UNIVERSIDADE FEDERAL DE MINAS GERAIS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO
DCC011: INTRODUÇÃO A BANCO DE DADOS

GABRIEL VIEIRA
GUSTAVO MATTOS LOPES
HEITOR GONÇALVES LEITE
LUISA LOPES CARVALHAES
MATHEUS TORRES PRATES

TRABALHO PRÁTICO II

"ACESSO, COLETA, GESTÃO E ANÁLISE DE DADOS PÚBLICOS"

BELO HORIZONTE
1º SEMESTRE DE 2024

Sumário

1 Introdução

A Lei de Acesso à Informação (Lei nº 12.527/2011) determina a transparência de toda informação produzida, coletada, organizada, armazenada, disponibilizada ou gerenciada por órgãos e entidades públicas. Em virtude dessa diretriz, quaisquer dados referentes às ações do poder público devem ser amplamente acessíveis, propiciando o alcance universal perante os cidadãos.

A disponibilidade da informação pode ocorrer mediante *transparência ativa*, em que dados de interesse do público em geral (documentos, relatórios, estudos, contratos, etc) são compilados e periodicamente publicados. Através de tal forma de divulgação, fontes de dados públicos se encontram à disposição do corpo social como um importante mecanismo de defesa da democracia, pois permite a participação cívica e o controle da atuação do governo.

O presente trabalho tem como objetivo conduzir um estudo baseado em bancos de dados públicos, disponíveis conforme a Lei de Acesso à Informação

(LAI). Nesse sentido, apresentaremos adiante o processo de acesso, coleta, gerenciamento, integração e análise de conjuntos de dados públicos, bem como uma revisão crítica das fontes obtidas.

2 Coleta dos Dados e Análise das Fontes

A etapa de coleta dos dados ilustrou diversos problemas com relação às fontes públicas: durante a investigação entre as opções disponíveis, encontramos bases de dados incompletas (traziam informações limitadas a um conjunto restrito de municípios), em formato inadequado para processamento por máquina (planilhas formatadas, principalmente), ausência de dicionário a respeito dos dados (explanação sobre o conteúdo do arquivo e significado dos valores nas colunas) e dados desatualizados e/ou inexistentes.

Os bancos de dados finalmente adotados contêm informações acerca do nível socioeconômico das escolas brasileiras, que serão correlacionadas aos dados do Exame Nacional do Ensino Médio (Enem) no mesmo período.

Os dados do Indíce de Nível Socioeconômico (Inse) classificam as escolas públicas do país em 8 níveis, conforme a distribuição dos estudantes por nível socioeconômico (explícita nos dados) e critérios relativos à instituição. No banco de dados, cada instância na tabela representa um colégio - identificado de forma única pelo seu código no CENSO Escolar - e inclui informações sobre o município ao qual pertence e sua classificação de acordo com o Inse.

A base de dados do Enem fornece as notas e o questionário socioeconômico respondido pelos inscritos no exame. Os nomes dos candidatos não são divulgados, uma vez que a Lei de Acesso à Informação protege a privacidade dos indivíduos quanto à exposição de informações pessoais. Entretanto, para cada número de inscrição, existe informações disponíveis sobre gênero, raça, nacionalidade e escolaridade.

Ambos os registros estão disponíveis no portal do Governo Federal e possuem informações (dicionário) sobre os dados contidos neles. As versões selecionadas referem-se às últimas edições do Inse - 2019 e 2021.

Não obstante, na tentativa de integração dos dados do Enem frente aos dados das escolas brasileiras, nota-se a ausência do identificador único da instituição acadêmica à qual o candidato está vinculado: esse atributo foi omitido nos dados mais recentes do Enem. Além disso, há um grande número de dados nulos em relação à instituição de conclusão do Ensino Médio (tipo, dependência administrativa, nome, localização e situação de funcionamento).

3 Esquema, Dicionário e Metadados

Os materiais produzidos em paralelo a este relatório podem ser encontrados no repositório do GitHub. [1].

4 Análise Exploratória

4.1 Preparação dos Dados

4.1.1 Importação

Inicialmente foram importados os dados do Enem dos anos de 2019 e 2021. Ambos estão disponíveis no formato CSV, e puderam ser importados utilizando as ferramentas nativas do SBGD SQLite, através da interface SQLite Browser, como ilustrado na Figura 1.

Nessa etapa, o principal desafio foi encontrar a codificação correta do arquivo, para que os acentos e caracteres especiais fossem importados corretamente. Devido a dimensão do arquivo, escolher uma nova codificação era um processo lento e escolher a codificação poderia fazer o SGBD *crashar*.

Foram então importados os dados do INSE, que estavam no formato proprietário de planilha XLSX. Para tal, os dados foram primeiramente convertidos para o formato CSV e então importados no SGBD.

4.1.2 Tratamento

Os dados importados continham muitas informações desnecessárias para nossas análises, e estes foram removidos a fim de diminuir o tamanho do banco e facilitar o compartilhamento e análise.

O banco de dados do Enem por exemplo, incluia para cada aluno o gabarito marcado em cada uma das áreas do exame (Ciências Humanas, Linguagens, Ciências da Natureza e Matemática), bem como o gabarito esperado para cada aluno (que poderia ser diferente, já que as questões são permutadas entre as provas), como visto na Figura 2. Esses dados são armazenados como strings e contribuiam em grande parte para o tamanho do BD. Além disso, os dados do Enem incluiam o ano do exame de maneira reduntante em todas as linhas do banco de dados.

[1] `https://github.com/lumiis2/TP2-IBD/tree/main`

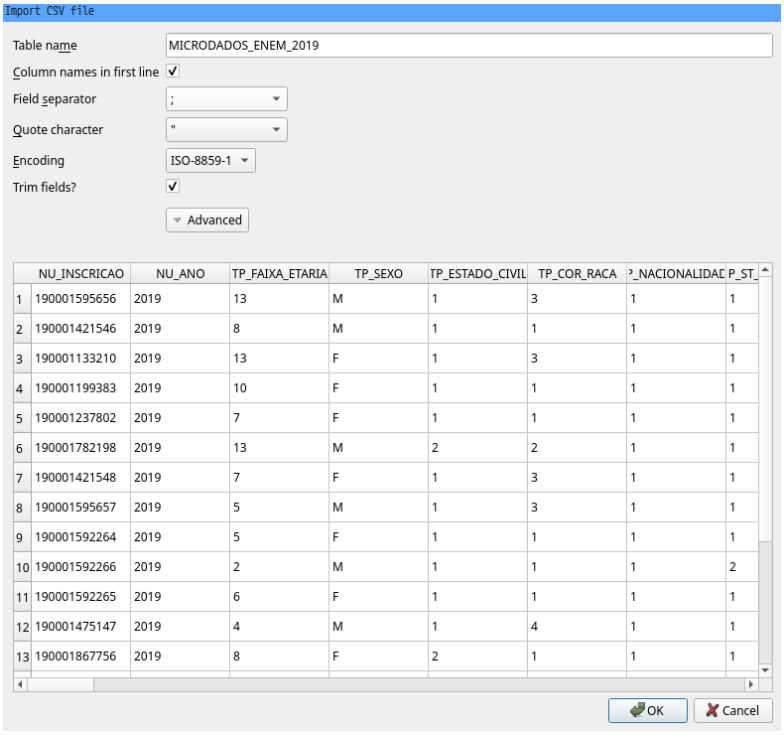

Figura 1: Importação de dados CSV no SQLite Browser.

