

UNIVERSIDADE PRESBITERIANA MACKENZIE  
Faculdade de Computação e Informática

CAROLINA MOLINARI MONTEFORTE  
LEANDRO DA CRUZ CIRQUEIRA  
LEVY SALLES BISPO DE OLIVEIRA  
WILLIAM SILVA VEÇOSO

Consumo e produção responsáveis através da educação  
Análise de dados do ensino básico e médio do Brasil visando a melhoria de indicadores  
educacionais.

## RESUMO

**Palavras-chave:**

## LISTA DE ILUSTRAÇÕES

Gráfico 1 - Determinação de valores faltantes (ENEM 2021).....	25
Gráfico 2 - Heatmap de correlação entre as variáveis (ENEM 2021).....	26
Gráfico 3 - Determinação de valores faltantes (Educação Básica).....	30
Gráfico 4 - Heatmap de correlação entre as variáveis (Educação Básica).....	31

## LISTA DE TABELAS

Tabela 1 -	Cronograma de tarefas.....	14
Tabela 2 -	Dicionário de metadados - Censo escolar 2020.....	16
Tabela 3 -	Dicionário de metadados - Enem 2021.....	19
Tabela 4 -	Resumo estatístico - Enem 2021.....	22
Tabela 5 -	Amostra frequência absoluta - Enem 2021.....	23
Tabela 6 -	Amostra frequência relativa - Enem 2021.....	24
Tabela 7 -	Resumo estatístico - Educação Básica.....	27
Tabela 8 -	Amostra frequência absoluta - Educação Básica.....	28
Tabela 9 -	Amostra frequência relativa - Educação Básica.....	29

## **LISTA DE ABREVIATURAS E SIGLAS**

INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
ODS	Objetivo de Desenvolvimento Sustentável
MEC	Ministério da Educação

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>12</b>
<b>2 PROPOSTA PARA SOLUÇÃO ANALÍTICA</b>	<b>12</b>
2.1 Apresentação da empresa/organização	12
2.2 Objetivos	13
2.3 Resultados Esperados	13
2.4 Metodologia	13
<b>3 CRONOGRAMA</b>	<b>14</b>
<b>4 ANÁLISE EXPLORATÓRIA DE DADOS</b>	<b>16</b>
4.1 Metadados	16
4.2 Análise descritiva (Estatística descritiva)	22
4.2.1. Enem 2021	22
4.2.2 Educação Básica	26
<b>5 DESENVOLVIMENTO</b>	<b>32</b>
5.1 Exemplos de ilustrações	32
<b>6 METODOLOGIA</b>	<b>33</b>
<b>7 RESULTADOS</b>	<b>34</b>
<b>8 CONCLUSÃO</b>	<b>35</b>
<b>9 REFERÊNCIAS</b>	<b>36</b>

## **1 INTRODUÇÃO**

### **Análise de dados do ensino básico e médio do Brasil visando a melhoria de indicadores educacionais.**

Este projeto tem como objetivo analisar os dados do ensino fundamental e médio no Brasil em 2020, a fim de verificar a existência de correlações entre as diferentes regiões do país e os resultados do Enem de 2021. Com base nessa análise, pretende-se propor melhorias nos indicadores educacionais.

Este projeto está vinculado ao Objetivo de Desenvolvimento Sustentável (ODS) número 12 da Organização das Nações Unidas (ONU), que visa assegurar padrões de produção e de consumo sustentáveis até 2030. Especificamente, o projeto se relaciona com as metas:

12.8: Integrar a educação para a cidadania global e o desenvolvimento sustentável nas políticas nacionais de educação, currículos escolares, formação de professores e avaliação de estudantes;

12.8.1: Avaliar o grau em que a educação para a cidadania global e o desenvolvimento sustentável são integrados nas políticas nacionais de educação, currículos escolares, formação de professores e avaliação de estudantes;

12.a: Apoiar países em desenvolvimento na melhoria de suas capacidades científicas e tecnológicas para mudar para padrões mais sustentáveis de produção e consumo.

## **2 PROPOSTA PARA SOLUÇÃO ANALÍTICA**

### **2.1 Apresentação da empresa/organização**

**Organização:** Ministério da Educação (MEC)

**Área de atuação:** Educação

**Missão:** Promover o desenvolvimento humano, social e econômico do país por meio da educação

**Importância para o país:** Ministério de extrema importância nacional, pois é responsável por formular e implementar políticas públicas educacionais que visam garantir o acesso à educação de qualidade para todos os brasileiros.

**Número de colaboradores:** 391.312( em 2023)

**Repositório:**[https://github.com/CirqueiraLeandro/MACKENZIE\\_DATA\\_SCIENCE/tree/main/projeto\\_aplicado\\_1](https://github.com/CirqueiraLeandro/MACKENZIE_DATA_SCIENCE/tree/main/projeto_aplicado_1)

## 2.2 Objetivos

O objetivo principal deste projeto é realizar uma análise exploratória dos dados do ensino fundamental e médio no Brasil em 2020, a fim de verificar a existência de correlações entre as diferentes regiões do país e os resultados do Enem de 2021. E com isso é esperado que seja possível:

- Identificar padrões e tendências nos dados educacionais do país;
- Propor melhorias nos indicadores educacionais, visando contribuir para o desenvolvimento sustentável do país;
- Contribuir para o alcance das metas estabelecidas pelo ODS 12, em especial as metas 12.8, 12.8.1 e 12.a, que visam desenvolvimento, produção e consumo responsáveis através da educação.

## 2.3 Resultados Esperados

Espera-se que esta análise traga à tona informações relevantes sobre o desempenho do ensino fundamental e médio em diferentes regiões do Brasil, bem como sobre a correlação entre esses resultados e o Enem de 2021. A partir disso, espera-se identificar oportunidades de melhoria nos indicadores educacionais e propor soluções que possam contribuir para um desenvolvimento mais sustentável do país.

## 2.4 Metodologia

Para alcançar os objetivos propostos, serão utilizados dados públicos disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Esses dados serão analisados utilizando técnicas de análise exploratória de dados e correlação, a fim de identificar possíveis relações entre as variáveis estudadas.



