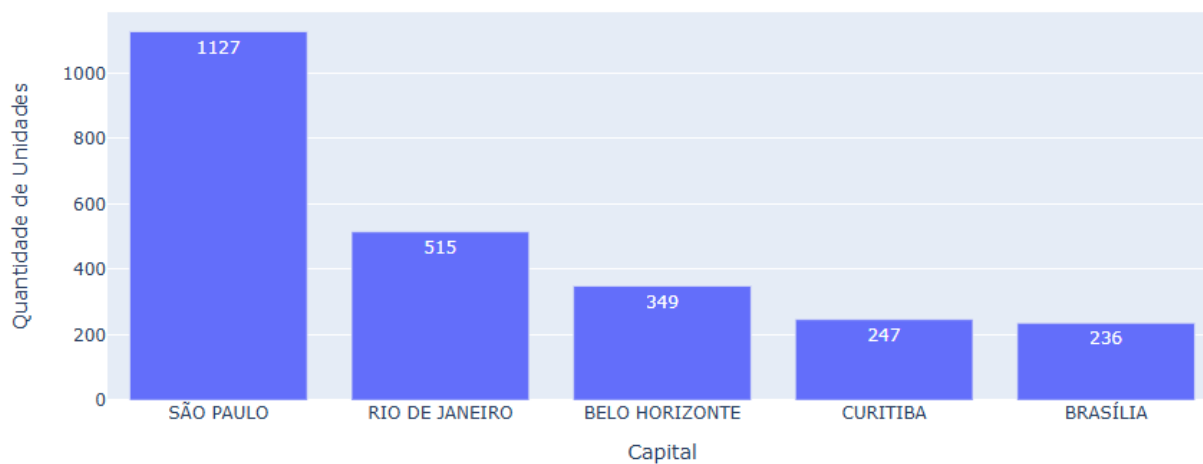
Ranking de Quantidade de Unidades de Ensino por UF (5 menores)



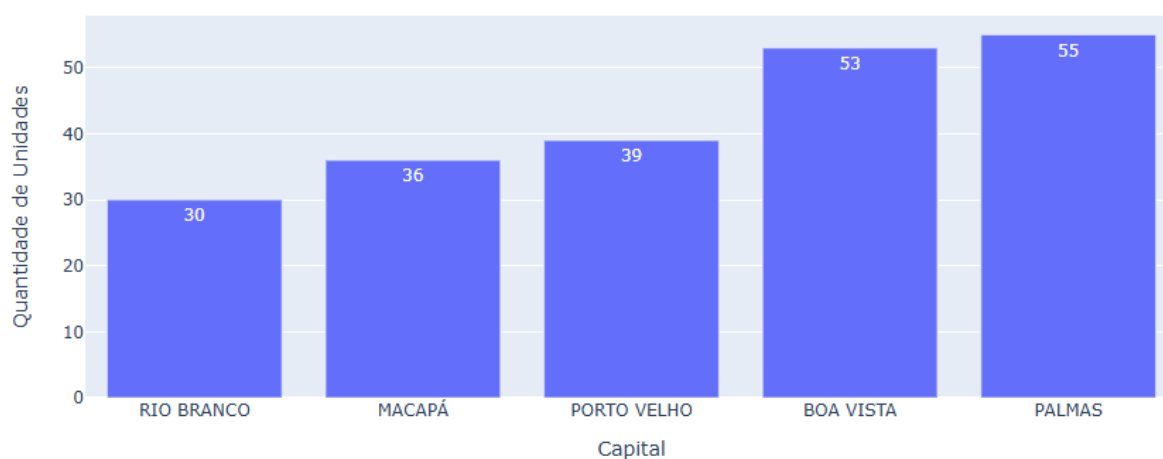
No ranking dos 5 estados com maior quantidade de Unidades de Ensino, os 4 primeiros estão nas regiões Sudeste e Sul (SP, MG, PR, RJ, BA). No ranking dos 5 estados com menor quantidade de Unidades de Ensino, os 4 primeiros estão na região Norte (AP, AC, RR, RO, AL).

2) Quais as 5 capitais com maior e menor quantidade de unidades de ensino?

Ranking de Quantidade de Unidades de Ensino por Capital (5 maiores)



Ranking de Quantidade de Unidades de Ensino por Capital (5 menores)



No ranking das 5 capitais com maior quantidade de Unidades de Ensino, as 4 primeiras estão nas regiões Sudeste e Sul. No ranking dos 5 capitais com menor quantidade de Unidades de Ensino, as 4 primeiras estão na região Norte.

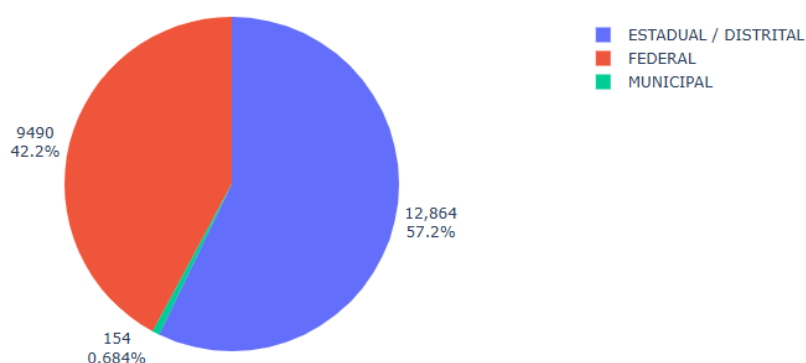
3) Quais os 5 estados com maior e menor quantidade de unidades de ensino em municípios diferentes de capitais?

