

# Projeto Final Big Data Science - Enem 2021

Alunos: Cayo Eduardo da Silva Neves

## Dados

Utilizando os microdados do enem do ano de 2021, pelo link

[https://download.inep.gov.br/microdados/microdados\\_enem\\_2021.zip](https://download.inep.gov.br/microdados/microdados_enem_2021.zip)

## Introdução

Iniciei a pesquisa e análise exploratória dos dados do enem 2021 com a seguinte perspectiva. Identificar se os dados mostram alguma variação de notas das provas em relação a cor/raça dos participantes ou renda familiar. Sabemos na sociedade atual quão o abismo entre ricos e pobres crescem a cada dia e como esse abismo interfere na vida comum, seja em questão de acesso a educação, oportunidades, qualidade de ensino. Não é novidade que classes com rendas mais altas possuem diversas vantagens daqueles com rendas menores e muitas vezes sem renda, e também como a cor e raça de uma pessoa, mesmo em um país onde muitos dizem que não existem preconceito e desigualdade, interferem nos resultados de um exame prestado para os alunos terem acesso ao ensino superior.

Com essas informações já em mente, e como elas não são novidade para ninguém, eu decidi ver se os resultados do enem, com base no dataset de 2021, demonstrava de alguma forma essas diferenças.

E como já previsto e analisado por outras pesquisas, vemos em todos os resultados, como pessoas de classes mais privilegiadas e de cor/raça mais privilegiada também, tal como a branca e a parda.

Nessa pesquisa eu listei algumas respostas que eu gostaria de obter para nortear o estudo, pois como estudado em diversos cursos de análise de big data e data science, tudo começa com o problema a ser resolvido, pois dependendo do problema, toda a estrutura dos dados e metodologia devem ser mudados. E as perguntas foram a seguinte:

- 1) Cor e renda possuem alguma correlação?
- 2) Qual a relação da renda e cor para os alunos que não concluíram a prova ?
- 3) As maiores notas pertencem a que classe de renda ?
- 4) As menores notas pertencem a que classe de renda ?
- 5) As maiores notas pertencem a que classe de cor ?
- 6) As menores notas pertencem a que classe de cor ?
- 7) Qual a relação de homens e mulheres que se inscreveram na prova e que desistiram ? A porcentagem de desistência em relação a inscrição é maior em qual sexo ?

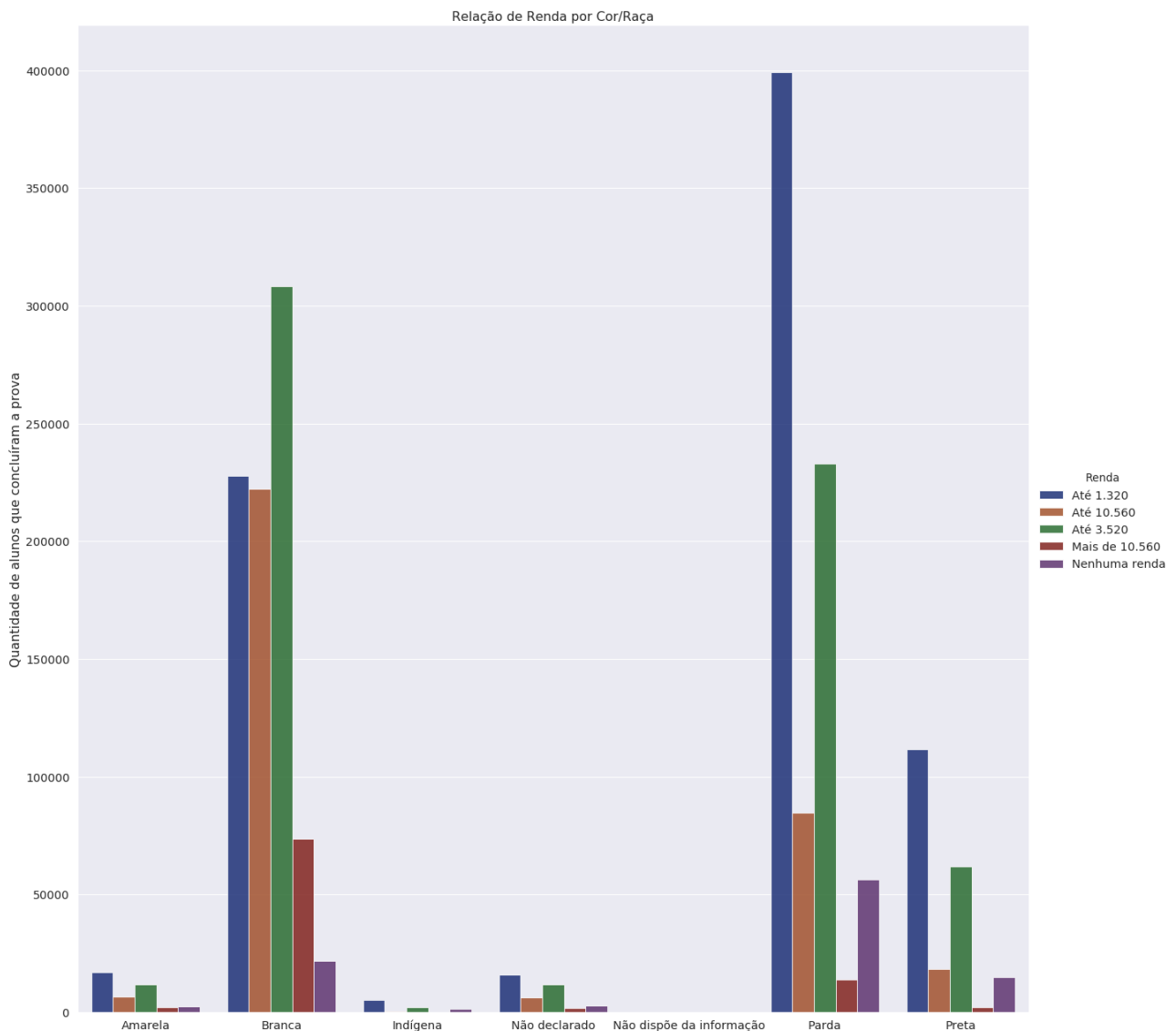
Por ultimo também inclui algumas questão a respeito dos registros faltosos no dataset. Quando um aluno tira zero em alguma área da prova, ele recebe 0, mas como foi identificado no dataset alguns registros possuem a nota como ausente, ou seja, o aluno não estava presente na prova para entregar as respostas, e com isso eu decidi também analisar o impacto que a desistência tem sobre os alunos tanto a respeito de quais cores/raça possuem a maior taxa de desistência, como

também a renda desses alunos. Analisando que possivelmente alunos de baixa renda e classes mais desprivilegiadas desistiam mais das provas, tanto por questão de acesso aos locais de prova, como por outras condições.