### 3 CRONOGRAMA

Tabela 1 – Cronograma de atividades

<b>Etapa</b>	<b>Atividade</b>	<b>Responsável</b>	<b>Data de entrega</b>	<b>Status</b>
<b>Etapa 1 – Apresentação do projeto, objetivos, metas e milestones</b> <i>Data de entrega: 28/08/2023</i>	Definição do tema	Todos	18/08/2023	Concluído
	Cronograma de atividades	Carolina Molinari Monteforte	23/08/2023	Concluído
	Pesquisa de bases	Levy Salles Bispo de Oliveira	25/08/2023	Concluído
	Criação do repositório	Leandro da Cruz Cirqueira	25/08/2023	Concluído
	Criação do documento descritivo do projeto	William Silva Veçoso	25/08/2023	Concluído
<b>Etapa 2 – Definição do produto analítico</b> <i>Data de entrega: 25/09/2023</i>	Preparação dos dados	Leandro da Cruz Cirqueira	02/09/2023	Concluído
	Análise descritiva do comportamento das variáveis	William Silva Veçoso	09/09/2023	Concluído
	Elaboração da proposta de solução analítica	Todos	20/09/2023	Concluído
	Documentação das etapas no documento descritivo do projeto e revisão	Todos	23/09/2023	Concluído

<b>Etapa 3 – Apresentação de produtos</b> <i>Data de entrega:</i> 30/10/2023	Identificação de padrões e comportamento das variáveis em relação umas com as outras	Carolina Molinari Monteforte Levy Salles Bispo de Oliveira	15/09/2023	Não iniciado
	Levantamento de hipóteses	Todos	16/09/2023	Não iniciado
	Criação das visualizações necessárias	Carolina Molinari Monteforte	20/10/2023	Não iniciado
	Revisão dos scripts e do conteúdo do repositório	William Silva Veçoso	25/10/2023	Não iniciado
	Organização da narrativa e criação de roteiro do vídeo	Levy Salles Bispo de Oliveira Leandro da Cruz Cirqueira	25/10/2023	Não iniciado
	Documentação das etapas no documento descritivo do projeto	Todos	27/10/2023	Não iniciado
<b>Etapa 4 – Encerramento do projeto</b> <i>Data de entrega:</i> 20/11/2023	Gravação do vídeo de apresentação do trabalho	Todos	10/11/2023	Não iniciado
	Revisão do documento final do projeto	Todos	17/11/2023	Não iniciado

Fonte: Os autores.

## 4 ANÁLISE EXPLORATÓRIA DE DADOS

**Objetivo:** Descobrir quais são os dados disponíveis nesse estudo, analisando os mesmos e resumizando suas principais características.

### 4.1 Metadados

**Nome da tabela:** Censo escolar 2020

**Descrição:** Base de dados do censo escolar da educação básica do ano de 2020 com ajustes (remoção de colunas com informação desnecessária para análise)

**Origem:** [https://github.com/CirqueiraLeandro/MACKENZIE\\_DATA\\_SCIENCE/blob/main/projeto\\_aplicado\\_1/datasets/ed\\_basica\\_2020.zip](https://github.com/CirqueiraLeandro/MACKENZIE_DATA_SCIENCE/blob/main/projeto_aplicado_1/datasets/ed_basica_2020.zip)

**Tipo de arquivo:** csv

**Sensibilidade:** Não há dados sensíveis

**Validade:** Dados relacionados ao ano de 2020

**Proprietário do dado:** INEP

**Restrições de uso:** Não há restrições de uso

**Número de colunas:** 46

**Número de linhas :** 224229

Tabela 2 – Dicionário de metadados - Censo escolar 2020

DICIONÁRIO DE METADADOS - CENSO ESCOLAR 2020				
NOME DA VARIÁVEL	Descrição	Variáveis Categóricas		Tipo
		Categoria	Descrição	
NU_ANO_CENSO	Ano do Censo			Numérica
NO_REGIAO	Nome da região geográfica			Alfanumérica
SG_UF	Sigla da Unidade da Federação			Alfanumérica
NO_MUNICIPIO	Nome do Município			Alfanumérica
NO_MICRORREGIAO	Nome da Microrregião			Alfanumérica
CO_ENTIDADE	Código da Escola			Numérica
NO_ENTIDADE	Nome da Escola			Alfanumérica

TP_DEPENDENCIA	Dependência Administrativa	1 2 3 4	Federal Estadual Municipal Privada	Numérica
TP_CATEGORIA_ESCOLA_PRIVADA	Categoria da escola privada	1 2 3 4 -	Particular Comunitária Confessional Filantrópica Não aplicável para escolas públicas	Numérica
TP_LOCALIZACAO	Localização	1 2	Urbana Rural	Numérica
IN_VINCULO_SECRETARIA_EDUCACAO	Órgão que a escola pública está vinculada - Secretaria de Educação/Ministério da Educação	0 1 -	Não Sim Não aplicável para escolas privadas	Numérica
IN_AGUA_FILTADA	Água consumida pelos alunos	0 1	Não Sim	Numérica
IN_AGUA_POTAVEL	Fornece água potável para o consumo humano	0 1	Não Sim	Numérica
IN_ENERGIA_EXISTENTE	Abastecimento de energia elétrica - Não há energia elétrica	0 1	Não Sim	Numérica
IN_ESGOTO_INEXISTENTE	Esgoto sanitário - Não há esgotamento sanitário	0 1	Não Sim	Numérica
IN_BANHEIRO	Dependências físicas existentes e utilizadas na escola - Banheiro	0 1	Não Sim	Numérica
IN_BIBLIOTECA	Dependências físicas existentes e utilizadas na escola - Biblioteca	0 1	Não Sim	Numérica
IN_BIBLIOTECA_SALA_LEITURA	Dependências físicas existentes e utilizadas na escola - Biblioteca e/ou Sala de leitura	0 1	Não Sim	Numérica
IN_COZINHA	Dependências físicas existentes e utilizadas na escola - Cozinha	0 1	Não Sim	Numérica
IN_LABORATORIO_Ciencias	Dependências físicas existentes e utilizadas na escola - Laboratório de ciências	0 1	Não Sim	Numérica
IN_LABORATORIO_INFORMATICA	Dependências físicas existentes e utilizadas na escola - Laboratório de informática	0 1	Não Sim	Numérica
IN_REFEITORIO	Dependências físicas existentes e utilizadas na escola - Refeitório	0 1	Não Sim	Numérica
QT_SALAS_EXISTENTES	Número de salas de aula existentes na escola			Numérica
QT_SALAS_UTILIZADAS	Número de salas de aula utilizadas na escola (dentro e fora do prédio)			Numérica

