

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
PROGRAMAÇÃO BÁSICA DE COMPUTADORES

**JULIA RIZZI BUSATTO, JULIANA RODRIGUES DE SOUZA E PEDRO BELUMAT
LOSER QUEIROZ**

ANÁLISE DE MICRODADOS DO ENEM 2022

Relatório

VITÓRIA

2023

Julia Rizzi Busatto, Juliana Rodrigues de Souza e Pedro Belumat Loser Queiroz

Relatório a respeito do trabalho sobre análise de dados

Trabalho apresentado à Universidade Federal do Espírito Santo (UFES), no curso de graduação Engenharia de Produção, na matéria Programação básica de computadores, ministrada pelo professor Luis Antonio de Souza Junior.

VITÓRIA

2023

1. Introdução

O Exame Nacional do Ensino Médio (ENEM) é um dos passos mais importantes da trajetória de qualquer estudante no nosso país, porém é de conhecimento comum que nem todos têm as mesmas oportunidades, assim gerando uma desigualdade entre eles.

Para combater esse problema foram criadas diversas cotas para todas as categorias de estudantes, sendo elas baseadas em raça, em renda ou em deficiências. Assim, pessoas que possam possivelmente ter tido seu aprendizado prejudicado por fatores diversos ganham o direito de competir somente com pessoas com situações semelhantes pelas vagas separadas para o seu grupo de cotas.

Ademais, é importante ressaltar a presença dos dados referentes às escolas em que cada aluno se formou, assim abrindo margem para uma comparação entre os formandos de escolas públicas e particulares, dentro e fora das cotas.

Portanto, dadas informações citadas acima, o assunto em questão possui como objetivo apresentar dados, a fim de fornecer uma análise abrangente e detalhada sobre os resultados do ENEM 2022, examinando minuciosamente as informações coletadas, esperando com que tais contribuam para a compreensão das possíveis correlações entre variáveis socioeconômicas presentes na realização do exame e em sua obtenção de resultados.

Logo, diante da nossa hipótese, pode-se ser esperado que ao decorrer das análises, se obtenha um resultado que mostre como as diferenças avaliadas beneficiem o grupo social cujas condições estivessem mais favoráveis ao sucesso das notas do que a obtenção de um resultado negativo, visto que atualmente, dentre as diversas esferas da sociedade, os grupos que apresentam dificuldades estruturais e/ou econômicas, são mais suscetíveis à falha por não apresentarem as mesmas oportunidades que os demais.

2. Descrição da base de dados

A base de dados a ser utilizada no trabalho será os Microdados do Enem 2022, que podem ser encontrados em: [Enem - Inep](#).

Os dados presentes na base, englobam o número de inscrição do participante, o qual é apenas uma máscara, logo não representa o número original, informações básicas sobre os inscritos, como a faixa etária, que varia de menor de 17 anos até maior de 70 anos, o sexo, estado civil, cor/raça, nacionalidade, o ano de conclusão do ensino médio, o tipo de escola que estudou no ensino médio (pública ou privada) e se o inscrito é considerado treineiro, no caso de não concluir o ensino médio em 2022.

Também estão presentes dados sobre a escola do participante, como a sigla do estado e o nome do município em que a escola está localizada, a dependência administrativa (federal, estadual, municipal ou privada) e localização (urbana ou rural). E a sigla do estado e o nome do município de aplicação da prova.

Além disso, pode-se encontrar os dados da prova objetiva e da redação, o primeiro inclui a presença nas provas dos dois dias (ciências da natureza, ciências humanas, linguagens e códigos e matemática), a língua estrangeira escolhida pelo participante, o código das provas aplicadas (cor, libras, digital, reaplicação), a nota que o inscrito obteve em cada prova, vetor com as respostas dos participantes e vetor com o gabarito de cada prova. O segundo, a nota total obtida na redação, a nota de cada competência (da 1ª a 5ª) e a situação da redação (anulada, em branco, fuga do tema, texto insuficiente, sem problemas).

O último item da base de dados, apresenta as questões e as respostas do questionário socioeconômico, o qual é composto por 25 perguntas, que contêm o nível de escolaridade e a ocupação dos pais do participante, a quantidade de pessoas que moram com o inscrito, a renda mensal da família, características da residência e se tem acesso à internet.

Por fim, a base de dados a ser utilizada apresenta 76 colunas e 1048576 linhas.

