

DESAFIO SEMANTIX 2022 – BIG DATA SCIENCE

ANÁLISE DOS DADOS DO ENEM 2021

Dezembro/2022
RODRIGO MENDES BESSONE
rmbessone@hotmail.com

SUMÁRIO

SUMÁRIO	
INTRODUÇÃO	
ANÁLISE DESCRITIVA.....	
VARIÁVEIS CATEGÓRICAS.....	
VARIÁVEIS QUANTITATIVAS.....	
PERGUNTAS	
ANÁLISE GERAL	
CORRELAÇÃO.....	
CONCLUSÃO	
BIBLIOGRAFIA	

Introdução

Como desafio, A **SEMANTIX** propôs aos alunos que analisem os dados disponibilizados do Enem 2021

(<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>).

O Exame Nacional do Ensino Médio (**ENEM**) foi instituído em 1998, com o objetivo de avaliar o desempenho escolar dos estudantes ao término da educação básica. Em 2009, o exame aperfeiçoou sua metodologia e passou a ser utilizado como mecanismo de acesso à educação superior. Desde 2020, o participante pode escolher entre fazer o exame impresso ou o Enem Digital, com provas aplicadas em computadores, em locais de prova definidos pelo Inep.

As notas podem ser usadas para acesso ao Sistema de Seleção Unificada (Sisu) e ao Programa Universidade para Todos (ProUni). Elas também são aceitas em mais de 50 instituições de educação superior. Além disso, os participantes do Enem podem pleitear financiamento estudantil em programas do governo, como o Fundo de Financiamento Estudantil (Fies). Os resultados do Enem possibilitam, ainda, o desenvolvimento de estudos e indicadores educacionais.

Com a finalidade de utilizarmos as habilidades aprendidas e desenvolvidas ao longo do treinamento.

O Enem é composto de 180 questões, distribuídas em quatro provas objetivas: Ciências Humanas e suas Tecnologias (História, Geografia, Filosofia e Sociologia); Ciências da Natureza e suas Tecnologias (Química, Física e Biologia); Linguagens, Códigos e suas Tecnologias (Língua Portuguesa, Literatura, Língua Estrangeira – Inglês ou Espanhol, Artes, Educação Física e Tecnologias da Informação e Comunicação); Matemática e suas Tecnologias (Matemática).

Além disso, os alunos devem fazer uma redação e responder a um questionário socioeconômico-cultural. O INEP disponibiliza todas as respostas do questionário bem como as notas em cada uma das áreas de conhecimento por meio dos microdados do ENEM.

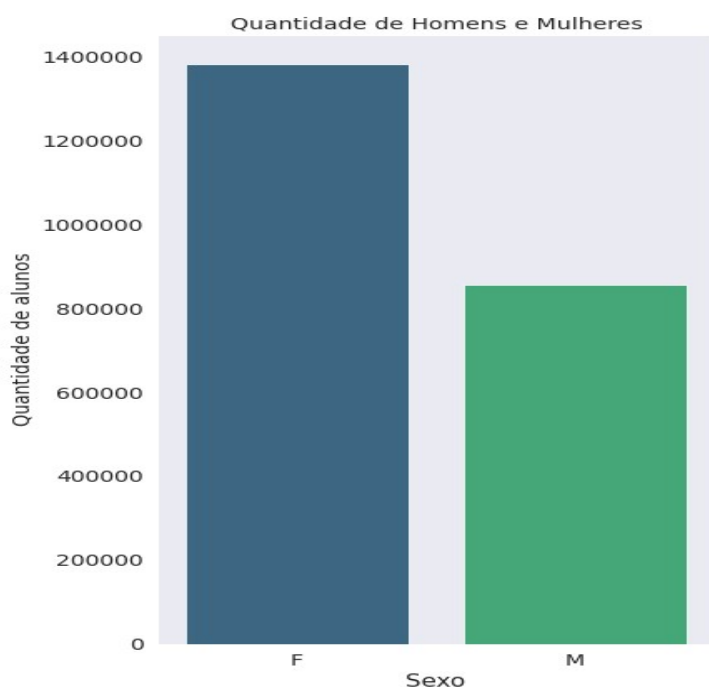
Para desempenhar esta análise me utilizei de dados de 3389831 alunos espalhados por 5 regiões do país, de escolas públicas e privadas, de todas as raças e sexo.

