

# Análise Exploratória - Data Education

**Tema:** Análises dos dados do ENEM em anos pré pandemia e durante a pandemia.

## **Membros:**

- Gabriel Vieira Rodrigues, 1759620, gabrielvr97, EC, UTFPR
- Maycom Deus, 2137089, maycom-data-science, BSI, UTFPR
- Rodrigo William Chavoni, 1996169, rodrigochavoni, BSI, UTFPR

## ● **Obtenção e processamento de dados**

**Fonte dos dados:** Estão sendo utilizados os dados disponibilizados pelo INEP (Instituto Nacional de Estudos e Pesquisas) referentes aos exames do ENEM dos anos de 2019 a 2021, que podem ser acessados através do link:

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

**Limpeza e organização:** após analisar o dicionário de dados do dataset, foram descartadas algumas colunas como número de inscrição, códigos de município, estado e colégio e gabaritos das questões.

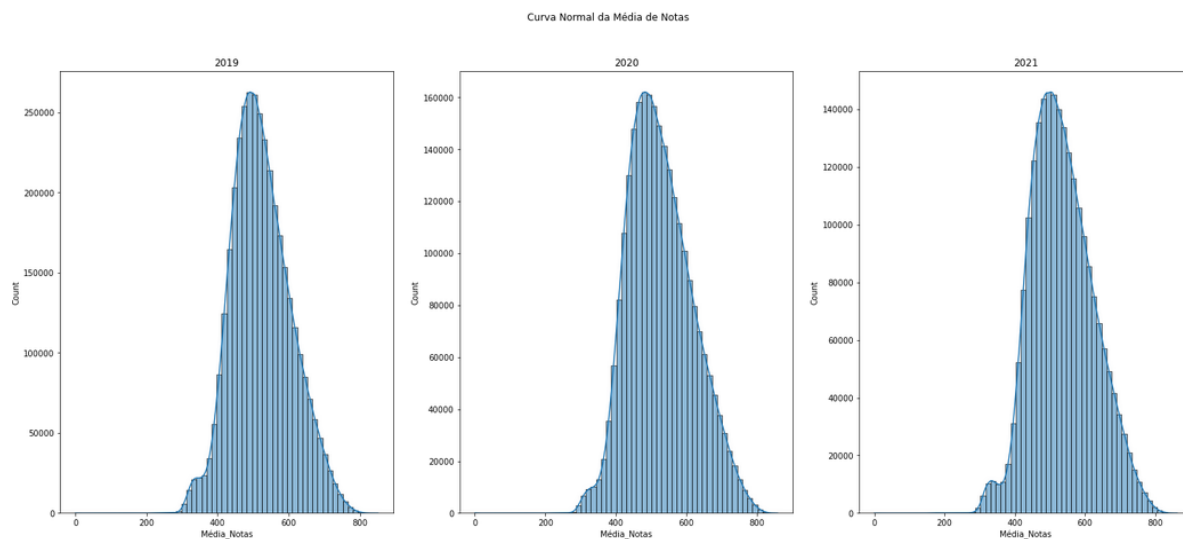
Após isso foram renomeadas algumas colunas referentes ao questionário socioeconômico, com o objetivo de facilitar as futuras análises. Por exemplo: a coluna "Q006" foi renomeada para "Renda\_Familiar".

O próximo passo foi separar os alunos em datasets separados: os de alunos presentes, os de alunos faltantes e os de alunos anulados.

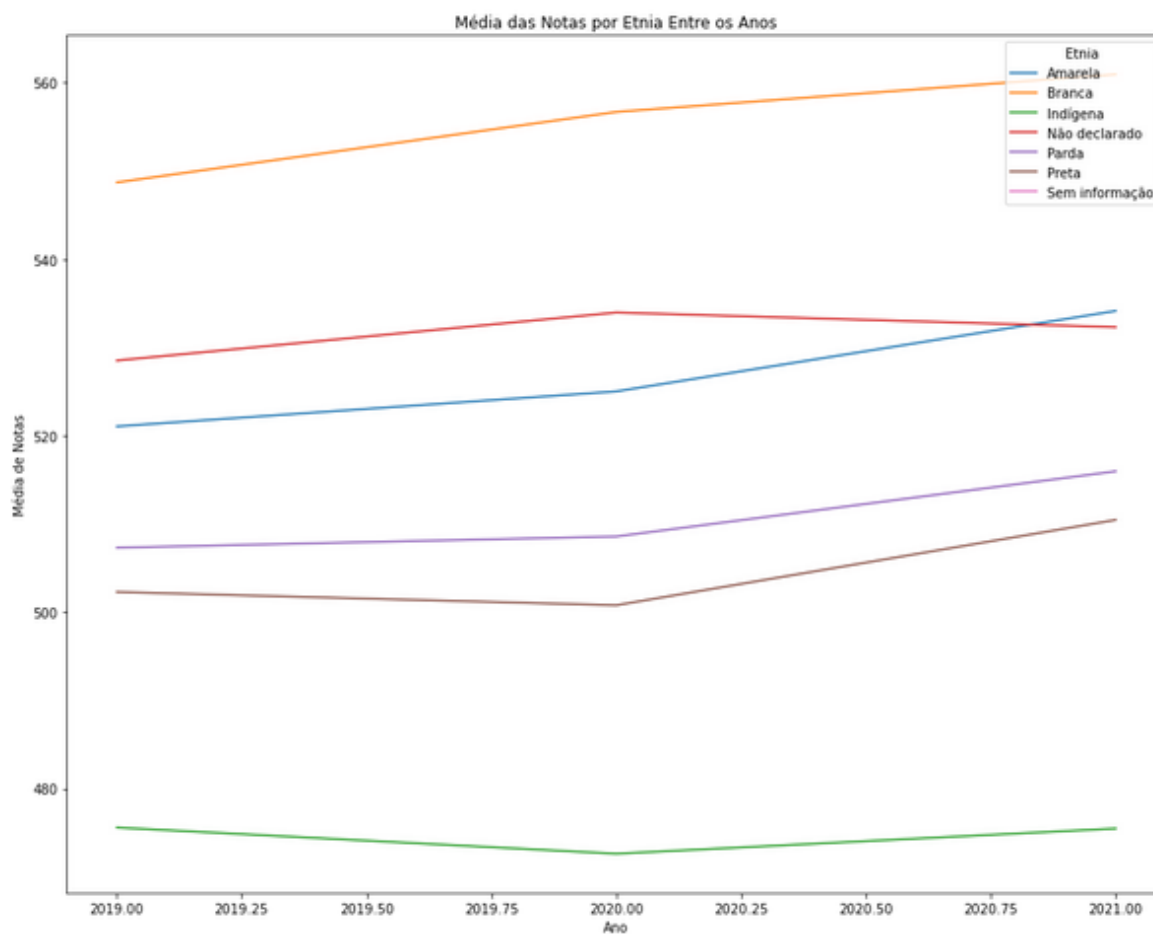
E por fim, foram mapeadas e adicionadas algumas colunas referentes a dados dos alunos como faixa etária, raça e tipo de escola que estudou.