2.1. Características da base de dados que o grupo irá focar na análise

Tabela 1 - Dados principais

Nome da coluna	Descrição	Tipo	Valores esperados
TP_FAIXA_ETARIA	Faixa etária	Inteiro	1 a 20
TP_SEXO	Sexo	Caractere	“F” e “M”
TP_COR_RACA	Cor/raça	Inteiro	0 a 6
TP_ST_CONCLUSAO	Situação de conclusão do Ensino Médio	Inteiro	1 a 4
TP_ANO_CONCLUIU	Ano de Conclusão do Ensino Médio	Inteiro	0 a 16
TP_ENSINO	Tipo de instituição que concluiu ou concluirá o Ensino Médio	Inteiro	1 ou 2
IN_TREINEIRO	Indica se o inscrito fez a prova com intuito de apenas treinar seus conhecimentos	Inteiro	0 ou 1
NO_MUNICIPIO_ESC	Nome do município da escola	String	“Vitória”, “São Paulo”, “Rio de Janeiro” etc
SG_UF_ESC	Sigla da Unidade da Federação da escola	String	“ES”, “SP”, “RJ”, “AM” etc
TP_DEPENDENCIA_ADM_ESC	Dependência administrativa (Escola)	Inteiro	1 a 4
TP_PRESENCA_CN	Presença na prova objetiva de Ciências da Natureza	Inteiro	0 a 2
TP_PRESENCA_CH	Presença na prova objetiva de Ciências Humanas	Inteiro	0 a 2
TP_PRESENCA_LC	Presença na prova objetiva de Linguagens e Códigos	Inteiro	0 a 2
TP_PRESENCA_MT	Presença na prova objetiva de Matemática	Inteiro	0 a 2
NU_NOTA_CN	Nota da prova de Ciências da Natureza	Float	0 a 1000
NU_NOTA_CH	Nota da prova de Ciências Humanas	Float	0 a 1000
NU_NOTA_LC	Nota da prova de Linguagens e Códigos	Float	0 a 1000
NU_NOTA_MT	Nota da prova de Matemática	Float	0 a 1000
TP_LINGUA	Língua Estrangeira	Inteiro	0 ou 1

TP_STATUS_REDACAO	Situação da redação do participante	Inteiro	1 a 9
NU_NOTA_COMP1	Nota da competência 1 - Demonstrar domínio da modalidade escrita formal da Língua Portuguesa.	Inteiro	0 a 200
NU_NOTA_COMP2	Nota da competência 2 - Compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa.	Inteiro	0 a 200
NU_NOTA_COMP3	Nota da competência 3 - Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista.	Inteiro	0 a 200
NU_NOTA_COMP4	Nota da competência 4 - Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação.	Inteiro	0 a 200
NU_NOTA_COMP5	Nota da competência 5 - Elaborar proposta de intervenção para o problema abordado, respeitando os direitos humanos.	Inteiro	0 a 200
NU_NOTA_REDACAO	Nota da prova de redação	Inteiro	0 a 1000
Q006	Corresponde à questão 6 do questionário socioeconômico, o aluno deve informar a renda mensal de sua família	Caractere	“A” a “Q”

3. Análise dos resultados

Nesta seção discutimos a relação entre diversos fatores que podem influenciar a nota final do aluno que participou do ENEM 2022.

A figura 1 mostra uma relação entre a renda mensal do participante e a nota obtida, percebe-se que nas quatro provas objetivas quanto maior a renda maior é a nota, sendo que essa tendência é mais perceptível na prova de matemática, o que pode indicar onde está o maior déficit no ensino brasileiro.