ANÁLISE DESCRITIVA

Variáveis categóricas

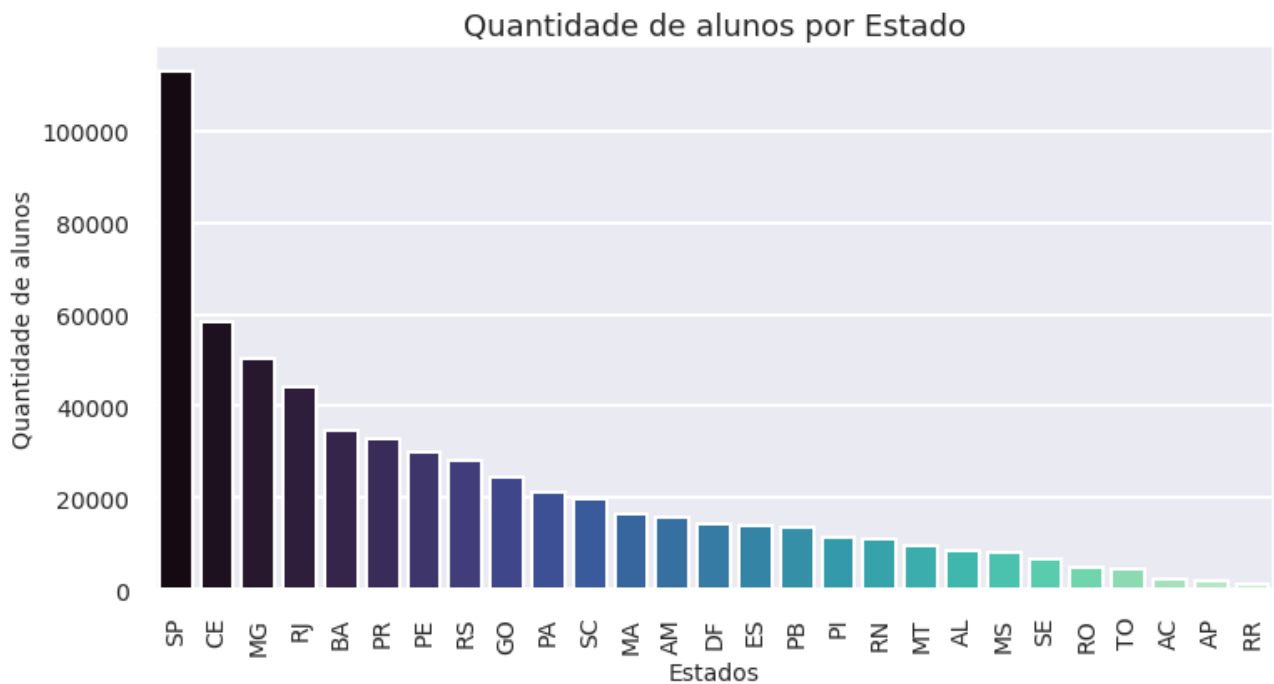
Iremos iniciar a análise descritiva dos dados com as variáveis categóricas. Em estatística, uma variável categórica é uma variável que pode assumir apenas um número limitado, e geralmente fixo, de valores possíveis, atribuindo cada indivíduo ou outra unidade de observação a um determinado grupo ou categoria nominal com base em alguma propriedade qualitativa.

Distribuição de alunos por sexo



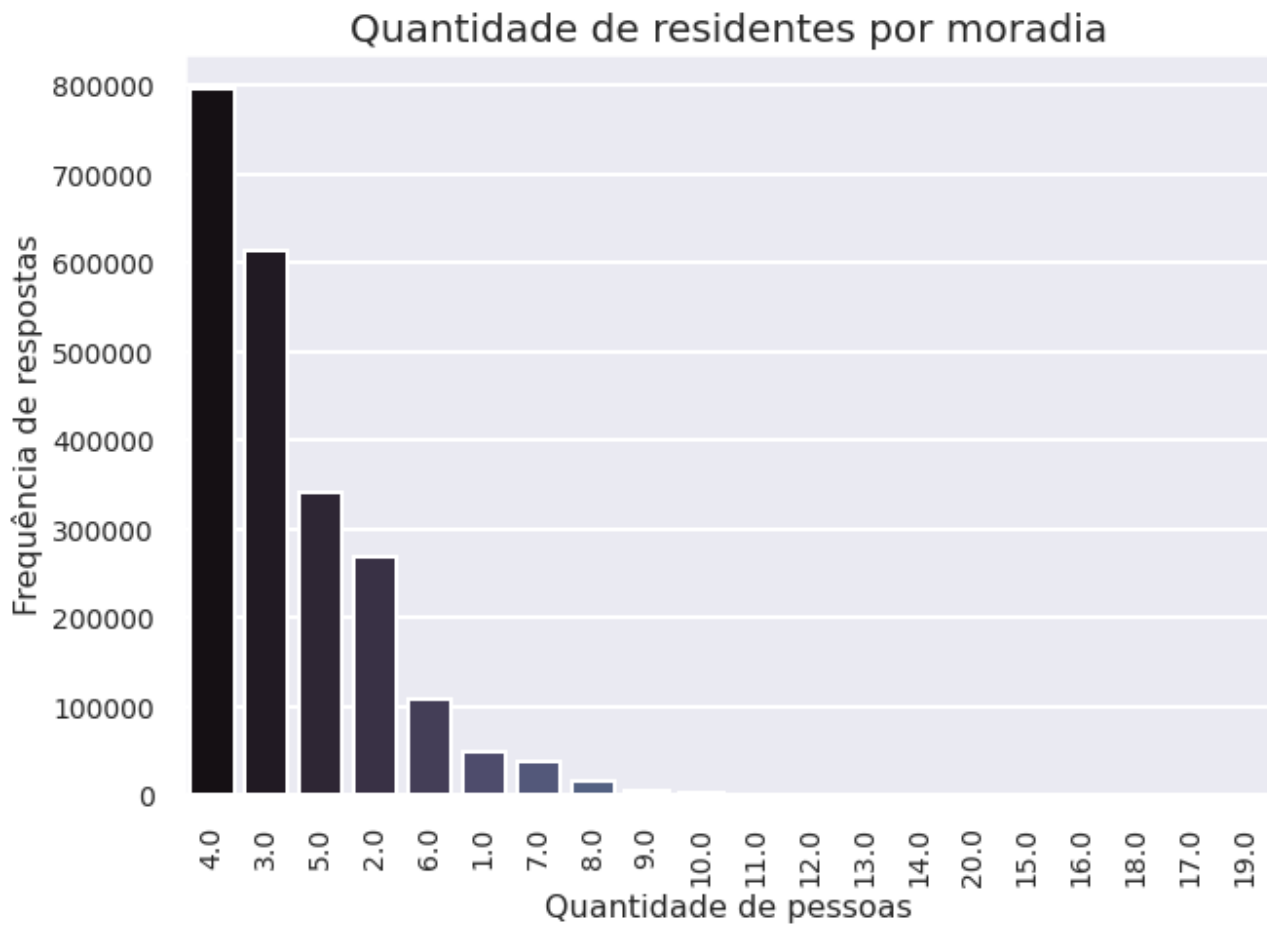
Mulheres representam 61.7% dos candidatos, contra 38.3% de homens.