^A _C UF	¹ ₂ QTD_UNIDADES	^A _C UF	¹ ₂ QTD_UNIDADES
SP	3769	RR	21
MG	3455	AP	25
PR	1481	AC	34
RS	1039	AM	86
BA	1008	AL	100

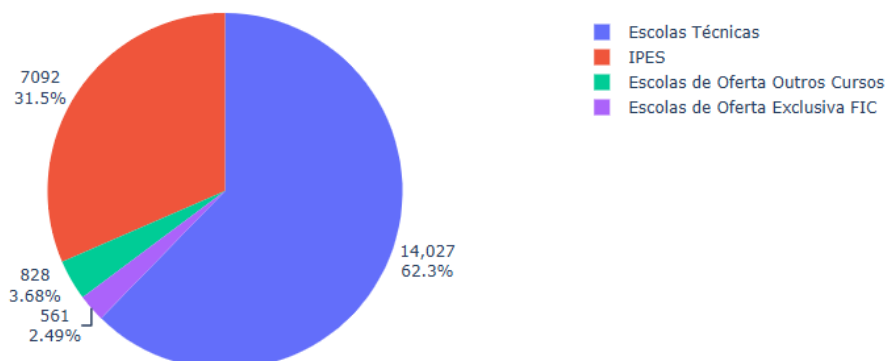
No ranking dos 5 estados com maior quantidade de Unidades de Ensino em municípios diferentes de capitais (tabela à esquerda), os 4 primeiros estão nas regiões Sudeste e Sul (SP, MG, PR, RS, BA). No ranking dos 5 estados com menor quantidade de Unidades de Ensino em municípios diferentes de capitais (tabela à direita), os 4 primeiros estão na região Norte (RR, AP, AC, AM, AL).

4) Como está a distribuição da quantidade de unidades de ensino por sistema de ensino, tipo de escola, dependência administrativa e sub-dependência administrativa?

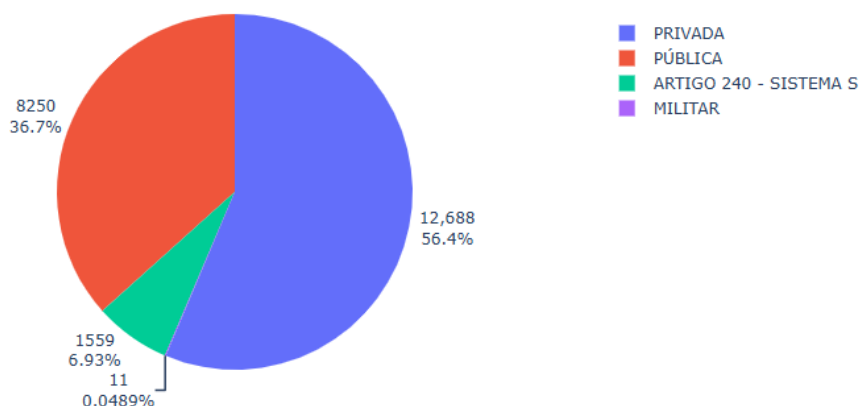
Distribuição da quantidade de unidades de ensino por Sistema de Ensino



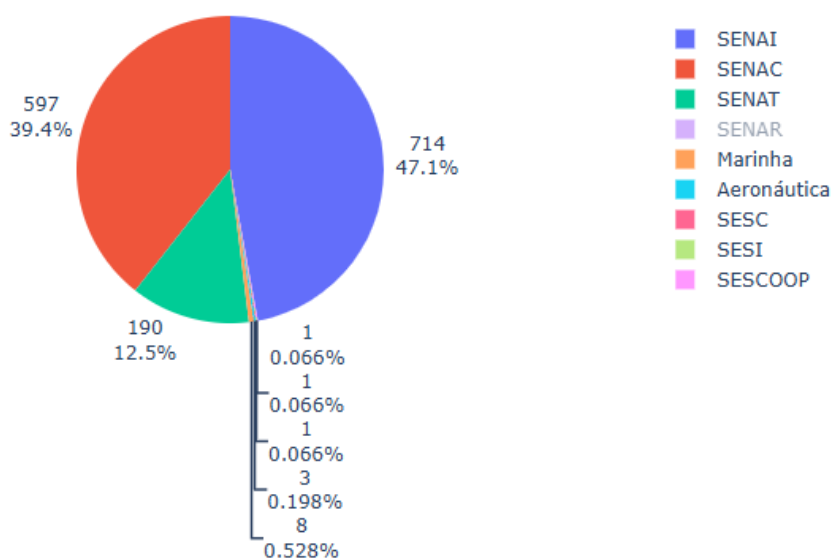
Distribuição da quantidade de unidades de ensino por Tipo de Escola



Distribuição da quantidade de unidades de ensino por Dependência Administrativa



Distribuição da quantidade de unidades de ensino por Dependência Administrativa



5) Quais as 5 unidades de ensino com maior e menor quantidade de cursos presenciais?
Em quais municípios e estados essas unidades estão?