A seguir os gráficos de cada uma das perguntas a ser respondidas e analisadas:

### 1) Cor e renda possuem alguma correlação?

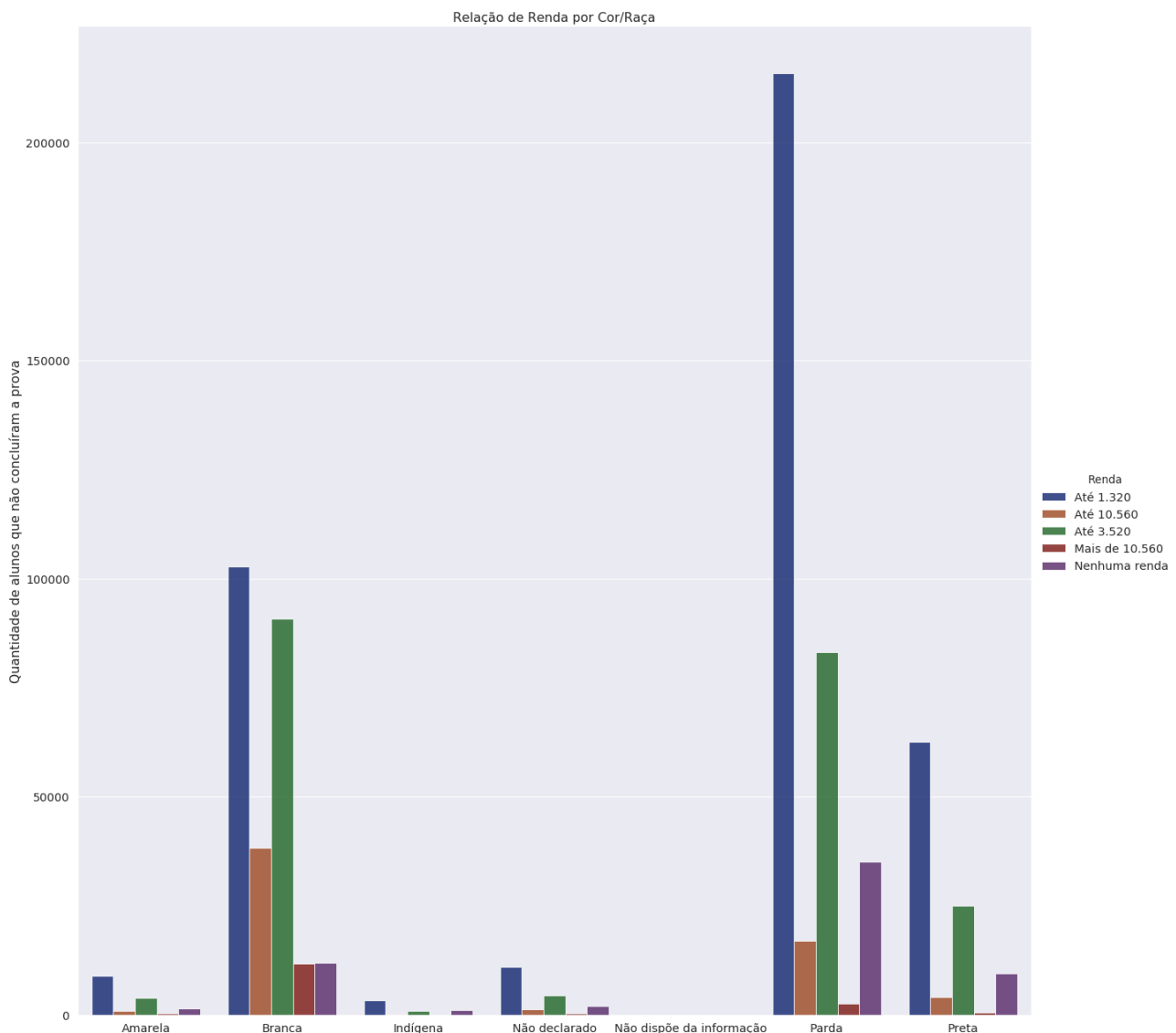
A maioria dos alunos que concluíram a prova estão nos grupo de raça/cor brancos e pardos, e consequentemente a quantidade de registros com maior renda está concentrada mais nesses duas classes.



### 2) Qual a relação da renda e cor para os alunos que não concluíram a prova ?

A redução de alunos autodeclarados brancos entre os alunos que não conseguiram de alguma forma concluir a prova foi grande, e o aumento em ambas as cores/raças de renda "Até R\$ 1.320,00"

comprova a desigualdade que famílias de renda alta e baixa possuem na hora de concluir a prova, mostrando também que pessoas autodeclaradas brancas, por mais que seja um número expressivo na amostra como um todo, teve uma redução entre os números de desistentes.