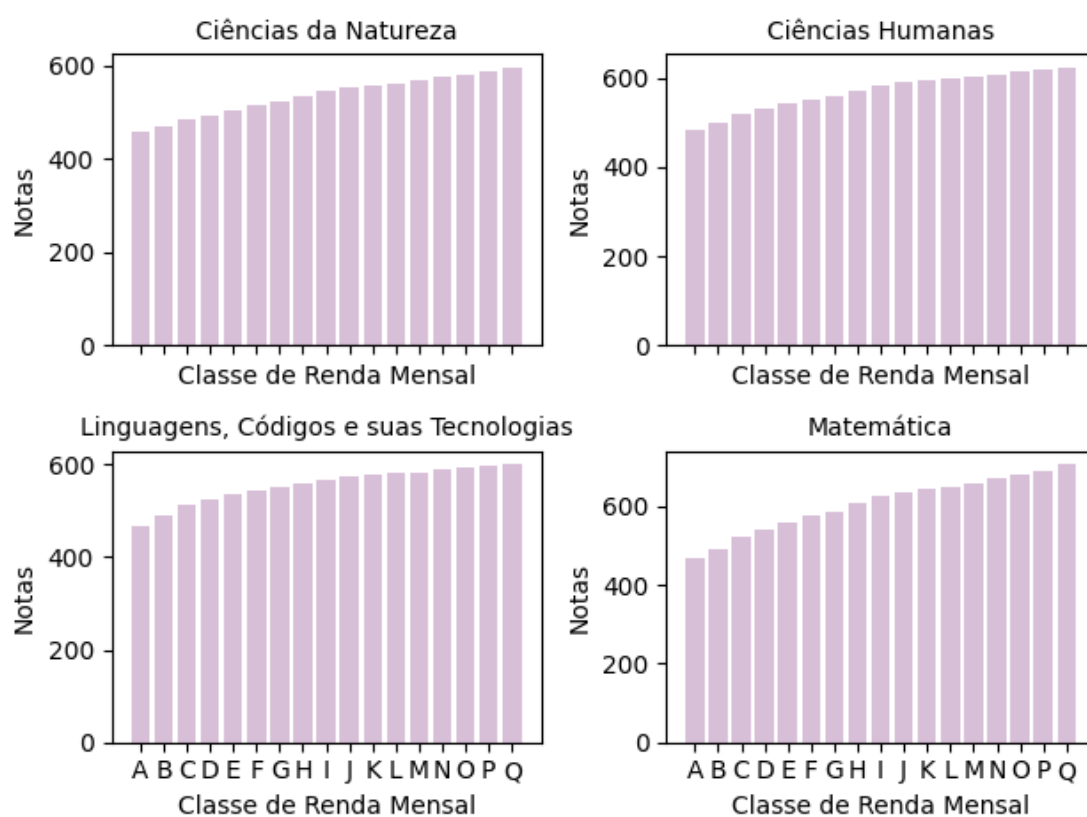


Figura 1. Relação entre renda mensal e média de nota

Tabela 2 - Legenda - Figura 1

Legenda - Classe de Renda Mensal					
A	Nenhuma Renda	G	De R\$ 3.636,01 até R\$ 4.848,00.	M	De R\$ 10.908,01 até R\$ 12.120,00.
B	Até R\$ 1.212,00	H	De R\$ 4.848,01 até R\$ 6.060,00.	N	De R\$ 12.120,01 até R\$ 14.544,00.
C	De R\$ 1.212,01 até R\$ 1.818,00.	I	De R\$ 6.060,01 até R\$ 7.272,00.	O	De R\$ 14.544,01 até R\$ 18.180,00.

D	De R\$ 1.818,01 até R\$ 2.424,00.	J	De R\$ 7.272,01 até R\$ 8.484,00.	P	De R\$ 18.180,01 até R\$ 24.240,00.
E	De R\$ 2.424,01 até R\$ 3.030,00.	K	De R\$ 8.484,01 até R\$ 9.696,00.	Q	Acima de R\$ 24.240,00.
F	De R\$ 3.030,01 até R\$ 3.636,00.	L	De R\$ 9.696,01 até R\$ 10.908,00.	-	-

A figura 2 mostra uma relação entre o tipo de escola e a nota média do participante, percebe-se que as escolas federais e privadas, obtêm notas maiores em comparação com as municipais e estaduais. A diferença é mais perceptível na prova de matemática.

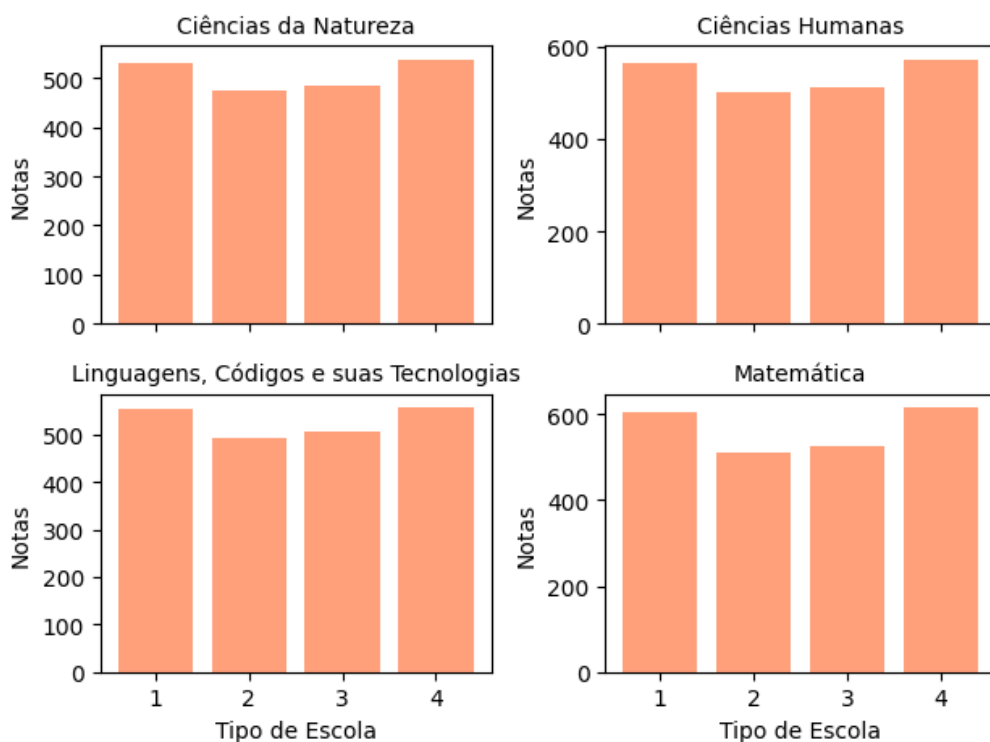


Figura 2. Relação entre média de nota e tipo de escola (privada, federal, estadual e municipal)

Tabela 3 - Legenda - Figura 2

Legenda - Tipo de escola			
1	2	3	4
Federal	Estadual	Municipal	Privada

Na figura 3 é mostrada uma relação entre cor/raça e nota média nas provas objetivas, percebe-se que os participantes de cor/raça branca obtêm as maiores notas nas quatro provas e em seguida quem não declarou a cor/raça. Os participantes pretos, pardos e indígenas, obtêm as menores notas, assim justificando a presença das cotas raciais para esse grupo de pessoas.