## ● **Cobertura e distribuição dos dados**

Para os dados dos alunos presentes, foi verificada a distribuição dos dados das notas nos anos de 2019 a 2021. Com o gráfico abaixo é possível notar, sutilmente, que a dispersão da média das notas vem aumentando a cada ano.



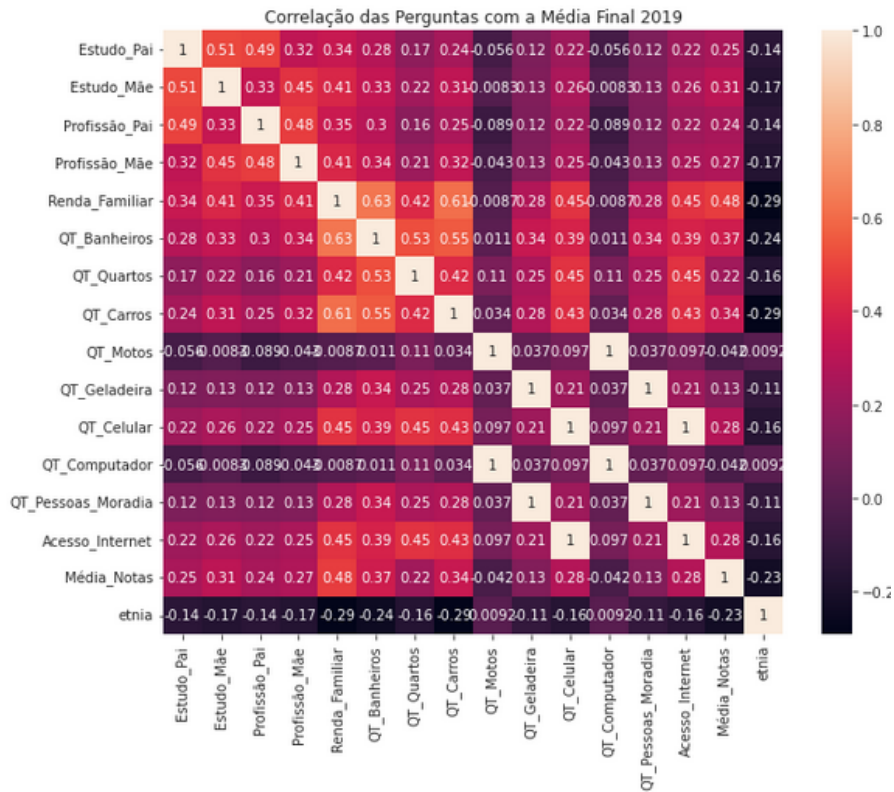
Além disso, é possível perceber um aumento nas notas acima da média. Com o gráfico de série temporal dos 3 anos relacionando a média de notas com a etnia, é visível a melhora da média de notas. Embora algumas etnias tenham reduzido a média em 2020, em 2021 a melhora foi significativa.



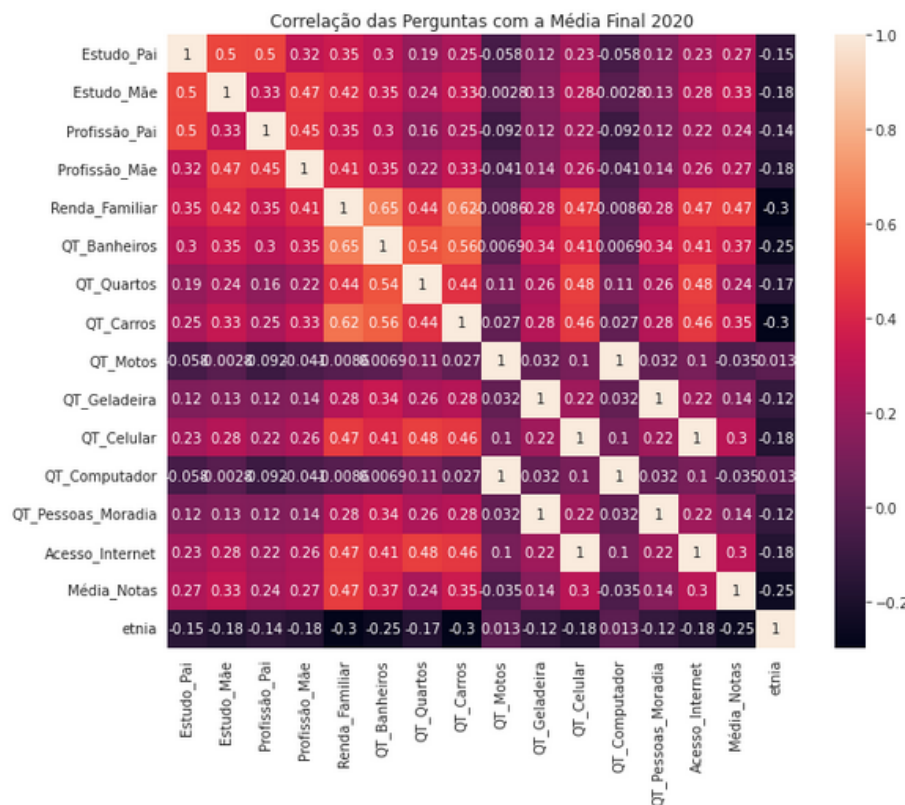
- **Identificação de Padrões Iniciais**

Com um mapa de calor é possível entender a correlação entre as perguntas socioeconômicas e as médias de nota.