IN_DESKTOP_ALUNO	Computadores em uso pelos alunos - Computador de mesa (desktop)	0 1	Não Sim	Numérica
IN_INTERNET_ALUNOS	Acesso à Internet - Para uso dos alunos	0 1	Não Sim	Numérica
IN_ACESSO_INTERNET_COMPUTADOR	Equipamentos que os alunos usam para acessar a internet da escola - Computadores de mesa, portáteis e tablets da escola (no laboratório de informática, biblioteca, sala de aula, etc.)	0 1 9	Não Sim Não informado	Numérica
QT_FUNCIONARIOS	Total de funcionários da escola (inclusive profissionais escolares em sala de aula)			Numérica
IN_EXAME_SELECAO	A escola faz exame de seleção para ingresso de seus alunos (Avaliação por prova e/ou análise curricular)	0 1 9	Não Sim Não informado	Numérica
QT_MAT_BAS	Número de Matrículas na Educação Básica			Numérica
QT_MAT_FUND	Número de Matrículas no Ensino Fundamental			Numérica
QT_MAT_MED	Número de Matrículas no Ensino Médio			Numérica
QT_MAT_PROF	Número de Matrículas na Educação Profissional			Numérica
QT_MAT_BAS_FEM	Número de Matrículas na Educação Básica - Feminino			Numérica
QT_MAT_BAS_MASC	Número de Matrículas na Educação Básica - Masculino			Numérica
QT_MAT_INF_INT	Número de Matrículas na Educação Infantil - Tempo Integral			Numérica
QT_MAT_FUND_INT	Número de Matrículas no Ensino Fundamental - Tempo Integral			Numérica
QT_MAT_MED_INT	Número de Matrículas no Ensino Médio - Tempo Integral			Numérica
QT_DOC_BAS	Número de Docentes da Educação Básica			Numérica
QT_DOC_FUND	Número de Docentes do Ensino Fundamental			Numérica
QT_DOC_MED	Número de Docentes do Ensino Médio			Numérica
QT_DOC_PROF	Número de Docentes da Educação Profissional			Numérica
QT_TUR_BAS	Número de Turmas de Educação Básica			Numérica
QT_TUR_FUND	Número de Turmas de Ensino Fundamental			Numérica
QT_TUR_MED	Número de Turmas de Ensino Médio			Numérica
QT_TUR_PROF	Número de Turmas de Educação			Numérica

	Profissional			
--	--------------	--	--	--

Fonte: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira | Inep (editado).

**Nome da tabela:** Enem 2021

**Descrição:** Base de dados do Enem do ano de 2021 com ajustes (remoção de colunas com informação desnecessária para análise)

**Origem:** [https://github.com/CirqueiraLeandro/MACKENZIE\\_DATA\\_SCIENCE/blob/main/projeto\\_aplicado\\_1/datasets/enem\\_2021.zip](https://github.com/CirqueiraLeandro/MACKENZIE_DATA_SCIENCE/blob/main/projeto_aplicado_1/datasets/enem_2021.zip)

**Tipo de arquivo:** csv

**Sensibilidade:** Não há dados sensíveis

**Validade:** Dados relacionados ao ano de 2021

**Proprietário do dado:** INEP

**Restrições de uso:** Não há restrições de uso

**Número de colunas:** 21

**Número de linhas :** 3389832

Tabela 3 – Dicionário de metadados - Enem 2021

DICIONÁRIO DE VARIÁVEIS - ENEM				
NOME DA VARIÁVEL	Descrição	Variáveis Categóricas		Tipo
		Categoria	Descrição	
NU_ANO	Ano do Enem			Númerica
TP_ST_CONCLUSAO	Situação de conclusão do Ensino Médio	1	Já concluí o Ensino Médio	Númerica
		2	Estou cursando e concluirei o Ensino Médio em 2021	
		3	Estou cursando e concluirei o Ensino Médio após 2021	
		4	Não concluí e não estou cursando o Ensino Médio	
TP_ANO_CONCLUSIU	Ano de Conclusão do Ensino Médio	0	Não informado	Númerica
		1	2020	
		2	2019	

		3	2018	
		4	2017	
		5	2016	
		6	2015	
		7	2014	
		8	2013	
		9	2012	
		10	2011	
		11	2010	
		12	2009	
		13	2008	
		14	2007	
		15	Antes de 2007	
TP_ESCOLA	Tipo de escola do Ensino Médio	1	Não Respondeu	Numérica
		2	Pública	
		3	Privada	
IN_TREINEIRO	Indica se o inscrito fez a prova com intuito de apenas treinar seus conhecimentos <sup>3</sup>	1	Sim	Numérica
		0	Não	
NO_MUNICIPIO_ESC	Nome do município da escola			Alfanumérica
SG_UF_ESC	Sigla da Unidade da Federação da escola			Alfanumérica
TP_DEPENDENCIA_ADM_ESC	Dependência administrativa (Escola)	1	Federal	Numérica
		2	Estadual	
		3	Municipal	
		4	Privada	
TP_LOCALIZACAO_ESC	Localização (Escola)	1	Urbana	Numérica
		2	Rural	
TP_PRESENCIA_CN	Presença na prova objetiva de Ciências da Natureza	0	Faltou à prova	Numérica
		1	Presente na prova	
		2	Eliminado na prova	
TP_PRESENCIA_CH	Presença na prova objetiva de Ciências Humanas	0	Faltou à prova	Numérica
		1	Presente na prova	
		2	Eliminado na prova	
TP_PRESENCIA_LC	Presença na prova objetiva de Linguagens e Códigos	0	Faltou à prova	Numérica
		1	Presente na prova	
		2	Eliminado na prova	
TP_PRESENCIA_MT	Presença na prova objetiva de Matemática	0	Faltou à prova	Numérica
		1	Presente na prova	
		2	Eliminado na prova	
NU_NOTA_CN	Nota da prova de Ciências da Natureza			Numérica
NU_NOTA_CH	Nota da prova de Ciências			Numérica

	Humanas			
NU_NOTA_LC	Nota da prova de Linguagens e Códigos			Númerica
NU_NOTA_MT	Nota da prova de Matemática			Númerica
NU_NOTA_REDACAO	Nota da prova de redação			Númerica
Q001	Até que série seu pai, ou o homem responsável por você, estudou?	A	Nunca estudou.	Alfanumérica
		B	Não completou a 4ª série/5º ano do Ensino Fundamental.	
		C	Completou a 4ª série/5º ano, mas não completou a 8ª série/9º ano do Ensino Fundamental.	
		D	Completou a 8ª série/9º ano do Ensino Fundamental, mas não completou o Ensino Médio.	
		E	Completou o Ensino Médio, mas não completou a Faculdade.	
		F	Completou a Faculdade, mas não completou a Pós-graduação.	
		G	Completou a Pós-graduação.	
		H	Não sei.	
Q002	Até que série sua mãe, ou a mulher responsável por você, estudou?	A	Nunca estudou.	Alfanumérica
		B	Não completou a 4ª série/5º ano do Ensino Fundamental.	
		C	Completou a 4ª série/5º ano, mas não completou a 8ª série/9º ano do Ensino Fundamental.	
		D	Completou a 8ª série/9º ano do Ensino Fundamental, mas não completou o Ensino Médio.	
		E	Completou o Ensino Médio, mas não completou a Faculdade.	
		F	Completou a Faculdade, mas não completou a Pós-graduação.	
		G	Completou a Pós-graduação.	
		H	Não sei.	
Q025	Na sua residência tem acesso à Internet?	A	Não.	Alfanumérica
		B	Sim.	