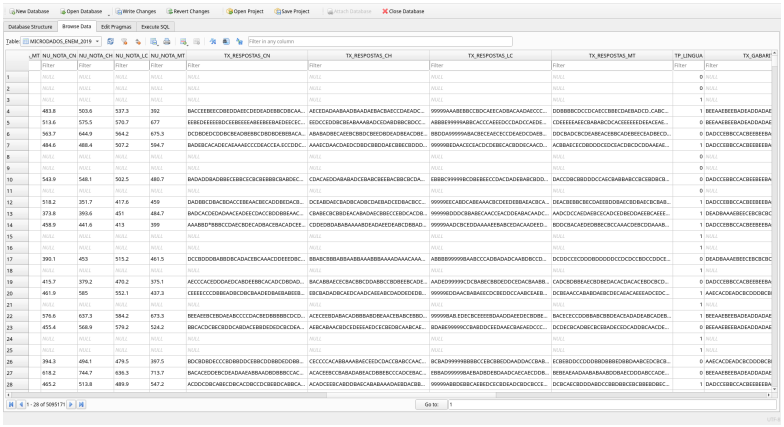

Figura 2: Dados desnecessários na base dados do Enem.

4.1.3 Normalização

Ambas as bases de dados utilizados contém dados sobre municípios e unidades da federação, inicialmente de maneira não normalizada, com alta redundância, o que aumentava desnecessariamente o tamanho do banco. Para resolver esse problema, foram criadas duas tabelas adicionais: MUNICIPIO

e UF, cujas tabelas existentes foram modificadas para referencia-las.

4.2 Definição dos Objetivos

Os objetivos da análise exploratória dirigida a seguir estão centrados na relação entre as condições socioeconômicas e o desempenho dos candidatos no Exame Nacional do Ensino Médio. Busca-se compreender a influência do status socioeconômico dos cidadãos no que tange à qualidade educacional, mensurada pela média da nota do Enem, refletindo assim que a desigualdade social reverbera inclusive na educação.

Nesse sentido, por meio dos dados obtidos, almejamos responder as seguintes perguntas:

- Como as condições socioeconômicas relacionadas à educação (dados INSE) a nível municipal e estadual se correlacionam com a média das notas do ENEM nesses locais?
- Os níveis de condições socioeconômicas relacionadas à educação (dados INSE) refletem os níveis socioeconômicos do questionário do ENEM?
- Como o período da pandemia pode ter influenciado nos dados de 2019 para 2021?

4.3 Análise Descritiva

4.3.1 Dados do INSE

Em primeiro lugar, cabe analisar a distribuição das escolas brasileiras entre os níveis estabelecidos pelo INSE (Figura 3). Percebe-se, de imediato, que a quantidade de escolas abordadas pelos dados cresceu de 2019 para 2021 (68868 para 69820). O nível geral das escolas também aumentou: em 2019, o INSE médio era de 4.70, enquanto, em 2021, havia crescido para 4.76.

Nesse contexto, também é importante avaliar a quantidade de alunos por escola. Esses dados estão sumarizados na Figura 4. É possível notar uma mudança relevante: apesar de o número de escolas abordadas pela avaliação do INSE ter aumentado, a quantidade de alunos por escola diminuiu substancialmente. Em 2019, a média era de 77.86 alunos por escola. Em 2021, esse número caiu para 70.82. Essa queda pode ser explicada pela evasão escolar causada pela pandemia.

Uma vez que o Brasil é um país de magnitude continental, as disparidades entre as distintas regiões geográficas são bastante salientes. As figuras 5 e 6 apresentam o INSE médio por estado nos dois anos analisados. Nelas, é

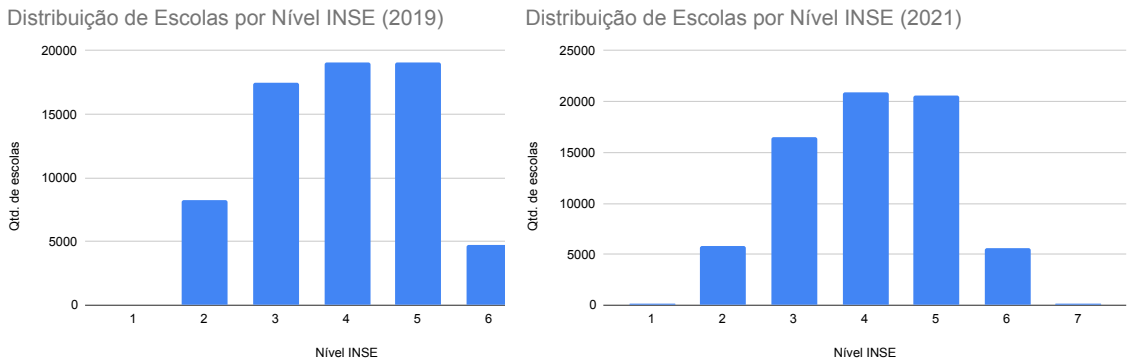

Figura 3: Distribuição de escolas entre os níveis INSE em 2019 e em 2021.

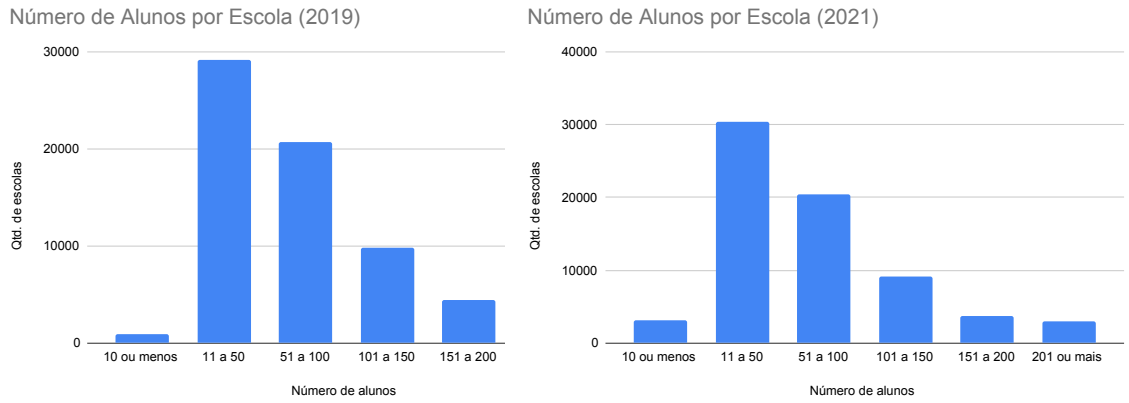

Figura 4: Distribuição de alunos entre escolas em 2019 e em 2021.

possível notar que os estados das regiões Sul, Sudeste e Centro-Oeste têm índices mais elevados que aqueles das regiões Norte e Nordeste.