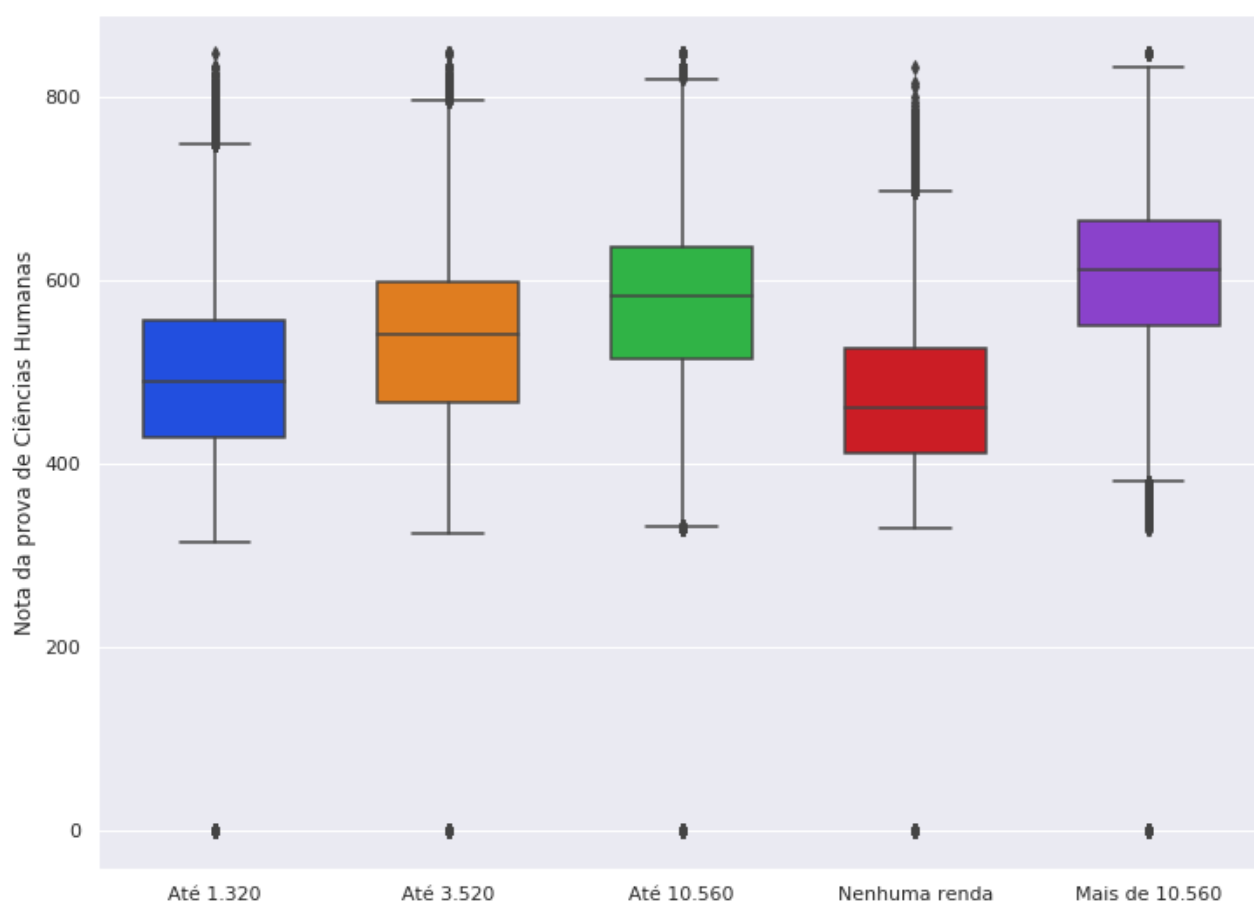
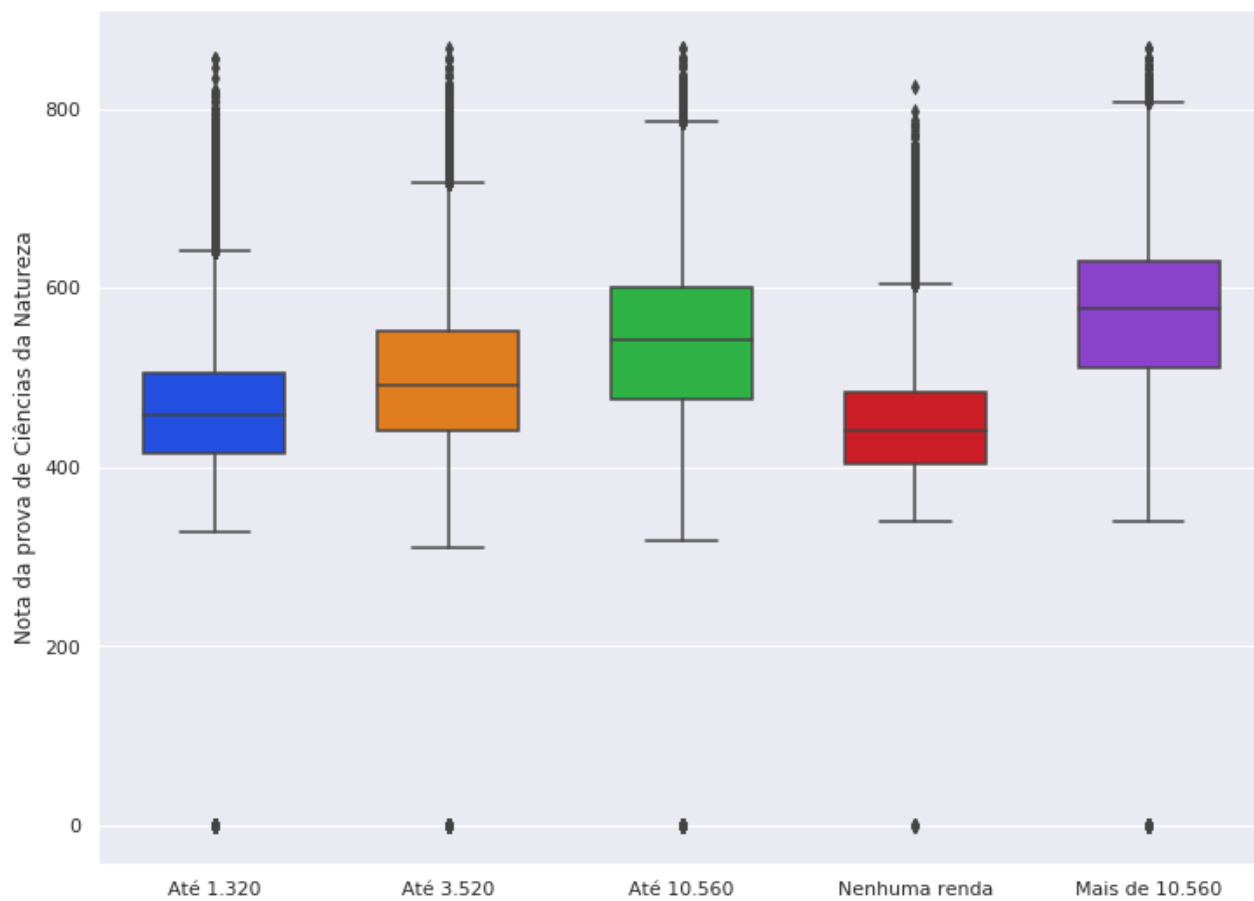