Fonte: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira | Inep (editado).



## 4.2 Análise descritiva (Estatística descritiva)

### 4.2.1. Enem 2021

- **Resumo estatístico:** Apresentação de medidas de tendência central e medidas de dispersão

Tabela 4 – Resumo estatístico - Enem 2021

% Nulos	Origem	Variável	count	mean	std	min	25%	50%	75%	max
0.00	enem_2021	Unnamed: 0	3389832.00	1694915.50	978560.35	0.00	847457.75	1694915.50	2542373.25	3389831.00
0.00	enem_2021	NU_ANO	3389832.00	2021.00	0.00	2021.00	2021.00	2021.00	2021.00	2021.00
0.00	enem_2021	TP_ST_CONCLUSAO	3389832.00	1.61	0.72	1.00	1.00	1.00	2.00	4.00
0.00	enem_2021	TP_ANO_CONCLUIU	3389832.00	2.58	4.28	0.00	0.00	0.00	3.00	15.00
0.00	enem_2021	TP_ESCOLA	3389832.00	1.40	0.59	1.00	1.00	1.00	2.00	3.00
0.00	enem_2021	IN_TREINEIRO	3389832.00	0.13	0.34	0.00	0.00	0.00	0.00	1.00
75.99	enem_2021	TP_DEPENDENCIA_ADM_ESC	813806.00	2.47	0.94	1.00	2.00	2.00	4.00	4.00
75.99	enem_2021	TP_LOCALIZACAO_ESC	813806.00	1.03	0.18	1.00	1.00	1.00	1.00	2.00
75.99	enem_2021	TP_PRESENCIA_CN	3389832.00	0.66	0.47	0.00	0.00	1.00	1.00	2.00
75.99	enem_2021	TP_PRESENCIA_CH	3389832.00	0.70	0.46	0.00	0.00	1.00	1.00	2.00
0.00	enem_2021	TP_PRESENCIA_LC	3389832.00	0.70	0.46	0.00	0.00	1.00	1.00	2.00
0.00	enem_2021	TP_PRESENCIA_MT	3389832.00	0.66	0.47	0.00	0.00	1.00	1.00	2.00
0.00	enem_2021	NU_NOTA_CN	2245844.00	491.79	80.57	0.00	431.00	482.00	545.80	867.10
0.00	enem_2021	NU_NOTA_CH	2378379.00	519.94	95.79	0.00	445.40	521.90	589.30	846.90
33.75	enem_2021	NU_NOTA_LC	2378379.00	502.56	79.16	0.00	447.60	505.90	558.80	826.10
29.84	enem_2021	NU_NOTA_MT	2245844.00	535.08	110.63	0.00	444.70	515.40	613.90	953.10
29.84	enem_2021	NU_NOTA_REDACAO	2378379.00	613.50	190.71	0.00	520.00	600.00	740.00	1000.00
33.75	enem_2021	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
29.84	enem_2021	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
0.00	enem_2021	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
0.00	enem_2021	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
0.00	enem_2021	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Fonte: Autores

Nos atributos NU\_NOTA\_CN, NU\_NOTA\_CH, NU\_NOTA\_LC, NU\_NOTA\_MT, NU\_NOTA\_REDACAO percebemos um desvio padrão considerável entre os resultados, o que indica que os dados dessas variáveis estão bastante espalhados em torno das médias (os dados completos podem ser encontrados no [link](#)). Isso pode indicar que há uma grande variedade de níveis de habilidade dos estudantes.

- **Distribuição de frequências:** Apresentação da frequência absoluta e relativa de cada variável. Por conta do volume de informação, só será apresentado uma amostra, podendo verificar o resultado completo no [link](#)

Tabela 5 – Amostra frequência absoluta - Enem 2021

Distribuição de Frequência da Variável TP_DEPENDENCIA_ADM_ESC - Frequência Absoluta:					
	Valor	Frequência Absoluta	Variância	Desvio Padrão	Amplitude
2	1.00	46789	1.67	1.29	3.00
0	2.00	548998	1.67	1.29	3.00
3	3.00	7254	1.67	1.29	3.00
1	4.00	210765	1.67	1.29	3.00

Distribuição de Frequência da Variável TP_LOCALIZACAO_ESC - Frequência Absoluta:					
	Valor	Frequência Absoluta	Variância	Desvio Padrão	Amplitude
0	1.00	785431	0.50	0.71	1.00
1	2.00	28375	0.50	0.71	1.00

Distribuição de Frequência da Variável TP_PRESENCA_CN - Frequência Absoluta:					
	Valor	Frequência Absoluta	Variância	Desvio Padrão	Amplitude
1	0	1142539	1.00	1.00	2
0	1	2245844	1.00	1.00	2
2	2	1449	1.00	1.00	2

Distribuição de Frequência da Variável TP_PRESENCA_CH - Frequência Absoluta:					
	Valor	Frequência Absoluta	Variância	Desvio Padrão	Amplitude
1	0	1007397	1.00	1.00	2
0	1	2378379	1.00	1.00	2
2	2	4056	1.00	1.00	2

Fonte: Autores

Tabela 6 – Amostra frequência relativa - Enem 2021

Distribuição de Frequência da Variável TP_DEPENDENCIA_ADM_ESC - Frequência Relativa:					
Valor	Frequência	Relativa	Média	Desvio Padrão	Coeficiente de Variação
2	1.00	0.06	2.50	1.29	51.64
0	2.00	0.67	2.50	1.29	51.64
3	3.00	0.01	2.50	1.29	51.64
1	4.00	0.26	2.50	1.29	51.64

Distribuição de Frequência da Variável TP_LOCALIZACAO_ESC - Frequência Relativa:					
Valor	Frequência	Relativa	Média	Desvio Padrão	Coeficiente de Variação
0	1.00	0.97	1.50	0.71	47.14
1	2.00	0.03	1.50	0.71	47.14

Distribuição de Frequência da Variável TP_PRESENCA_CN - Frequência Relativa:					
Valor	Frequência	Relativa	Média	Desvio Padrão	Coeficiente de Variação
1	0	0.34	1.00	1.00	100.00
0	1	0.66	1.00	1.00	100.00
2	2	0.00	1.00	1.00	100.00