Maior quantidade de cursos presenciais:

- UNIVERSIDADE FEDERAL DE MINAS GERAIS - COLEGIO TÉCNICO DA UFMG; MG; BELO HORIZONTE; 176 cursos.
- CETAM - UNIDADE DE MANAUS; AM; MANAUS; 60 cursos.
- IETAAM - INSTITUTO DE EDUCAÇÃO TECNOLÓGICA AVANÇADA DA AMAZÔNIA; PA; BELÉM; 54 cursos.
- SENAC CAMPINAS CENTRO; SP; CAMPINAS; 48 cursos.
- UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE - ESCOLA AGRÍCOLA DE JUNDIAÍ DA UFRN; RN; MACAÍBA; 45.

Menor quantidade de cursos presenciais:

- POLICURSOS COLEGIO; SP; MOGI DAS CRUZES; 1.
- COLÉGIO ESTADUAL FRANCISCO ASSUMPÇÃO; RJ; NOVA IGUAÇU; 1.
- EE JOÃO PEDRO PEDROSSIAN; MS; BODOQUENA; 1.
- SENAC- PITANGA; PR; PITANGA; 1.
- UNIVERSIDADE DO VALE DO ITAJAÍ; SC; BALNEÁRIO CAMBORIÚ; 1.

6) Quais as 5 unidades de ensino com a maior e menor quantidade de cursos à distância?
Em quais municípios e estados essas unidades estão?

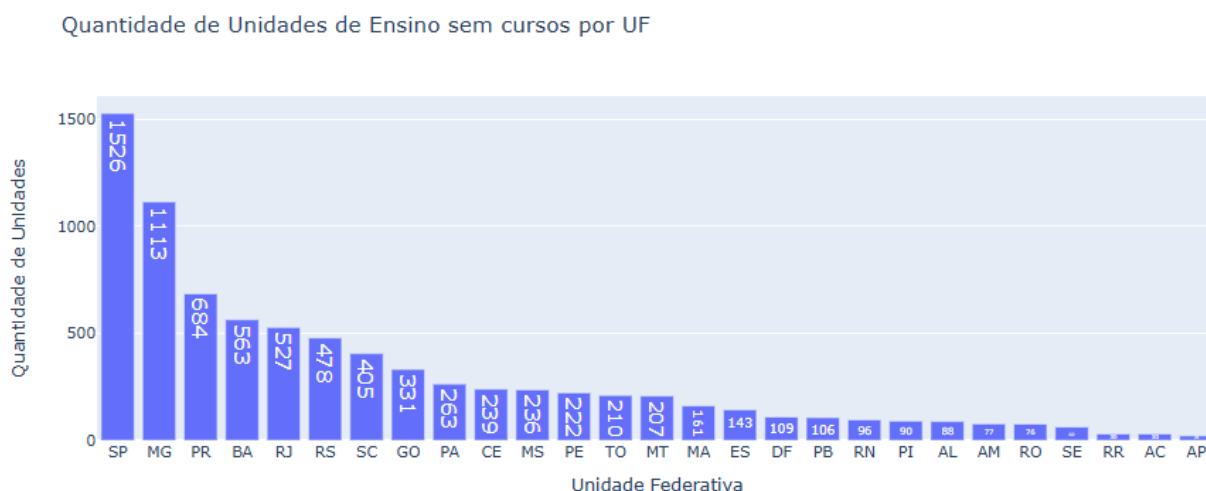
Maior quantidade de cursos à distância:

- FORMA CURSOS ENSINO A DISTÂNCIA LTDA; PB; JOÃO PESSOA; 65.
- ESCOLA RESIDÊNCIA SAÚDE; AL; MACEIÓ; 48.
- CIEBJA UNINTER - EF M P; PR; CURITIBA; 39.
- CENTRO DE EDUCAÇÃO E QUALIFICAÇÃO POTÊNCIA MASTER; PA; CASTANHAL; 36.
- ESCOLA TÉCNICA MERCURY; RJ; RIO DE JANEIRO; 34.

Menor quantidade de cursos à distância:

- COLEGIO ESTADUAL BARÃO DE ANTONINA - ENSINO FUNDAMENTAL-MEDIO- NORMAL E PROFISSIONAL; PR; RIO NEGRO; 1.
- CENTRO DE FORMACAO DE CORRETORES DE IMOVEIS / SINDIMOVEIS; CE; FORTALEZA; 1.
- CENTRO UNIVERSITÁRIO LEONARDO DA VINCI - UNIASSELVI - POLO DE APOIO PRESENCIAL DE CARAZINHO/RS; RS; CARAZINHO; 1.
- ESCOLA TÉCNICA MRX CURSOS; AL; MACEIÓ; 1.
- COLEGIO MONTES BELOS; GO; SÃO LUÍS DE MONTES BELOS; 1.

7) Existem unidades de ensino sem oferta de cursos? Em que estados elas estão?



Sim. Em todos os 26 estados e no Distrito Federal.

8) Qual o percentual de unidades de ensino com e sem cursos por UF?

summary	UF	QTD_UNIDADES	QTD_UNIDADES_COM_CURSOS	QTD_UNIDADES_SEM_CURSOS	PERC_COM_CURSOS	PERC_SEM_CURSOS
count	27	27	27	27	27	27
mean	null	833.6296296296297	533.8888888888889	299.74074074074076	61.25925925925926	39.7037037037037
stddev	null	1115.162945946195	774.4168194560408	347.34283580354094	8.088893585046067	8.037483412184413
min	AC	61	34	21	41	23
max	TO	4896	3370	1526	78	60

A média dos percentuais de unidades com oferta de cursos por estado é 61%, enquanto que a média dos percentuais de unidades sem oferta de cursos é 39%.