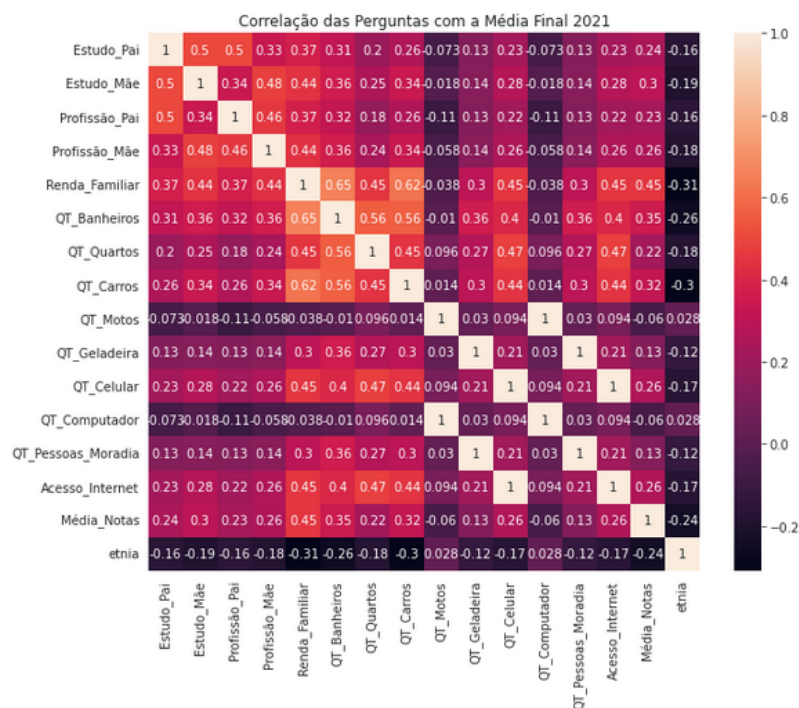
2019:



2020:



2021:

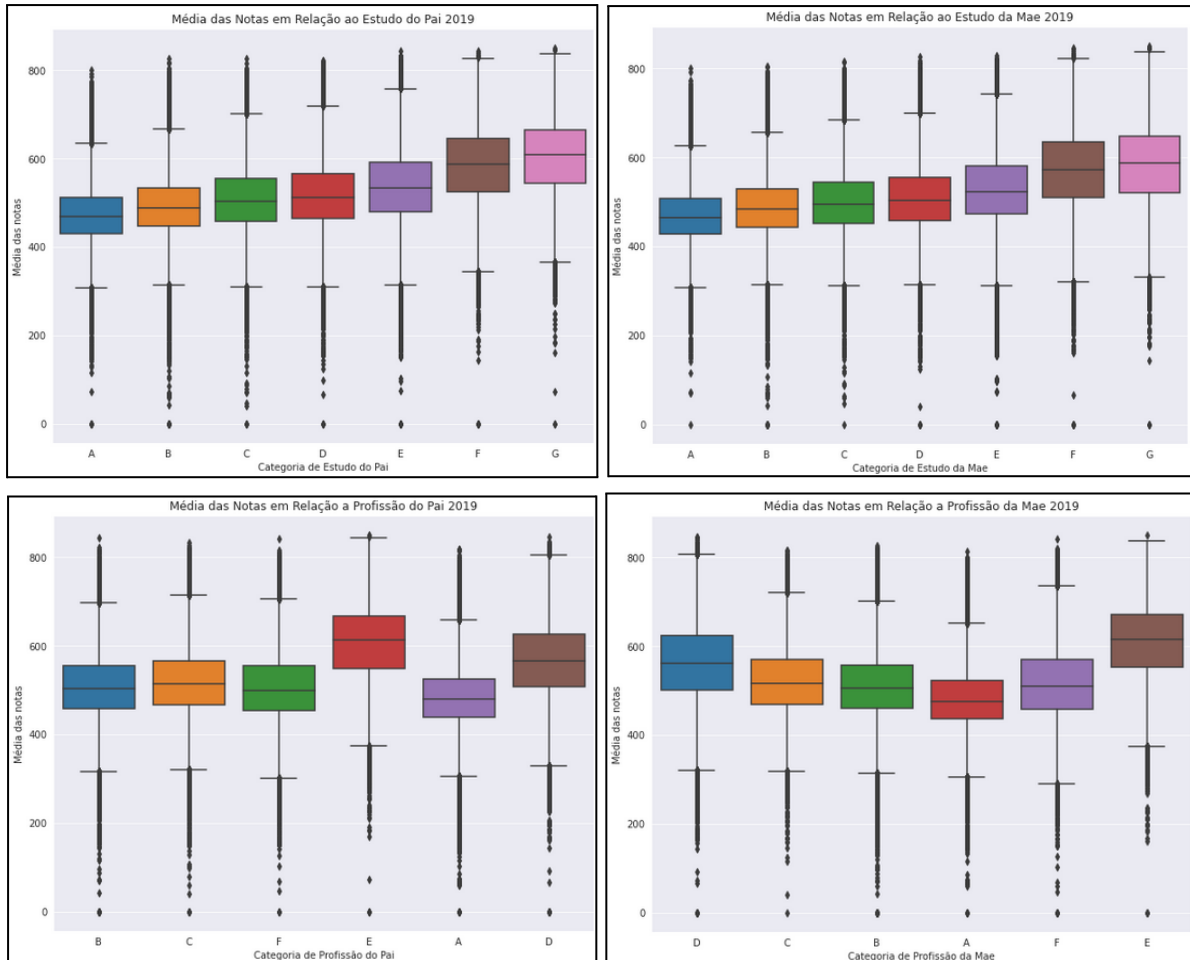


Os gráficos mostram que algumas perguntas socioeconômicas têm correlação significativa com as notas, principalmente Renda Familiar.