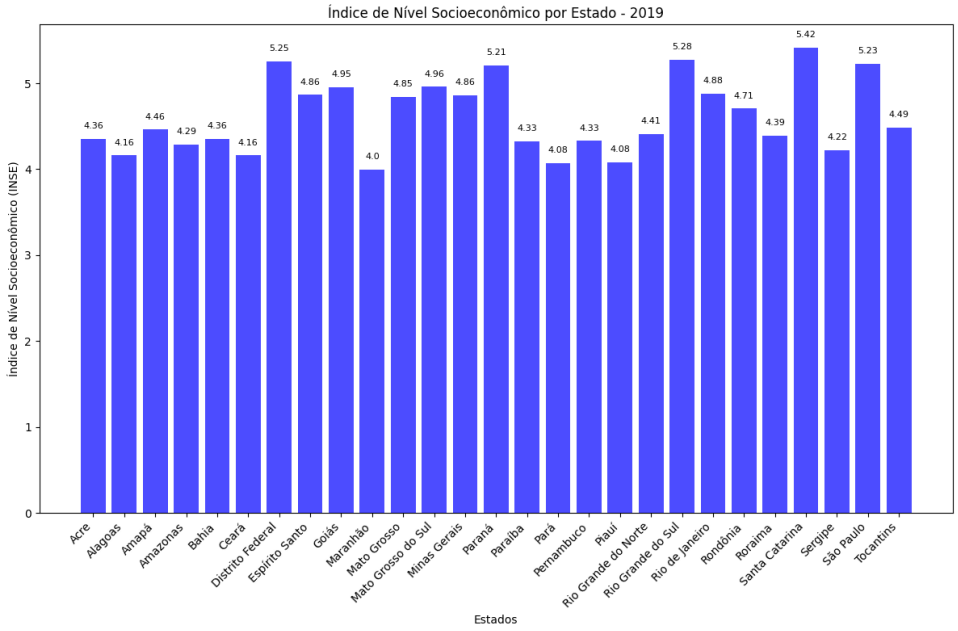

Figura 5: INSE médio por estado em 2019.

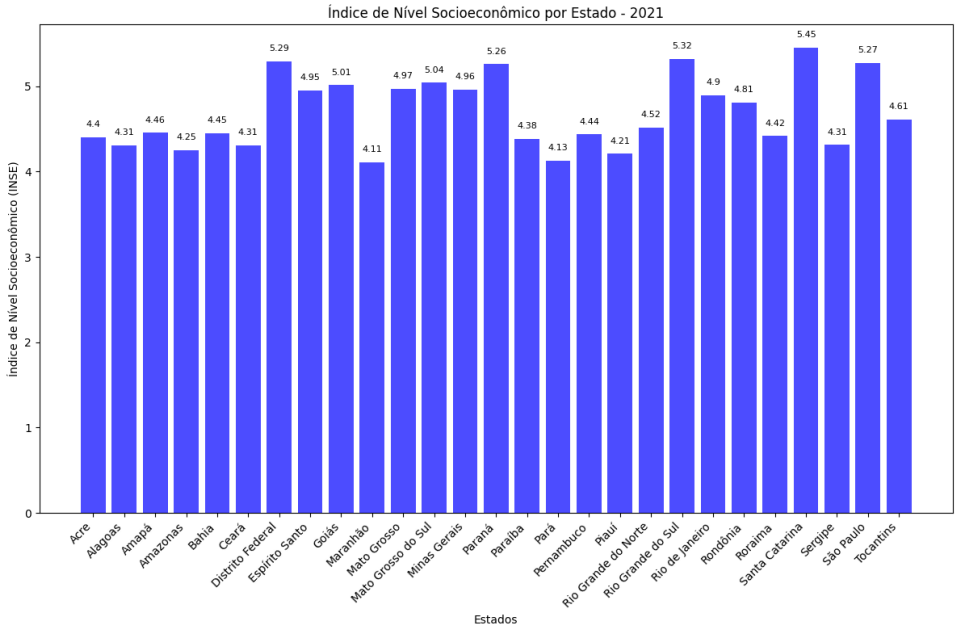

Figura 6: INSE médio por estado em 2021.

4.3.2 Exame do ENEM

Nesta seção, serão analisadas as distribuições de notas dos candidatos do ENEM.

	2019	2021
Linguagens	520.21	502.56
Humanas	507.25	519.94
Natureza	477.82	491.79
Matemática	523.12	535.08

Figura 7: Notas médias do ENEM em 2019 e em 2021.

A nota média de cada prova do ENEM nos dois anos em análise está representada na Figura 7. Não é ideal compararar diretamente as médias entre edições diferentes do exame, visto que o modelo de TRI usado para calculá-las não foi o mesmo. Vale destacar que as médias obtidas nesta análise exploratória são ligeiramente diferentes daquelas reportadas por veículos de informação [2]. Essa diferença pode ser explicada por correções feitas nos dados do INEP após sua publicação inicial.

	M	F
Linguagens	521.63	519.76
Humanas	515.13	501.89
Natureza	489.32	469.98
Matemática	547.97	506.17

	M	F
Linguagens	507.52	499.49
Humanas	530.72	513.28
Natureza	505.76	483.12
Matemática	559.70	519.81

Figura 8: Notas médias do ENEM por gênero em 2019 (acima) e em 2021 (abaixo).

[2] `https://g1.globo.com/educacao/noticia/2020/01/17/notas-medias-do-enem-2019-caem-em-todas-as-provas-objetivas.ghtml`

A figura 8 apresenta as notas médias por gênero. As pontuações maiores dos candidatos masculinos indicam um grave problema na sociedade brasileira. Dado o cenário machista que ainda prevalece no Brasil no século XXI, a falta de incentivo à educação das mulheres está refletida nos dados do exame.

	Linguagens	Humanas	Natureza	Matemática
Não-declarado	525.04	515.25	486.55	532.98
Branca	538.37	529.26	499.30	555.98
Preta	509.12	492.75	461.05	493.80
Parda	509.54	493.77	464.96	504.37
Amarela	518.43	502.41	477.35	525.55
Indígena	485.68	468.77	444.13	475.67

	Linguagens	Humanas	Natureza	Matemática
Não-declarado	502.51	522.38	493.83	532.78
Branca	523.23	541.65	511.66	565.46
Preta	486.07	501.73	471.42	501.14
Parda	485.94	502.61	476.25	512.25
Amarela	497.76	514.52	493.89	539.90
Indígena	452.51	470.17	450.36	474.23

Figura 9: Notas médias do ENEM por cor/raça autodeclarada em 2019 (acima) e em 2021 (abaixo).

Por sua vez, a figura 9 mostra as distribuições de nota por cor/raça autodeclarada. Tal como na análise por gênero, é perceptível a discrepância entre as notas, o que reflete o racismo estrutural ainda presente na sociedade brasileira.

Assim como feito com os dados do INSE, foram calculadas as médias das pontuações do ENEM por estado (Figuras 10 e 11). Novamente, são perceptíveis as diferenças entre o norte e o sul do país, evidenciando a desigualdade regional.