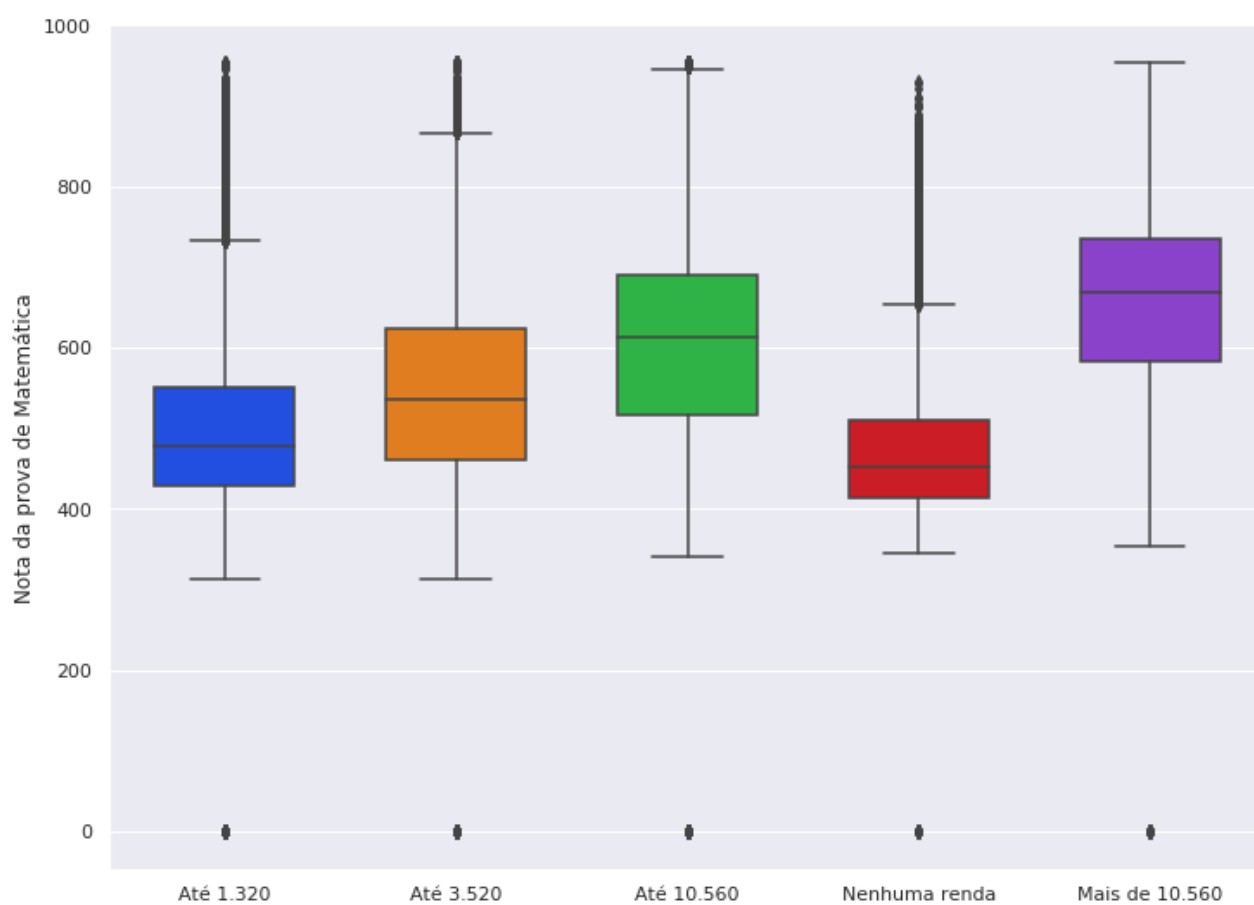
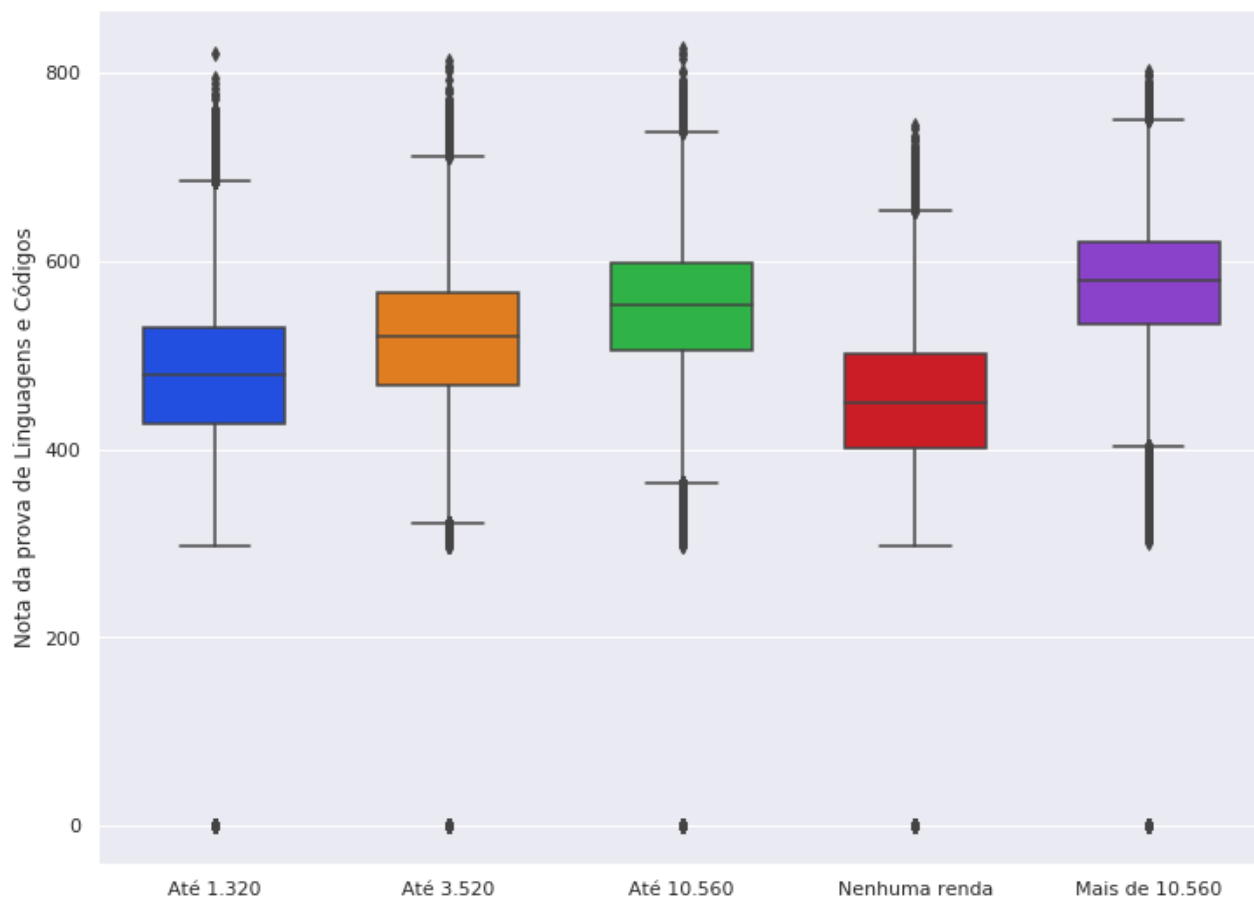
### 3) As maiores notas pertencem a que classe de renda ?