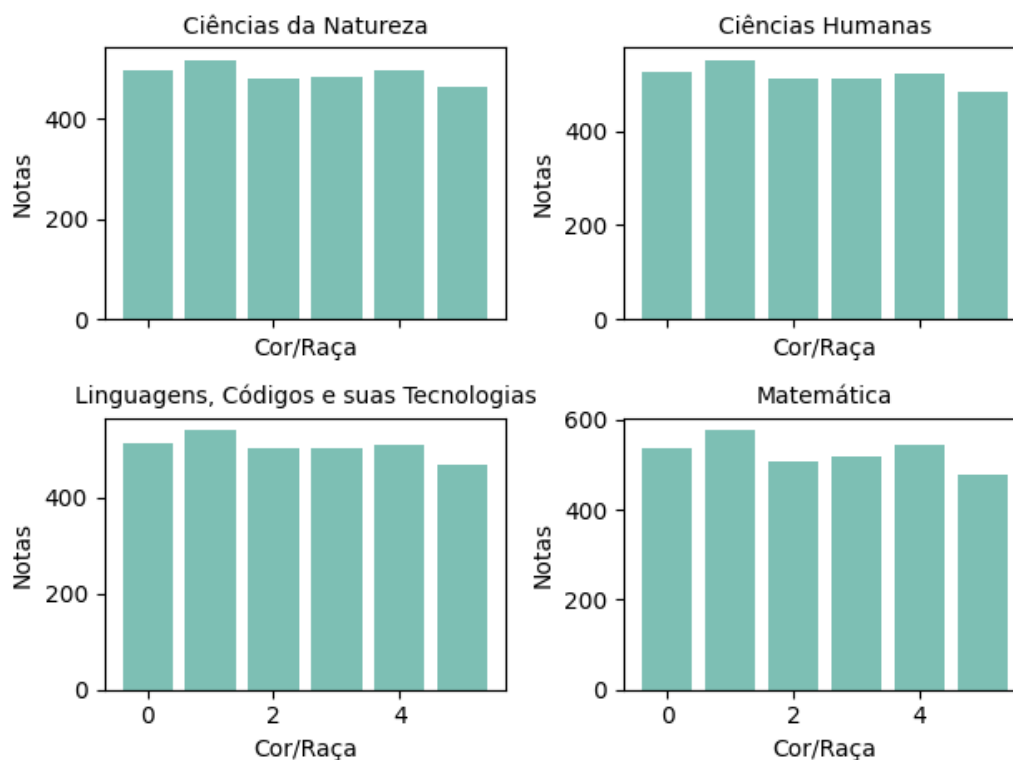


Figura 3. Relação entre média de nota e cor/raça

Tabela 4 - Legenda - Figura 3

Legenda - Cor/Raça					
0	1	2	3	4	5
Não declarado	Branca	Preta	Parda	Amarela	Indígena

A figura 4 mostra a % dos participantes de acordo com sua faixa etária, percebe-se que a maioria dos inscritos têm entre 17 e 19 anos, representando mais de 60% dos participantes, sendo que quem tem 17 e 18 anos, representam metade dos inscritos no ENEM 2022. A partir dos 19 anos, os participantes das próximas faixas etárias tendem a diminuir.