Distribuição de alunos por Estados



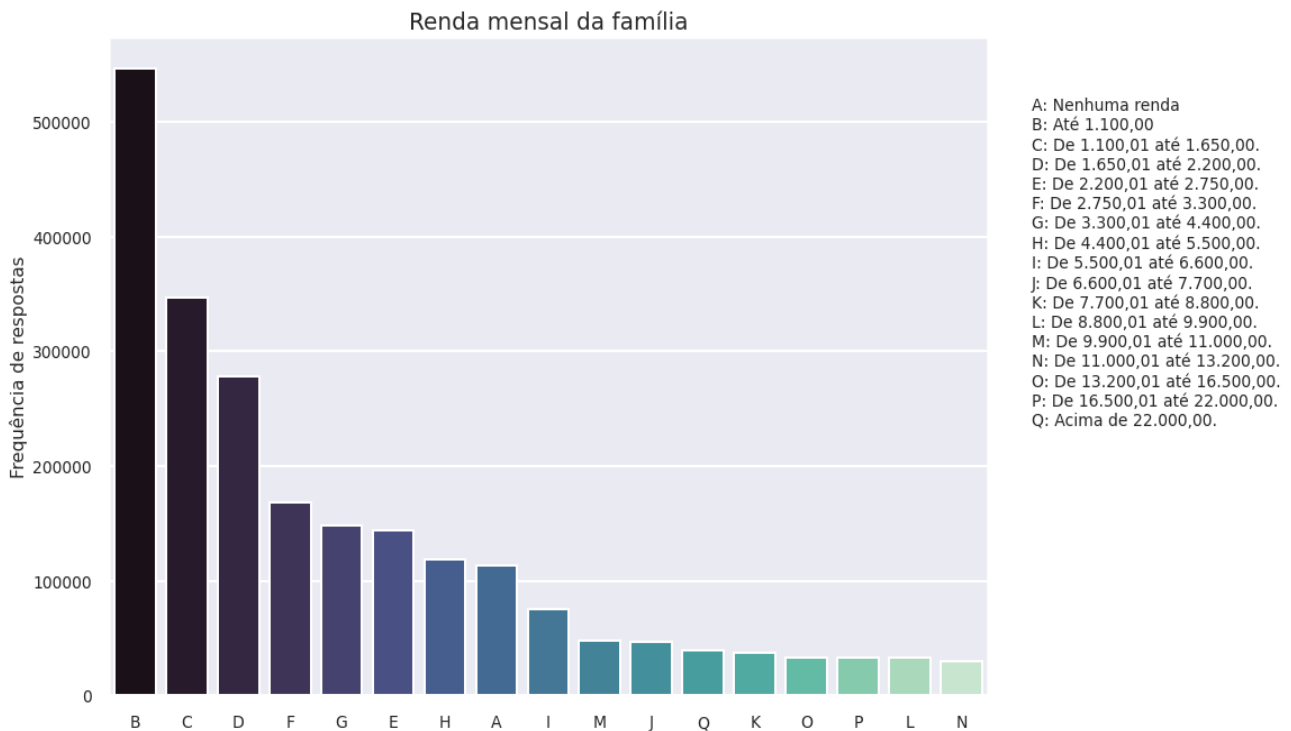
A região Sudeste representa 36.3% do total, seguida pela região Nordeste com 31.1% dos inscritos. A região Norte representa 8.4%, a região Centro-Oeste representa 15.9% dos participantes e por fim a região Sul representa 13.3% dos participantes do ENEM.