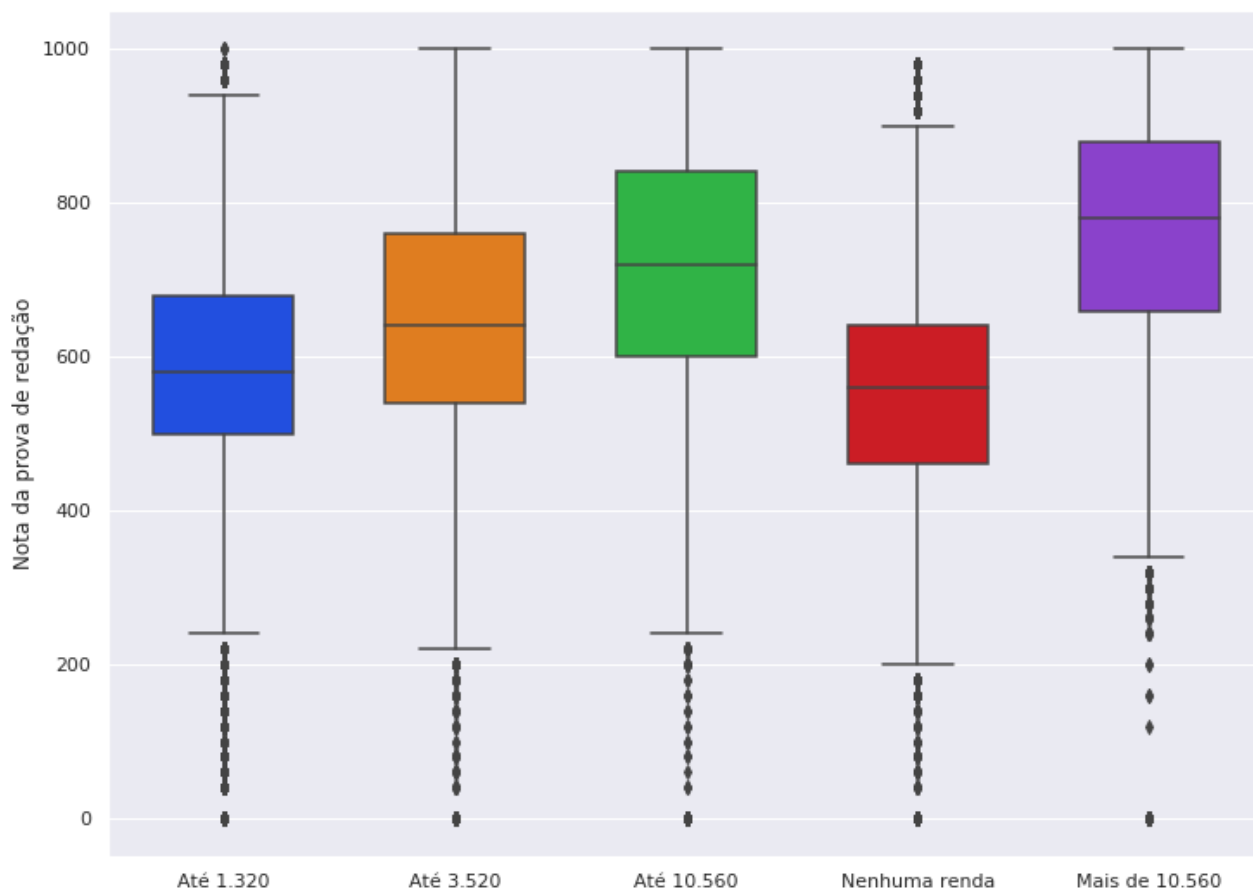
As maiores notas estão entre as duas classes de rendas mais altas, Até R\$ 10.560,00 e "Mais de R\$ 10.560,00".

### 4) As menores notas pertencem a que classe de renda ?

As menos notas estão entre as duas classes de rendas mais baixas, "Nenhuma renda" e "Até R\$ 1.320,00".





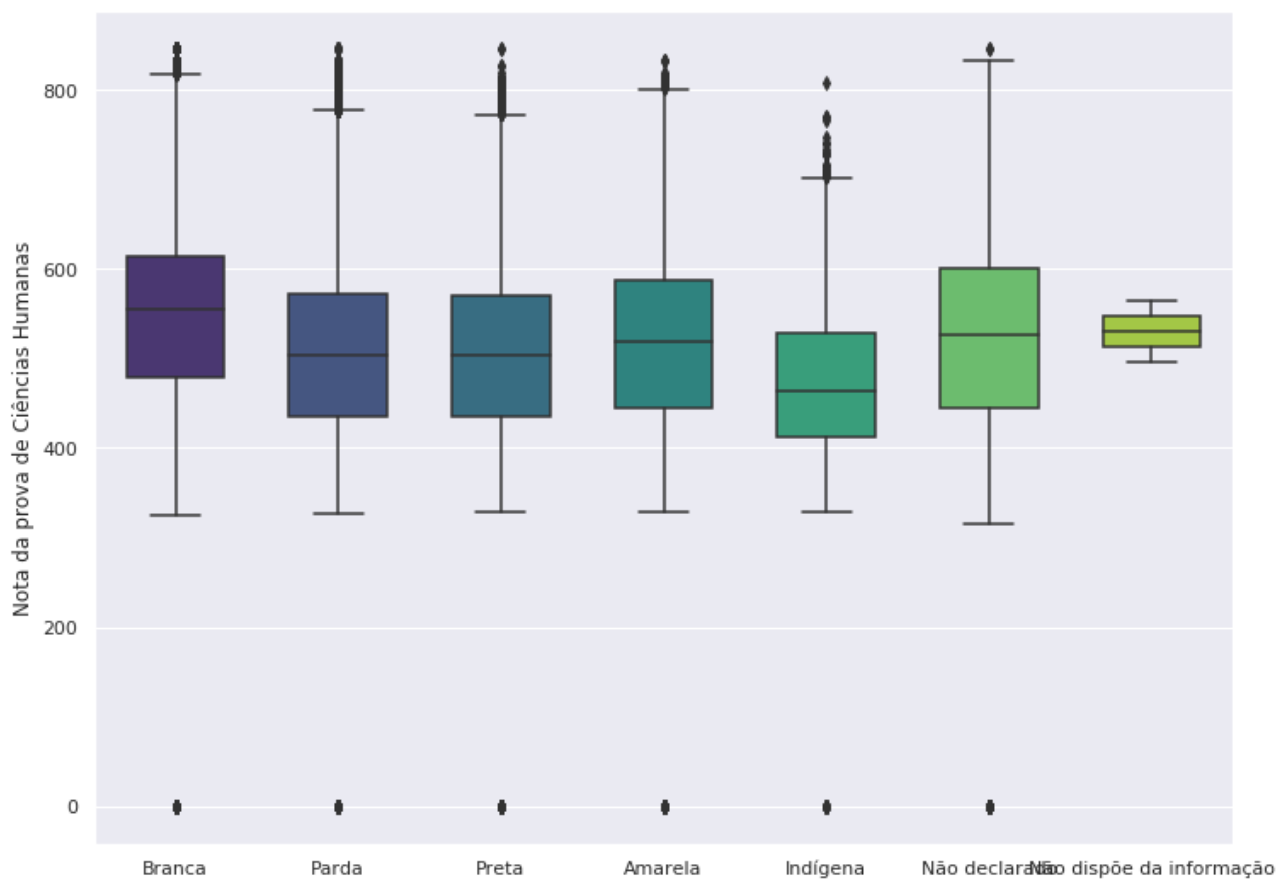
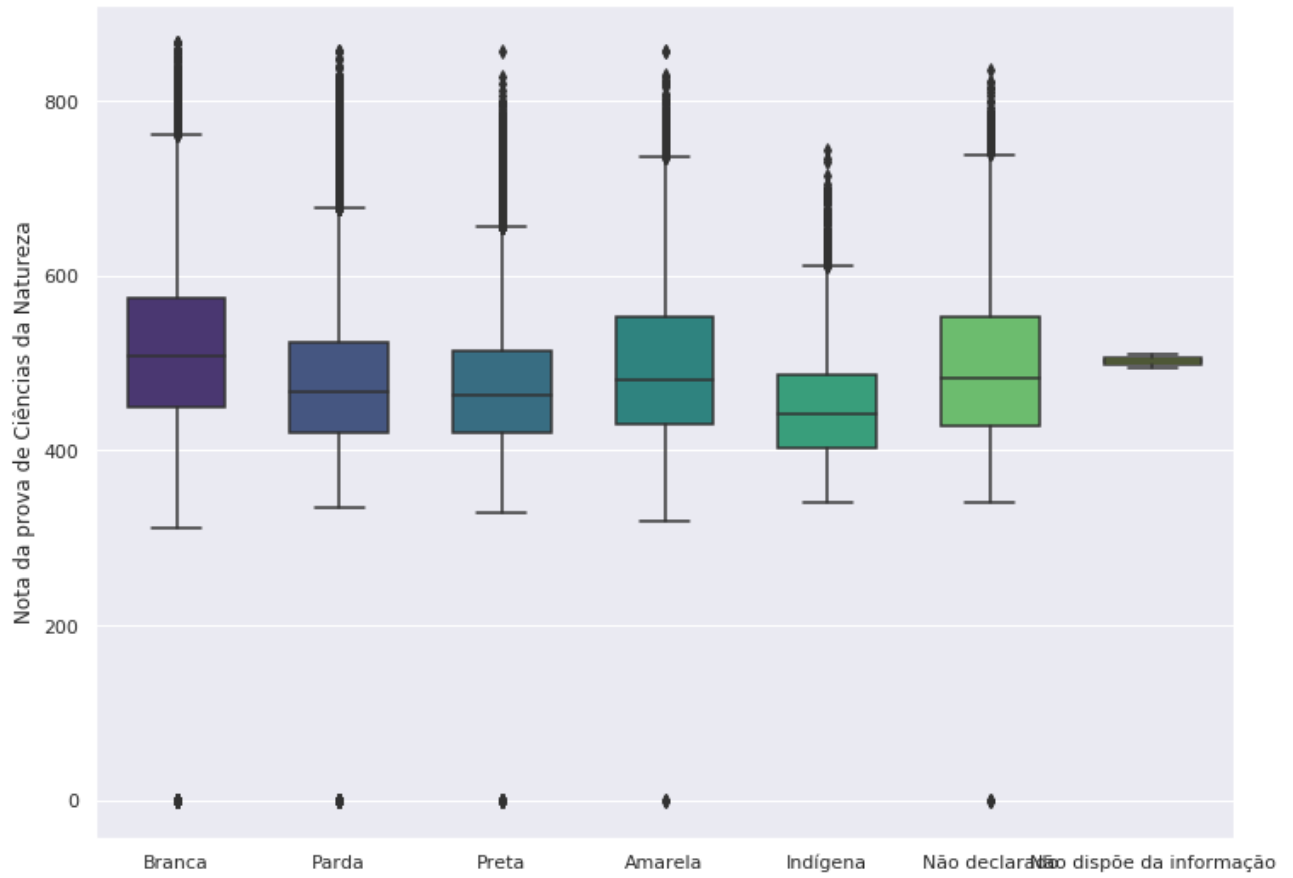