9) Qual o percentual de unidades de ensino dentro e fora de capitais por UF?

summary	UF	QTD_UNIDADES	QTD_UNIDADES_CAPITAIS	QTD_UNIDADES_NAO_CAPITAIS	PERC_CAPITAIS	PERC_NAO_CAPITAIS
count	27	27	27	27	27	27
mean	null	833.6296296296297	166.4814814814815	667.1481481481482	29.333333333333332	71.55555555555556
stddev	null	1115.162945946195	222.31480977180678	929.6091694778016	22.82879423346415	22.851246950210115
min	AC	61	0	0	0	0
max	TO	4896	1127	3769	100	100

A média dos percentuais de unidades em capitais por estado é 29%, enquanto que a média dos percentuais de unidades fora das capitais é 71%.

Resumo da Análise de Unidades de Ensino

A maior parte das análises realizadas sobre Unidades de Ensino procurou mapear sua distribuição geográfica. A partir dos resultados obtidos, percebe-se uma alta concentração de unidades nas regiões Sudeste e Sul, seja na análise por estado, por capitais ou por municípios diferentes de capitais. Ao mesmo tempo, pode-se observar baixa concentração de unidades na região Norte na análise por estado, por capitais e por municípios diferentes de capitais.

Desperta curiosidade o fato de que, na análise de unidades com maior oferta de cursos, verificamos a presença de unidades do Norte e do Nordeste na lista das maiores ofertas tanto na modalidade presencial, quanto à distância.

Quando observamos os percentuais de unidades dentro e fora de capitais por estado, percebemos que, em média, os estados possuem uma concentração maior de unidades de ensino fora das capitais, que é de 71%.

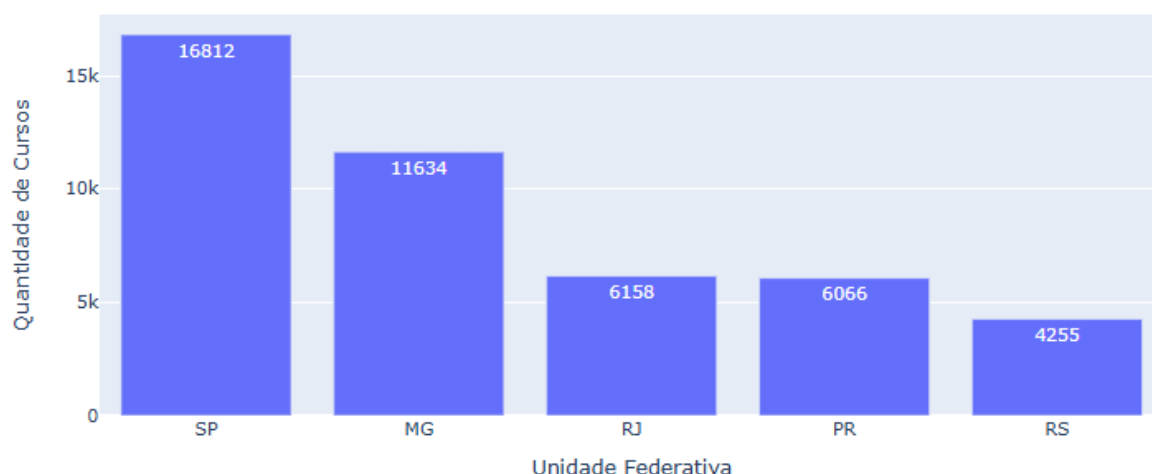
Um aspecto que merece atenção é o fato de que, em todos os estados, incluindo Distrito Federal, existem unidades sem oferta de cursos. Esse fato é comprovado pelo percentual de unidades sem oferta de cursos por estado, cuja média é 39%.

Ainda foram realizados outros agrupamentos para conhecer características dessas unidades de ensino e foi detectado que a maioria delas apresenta sistema de ensino Estadual ou Distrital (57,2%); tipo de Escola Técnica (62,3%); dependência administrativa Privada (56,4%); e para dependência Artigo 240 – Sistema S, temos subdependência administrativa SENAI (47,1%).

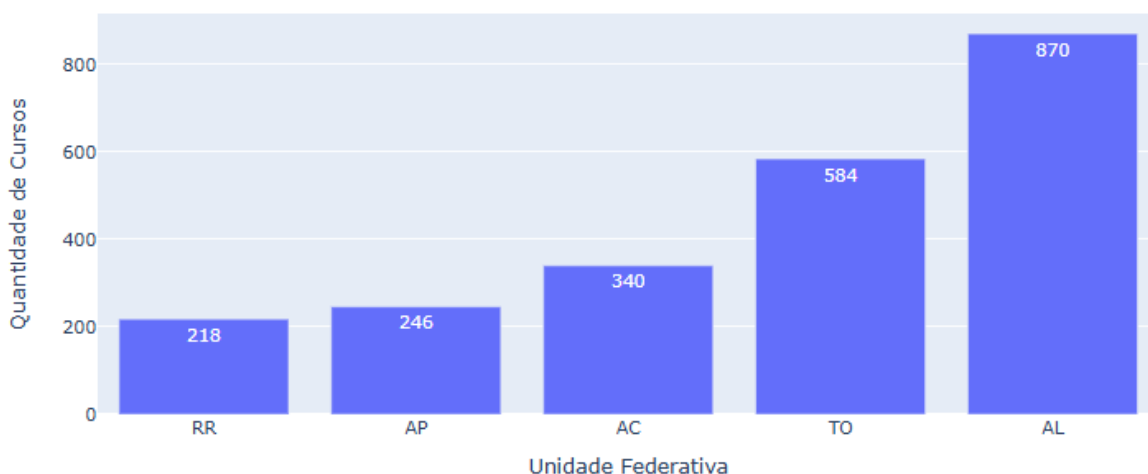
B) Sobre **Cursos**

1) Quais os 5 estados com maior e menor quantidade cursos?