Distribuição de Frequência da Variável TP_PRESENCA_CH - Frequência Relativa:					
Valor	Frequência	Relativa	Média	Desvio Padrão	Coeficiente de Variação
1	0	0.30	1.00	1.00	100.00
0	1	0.70	1.00	1.00	100.00
2	2	0.00	1.00	1.00	100.00

Distribuição de Frequência da Variável TP_PRESENCA_LC - Frequência Relativa:					
Valor	Frequência	Relativa	Média	Desvio Padrão	Coeficiente de Variação
1	0	0.30	1.00	1.00	100.00
0	1	0.70	1.00	1.00	100.00
2	2	0.00	1.00	1.00	100.00

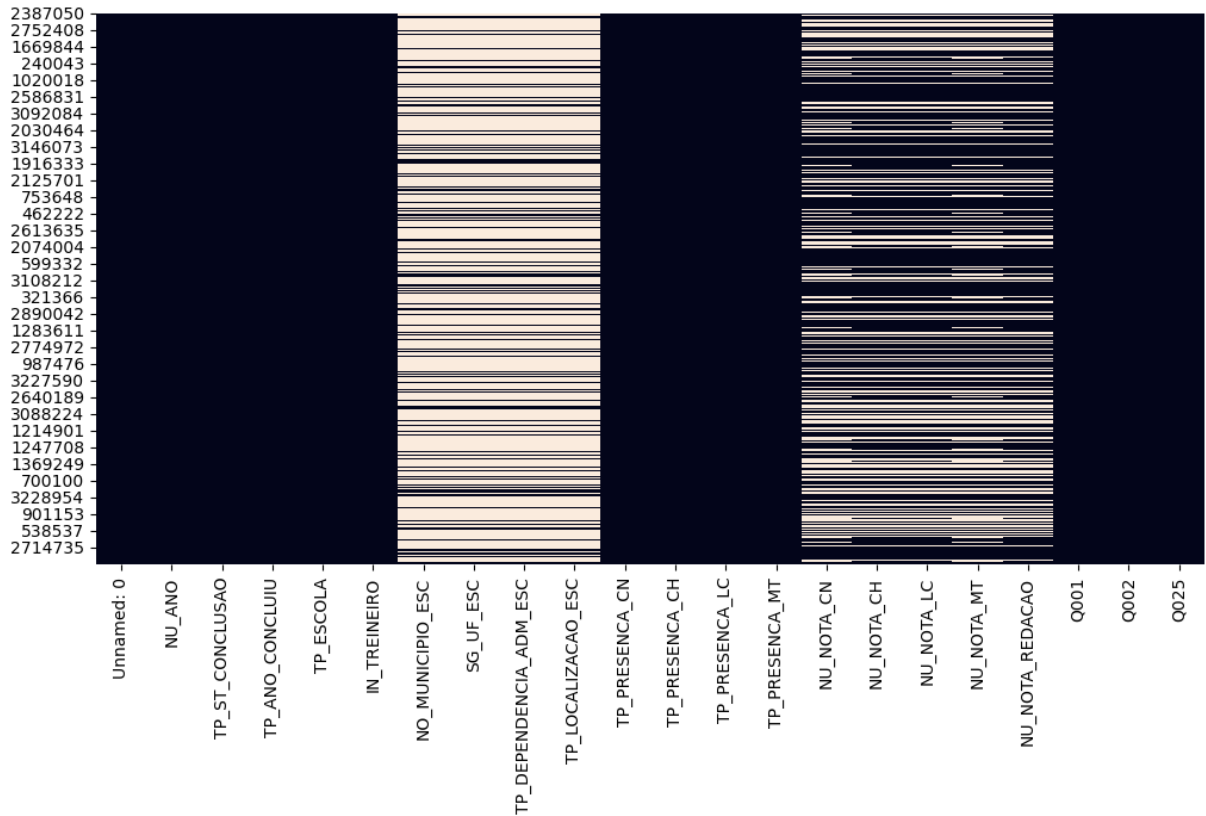
Distribuição de Frequência da Variável TP_PRESENCA_MT - Frequência Relativa:					
Valor	Frequência	Relativa	Média	Desvio Padrão	Coeficiente de Variação
1	0	0.34	1.00	1.00	100.00
0	1	0.66	1.00	1.00	100.00
2	2	0.00	1.00	1.00	100.00

Fonte: Autores

- **Análise de valores faltantes:** Observamos que a grande quantidade de campos nulos nos atributos NO\_MUNICIPIO\_ESC, SG\_UF\_ESC, TP\_DEPENDENCIA\_ADM\_ESC e TP\_LOCALIZACAO\_ESC pode indicar que esses dados não foram coletados ou não estão disponíveis para todas as entradas no conjunto de dados. Isso pode prejudicar a análise posterior dessas variáveis, uma vez

que as informações faltantes podem afetar a precisão dos resultados ou a capacidade de realizar análises mais detalhadas.

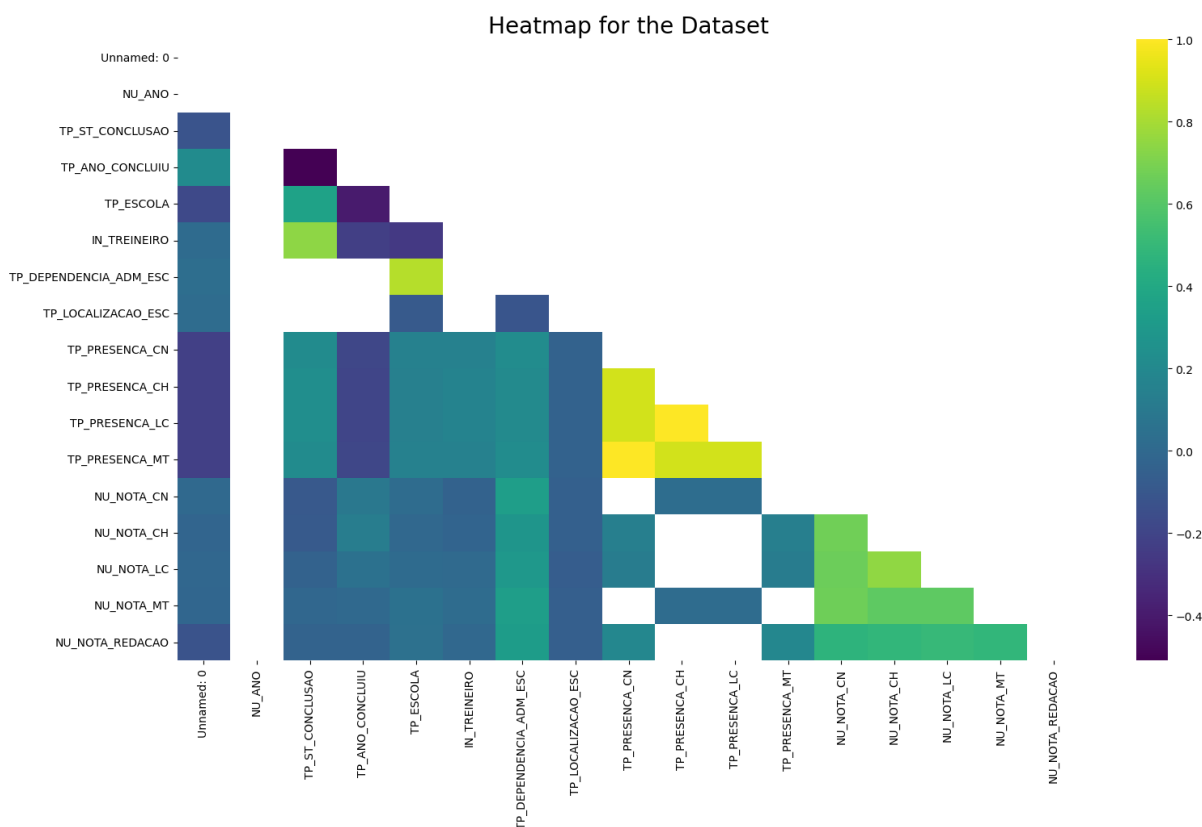
Gráfico 1 - Determinação de valores faltantes (ENEM 2021)