Moradores por residência



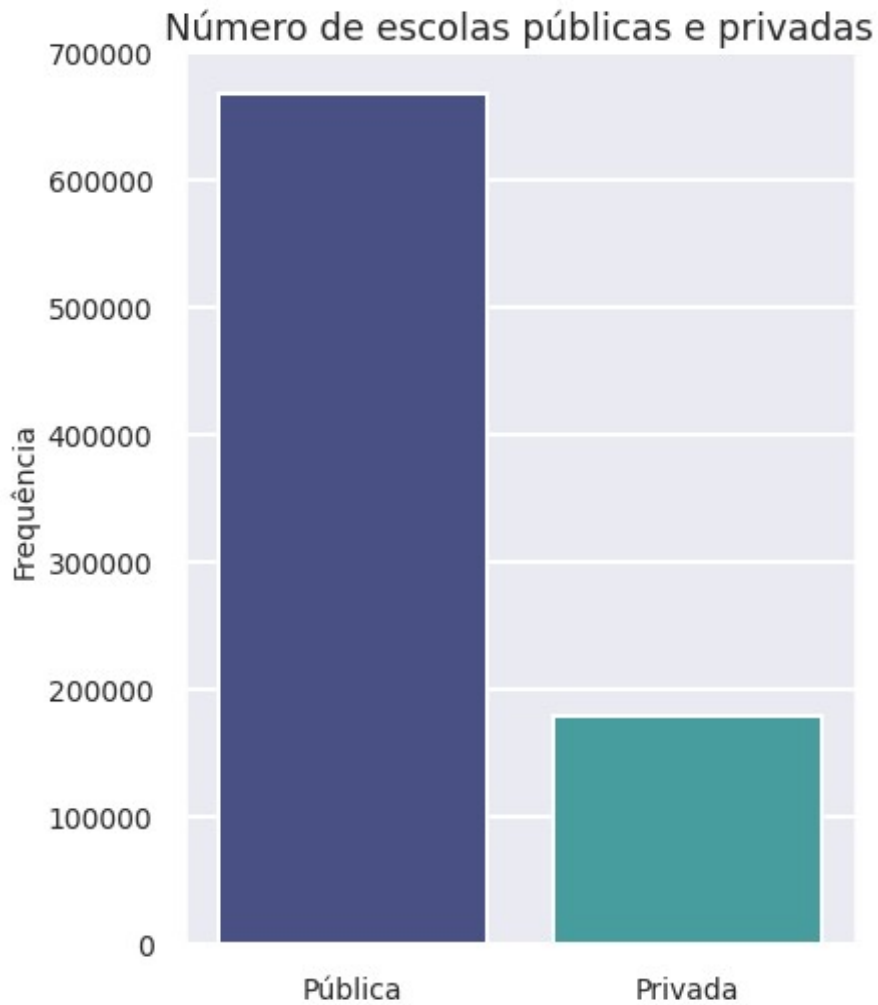
Pelo gráfico percebe-se que a grande maioria dos lares possuem de três a quatro moradores por residência.

Renda por família



Percebe-se que mais da metade das famílias de candidatos do ENEM vivem com até dois salários-mínimos. Este gráfico demonstra uma noção clara da imensa desigualdade social que existe no Brasil (em 2021), com a maioria da população estando nas condições B ou C (33% das respostas).

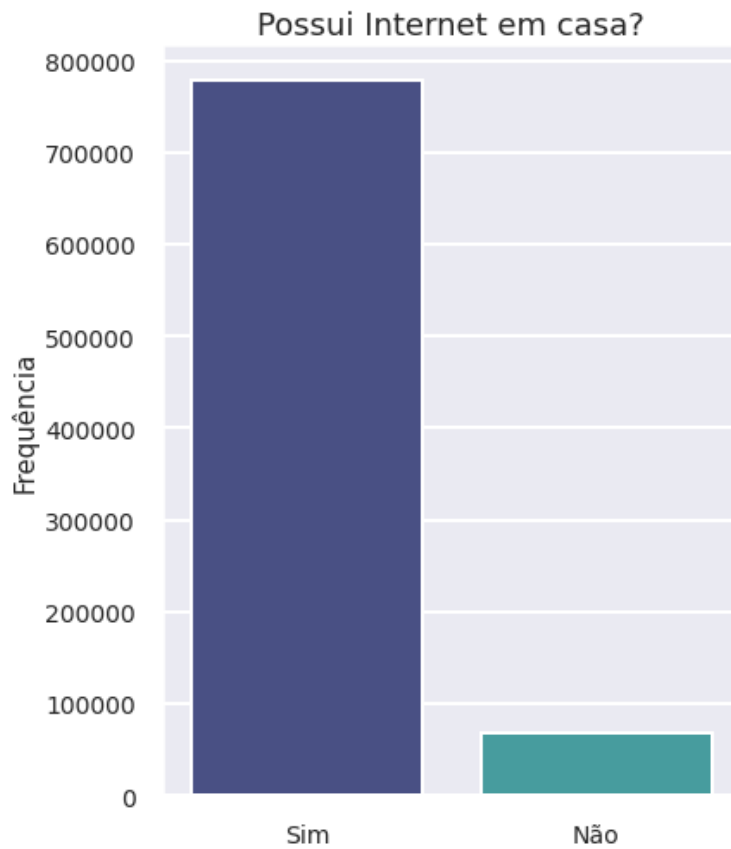
Tipo de escola



Do universo de alunos inscritos no ENEM, temos que 78.8% são oriundos de escolas públicas, ou seja, parte considerável depende de investimentos na educação pública.

PERGUNTAS

Possui internet em casa?

