

**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**

Pré-projeto

**VITÓRIA**

2023

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

**Pré-projeto 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_ESCOLA	Tipo de escola do Ensino Médio	Inteiro	1 a 3
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. Possíveis análises**

#### **Funções feitas para fazer a análise**

- Relação entre renda mensal e média de nota;
- Relação entre média de nota e tipo de escola (privada, federal, estadual e municipal);
- Relação entre tipo de escola e cor/raça;
- Relação entre média de nota e cor/raça;
- Relação entre ES, % de participantes de cor/raça, tipo de escola e média de nota;
- Faixa etária dos participantes;
- Relação entre faixa etária e média de nota;
- Participantes do sexo feminino e masculino;
- Evolução da nota média durante os anos (redação, CH, CN, LC, MAT);
- Média da nota por estados;
- Distribuição das notas da redação;
- Relação entre redação nota 1000, cor/raça, renda, UF;
- Em média qual das competências tem as menores notas.

### **4. Conclusão**

Portanto, referente aos dados analisados, [...]