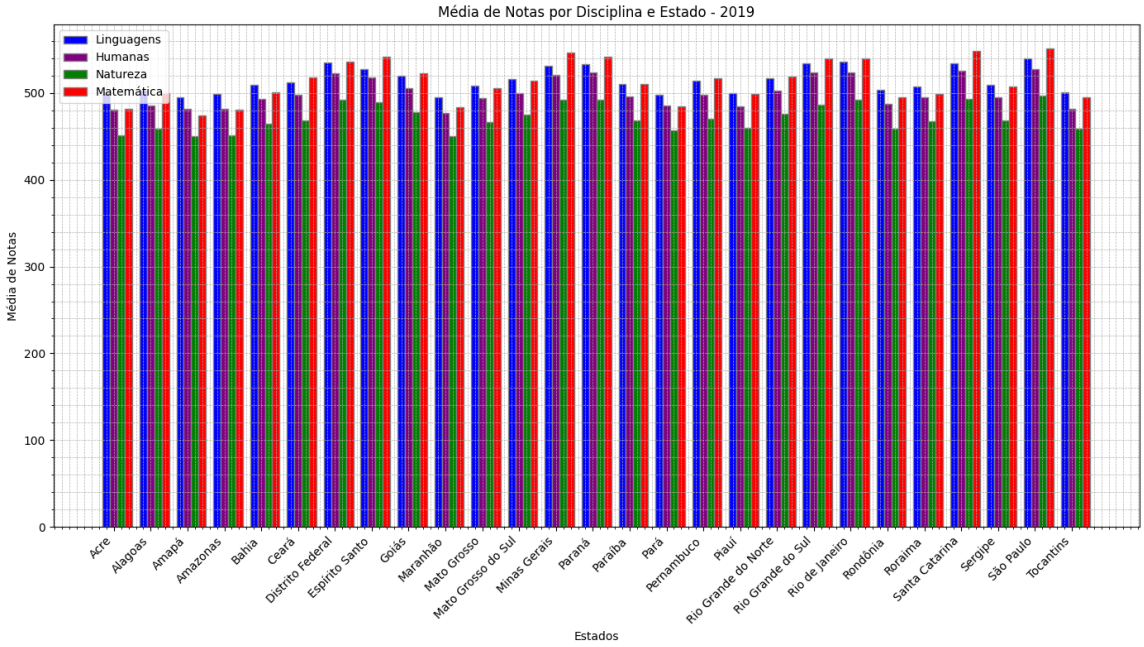

Figura 10: Notas médias por estado em 2019.

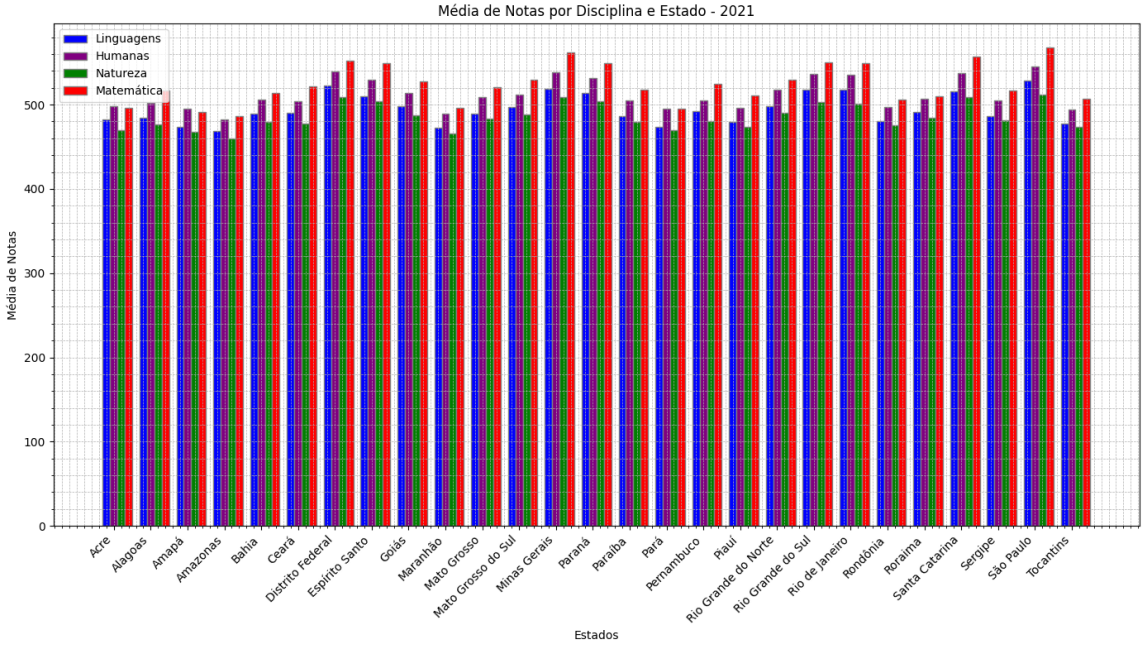

Figura 11: Notas médias por estado em 2021.

4.3.3 Respostas do Questionário Socioeconômico

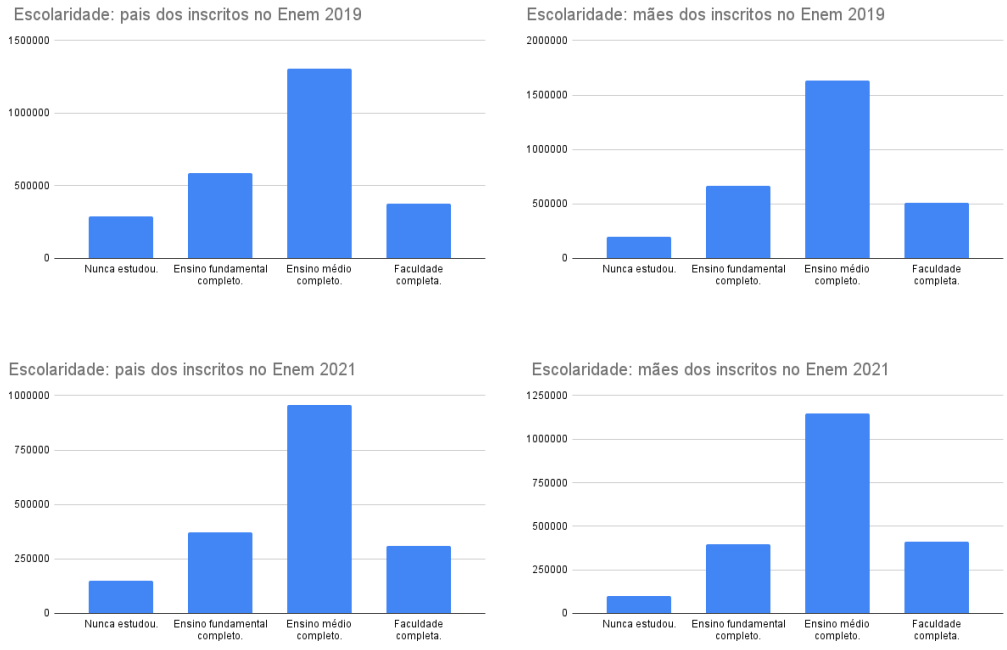

Figura 12: "Até que série seu pai/mãe estudou?"

Em uma visão geral, a partir das respostas fornecidas pelos candidatos ao Exame Nacional do Ensino Médio, é evidente a deficiência do sistema educacional brasileiro no que concerne ao grau de formação acadêmica da população.

A contestação acerca do grau de escolaridade dos pais dos inscritos no Enem produz uma boa amostragem do nível médio de educação da sociedade brasileira: percebe-se, mediante análise dos gráficos exibidos na figura 12, que a grande maioria dos cidadãos não possui o ensino superior profissionalizante completo. Esse fato traz à tona várias consequências ante o âmbito social, essencialmente relacionadas à baixa qualificação da força de trabalho e subsequente desigualdade econômica.

Os níveis insatisfatórios de educação no Brasil tendem a se perpetuar, uma vez que, relacionando o baixo grau de instrução dos indivíduos a condições socioeconômicas desfavoráveis, essas impactam o acesso a oportunidades de ascensão e a qualidade de ensino ofertada à população infanto-juvenil. Como vemos no gráfico da figura 13, a nota média dos candidatos do Enem diminui à medida que o grau de educação dos pais é menos elevado (nesse caso,

consideramos o desempenho apenas dos inscritos que possuem pai e mãe com o mesmo nível de formação).