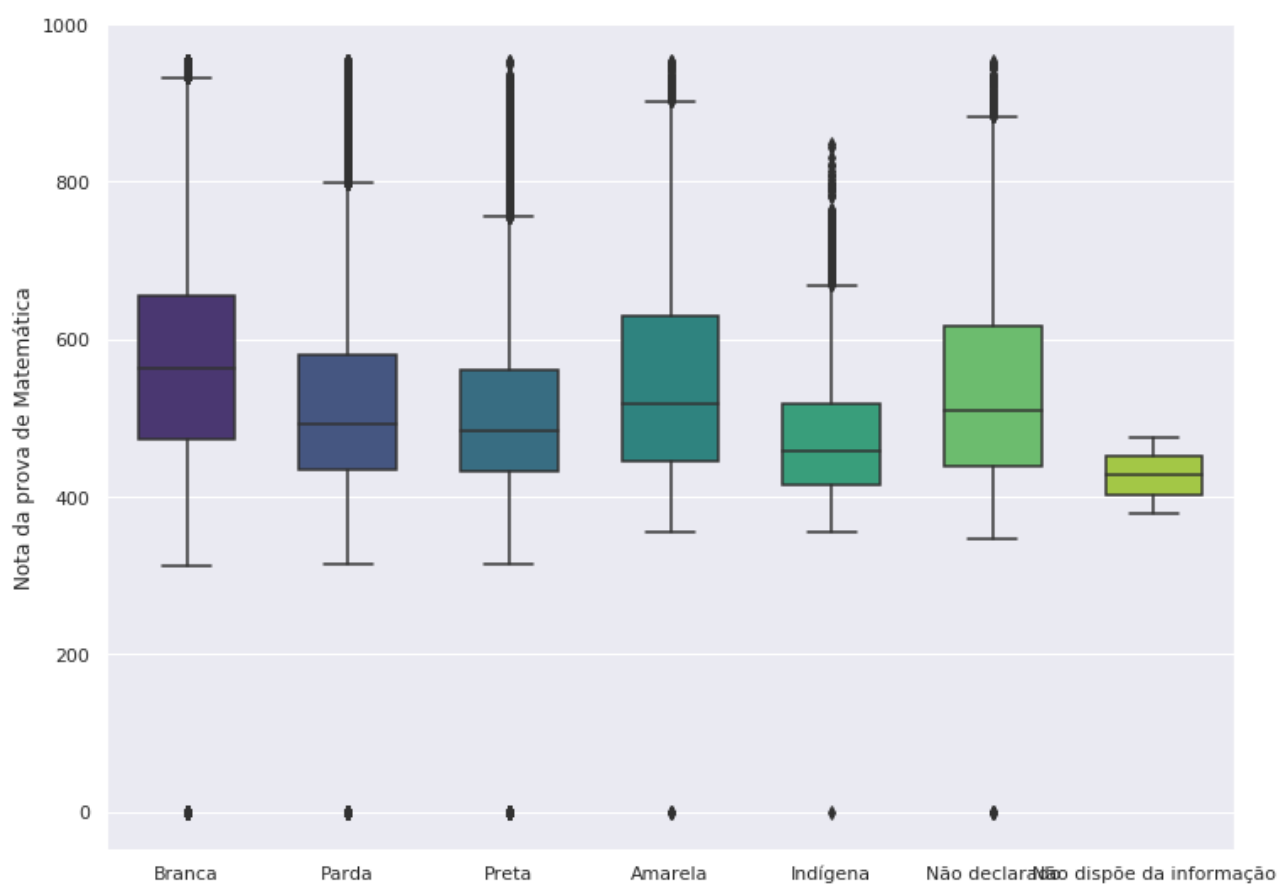
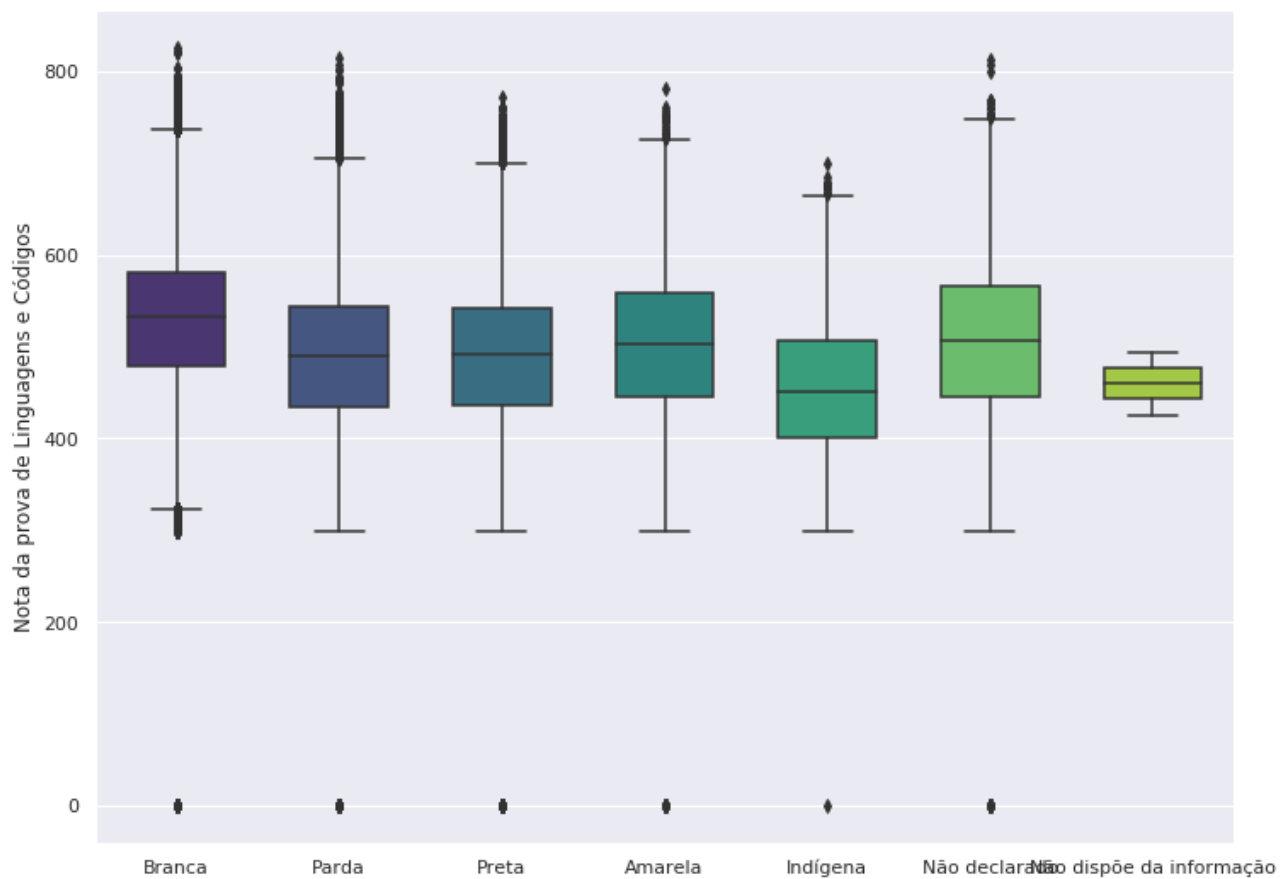
**5) As maiores notas pertencem a que classe de cor ?**