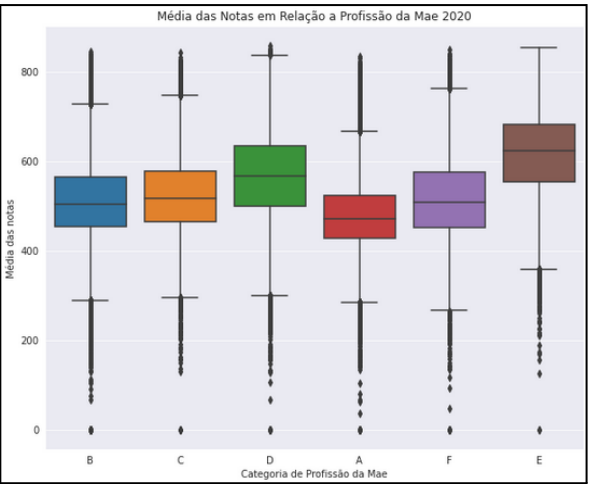
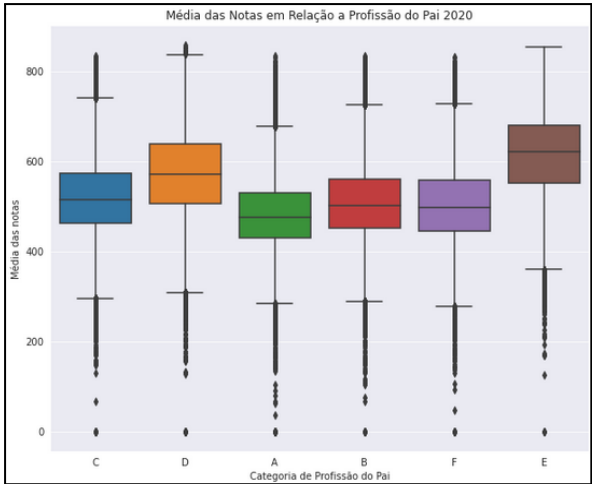
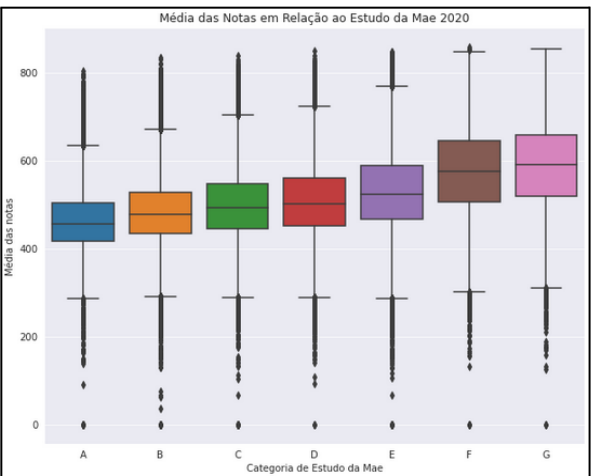
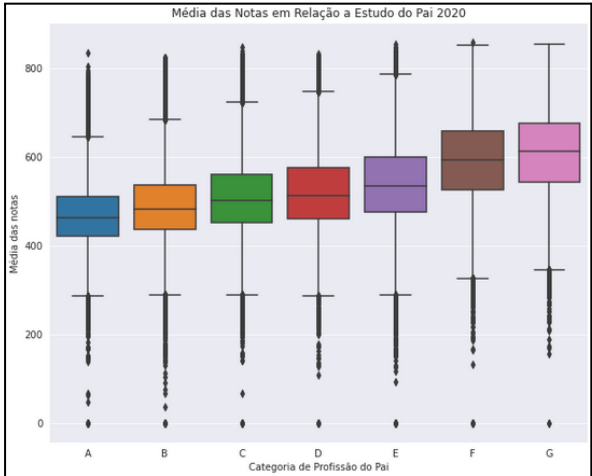
- **Perguntas de pesquisa e explorações iniciais**

As perguntas socioeconômicas, quando relacionadas com as notas, foi possível perceber alguns padrões. No caso, quando relacionamos as médias de notas dos alunos por categoria de ensino e profissão dos pais, conseguimos notar o aumento das médias conforme as categorias eram mais qualificadas (sendo a menos qualificada a categoria 'A' e vai crescendo de acordo com o alfabeto).

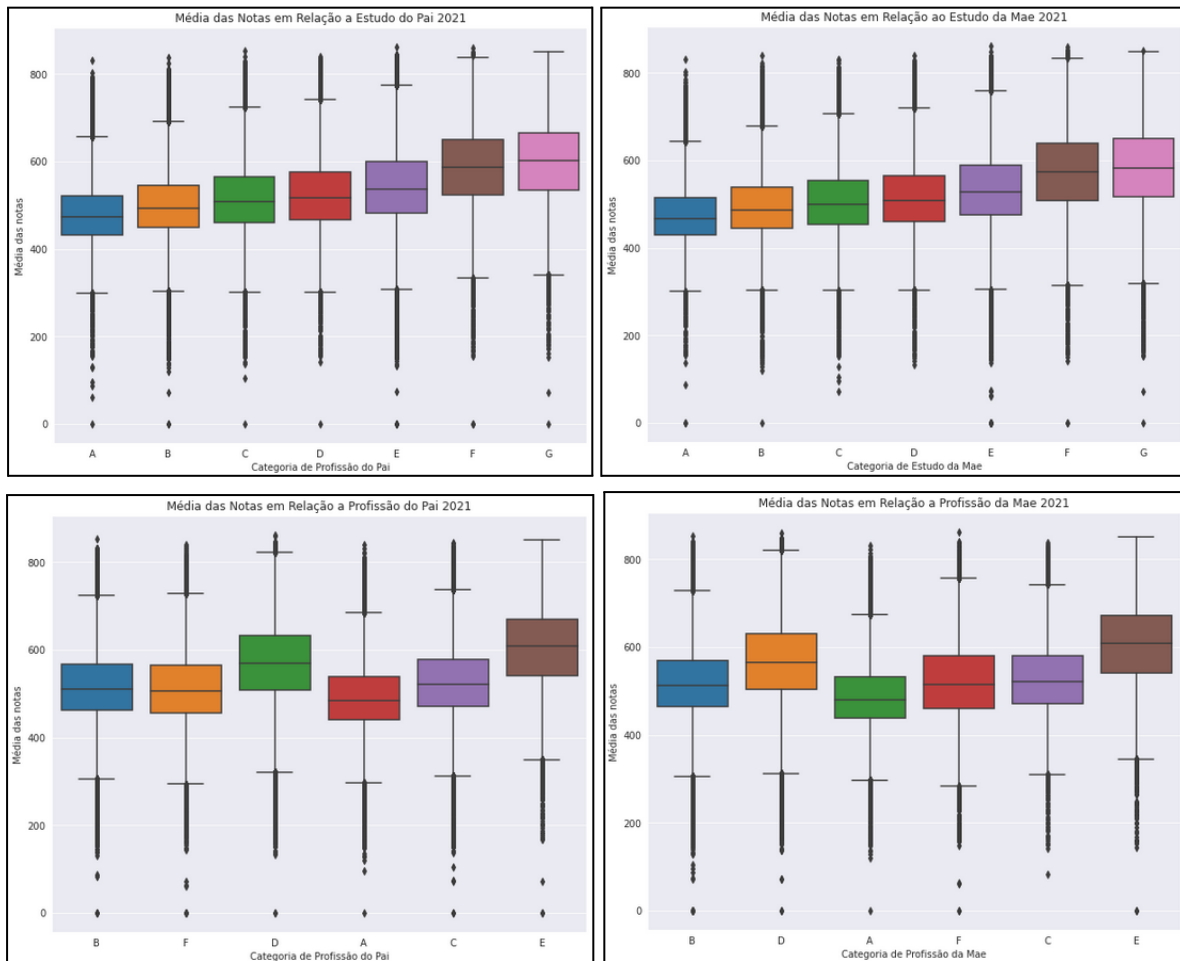
2019:



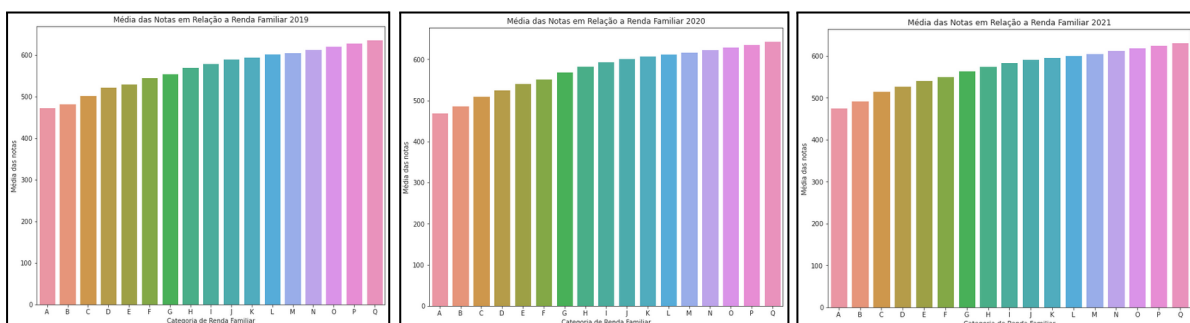
2020:



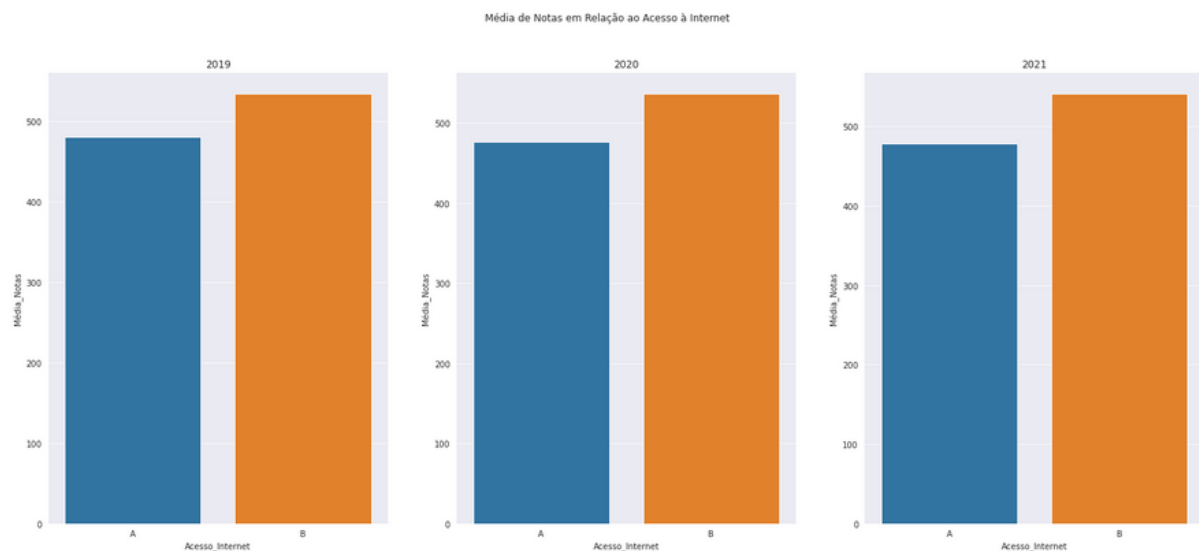
2021:



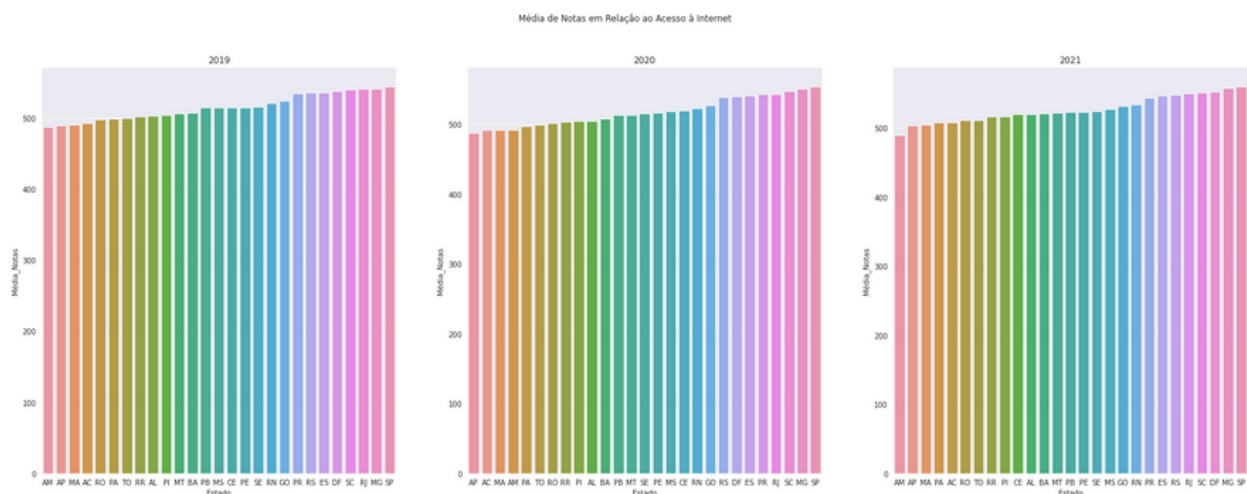
Além disso foi possível notar que a média de notas aumenta conforme a renda familiar da família (sendo a menor renda a categoria 'A' e vai crescendo de acordo com o alfabeto), em todos os anos, com o gráfico de barras abaixo:



Outro fator verificado foi o acesso à internet ou não dos alunos, e no gráfico abaixo conseguimos visualizar que alunos que têm acesso à internet (B) tem uma média de notas melhor.