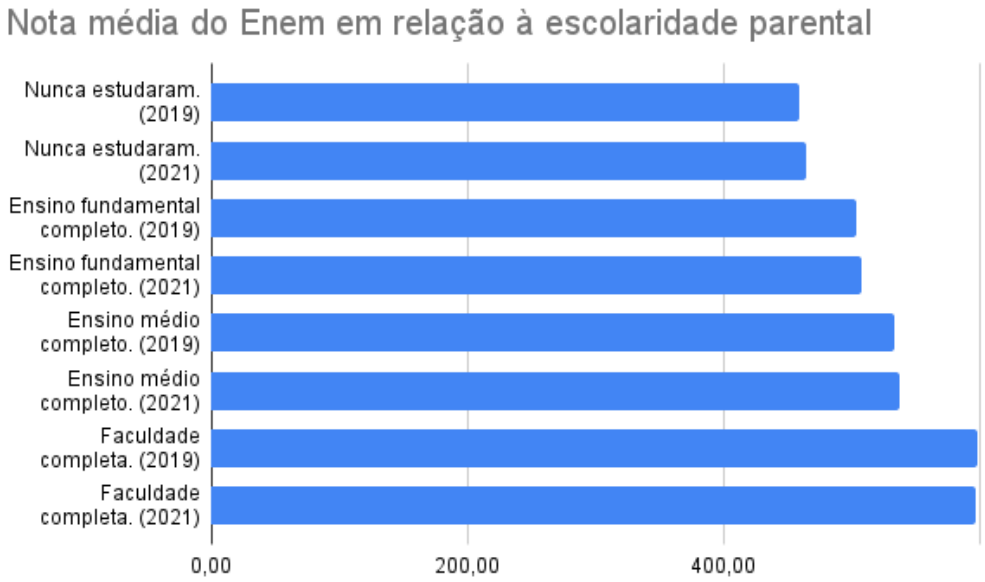

Figura 13: Correlação entre a educação dos pais e a nota do Enem

Agora, vamos analisar outro aspecto que potencialmente impacta a qualidade da educação do jovem brasileiro: o acesso à internet.

No panorama hodierno, em que diversos conteúdos e recursos educativos estão disponíveis online, a disposição de um computador e de internet podem desempenhar um papel importante no propósito acadêmico. A utilização de ferramentas digitais amplia o acesso ao conhecimento, a colaboração virtual e a facilidade de pesquisa. Nessa perspectiva, além de denotar a desigualdade socioeconômica, que, como vimos, está relacionada por si só a discrepantes condições de educação, a disparidade tecnológica pode acentuar o abismo entre os níveis educacionais do país.

Através das figuras 14 e 15, percebe-se que uma parcela significativa dos estudantes que prestaram o Enem não possuíam sequer um computador em casa, conforme indicado pelas respostas do questionário socioeconômico. Ademais, nota-se, ao contrário da detenção do equipamento físico, um aumento expressivo no acesso à internet entre os anos de 2019 e 2021. Existem duas hipóteses relacionadas a essa evolução, vinculadas ao período da pandemia: (1) houve um incremento real no acesso à internet, devido a iniciativas para expandir o suporte ao ensino remoto durante a pandemia; (2) as inscrições do Enem 2021 ocorreram majoritariamente por aqueles que conse-

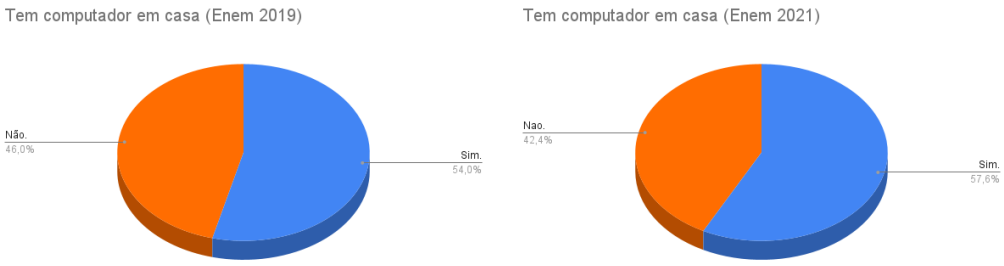

Figura 14: "Na sua residência tem computador?"

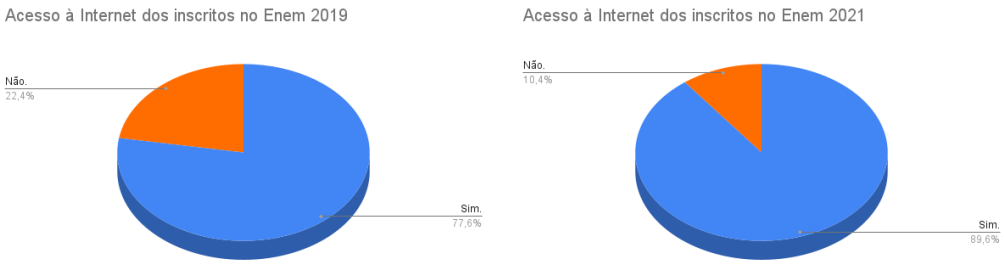

Figura 15: "Na sua residência tem acesso à Internet?"

guiram continuar seus estudos durante a pandemia, ou seja, tinham acesso à internet entre 2020-2021.

Na tabela 1, observa-se que, de fato, a deficiência no acesso à tecnologia repercute no desempenho do aluno na prova do Enem, tanto em decorrência da condição socioeconômica associada à carência de internet/computador, quanto à dificuldade em acessar materiais e recursos educacionais online visando estudos mais eficazes e preparatórios.

Questionário Socioeconômico	Resposta	Nota média no Enem
Na sua residência tem computador?	Sim. (2019)	588,98
	Não. (2019)	488,65
	Sim. (2021)	591,15
	Não. (2021)	492,29
Na sua residência tem acesso à Internet?	Sim. (2019)	533,91
	Não. (2019)	480,50
	Sim. (2021)	540,85
	Não. (2021)	478,13

Tabela 1: Correlação entre o acesso digital e a nota do Enem

Por fim, resta analisar um aspecto crítico para o cenário atual da educação brasileira: a renda *per capita*. A condição socioeconômica de uma família é uma característica determinante para o nível de acesso a informação e conteúdo de seus membros. Uma baixa renda *per capita* limita os potenciais meios de aprendizagem, uma vez que impossibilita o contato com tecnologias e cursos preparatórios. Esse recursos, com destaque para aparelhos eletrônicos, podem ser decisivos para a aprovação do candidato. A figura 16 ilustra essa correlação.

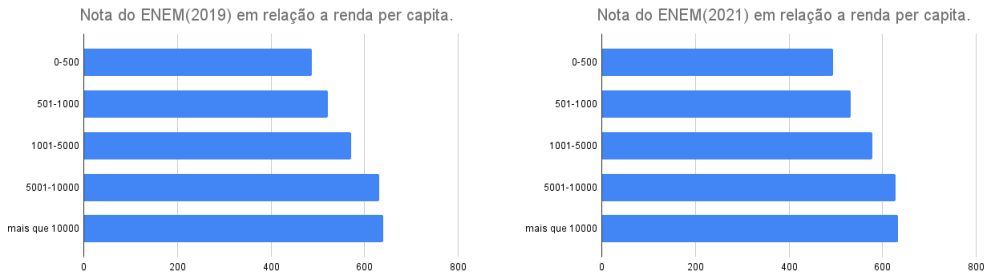

Figura 16: Distribuição de notas de acordo com a renda *per capita* em 2019 e em 2021.