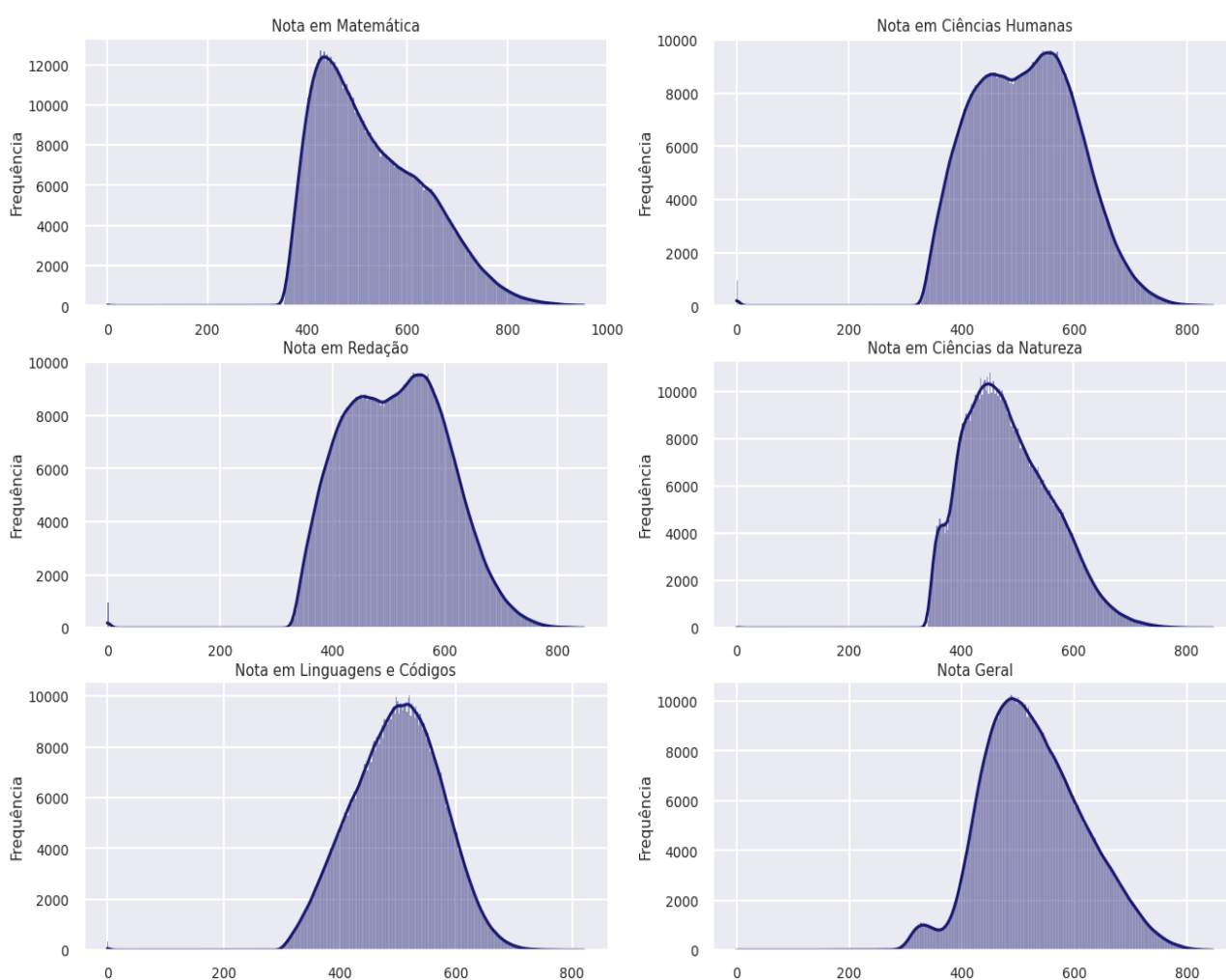


Dos alunos inscritos no ENEM o percentual de 91.9% possuem internet em casa. O que possibilita o crescimento de cursos EAD.