Ranking de Quantidade de Cursos por UF (5 maiores)



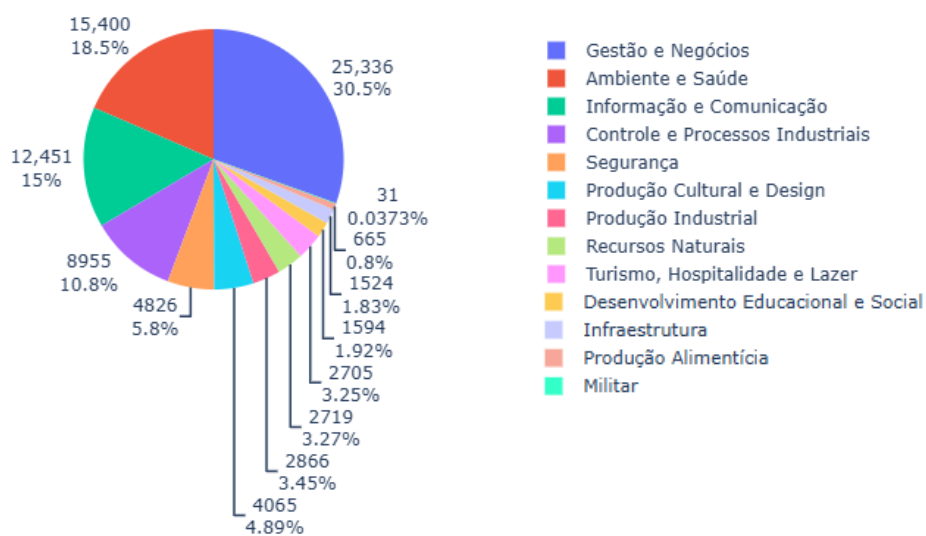
Ranking de Quantidade de Cursos por UF (5 menores)



No ranking dos 5 estados com maior quantidade de cursos, todos estão nas regiões Sudeste e Sul (SP, MG, RJ, PR, RS). No ranking dos 5 estados com menor quantidade de cursos, os 4 primeiros estão na região Norte (RR, AP, AC, TO).

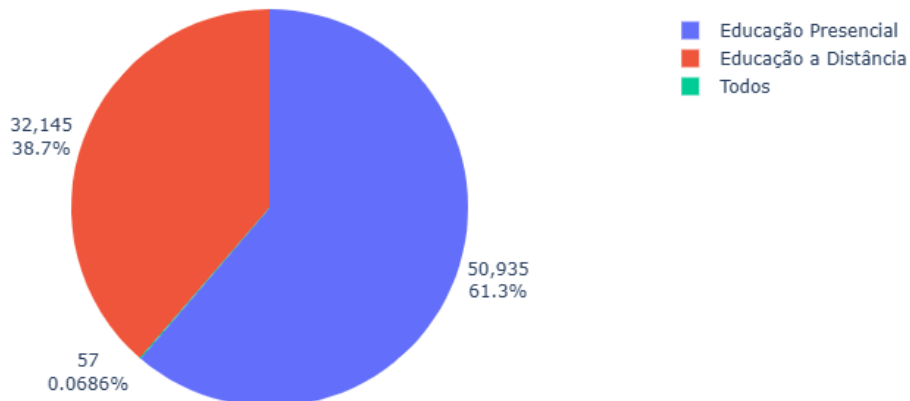
2) Como está a distribuição da quantidade de cursos por eixo tecnológico?

Distribuição da quantidade de cursos por Eixo Tecnológico



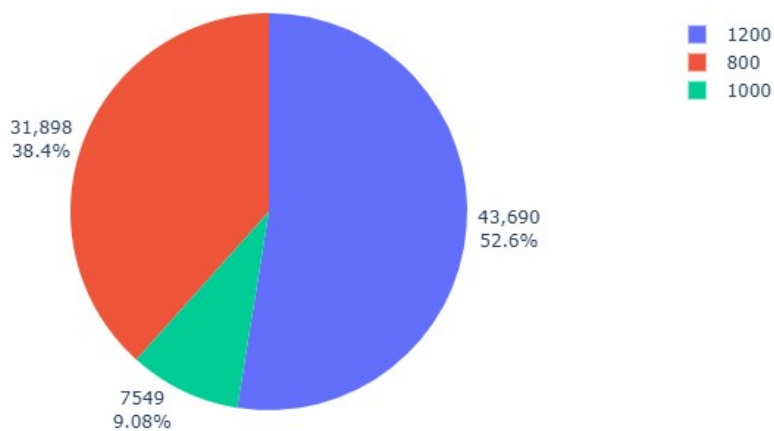
3) Como está a distribuição da quantidade de cursos por modalidade?