Como previsto, a renda influencia diretamente a nota média dos participantes. Esse impacto é um reflexo, também, da defasagem do ensino público brasileiro, dado que, quanto menor a renda, maior a probabilidade desse estudante cursar seu ensino médio em um instituto público de ensino. Essa discrepância é extremamente substâncial, uma vez que o aluno com maior vantagem econômica tem uma nota média 20% maior, quando comparado aos estudantes com condições econômicas precárias.

Os dados fornecidos pela imagem 16 se tornam ainda mais alarmantes quando levamos em consideração a densidade de estudantes em cada classe.

A figura 17 trás à tona a distribuição de renda discrepante dos participantes do ENEM.

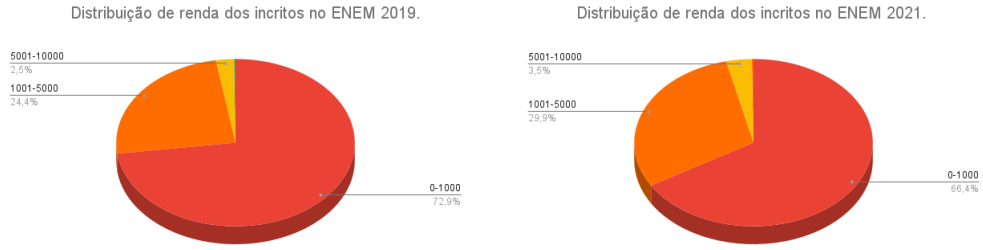

Figura 17: Distribuição de renda, em reais, dos candidatos do ENEM em 2019 e em 2021.

Uma análise desatenta desses dados poderia concluir que a desigualdade se ateunou em 2021. Entretanto, a realidade é que em 2021 houve uma redução de 30% de inscritos em relação a 2019. Essa reduçãõ pode ser vista como uma consequência direta da pandemia de COVID-19, uma vez que os cidadãos economicamente vulneráveis foram expressivamente mais afetados. Isso é condizente com a figura 17, pois a redução de alunos de baixa renda diminuiu substâncialmente, indicando a elevada evasão dessa classe.

4.4 Identificação de Valores Discrepantes

Durante a análise exploratória, encontramos uma instância interessante de dados discrepantes. Entre as opções no momento em que o candidato deve declarar cor/raça, há uma nova categoria adicionada em 2021. Em 2019, havia somente cinco opções: não-declarado, branca, preta, parda, amarela e indígena. Em 2021, foi adicionada uma sexta opção: sem informação. Somente 5 entre os 3389832 candidatos estavam associados a essa categoria no banco de dados, uma quantidade absurdamente pequena. A segunda menor categoria, indígena, tem 19175 estudantes vinculados. Portanto, decidimos ignorar as cinco linhas "sem informação"nas análises que dependiam desse atributo.

4.5 Análise de Correlação

A correlação entre os atributos das tabelas é bastante evidente no que tange ao questionário socioeconômico respondido pelos candidatos do Enem no momento da inscrição. Percebemos que, para cada instância, é possível inferir a situação socioeconômica de forma bastante precisa a partir de algumas

respostas específicas, visto que as respostas tendem a se assemelhar em decorrência da condição financeira do candidato e subsequente disponibilidade de recursos e oportunidades. Utilizamos as respostas relacionadas à renda familiar do candidato, à educação parental e ao acesso digital como parâmetros principais para determinar o nível socioeconômico do candidato e, assim, comparar com a média da nota do Enem.

Existiam algumas hipóteses que conseguimos verificar através desses dados: o acesso à internet está atrelado à posse de dispositivos móveis (98% dos candidatos com acesso à internet, possuem telefone celular) - essa correlação expressiva não se observa quanto à aquisição de computadores, como vimos nos gráficos da figura 14 e 15. Acerca da correlação referente à educação do pai e da mãe, concluímos que, majoritariamente, o nível educacional dos pais não é homogêneo, ou seja, um dos responsáveis tem um grau de formação mais elevado.

Outra observação se refere à presença do candidato nos dias do exame: a presença na prova de Linguagens e de Ciências Humanas é equivalente, bem como na prova de Ciências da Natureza e Matemática.

4.6 Conclusões

Com os resultados descritos acima em mente, podemos fazer algumas conclusões pertinentes acerca dos dois dados estudados.

O banco de dados do ENEM possui um grande volume de valores nulos em certos atributos importantes, como município de origem. Todavia, as colunas com notas, respostas do questionário e dados referentes à realização da prova estão completas, o que permite a realização das análises com segurança. Em geral, esse banco é um reflexo nítido das desigualdades sociais presentes no Brasil, sejam elas de gênero, de raça ou de renda.

Em contrapartida, o banco do INSE é mais sucinto, e não possui tantos valores nulos. Enquanto os dados do ENEM indicam problemas observados no desempenho dos estudantes, o INSE apresenta uma possível causa para eles. Na seção a seguir, discorremos sobre a relação entre a qualidade das escolas de cada município, medida pelo INSE, e a pontuação obtida pelos estudantes no ENEM.

5 Integração dos Dados

A integração realizada entre fontes diferentes visa compreender as interseções entre os dados coletados do ENEM e do Índice de Nível Socioeconômico (INSE). Ambas as tabelas oferecem diferentes perspectivas de um aspecto

crucial sobre a educação e a situação socioeconômica no Brasil. O INSE se concentra em avaliar as condições das escolas e de seus alunos em todo o país, abordando questões relacionadas ao acesso a educação e à qualidade do ensino. Por outro lado, o ENEM reflete diretamente a realidade educacional dos estudantes brasileiros por meio de suas notas.

O cruzamento de dados entre essas tabelas revelou-se extremamente produtivo para nossa análise.

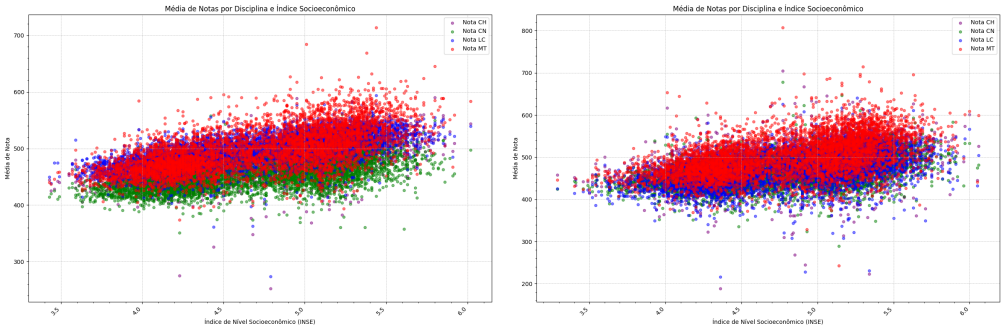

Figura 18: "ENEM X INSE - 2019 e 2021"

Por meio do cruzamento das tabelas do INSE e do ENEM, foi produzida essa relação entre desempenho no ENEM e Nível INSE, agrupada por município. Nessa análise, foi considerado o município onde cada candidato realizou a prova, pois há um menor número de valores nulos. Para ter um primeiro panorama dos dados, o gráfico de dispersão acima foi feito. É possível analisar que a distribuição de notas ocorre de forma heterogênea, porém condensada, e segue um padrão de dispersão em que podemos capturar uma certa tendência. Em geral, quanto maior o INSE, maior a média das notas por município. Além disso, desses gráficos, é possível notar uma certa discrepância entre as áreas de conhecimento. No ano de 2019, a área de Ciências da Natureza (pontos verdes) apresenta notas visivelmente mais baixas que as outras áreas de conhecimento. Além disso, ao comparar os anos de 2021 e 2019, em 2021 vemos uma queda geral das notas de Linguagens.