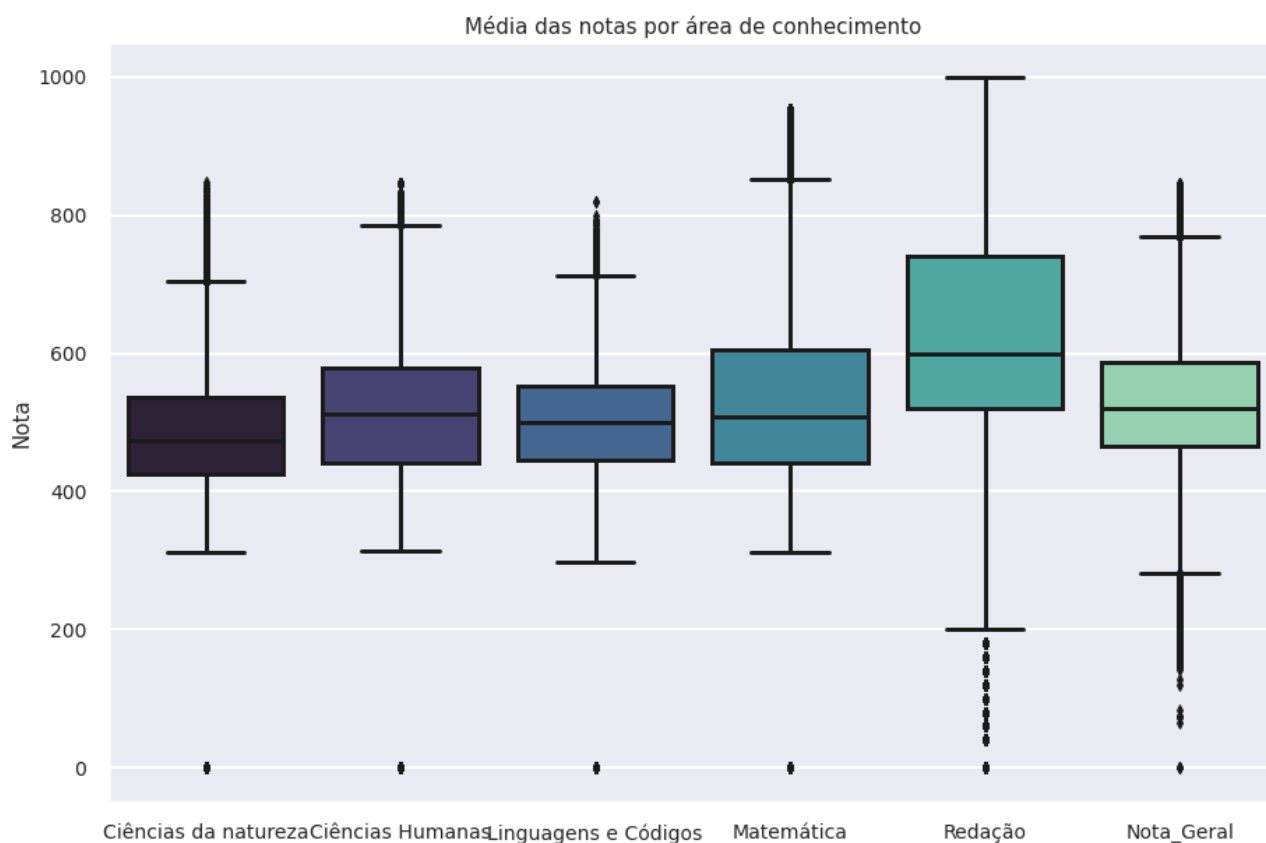
Variáveis quantitativas

São aquelas cujos valores são expressos em números. As variáveis quantitativas se subdividem, ainda, em: variáveis quantitativas discretas – quando resultam de um conjunto finito (ou enumerável) de valores possíveis.