As maiores notas estão entre as classes de raça/cor "Branca" e "Amarela", desconsiderando os não declarados.

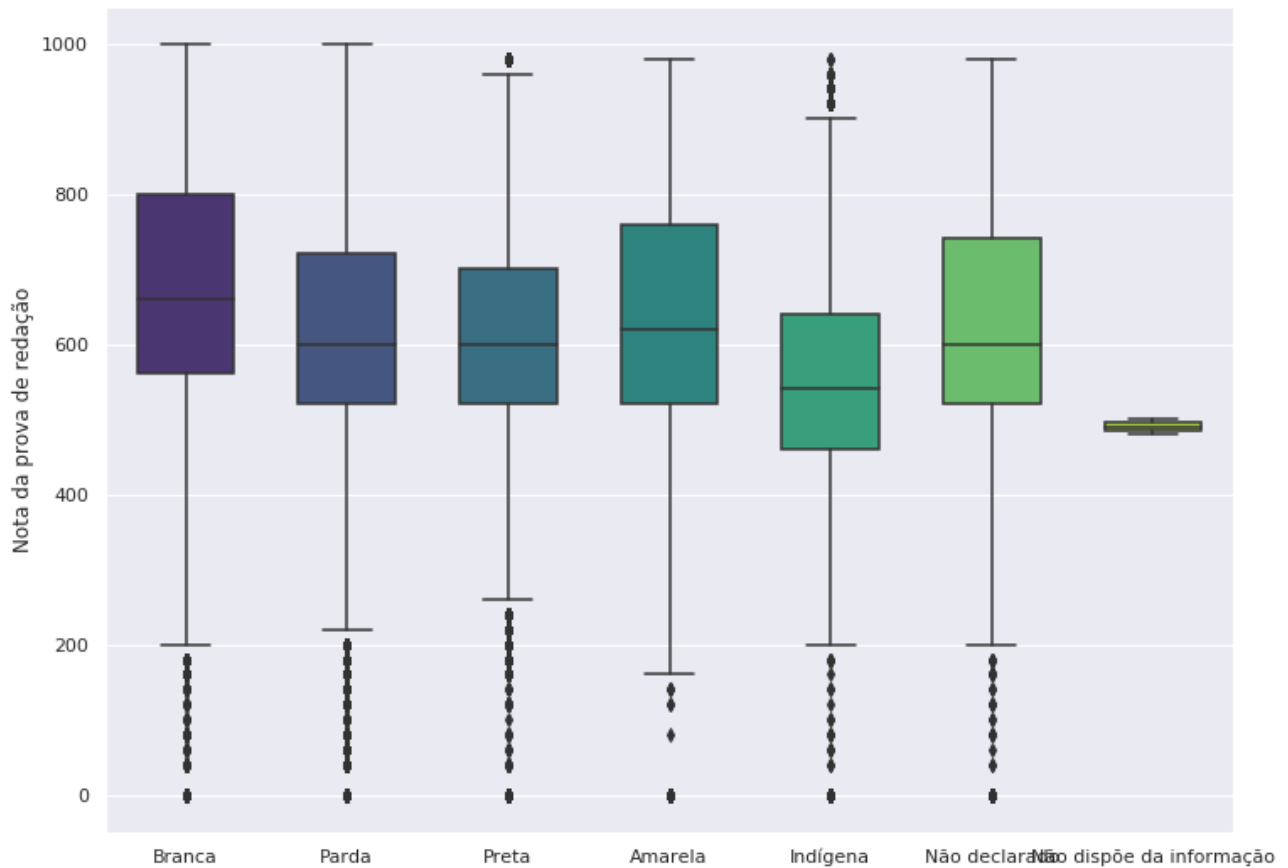
**6) As menores notas pertencem a que classe de cor ?**

As menores notas estão entre as classes de raça/cor "Indígena" e "Preto", desconsiderando os não declarados.