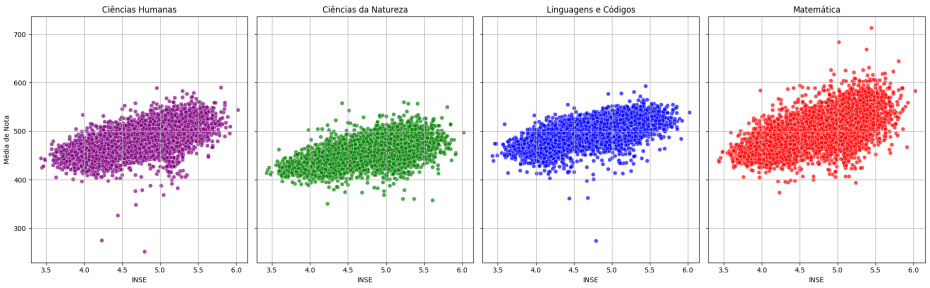

Figura 19: Distribuição de Notas por Disciplina em relação ao INSE - 2019

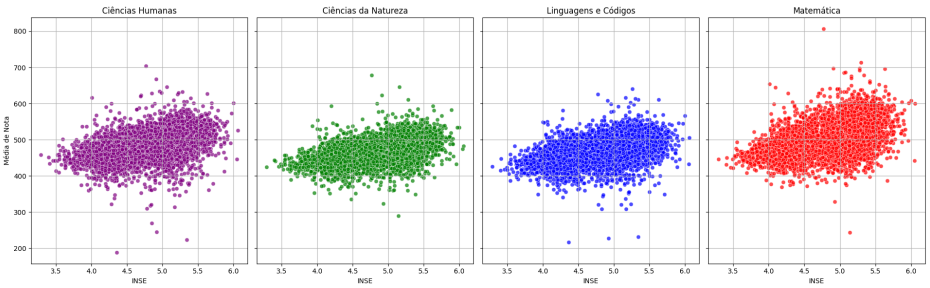

Figura 20: Distribuição de Notas por Disciplina em relação ao INSE - 2019

Os gráficos apresentados aqui demonstram de forma mais clara as tendências observadas anteriormente. Observamos que a dispersão das notas de Ciências da Natureza está mais baixa, enquanto Matemática e Português são mais consistentes, com uma média mais elevada. Quanto à relação entre os anos de 2021 e 2019, podemos ter um panorama mais claro no gráfico abaixo. Verificamos que, de maneira geral, o desempenho em 2019 foi ligeiramente superior ao desempenho em 2021 em termos de notas. Atribuímos essa correlação inicialmente à pandemia, que poderia ter afetado o desempenho dos alunos, no entanto, não observamos uma diferença de média tão significativa. Isso deve-se à forma de avaliação do ENEM pelo TRI, que de certa forma normaliza a prova com base no desempenho geral dos estudantes na prova.

A partir dos dados do INSE, também é possível calcular uma outra métrica, que nós chamamos de INSE *per capita*. Em suma, fazemos a média ponderada do INSE absoluto de cada escola pela quantidade de alunos que ela possui. Com essa nova medida, podemos separar os municípios de acordo com os níveis do INSE, levando em consideração a proporção de alunos entre suas escolas. As médias de notas usando essa métrica estão dispostas na Figura 21.

Claramente, as pontuações dos candidatos são influenciadas pelo INSE *per capita*. Além disso, a figura abaixo revela uma dinâmica intrigante em

relação ao desempenho dos estudantes em diferentes provas: as notas médias dos alunos de municípios Nível I são maiores em Linguagens do que em Matemática. No Nível V, o oposto ocorre. Isso indica que a prova de Matemática é mais eficaz em diferenciar os alunos, o que condiz com o cenário de defasagem no aprendizado dessa disciplina no Brasil.

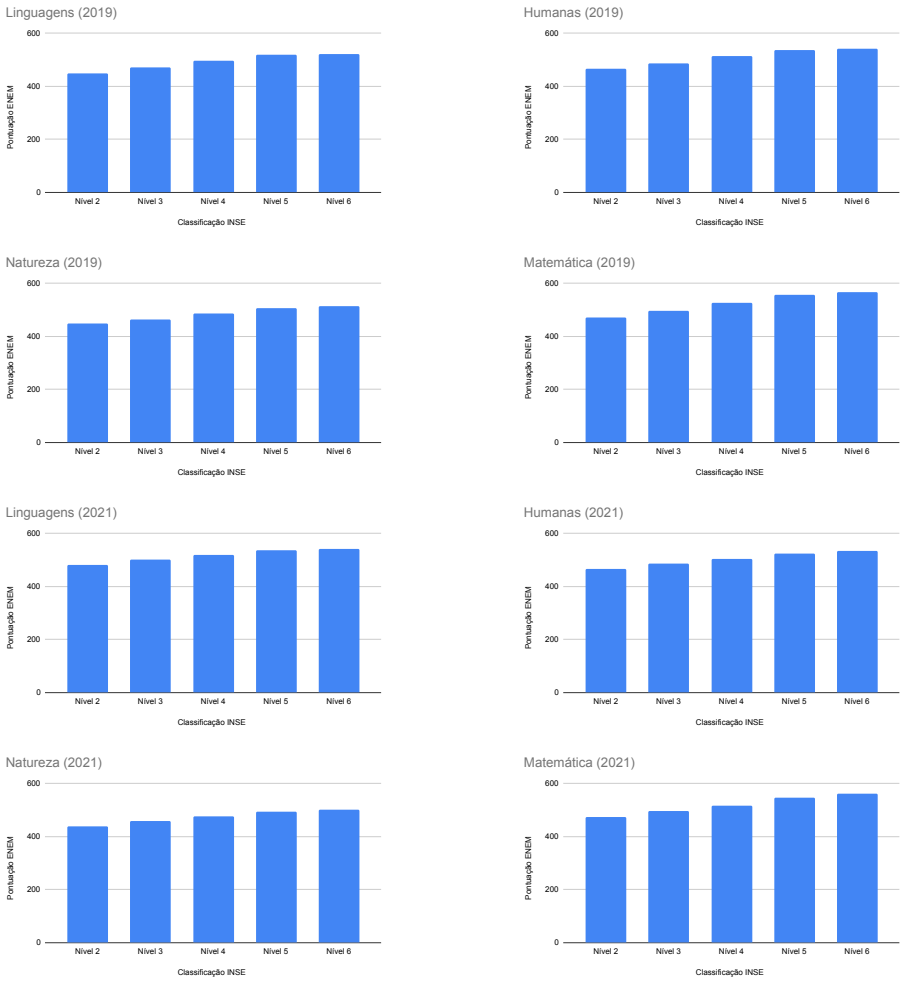

Figura 21: Médias de notas por municípios, agrupados pela métrica customizada INSE *per capita*.