Notas dos alunos

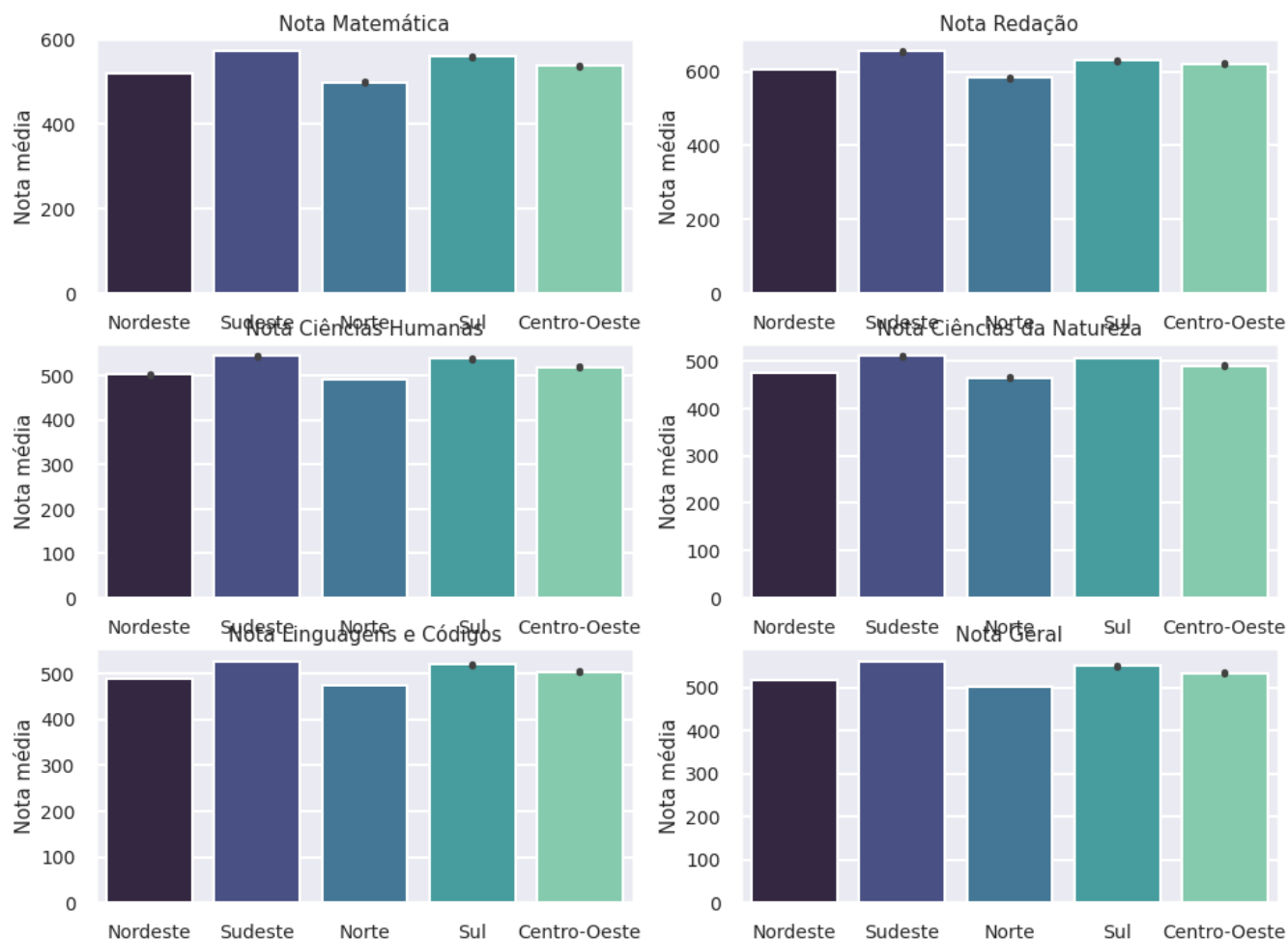


Ao analisar os gráficos das notas de cada disciplina e a nota geral. Podemos notar uma certa aproximação de uma distribuição normal em alguns dos gráficos. Me chamou a atenção o gráfico das notas de matemática, o qual nos oferece uma assimetria a direita, o que não é uma surpresa, haja vista o baixo desempenho educacional do Brasil na disciplina, como exemplo os dados do Programa Internacional de Avaliação de Estudantes (PISA), uma vez que o Brasil ocupa a 54^o colocação.



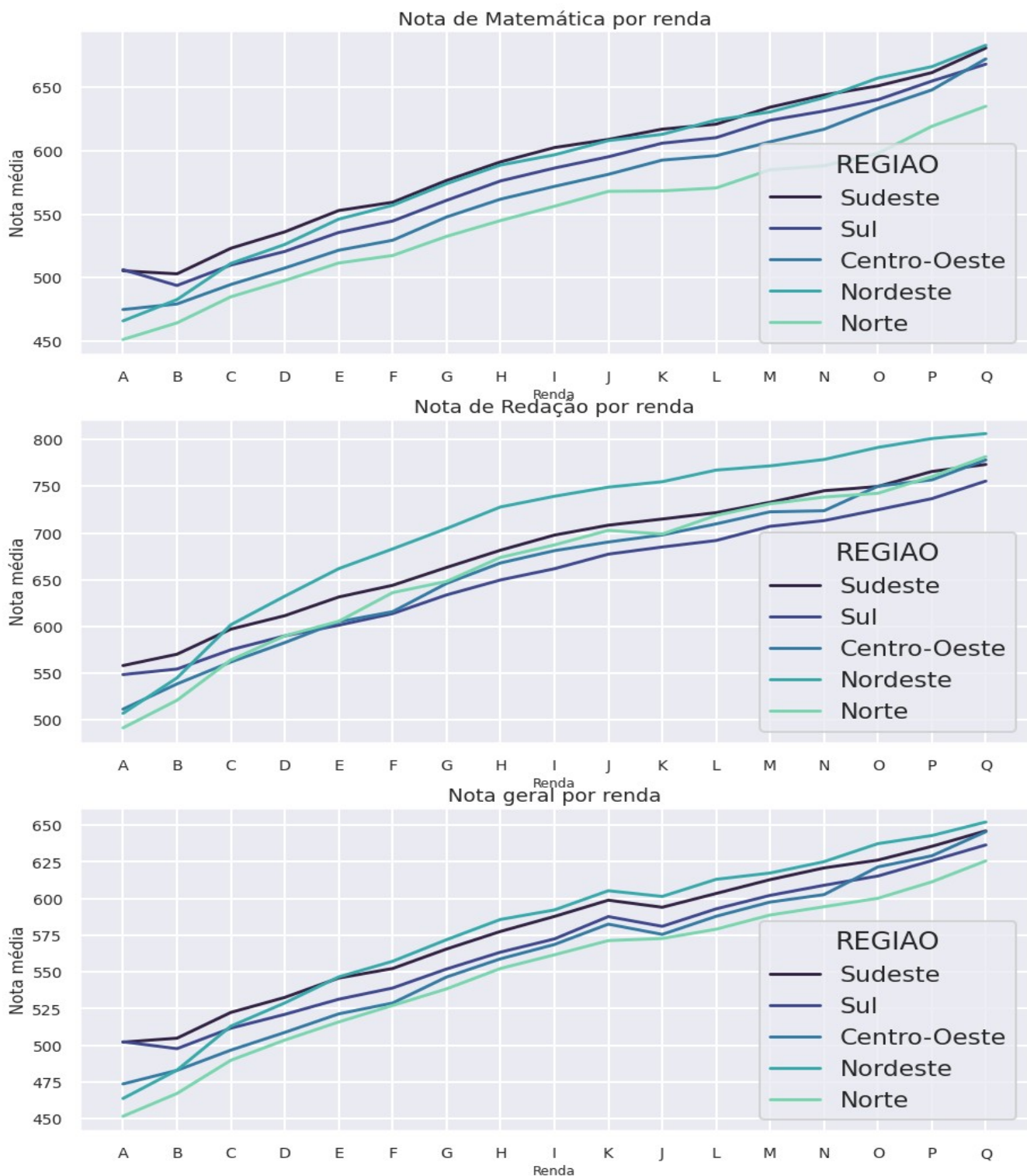
Note que a média é a curva que mais se aproxima de uma normal, o que pode ser descrito pelo Teorema central do limite, que diz que quanto maior o tamanho da amostra, a distribuição amostral da sua média tende a uma distribuição normal. Já que temos uma amostra bem grande, podemos ver esse teorema na prática. Há um destaque das notas de redação que produz grandes oscilações.