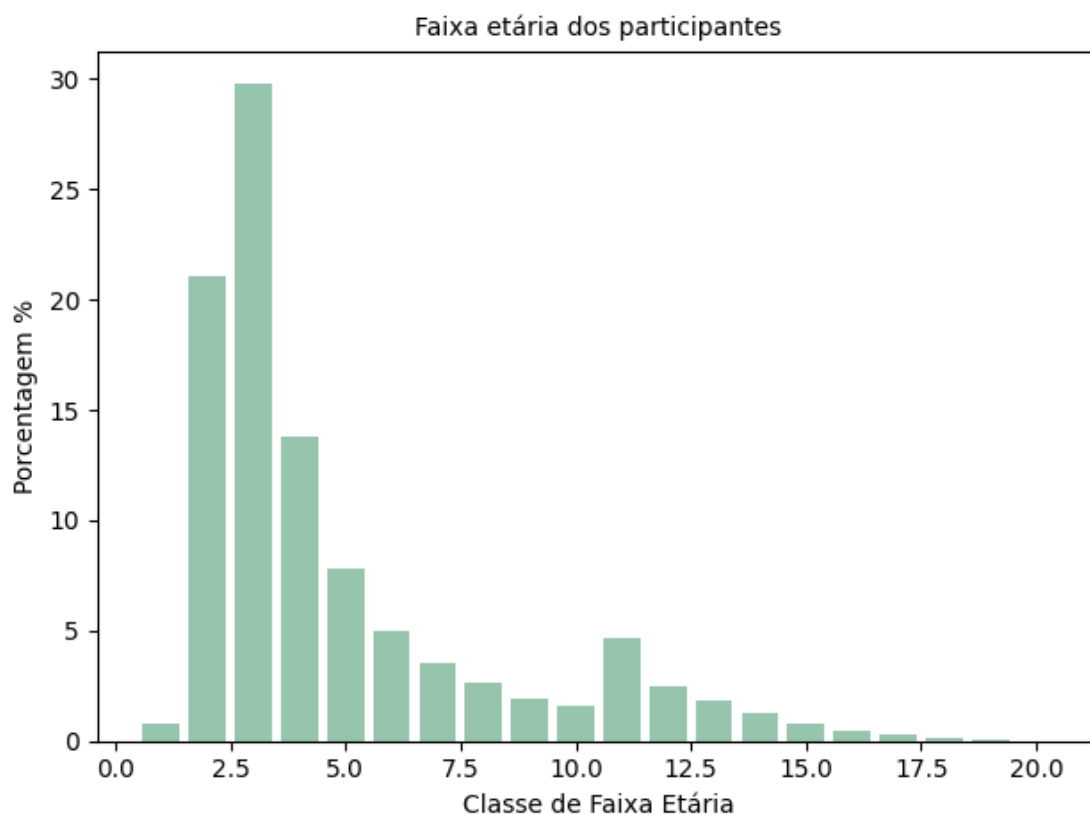


Figura 4. Faixa etária dos participantes

Tabela 5 - Legenda - Figura 4

Legenda - Faixa Etária							
1	Menor de 17 anos	6	21 anos	11	Entre 26 e 30 anos	16	Entre 51 e 55 anos
2	17 anos	7	22 anos	12	Entre 31 e 35 anos	17	Entre 56 e 60 anos
3	18 anos	8	23 anos	13	Entre 36 e 40 anos	18	Entre 61 e 65 anos
4	19 anos	9	24 anos	14	Entre 41 e 45 anos	19	Entre 66 e 70 anos
5	20 anos	10	25 anos	15	Entre 46 e 50 anos	20	Maior de 70 anos

A figura 5 mostra a relação entre média de nota nas provas objetivas e a faixa etária, percebe-se que a idade do participante não influencia de maneira significativa na nota final,

apenas na prova de matemática é possível visualizar uma nota mais baixa quando a faixa etária é mais avançada.

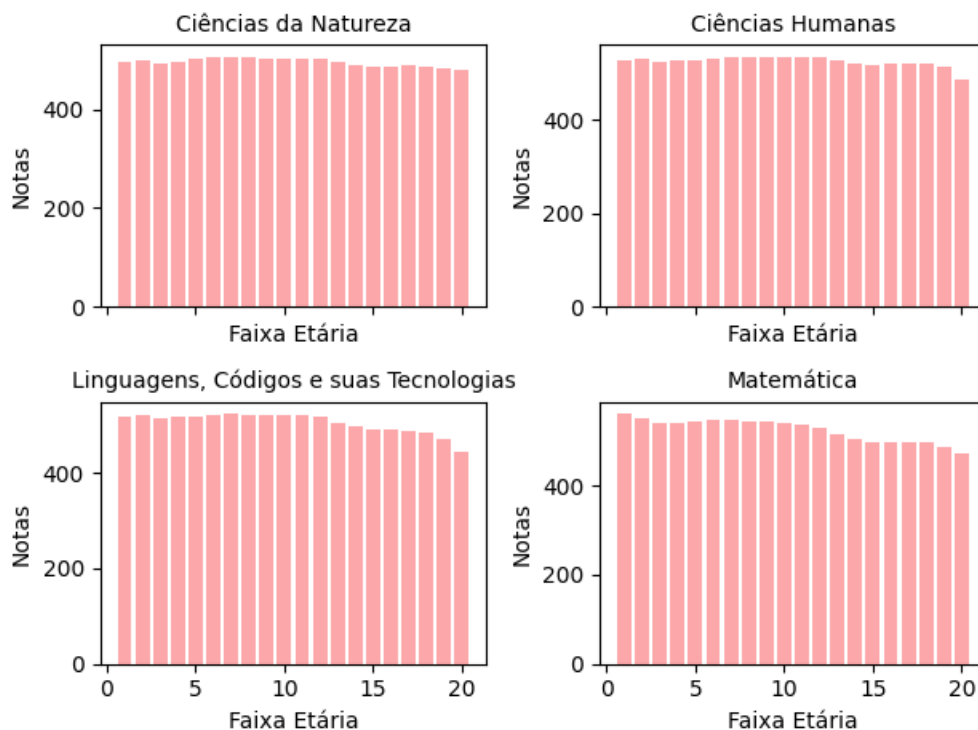


Figura 5. Relação entre faixa etária e média de nota

A figura 6 mostra a porcentagem de participantes de acordo com o sexo no estado do ES, é possível visualizar que mais de 50% dos participantes do ENEM 2022 no ES foram mulheres.