Distribuição da quantidade de cursos por Modalidade



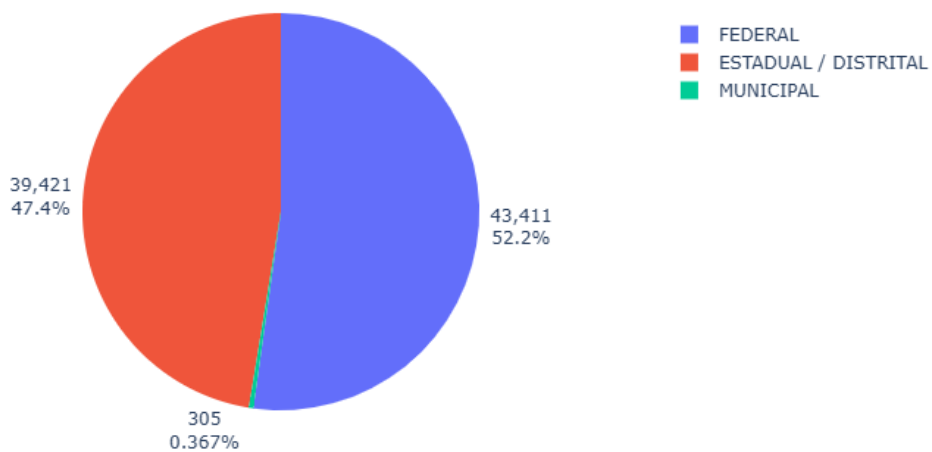
4) Como está a distribuição da quantidade de cursos por carga horária?

Distribuição da quantidade de cursos por Carga Horária



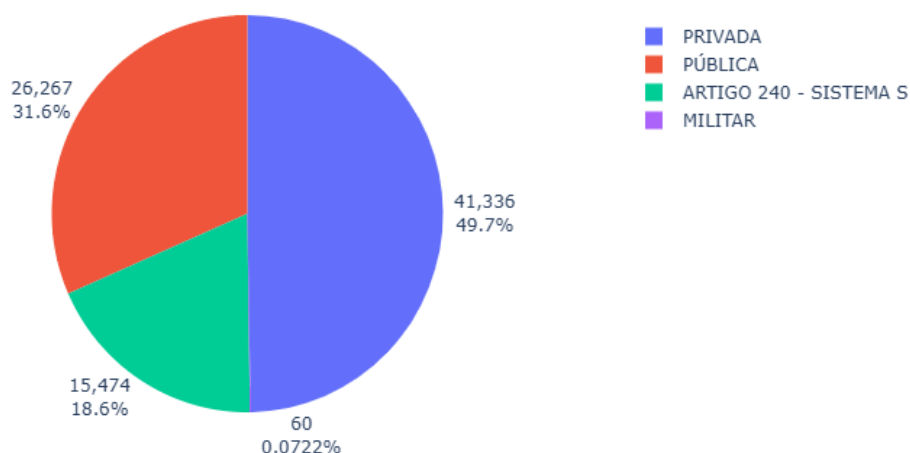
5) Como está a distribuição da quantidade de cursos por sistema de ensino?

Distribuição da quantidade de cursos por Sistema de Ensino



6) Como está a distribuição da quantidade de cursos por dependência administrativa?

Distribuição da quantidade de cursos por Dependência Administrativa



7) Quais os 5 cursos presenciais com maior e menor oferta?

A ^B _C CURSO	1 ² ₃ QTD_CURSOS
TÉCNICO EM ADMINISTRAÇÃO	5208
Técnico em Enfermagem	3899
Técnico em Informática	3615
Técnico em Segurança do Trabalho	2990
Técnico em Logística	2218

A ^B _C CURSO	1 ² ₃ QTD_CURSOS
Técnico em Pós-Colheita	1
Técnico em Eletricidade e Instrumentos Aeronáuticos	1
Técnico em Manobras e Equipamentos de Convés	1
Técnico em Processamento da Madeira	1
Técnico em Sensores de Aviação	1

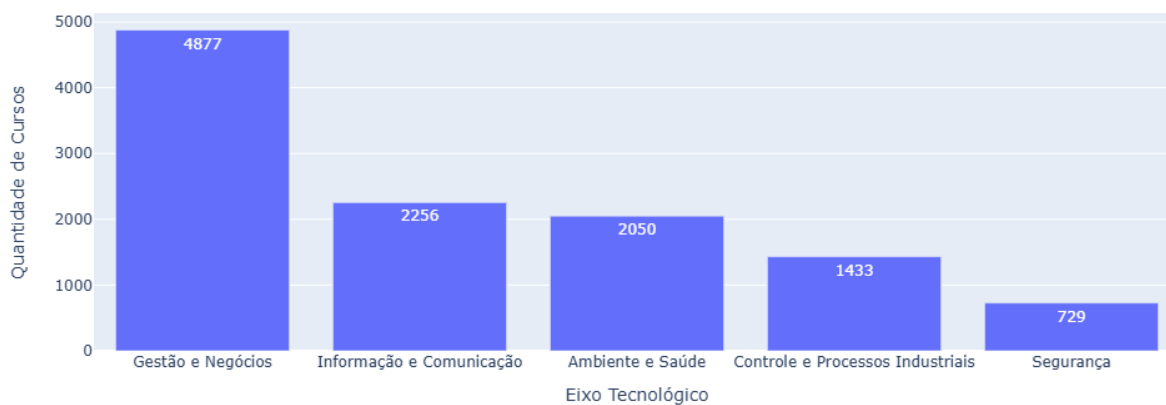
8) Quais os 5 cursos à distância com maior e menor oferta?