Valor das médias das notas por estados e regiões



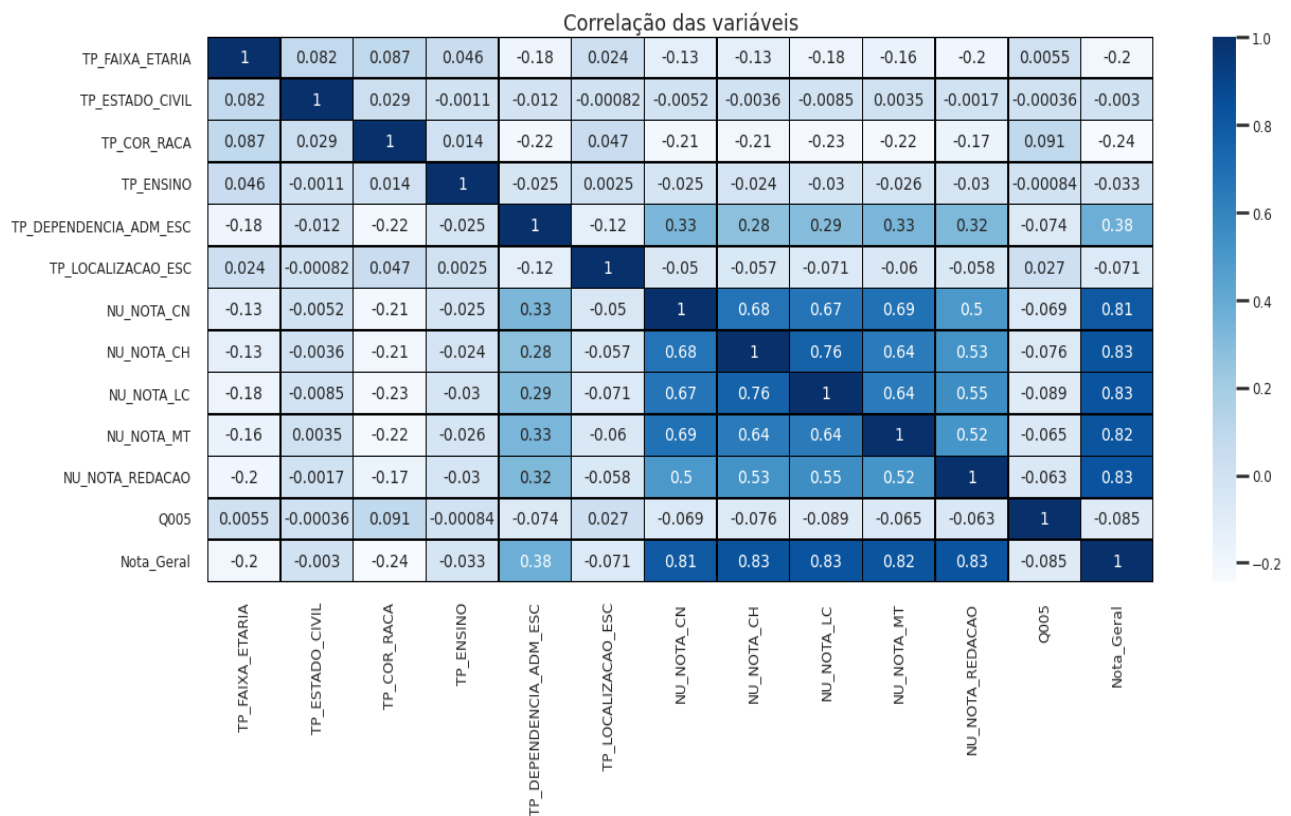
A análise dos gráficos evidencia a diferença entre as regiões brasileiras, onde a região Sudeste obteve as melhores médias e a região Norte as piores, em todas as áreas de conhecimento. O Sul e o Sudeste se sobressaiu e reflete hoje melhor qualidade econômica e social, o que influencia nas maiores médias de seus alunos em todas as áreas do conhecimento.

A renda familiar influencia na nota?



Os gráficos comprovam que a relação entre renda e nota é praticamente linear, uma correlação altamente positiva, quanto maior a renda, melhor o desempenho do aluno no ENEM.

Correlação



Note que quando se trata das notas nas áreas de conhecimento individuais (CH, CN, LC, MT e Redação) a correlação entre elas é considerada forte.

Conclusão

Podemos concluir que os candidatos participantes do Enem 2021 possuem idades, raça, escolaridade e classes sociais muito distintas.

Os candidatos são compostos por sua maioria por mulheres e possuem média geral de 538,40 pontos, valor que varia bastante quando analisamos os candidatos com base em sua escolaridade, renda, estado de origem, e tipo de escola.

Algumas tendências, como o aumento das notas na medida do aumento da renda, outras descobertas, como a diferença de escolaridade ser fator determinante para a nota.

De maneira geral, percebe-se mais necessidade de investimentos em educação nas regiões Norte e Nordeste.

Ademais, devido a grande relevância dessa problemática para o futuro educacional do ensino brasileiro, acredito que este pequeno estudo sirva de pontapé inicial para algumas contestações.

Bibliografia

TEOREMA CENTRAL DO LIMITE – WIKIPÉDIA, A ENCICLOPÉDIA LIVRE. Disponível em: https://pt.wikipedia.org/wiki/Teorema_central_do_limite.

ENEM 2021 Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem>

REFERÊNCIAS BIBLIOGRÁFICAS. [1] BUSSAB, W. O.; MORETTIN, P. A. – Estatística Básica. Atual Editora,. São Paulo, 1988