Uma outra correlação que podemos traçar é relacionando o desempenho das escolas públicas e privadas com seu nível no INSE. Primeiramente, é possível fazer um gráfico de dispersão que analisa a relação do INSE e o desempenho do ENEM das escolas privadas e públicas. Esses dados estão resumidos no gráfico abaixo:

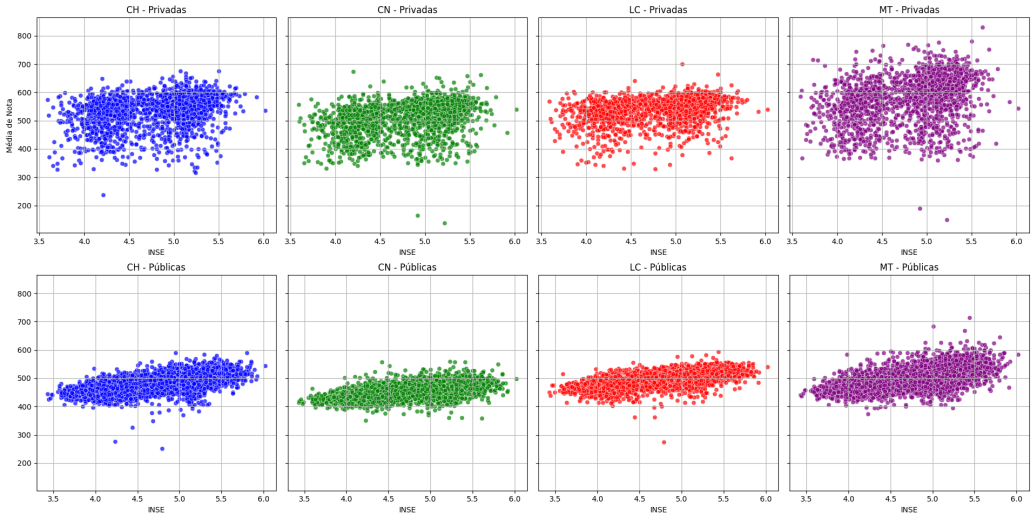

Figura 22: Distribuição - Escolas Públicas x Privadas - 2019

O gráfico mostra de maneira evidente a diferença de desempenho entre as escolas públicas e privadas. Notamos que as notas das escolas privadas são mais consistentes em relação ao INSE (não dependem tanto dele, aprofundaremos mais adiante), e têm uma distribuição mais espaçada e heterogênea por município, evidenciando municípios com escolas que geram bons desempenhos médios no ENEM e outras com desempenho inferior. Além disso, é evidente como em todas as disciplinas os resultados das escolas privadas superam os das escolas públicas. As escolas públicas dificilmente alcançam a média de 600 pontos, enquanto as escolas privadas frequentemente ultrapassam esse valor, especialmente na disciplina de Matemática, onde isso é bastante notável. Observando os dados das escolas públicas, vemos que são muito mais condensados, mostrando um nível comum às escolas públicas por municípios, sem grandes disparidades. Além disso, observamos a influência direta do INSE, já que os dados formam uma espécie de "disco"inclinado para cima à medida que o INSE cresce. Para entender melhor a correlação entre o INSE e o desempenho no ENEM neste cenário de escolas públicas e privadas, prosseguimos com a análise.

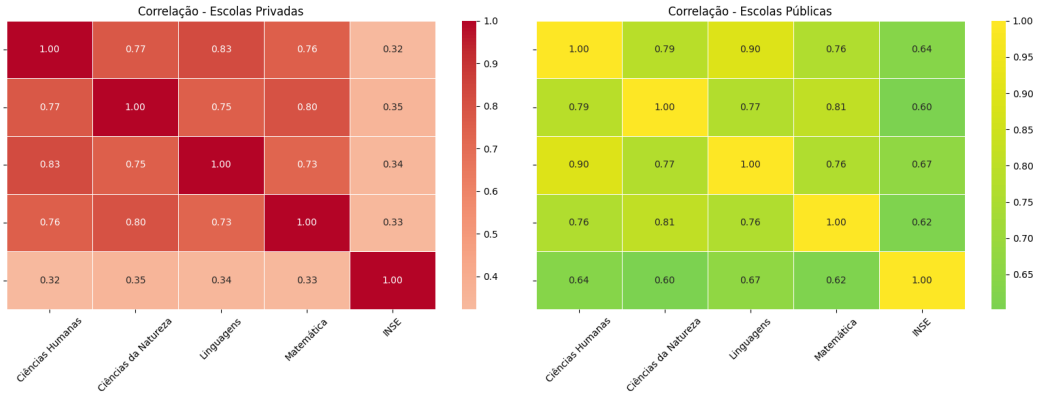

Figura 23: Heatmap - Escolas Públicas x Privadas - 2019

As matrizes de correlação da Figura 23 confirmam a hipótese de que as escolas públicas têm maior correlação/dependência do INSE para explicar o desempenho no ENEM. Observamos que a correlação em todas as disciplinas é maior nas escolas públicas do que nas escolas privadas.

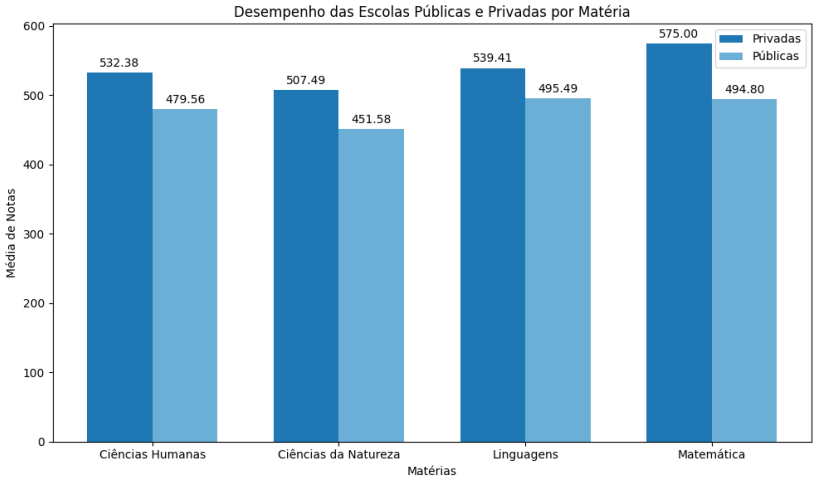

Figura 24: ENEM - Escolas Públicas x Privadas - 2019

Média do INSE das Escolas Públicas e Privadas

Tipo de Escola	Média do INSE
Privadas	4.79
Públicas	4.70

Figura 25: INSE - Escolas Públicas x Privadas - 2019

Por fim, temos as Figuras 24 e 25, que revelam a clara disparidade entre escolas públicas e privadas nos resultados do ENEM e na métrica INSE. É possível ver como as escolas particulares, mesmo não apresentando uma discrepância tão grande no INSE, por não serem tão correlacionadas com esse dado, conseguem desempenhos consideravelmente superiores às escolas públicas em todas as disciplinas. Este tipo de análise é fundamental ao considerarmos políticas como cotas e outras medidas similares, que visam equalizar essas disparidades.

6 Referências dos Dados

Disponível em: `https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem`

Acesso em: 19 de jun. de 2024.

Disponível em: `https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadores-educacionais/nivel-socioeconomico`

Acesso em: 19 de jun. de 2024.