Fonte: Os autores

- **Apresentação de Correlação:** As correlações vistas são entre variáveis que possuem semelhanças em sua criação ou natureza. No caso específico das variáveis TP\_DEPENDENCIA\_ADM\_ESC e TP\_ESCOLA, por exemplo, é possível dizer que elas estão relacionadas à natureza da escola em que o candidato realizou a prova do ENEM.

Gráfico 2 - Heatmap de correlação entre as variáveis (ENEM 2021)



Fonte: Os autores.

#### 4.2.2 Educação Básica

- **Resumo estatístico:** Apresentação de medidas de tendência central e medidas de dispersão

Tabela 7 – Resumo estatístico - Educação Básica

% Nulos	Origem	Variável	count	mean	std	min	25%	50%	75%	max
0.00	ed_basica	Unnamed: 0	224229.00	112114.00	64729.48	0.00	56057.00	112114.00	168171.00	224228.00
0.00	ed_basica	NU_ANO_CENSO	224229.00	2020.00	0.00	2020.00	2020.00	2020.00	2020.00	2020.00
0.00	ed_basica	CO_ENTIDADE	224229.00	30273997.97	9410084.53	11000023.00	23249650.00	31082651.00	35235246.00	53088034.00
0.00	ed_basica	TP_DEPENDENCIA	224229.00	3.07	0.82	1.00	3.00	3.00	3.00	4.00
0.00	ed_basica	TP_CATEGORIA_ESCOLA_PRIVADA	224229.00	0.30	0.82	0.00	0.00	0.00	0.00	4.00
0.00	ed_basica	TP_LOCALIZACAO	224229.00	1.36	0.48	1.00	1.00	1.00	2.00	2.00
0.00	ed_basica	IN_VINCULO_SECRETARIA_EDUCACAO	224229.00	0.62	0.49	0.00	0.00	1.00	1.00	1.00
0.00	ed_basica	IN_AGUA_FILTRADA	224229.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	ed_basica	IN_AGUA_POTAVEL	224229.00	0.76	0.43	0.00	1.00	1.00	1.00	1.00
0.00	ed_basica	IN_ENERGIA_INEXISTENTE	224229.00	0.02	0.13	0.00	0.00	0.00	0.00	1.00
0.00	ed_basica	IN_ESGOTO_INEXISTENTE	224229.00	0.04	0.19	0.00	0.00	0.00	0.00	1.00
0.00	ed_basica	IN_BANHEIRO	224229.00	0.78	0.41	0.00	1.00	1.00	1.00	1.00
0.00	ed_basica	IN_BIBLIOTECA	224229.00	0.29	0.46	0.00	0.00	0.00	1.00	1.00
0.00	ed_basica	IN_BIBLIOTECA_SALA_LEITURA	224229.00	0.42	0.49	0.00	0.00	0.00	1.00	1.00
0.00	ed_basica	IN_COZINHA	224229.00	0.74	0.44	0.00	0.00	1.00	1.00	1.00
0.00	ed_basica	IN_LABORATORIO_CIENCIAS	224229.00	0.10	0.30	0.00	0.00	0.00	0.00	1.00
0.00	ed_basica	IN_LABORATORIO_INFORMATICA	224229.00	0.28	0.45	0.00	0.00	0.00	1.00	1.00
0.00	ed_basica	IN_REFEITORIO	224229.00	0.38	0.49	0.00	0.00	0.00	1.00	1.00
0.00	ed_basica	QT_SALAS_EXISTENTES	224229.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	ed_basica	QT_SALAS_UTILIZADAS	224229.00	6.86	24.59	0.00	2.00	5.00	10.00	9999.00
0.00	ed_basica	IN_DESKTOP_ALUNO	224229.00	0.37	0.48	0.00	0.00	0.00	1.00	1.00
0.00	ed_basica	IN_INTERNET_ALUNOS	224229.00	0.25	0.43	0.00	0.00	0.00	1.00	1.00
0.00	ed_basica	IN_ACESSO_INTERNET_COMPUTADOR	224229.00	0.37	1.22	0.00	0.00	0.00	0.00	9.00
0.00	ed_basica	QT_FUNCIONARIOS	224229.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	ed_basica	IN_EXAME_SELECAO	224229.00	0.09	0.76	0.00	0.00	0.00	0.00	9.00
0.00	ed_basica	QT_MAT_BAS	224229.00	210.92	315.14	0.00	17.00	103.00	285.00	40619.00
0.00	ed_basica	QT_MAT_FUND	224229.00	119.16	203.72	0.00	0.00	14.00	163.00	5339.00
0.00	ed_basica	QT_MAT_MED	224229.00	33.67	130.51	0.00	0.00	0.00	0.00	3080.00
0.00	ed_basica	QT_MAT_PROF	224229.00	8.83	141.20	0.00	0.00	0.00	0.00	40619.00
0.00	ed_basica	QT_MAT_BAS_FEM	224229.00	103.80	159.82	0.00	7.00	49.00	139.00	24127.00
0.00	ed_basica	QT_MAT_BAS_MASC	224229.00	107.13	158.87	0.00	9.00	53.00	146.00	21211.00
0.00	ed_basica	QT_MAT_INF_INT	224229.00	12.00	38.96	0.00	0.00	0.00	0.00	2279.00
0.00	ed_basica	QT_MAT_FUND_INT	224229.00	8.86	47.16	0.00	0.00	0.00	0.00	1351.00
0.00	ed_basica	QT_MAT_MED_INT	224229.00	4.30	38.67	0.00	0.00	0.00	0.00	1828.00
0.00	ed_basica	QT_DOC_BAS	224229.00	12.30	14.81	0.00	1.00	8.00	18.00	332.00
0.00	ed_basica	QT_DOC_FUND	224229.00	7.34	10.86	0.00	0.00	1.00	12.00	228.00
0.00	ed_basica	QT_DOC_MED	224229.00	2.75	8.86	0.00	0.00	0.00	0.00	310.00
0.00	ed_basica	QT_DOC_PROF	224229.00	0.61	5.35	0.00	0.00	0.00	0.00	332.00
0.00	ed_basica	QT_TUR_BAS	224229.00	9.55	10.53	0.00	2.00	7.00	14.00	415.00
0.00	ed_basica	QT_TUR_FUND	224229.00	5.24	7.56	0.00	0.00	1.00	9.00	166.00
0.00	ed_basica	QT_TUR_MED	224229.00	1.11	4.00	0.00	0.00	0.00	0.00	164.00
0.00	ed_basica	QT_TUR_PROF	224229.00	0.31	2.92	0.00	0.00	0.00	0.00	209.00
0.00	ed_basica	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
0.00	ed_basica	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
0.00	ed_basica	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
0.00	ed_basica	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
0.00	ed_basica	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Fonte: Os autores