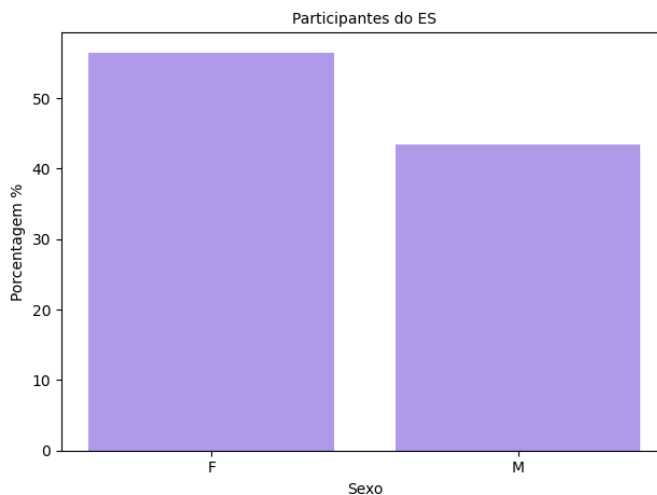


Figura 6. Participantes do sexo feminino e masculino

Tabela 6 - Legenda - Figura 6

Legenda - Sexo	
F	Feminino
M	Masculino

A figura 7 mostra a porcentagem de participantes do ES no ENEM 2022 com relação à cor/raça, percebe-se que mais de 80% dos participantes são brancos ou pardos e pouco mais de 10% são pretos.

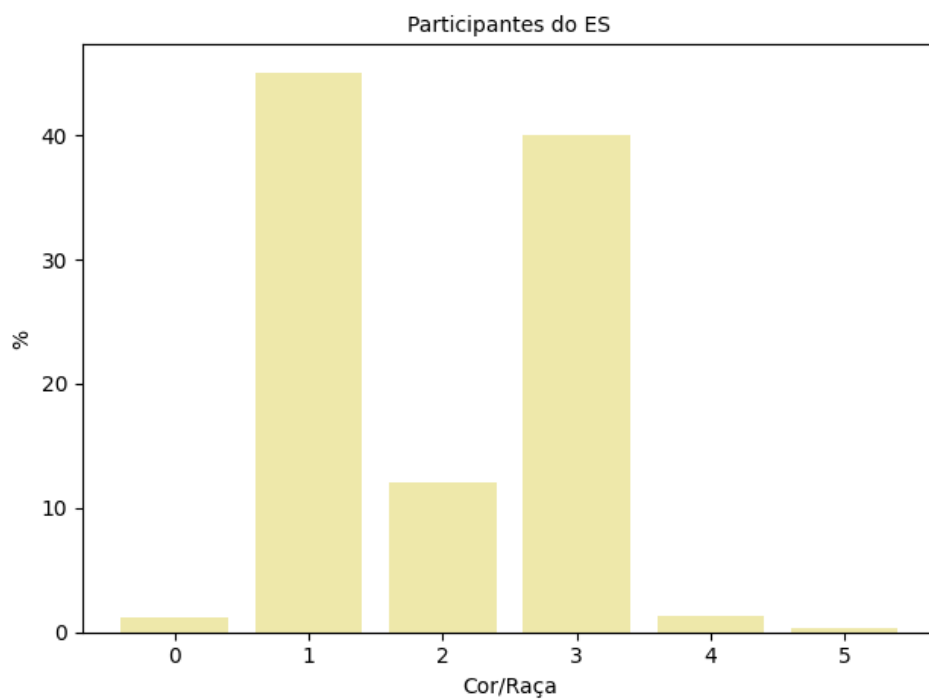


Figura 7. Cor/Raça dos participantes do estado do ES

Tabela 7 - Legenda - Figura 7

Legenda - Cor/Raça					
0	1	2	3	4	5
Não declarado	Branca	Preta	Parda	Amarela	Indígena

A figura 8, mostra a distribuição das notas de redação obtidas no ENEM 2022, percebe-se que a maioria dos participantes obtêm uma nota entre 500 e 700.