A ^B _C CURSO	1 ² ₃ QTD_CURSOS
TÉCNICO EM ADMINISTRAÇÃO	1834
Técnico em Segurança do Trabalho	1813
Técnico em Logística	1661
Técnico em Recursos Humanos	1526
Técnico em Transações Imobiliárias	1510

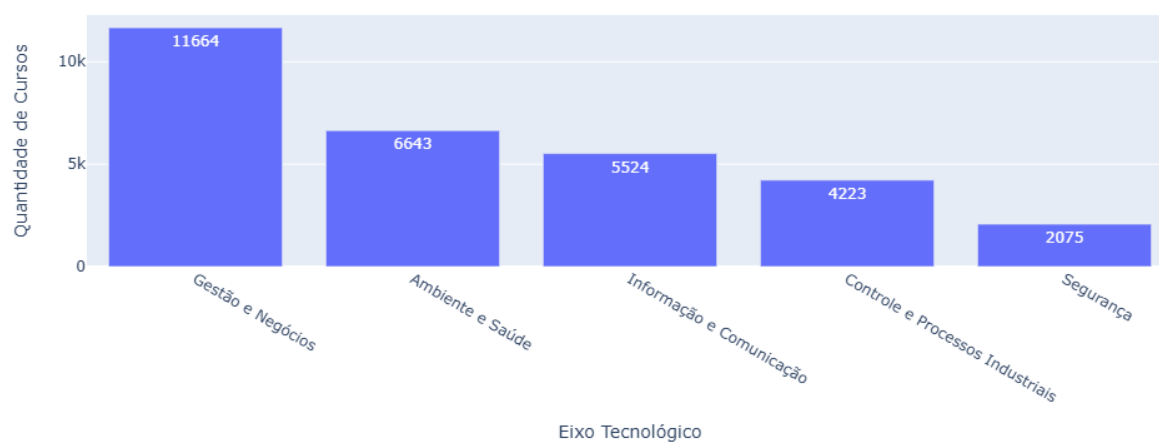
A ^B _C CURSO	1 ² ₃ QTD_CURSOS
Técnico em Ludoteca	1
Técnico em Cafeicultura	1
Técnico em Meio Ambiente	1
Técnico em Grãos	1
Técnico em Análises Clínicas	1

9) Qual a distribuição dos cursos por eixo e região?

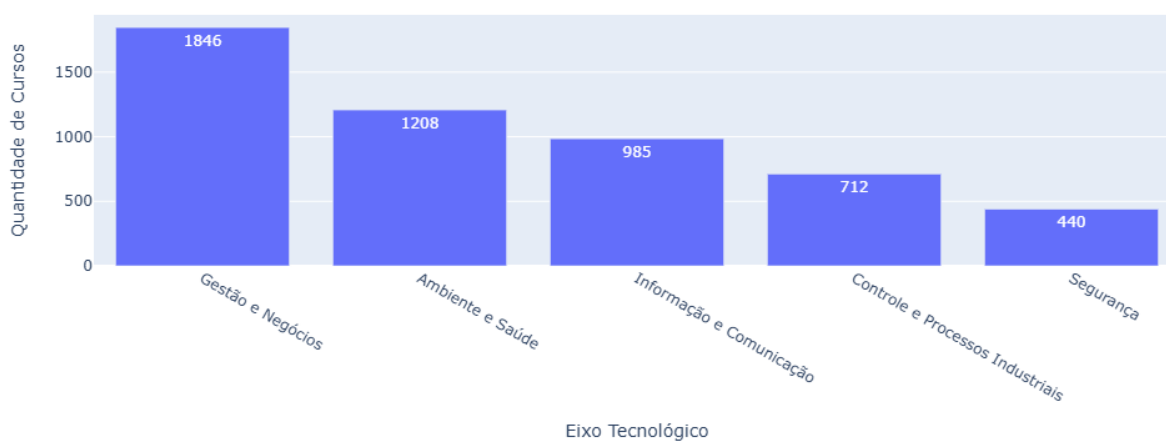
Ranking de Quantidade de Cursos por Eixo Tecnológico - Região Sul (5 maiores)



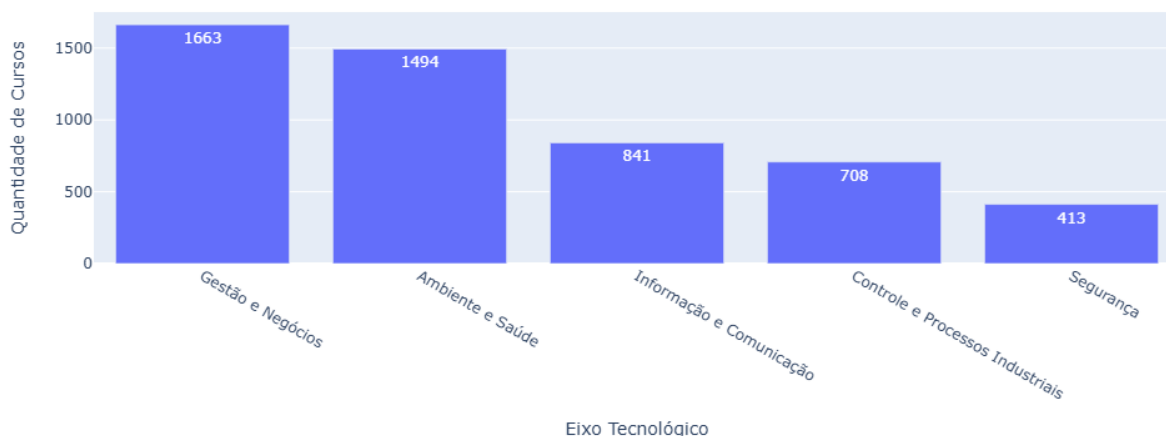
Ranking de Quantidade de Cursos por Eixo Tecnológico - Região Sudeste (5 maiores)