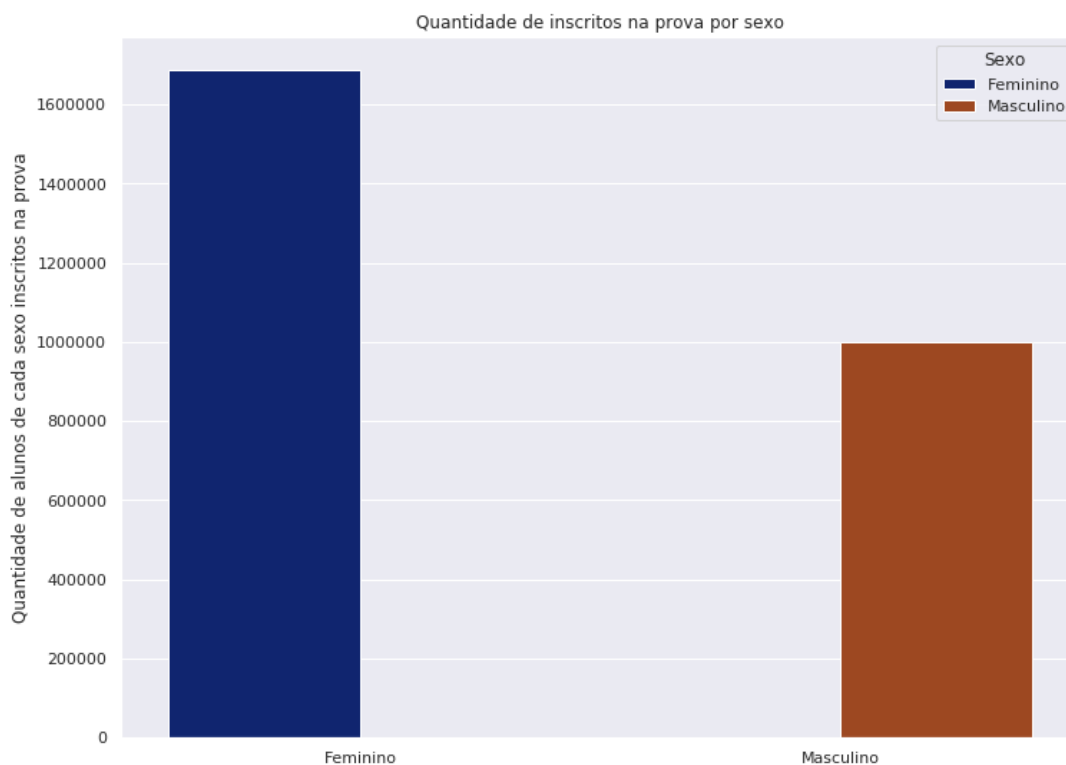


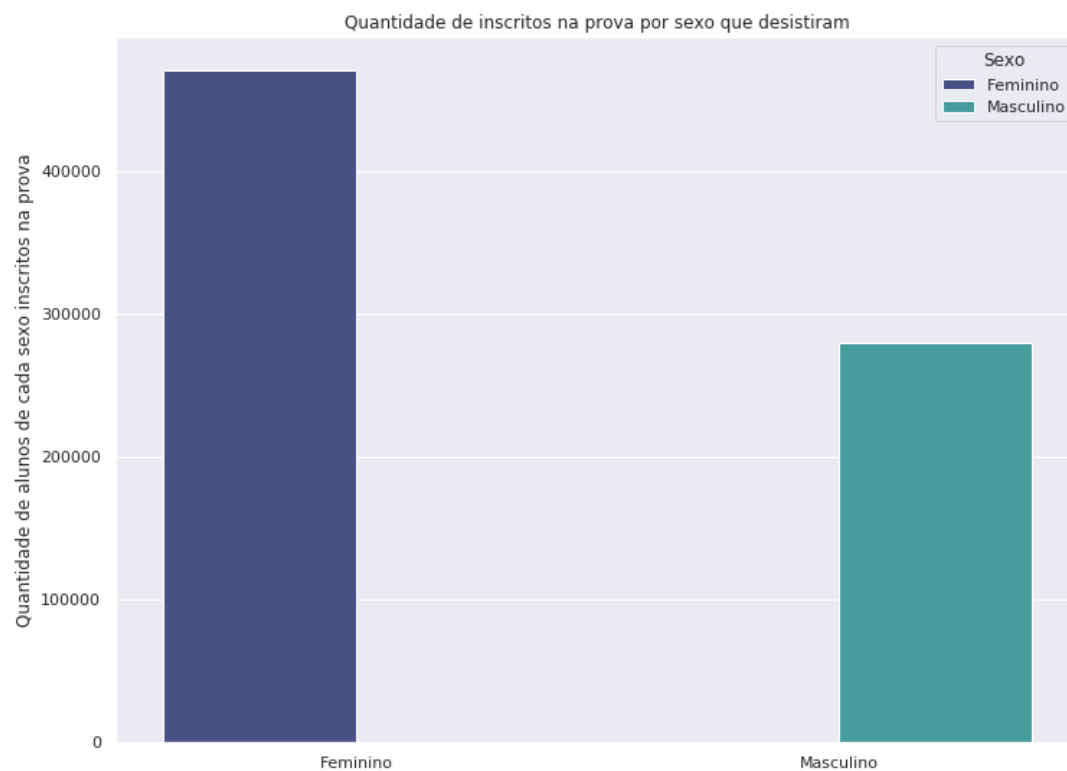




**7) Qual a relação de homens e mulheres que se inscreveram na prova e que desistiram ? A porcentagem de desistência em relação a inscrição é maior em qual sexo ?**

Tiveram um número bem maior de inscrições do sexo feminino, porem a taxa de desistência entre os sexos foi maior no sexo masculino.





**Porcentagem de desistentes no sexo masculino: 28.03%**

**Porcentagem de desistentes no sexo feminino: 27.90%**