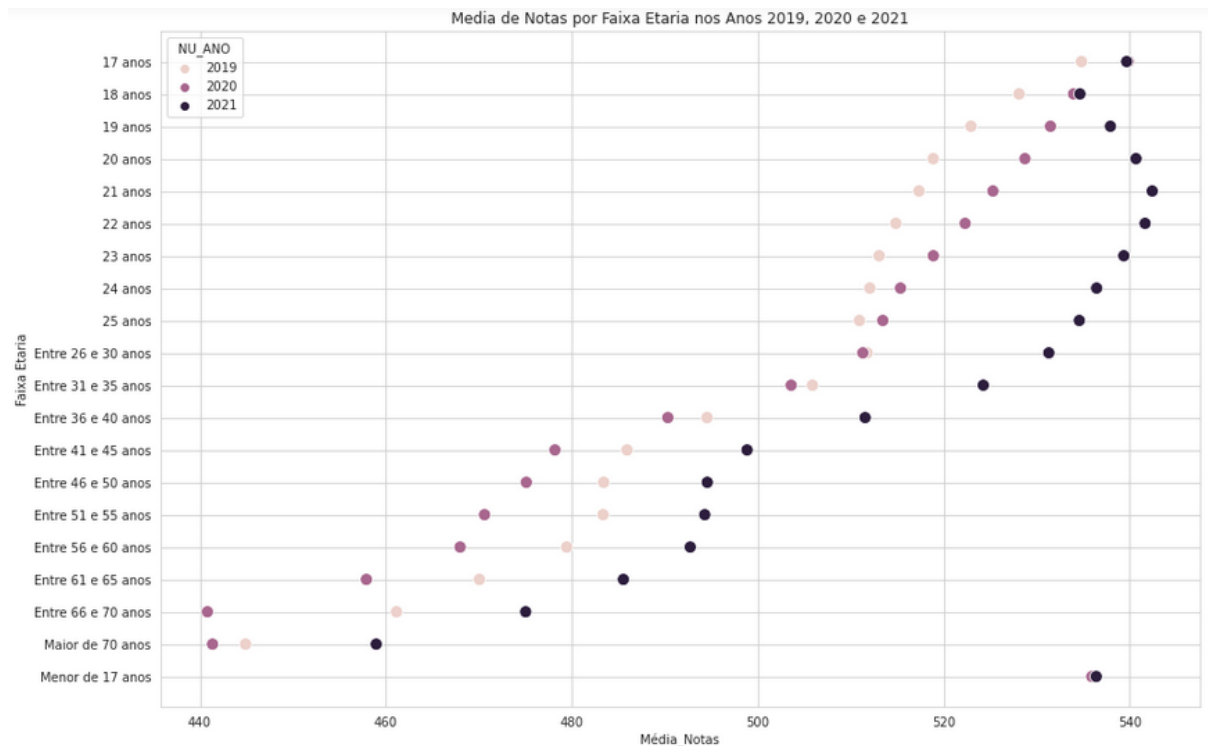


Olhamos também para as médias de notas de cada estado, onde foi constatado que as maiores médias pertencem aos estados mais desenvolvidos, nos 3 anos analisados. Entre os melhores estão os estados mais do sul, já entre os piores médias estão os estados do norte.



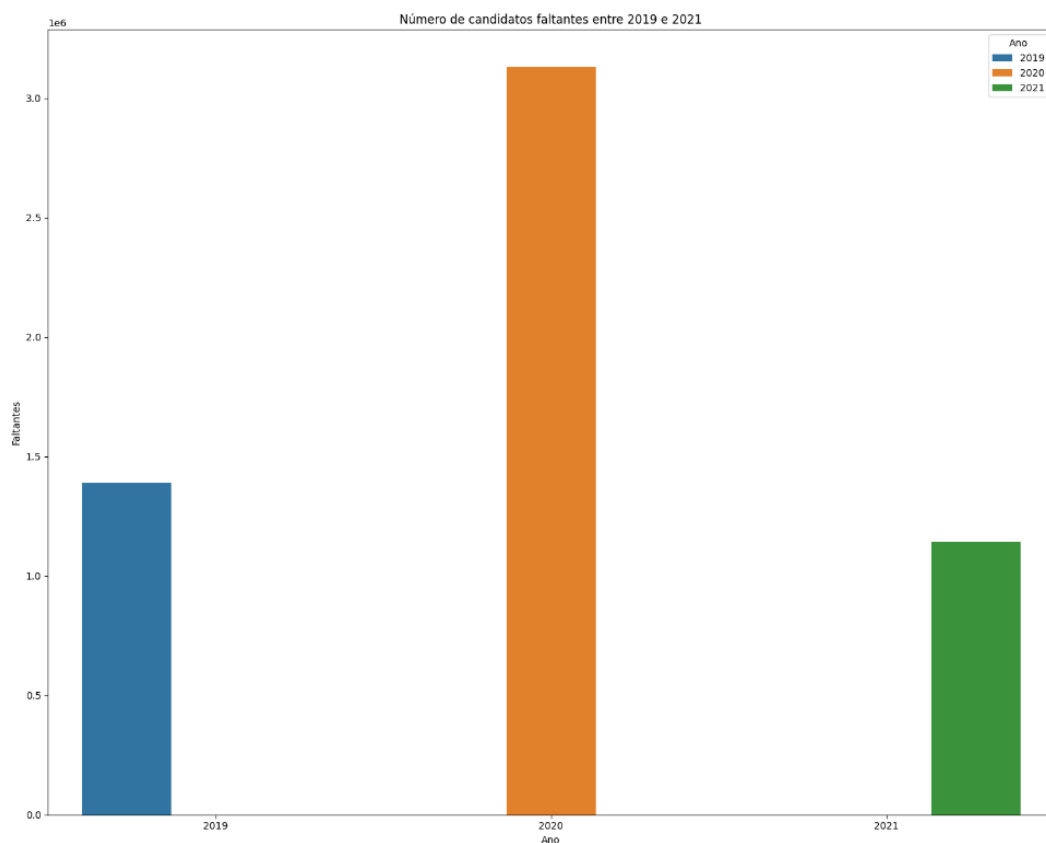
Através de um gráfico de pontos, conseguimos perceber como as médias de notas vem aumentando ao longo dos 3 anos em todas as faixas etárias. Com isso, é possível ter uma noção se a pandemia, que trouxe o ensino remoto, afetou ou não as média de notas.