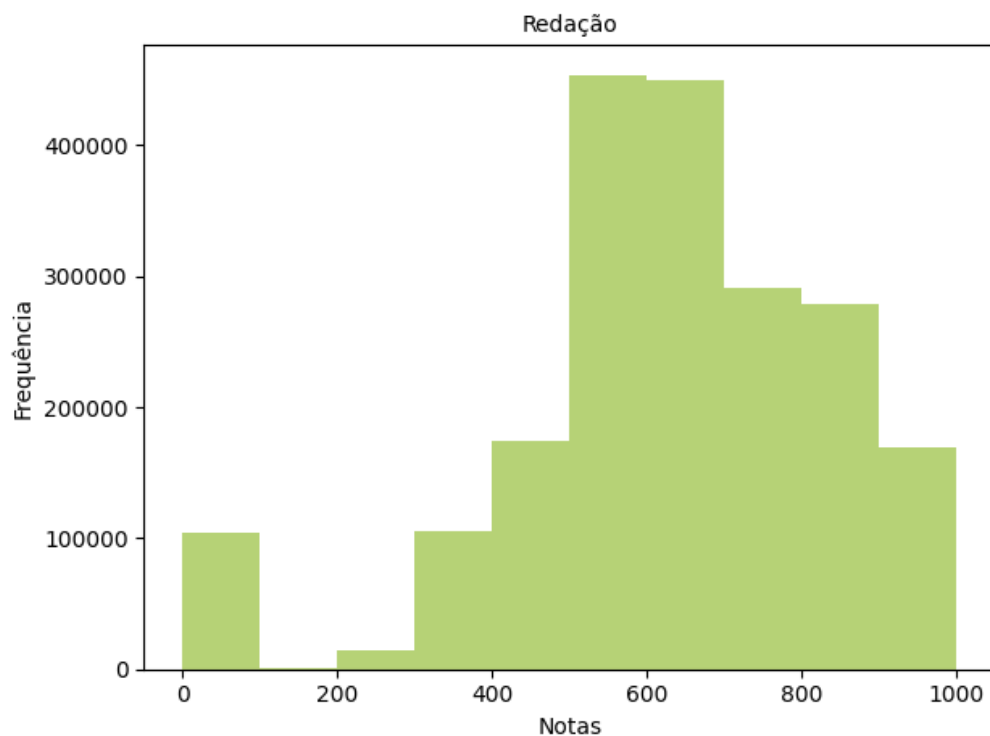


Figura 8. Distribuição das notas da redação

A figura 9, mostra relações de renda mensal, cor/raça, tipo de escola e estado com a nota 1000 na redação. Os participantes que obtiveram 1000 na redação são majoritariamente brancos e são da região nordeste.

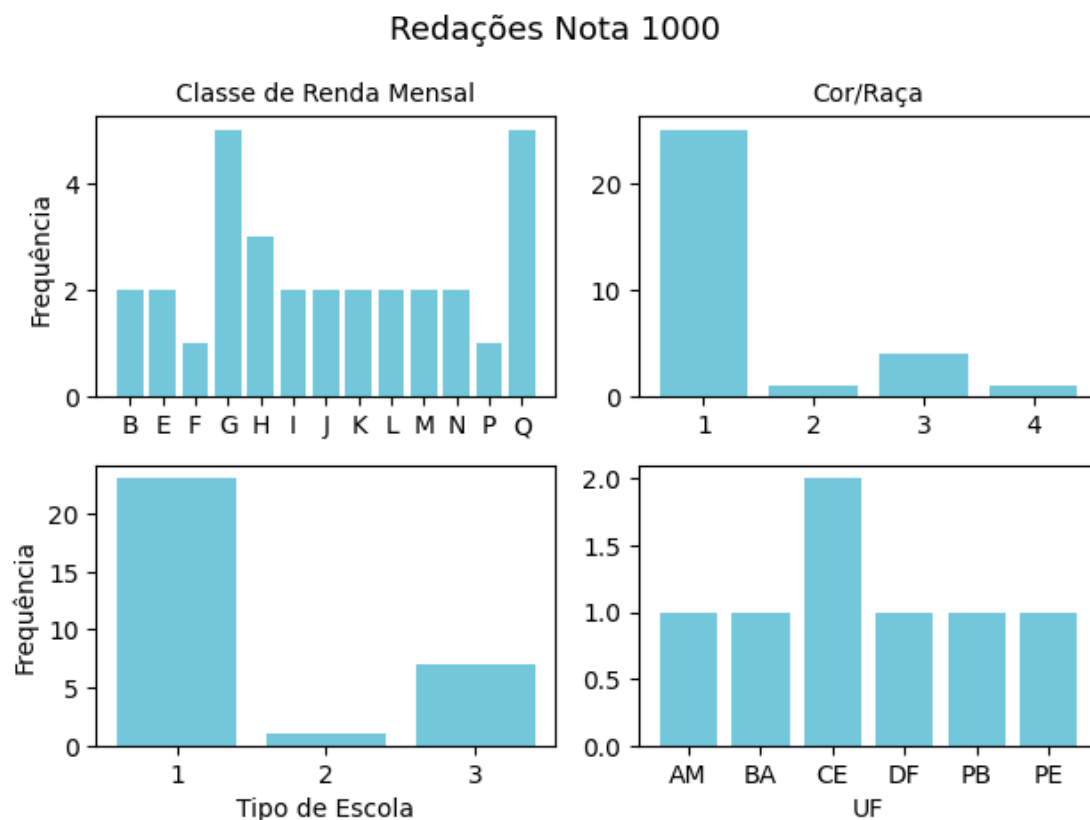


Figura 9. Relação entre redação nota 1000, cor/raça, renda, UF

Tabela 8 - Legenda - Figura 9

Legenda - Tipo de escola		
1	2	3
Não Respondeu	Pública	Privada

A figura 10, mostra a relação entre as notas médias das provas objetivas obtidas pelos participantes do ES de acordo com o tipo de escola.