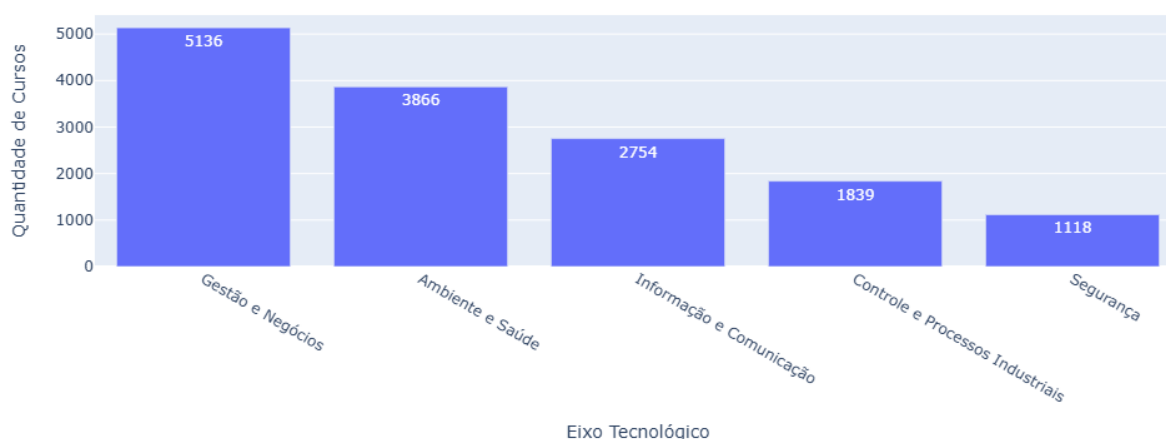
Ranking de Quantidade de Cursos por Eixo Tecnológico - Região Centro-Oeste (5 maiores)



Ranking de Quantidade de Cursos por Eixo Tecnológico - Região Norte (5 maiores)



Ranking de Quantidade de Cursos por Eixo Tecnológico - Região Nordeste (5 maiores)



Resumo da Análise de Cursos

A maior parte das análises realizadas sobre Cursos procurou mapear sua distribuição geográfica. No ranking dos 5 estados com maior quantidade de cursos, todos estão nas regiões Sudeste e Sul, enquanto que, no ranking dos 5 estados com menor quantidade de cursos, os 4 primeiros estão na região Norte. Esta análise segue a mesma tendência do ranking de maiores concentrações de unidades de ensino por estados, mostrado na seção anterior.

A análise de cursos por eixo tecnológico e região mostra um aspecto interessante: para todas as regiões (Sul, Sudeste, Centro-Oeste, Norte e Nordeste), os 5 maiores eixos tecnológicos são Gestão e Negócios; Ambiente e Saúde; Informação e Comunicação; Controle e Processos Industriais; e Segurança. Com exceção de uma pequena mudança de ordem entre o segundo e terceiro colocado na região Sul, todas as demais regiões seguem exatamente a ordem de cursos listada anteriormente. Essa mesma ordem pode ser encontrada no agrupamento global de cursos por eixo tecnológico.

Foram realizados outros agrupamentos para conhecer características desses cursos e foi detectado que a maioria deles está no eixo tecnológico Gestão e Negócios (32,4%); modalidade Educação Presencial (61,3%); carga horária 1.200 horas (52,6%); estão no sistema de ensino Federal (52,2%); e na dependência administrativa Privada (49,7%).

6. Autoavaliação

A princípio, considerando os conjuntos de dados escolhidos, considero que o objetivo de conhecer a distribuição das Unidades de Ensino e da oferta de Cursos Técnicos pelos estados e regiões do país foi atingido. Além disso, foi importante mapear alguns agrupamentos relacionados à natureza das Unidades de Ensino e dos Cursos, e também alguns fatos como proporção de Unidades sem Cursos e de Unidades fora das capitais dos estados.

Algumas consultas sobre identificação de Unidades de Ensino e de Cursos serviram mais como amostras de dados do que informações que colaborassem diretamente com o objetivo traçado.

Dados sobre ocupação populacional poderiam ter sido utilizados para validar se a quantidade de unidades e cursos por estado e região estaria adequada. Da mesma forma, dados sobre vagas de emprego seriam importantes para avaliar de forma qualitativa a distribuição de cursos por eixos tecnológicos.

7. Referências

[1] Ministério da Educação, SISTEC.

Disponível em: <https://www.gov.br/mec/pt-br/aceso-a-informacao/institucional/estrutura-organizacional/orgaos-especificos-singulares/secretaria-de-educacao-profissional/sistec>

[2] Portal Brasileiro de Dados Abertos.

Disponível em: <https://dados.gov.br/home>

[3] Databricks: O que é a arquitetura “medallion lakehouse”?

Disponível em: <https://docs.databricks.com/gcp/pt/lakehouse/medallion>