- **Distribuição de frequências:** Apresentação da frequência absoluta e relativa de cada variável. Por conta do volume de informação, só será apresentado uma amostra, podendo verificar o resultado completo no [link](#)

Tabela 8 – Amostra frequência absoluta - Educação Básica

Distribuição de Frequência da Variável TP\_DEPENDENCIA - Frequência Absoluta:

Valor	Frequência Absoluta	Variância	Desvio Padrão	Amplitude
3	1	714	1.67	1.29
2	2	33658	1.67	1.29
0	3	138033	1.67	1.29
1	4	51824	1.67	1.29

Distribuição de Frequência da Variável TP\_CATEGORIA\_ESCOLA\_PRIVADA - Frequência Absoluta:

Valor	Frequência Absoluta	Variância	Desvio Padrão	Amplitude
0	0	182466	2.50	1.58
1	1	31429	2.50	1.58
3	2	2046	2.50	1.58
4	3	323	2.50	1.58
2	4	7965	2.50	1.58

Distribuição de Frequência da Variável TP\_LOCALIZACAO - Frequência Absoluta:

Valor	Frequência Absoluta	Variância	Desvio Padrão	Amplitude
0	1	142615	0.50	0.71
1	2	81614	0.50	0.71

Distribuição de Frequência da Variável IN\_VINCULO\_SECRETARIA\_EDUCACAO - Frequência Absoluta:

Valor	Frequência Absoluta	Variância	Desvio Padrão	Amplitude
1	0	85070	0.50	0.71
0	1	139159	0.50	0.71

Distribuição de Frequência da Variável IN\_AGUA\_FILTRADA - Frequência Absoluta:

Valor	Frequência Absoluta	Variância	Desvio Padrão	Amplitude
0	0	224229	NaN	NaN

Fonte: Os autores

Tabela 9 – Amostra frequência relativa- Educação Básica

Distribuição de Frequência da Variável NO\_MUNICIPIO - Frequência Relativa:

	Valor	Frequência Absoluta	Frequência Relativa
0	São Paulo	8345	0.04
1	Rio de Janeiro	4235	0.02
2	Belo Horizonte	1944	0.01
3	Recife	1517	0.01
4	Salvador	1477	0.01
...	...	...	...
5292	Floreal	2	0.00
5293	Nova Pádua	2	0.00
5294	Oliveira de Fátima	2	0.00
5295	Queluzito	2	0.00
5296	Grupiara	2	0.00

[5297 rows x 3 columns]

Distribuição de Frequência da Variável NO\_MICRORREGIAO - Frequência Relativa:

	Valor	Frequência Absoluta	Frequência Relativa
0	São Paulo	9996	0.04
1	Rio de Janeiro	7873	0.04
2	Belo Horizonte	3643	0.02
3	Recife	3110	0.01
4	Porto Alegre	2679	0.01
..	...	...	...
549	Bananal	36	0.00
550	Meruoca	36	0.00
551	Alto Paraguai	34	0.00
552	Alto Araguaia	29	0.00
553	Fernando de Noronha	2	0.00

[554 rows x 3 columns]

Distribuição de Frequência da Variável CO\_ENTIDADE - Frequência Relativa:

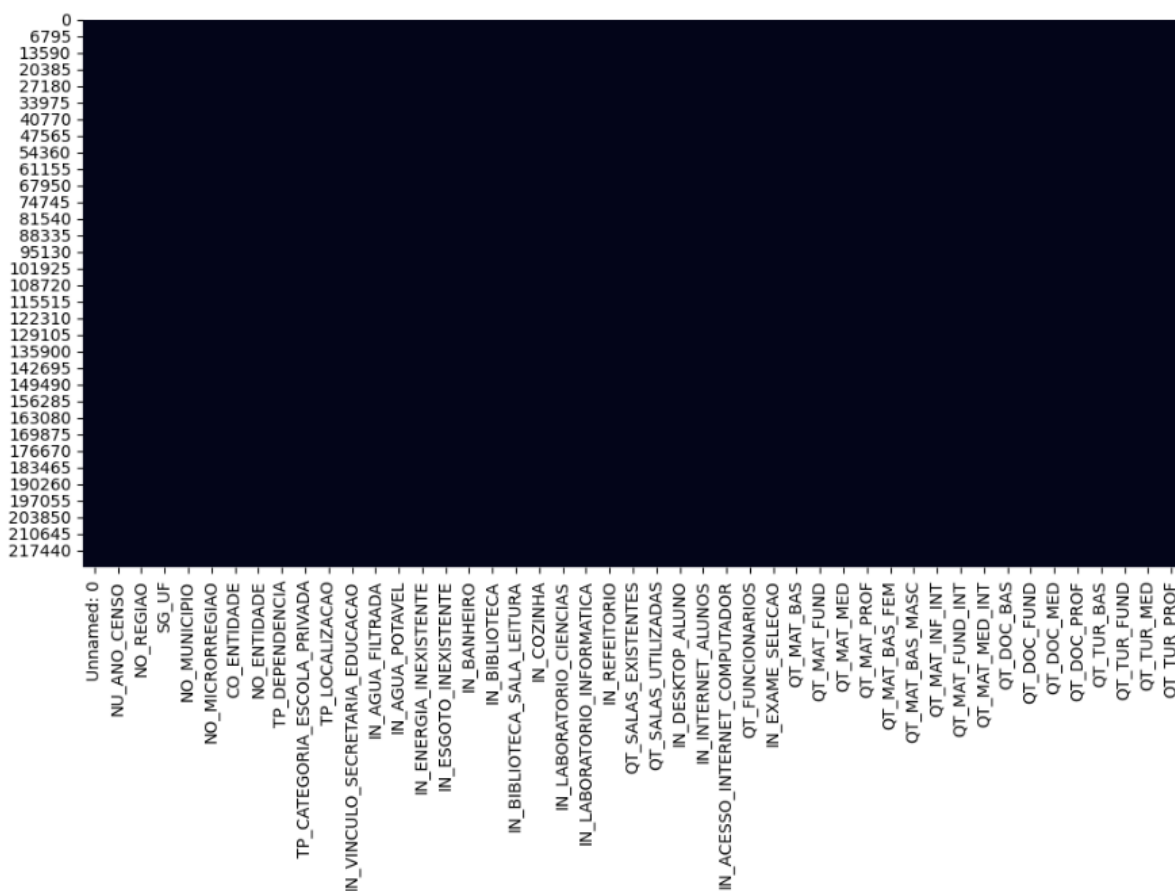
	Valor	Frequência Relativa	Média	Desvio Padrão \
148906	11000023	0.00	30273997.97	9410084.53
148907	11000040	0.00	30273997.97	9410084.53
148908	11000058	0.00	30273997.97	9410084.53
148909	11000082	0.00	30273997.97	9410084.53
148910	11000104	0.00	30273997.97	9410084.53
...	...	...	...	...
74749	53084071	0.00	30273997.97	9410084.53
74750	53085000	0.00	30273997.97	9410084.53
74751	53086007	0.00	30273997.97	9410084.53
74752	53087003	0.00	30273997.97	9410084.53
224228	53088034	0.00	30273997.97	9410084.53