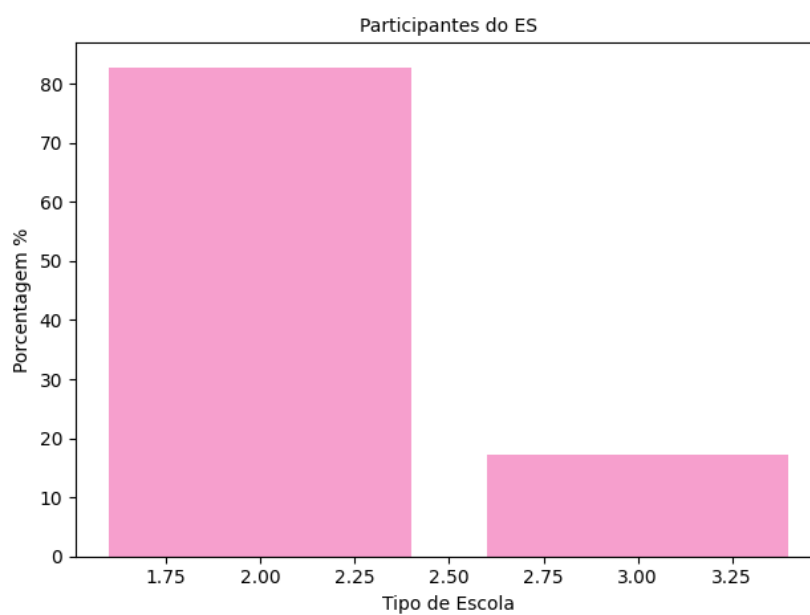


Figura 10. Média das notas no estado do ES de acordo com o tipo de escola

4. Metodologia

Neste trabalho utilizamos um estudo de caso, com dados providos de uma base de dados fornecida pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). As análises foram feitas de forma quantitativa. A base de dados fornecida pelo INEP abrange todos os alunos que tomaram parte no Exame Nacional do Ensino Médio (ENEM). Para este trabalho consideramos a média do aluno em cada uma das 4 seções da prova (matemática, língua portuguesa, ciências da natureza e ciências humanas) e a redação, a classificação do aluno perante as cotas (raciais, de escola pública e sociais), faixa etária e o estado de realização da prova.

A programação foi toda feita na linguagem Python, utilizando de bibliotecas como pandas, numpy e matplotlib. Toda a programação foi feita no software Google Colab e o upload foi feito para o repositório do github: https://github.com/pedroblq/P_enem.git.

Para a análise dos dados de 2022 direcionamos nossa atenção para as notas dos alunos e os possíveis limitantes presentes no contexto de cada um (tipo de educação, renda e sua cor de pele), para assim tentar localizar se existe alguma relação entre essas características com a média dos alunos, tanto nas questões objetivas quanto na redação.

5. Conclusão

Esse trabalho analisou os dados dos estudantes que prestaram o ENEM 2022, tendo como principal objetivo evidenciar as diferenças das notas dos alunos de diferentes etnias e classes sociais. Além disso, também foi avaliado o tipo de escola que o candidato estudou e sua idade ao fazer a prova.

Em primeiro momento já se torna evidente a necessidade das cotas raciais e sociais, uma vez que as maiores notas pertencem aos alunos que têm a maior renda, e quando analisamos por etnias vemos que os de pele branca tem destaque nas notas também.

Quando tratamos sobre os tipos de escola que os estudantes frequentaram, as particulares e federais tomam a liderança, o que em parte evidencia ainda mais a análise que liga a renda com o desempenho (no caso das particulares), e ressalta o grande valor da educação de qualidade das federais brasileiras. Outro fator a ser ressaltado é como a maior diferença das notas vem na prova de matemática, mostrando que é uma área onde o ensino público brasileiro deixa a desejar.

Ainda mais podemos perceber que a maioria dos candidatos estão na faixa de 17-19 anos, o que é esperado sendo essa a idade comum para os alunos terminarem o ensino médio.

Assim nossa hipótese que os grupos mais privilegiados teriam as maiores notas se concretiza, uma vez que os alunos brancos, com mais dinheiro e muitas vezes de escolas particulares ocupam as maiores notas.