Agora para analisar o caso de alunos que se inscreveram para o exame mas não compareceram, pode-se levantar algumas questões e hipóteses:

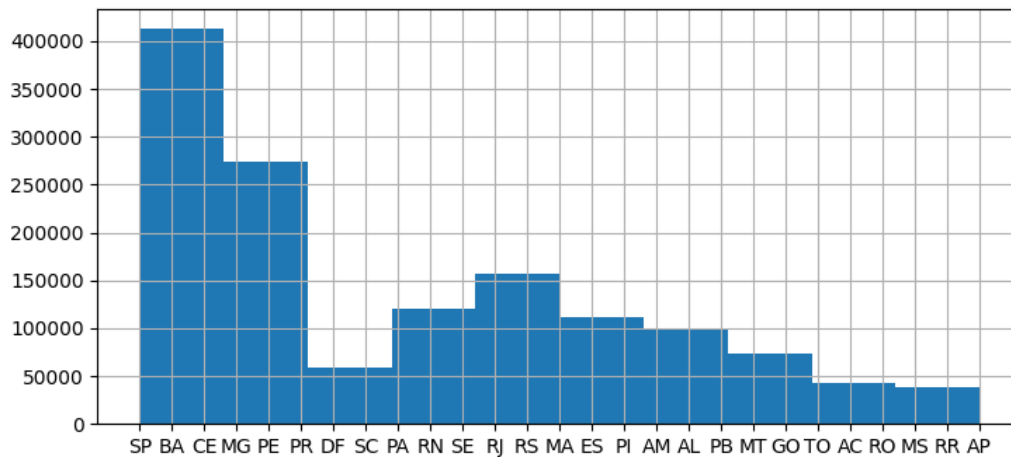
- 1) Teve um aumento de candidatos faltantes nos anos de 2020 e 2021 (anos de pandemia) em relação ao ano de 2019?



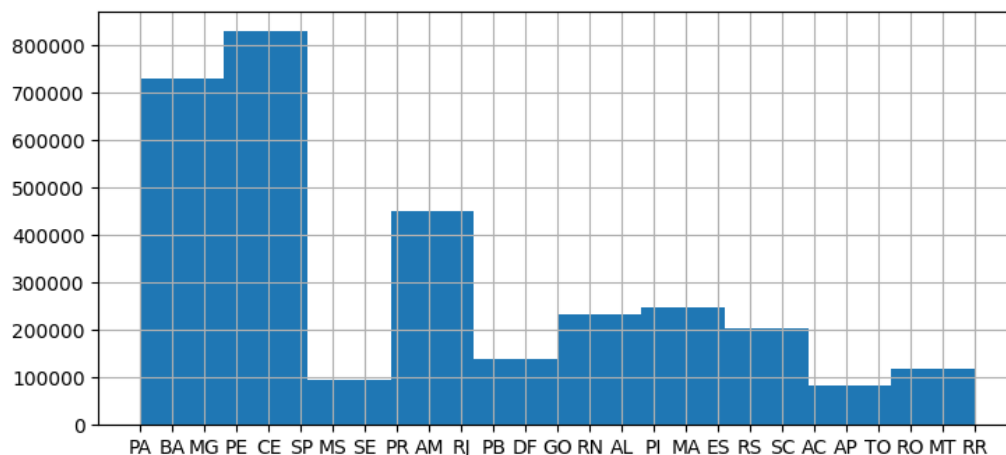
Pode-se observar que houve um aumento relevante de candidatos faltantes no ano de 2020 (primeiro ano da pandemia). Isso pode ter relação com fatores como um maior número de casos, *lockdowns* mais severos pelo país e ainda inexistência das vacinas.

- 2) Quais os estados que concentram o maior número de candidatos faltantes?

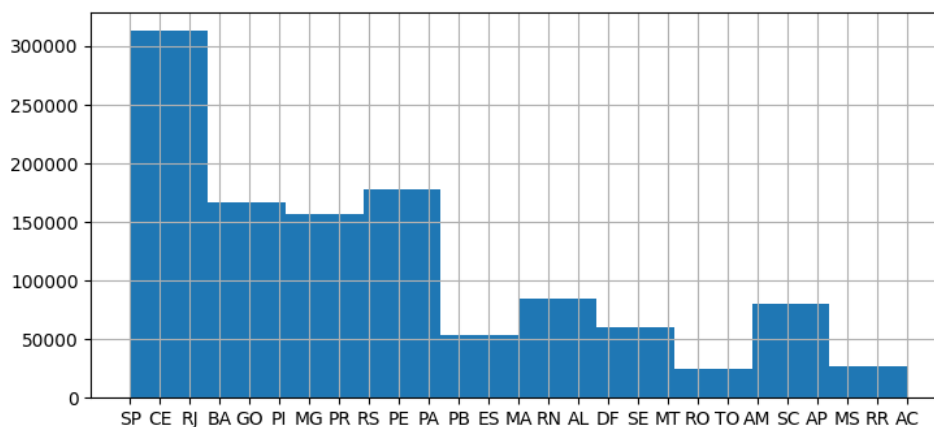
2019:



2020:



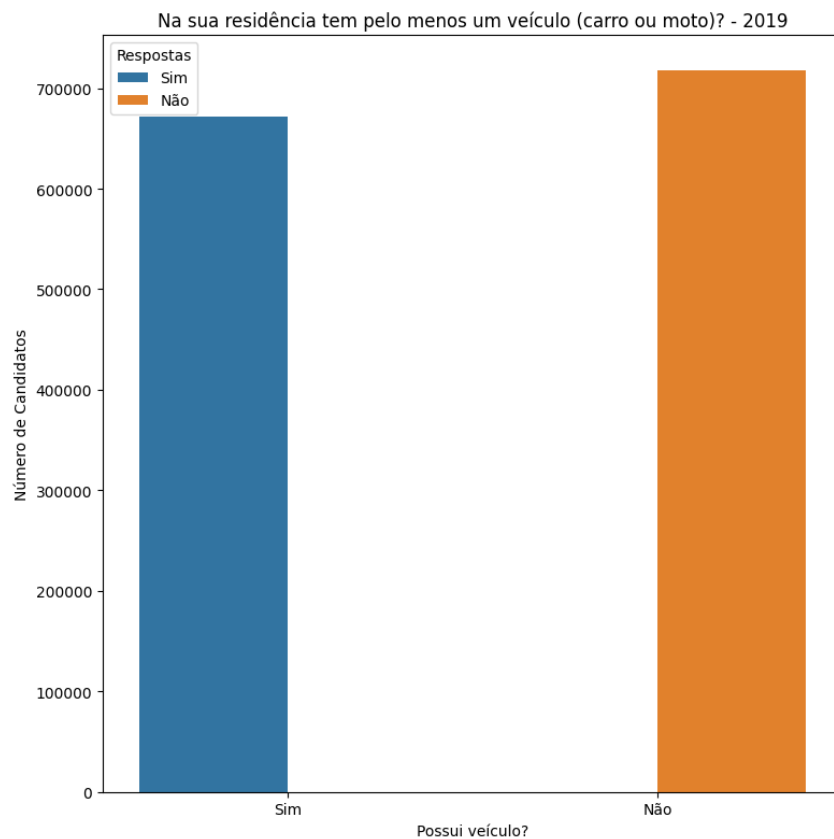
2021:



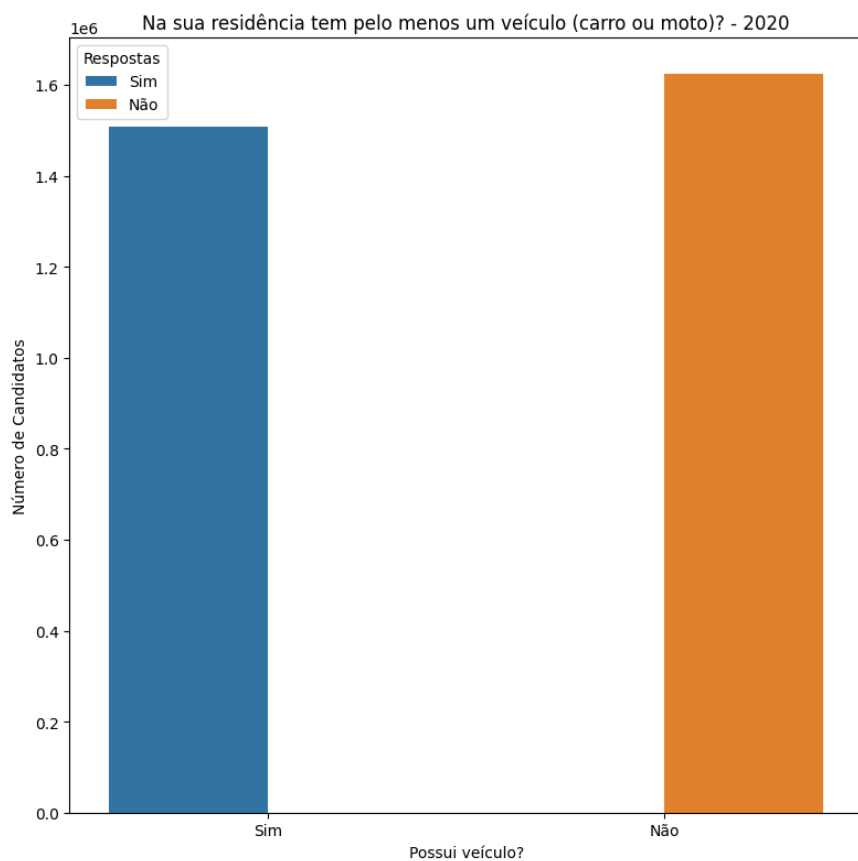
Pode-se observar que SP sempre está entre os primeiros em número de candidatos faltantes, isso se deve ao maior número de candidatos inscritos no Estado. Mas um caso que chama a atenção é o do CE, que também apareceu entre os três primeiros nos três anos, mas não tem um grande número de candidatos comparado à SP, por exemplo.

- 3) Não possuir um meio de transporte em casa pode influenciar no não comparecimento no exame?

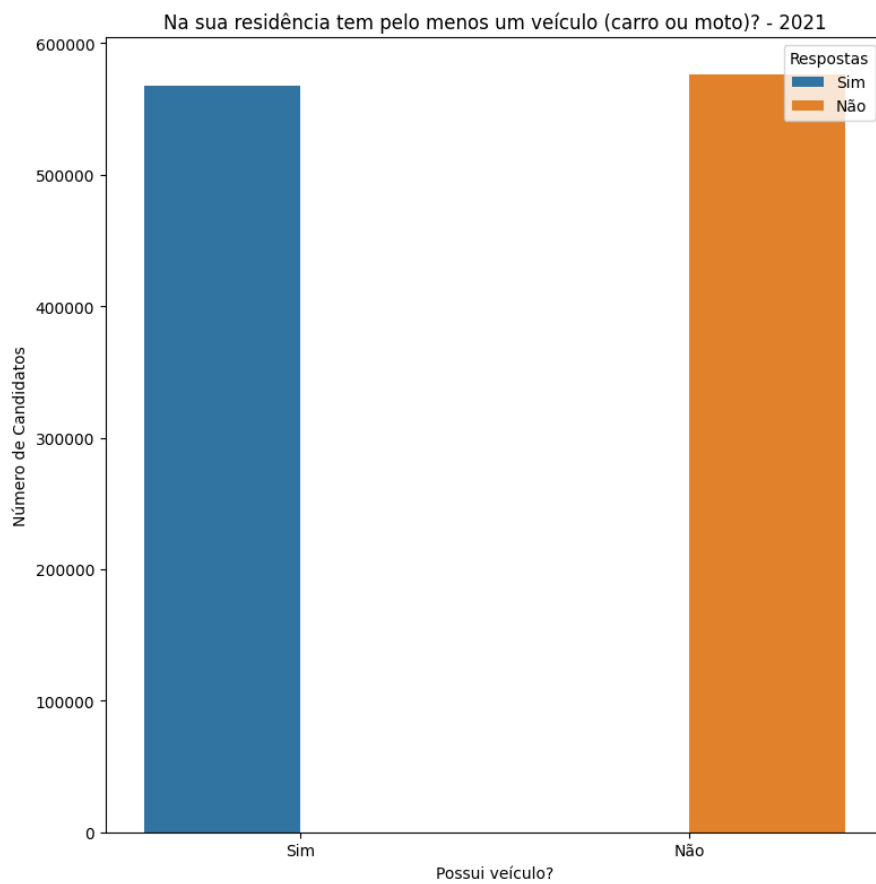
2019:



2020:



2021:



Pode-se observar que o número de candidatos que possuem e não possuem veículos em casa é bem parecido, então não é possível concluir que uma possível dificuldade de locomoção para o local da prova é um fator determinante para o não comparecimento.

- **Discussão e próximos passos**

As análises feitas até o momento trazem respostas para algumas hipóteses. A relação de um aluno ter uma renda familiar maior, a sua nota será melhor.

Outro ponto interessante nas questões socioeconômicas foi a relação de pais com formação mais completa, ou profissões mais rentáveis, com a média das notas. Isso mostrou que as médias são maiores para pais que têm mais formação e profissões mais rentáveis.

As análises feitas por etnias apontam que pessoas brancas têm um melhor desempenho no exame do enem.

Nos estados, pode-se perceber quais são os que têm resultados melhores no enem. Um dos nossos objetivos é tentar avaliar que, para estados mais desenvolvidos, teremos desempenhos melhores dos alunos.

Para as próximas etapas, precisamos definir melhor quais as perguntas socioeconômicas que mais interferem nas notas, além disso conseguir dados sobre investimentos em educação em cada estado, para entendermos melhor o que influencia no estado ter melhores notas.