Fonte: Os autores

- **Análise de valores faltantes:** Conforme gráfico a seguir, foi possível confirmar que não existem dados faltantes.



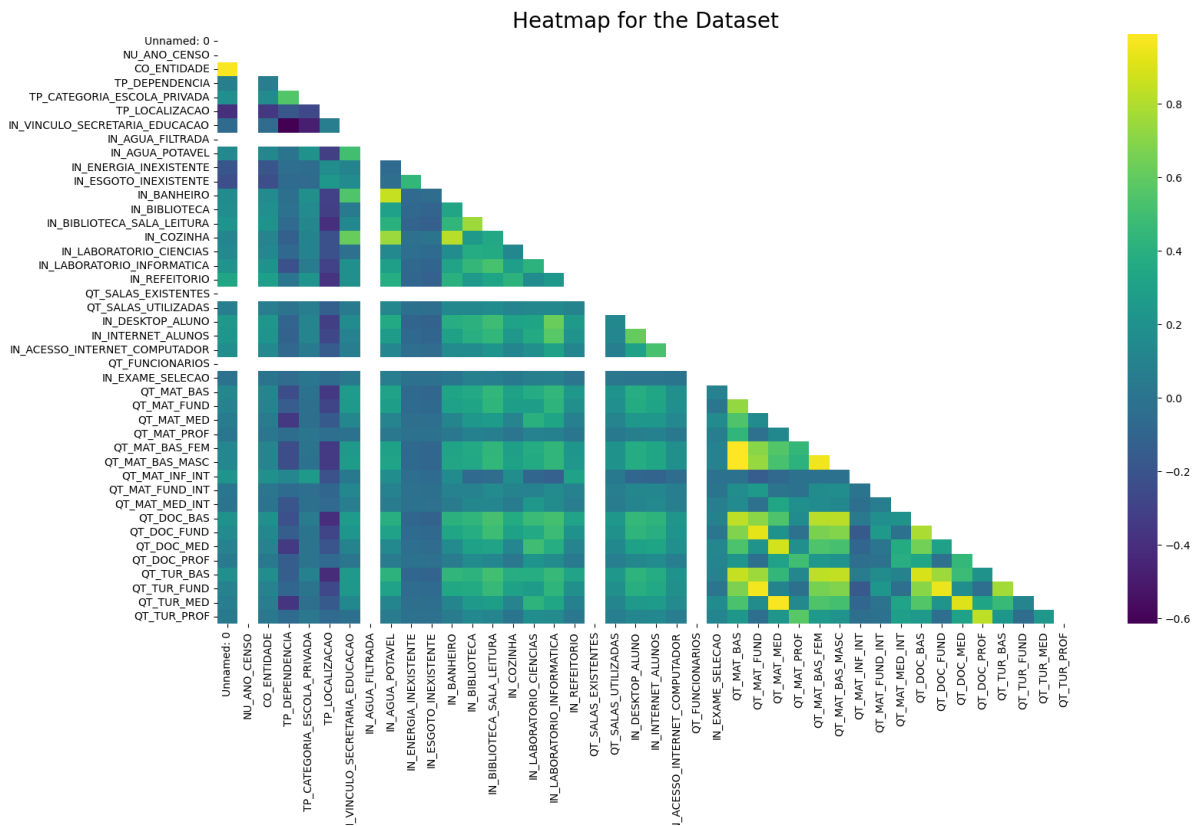
Gráfico 3 - Determinação valores faltantes (Educação Básica)



Fonte: Os autores.

- **Apresentação de correlação entre as variáveis:** As correlações vistas são entre variáveis que possuem semelhanças em sua criação ou natureza. Por exemplo, a variável QT\_MAT\_MED e QT\_TUR\_MED, indicam a relação entre número de matrículas no ensino médio e a quantidade de turmas no ensino médio.

Gráfico 4 - Heatmap de correlação entre as variáveis (Educação Básica)



Fonte: Os autores.

## **5        DESENVOLVIMENTO**

### **5.1 Exemplos de ilustrações**

Imagem 1 –

Fonte:

## **6      METODOLOGIA**

## **7      RESULTADOS**

## **8 CONCLUSÃO**

## 9 REFERÊNCIAS

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Microdados do Censo Escolar da Educação Básica 2020. Disponível em: <[https://download.inep.gov.br/dados\\_abertos/microdados\\_censo\\_escolar\\_2020.zip](https://download.inep.gov.br/dados_abertos/microdados_censo_escolar_2020.zip)>. Acesso em: 25 ago. 2023.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Microdados do Enem 2021. Disponível em: <[https://download.inep.gov.br/microdados/microdados\\_enem\\_2021.zip](https://download.inep.gov.br/microdados/microdados_enem_2021.zip)>. Acesso em: 25 ago. 2023.