



Projeto “Mousinho da Silveira”

Análise do desempenho de estudantes

Introdução ao Projeto “Mousinho da Silveira”

Neste estudo, analisamos o desempenho de estudantes das escolas **públicas Mousinho da Silveira e Gabriel Pereira, em Portugal, durante o ano letivo de 2005–2006**. O objetivo é investigar como fatores sociais e demográficos influenciam as notas nas disciplinas de Matemática e Português. **A análise é baseada em uma escala de 0 a 20 pontos, considerando as avaliações dos três períodos escolares: nota1, nota2 e nota3 (sendo esta última a nota final).**

Foram analisadas as notas dos três períodos (**nota1, nota2 e nota3**), o número de reprovações (**estudante_nr_reprovacoes**) e variáveis sociais como o estado civil dos pais (**estudante_status_pais**), acesso à internet (**estudante_acesso_internet**), tamanho da família (**estudante_tamanho_familia**), sexo do aluno (**estudante_sexo**) e se o estudante está em um relacionamento (**estudante_relacionamento**).

Com base nessas variáveis, buscamos compreender quais fatores estão mais associados ao sucesso ou à dificuldade acadêmica, utilizando bibliotecas como Pandas, Matplotlib e Seaborn para realizar a análise e visualização dos dados.

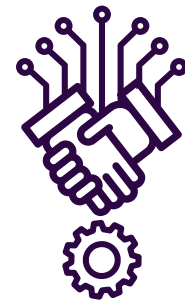


Fonte dos Dados, Tecnologias e Bibliotecas Utilizadas



UCI Machine Learning Repository.

<https://archive.ics.uci.edu/dataset/320/student+performance>



Tecnologias

- Python 3x
- Google Colab



Bibliotecas

- Pandas
- NumPy
- Matplotlib
- Seaborn

Transformações no DataFrame

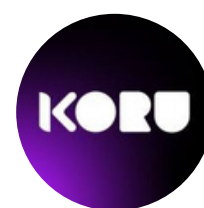
Nesta etapa, o DataFrame original (**df_math** ou **df_port**) foi copiado para um novo (**df_math_renomeado** ou **df_port_renomeado**) e passou por diversas transformações com o objetivo de facilitar a análise e a interpretação dos dados. Nesta etapa, o DataFrame original foi copiado e passaram a ocorrer duas principais transformações: a renomeação das colunas para nomes em português e a substituição dos valores categóricos por termos mais claros e traduzidos, facilitando a compreensão dos dados para a análise.

Substituir Valores Categóricos

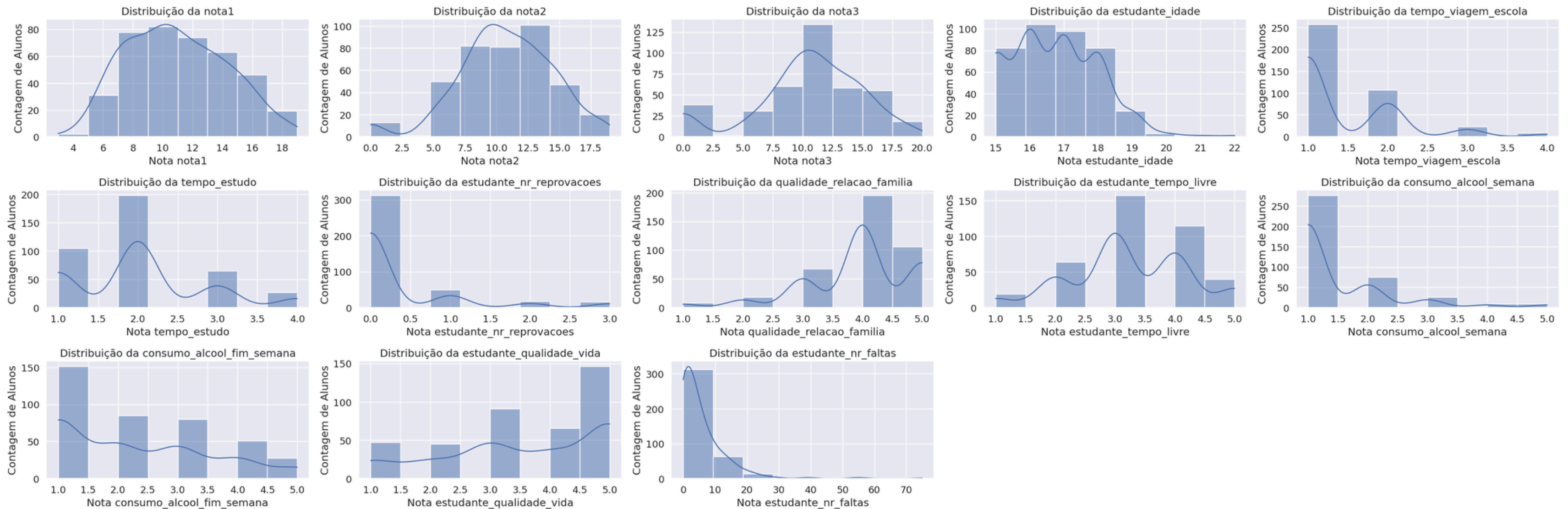
- Valores categóricos foram traduzidos para o português para facilitar a compreensão.
- Na coluna `estudante_tamanho_familia`:
 - 'GT3' → 'maior_que_3_pessoas'
 - 'LE3' → 'menor_que_3_pessoas'
- Na coluna `estudante_status_pais`:
 - 'T' → 'pais_casados'
 - 'A' → 'pais_divorciados'
- Tradução aplicada também nas colunas:
 - `mae_trabalho`
 - `pai_trabalho`
 - `motivo_escolha_escola`
 - `responsavel_estudante`
 - `estudante_casa_endereco`

Renomeação de Coluna

- `school` para `estudante_escola`,
- `sex` para `estudante_sexo`,
- `age` para `estudante_idade`,
- `famsize` para `estudante_tamanho_familia`,
- `Pstatus` para `estudante_status_pais`,
- `internet` para `estudante_acesso_internet`,
- `failures` para `estudante_nr_reprovacoes` e
- `G1`, `G2`, `G3` para `nota1`, `nota2` e `nota3`.

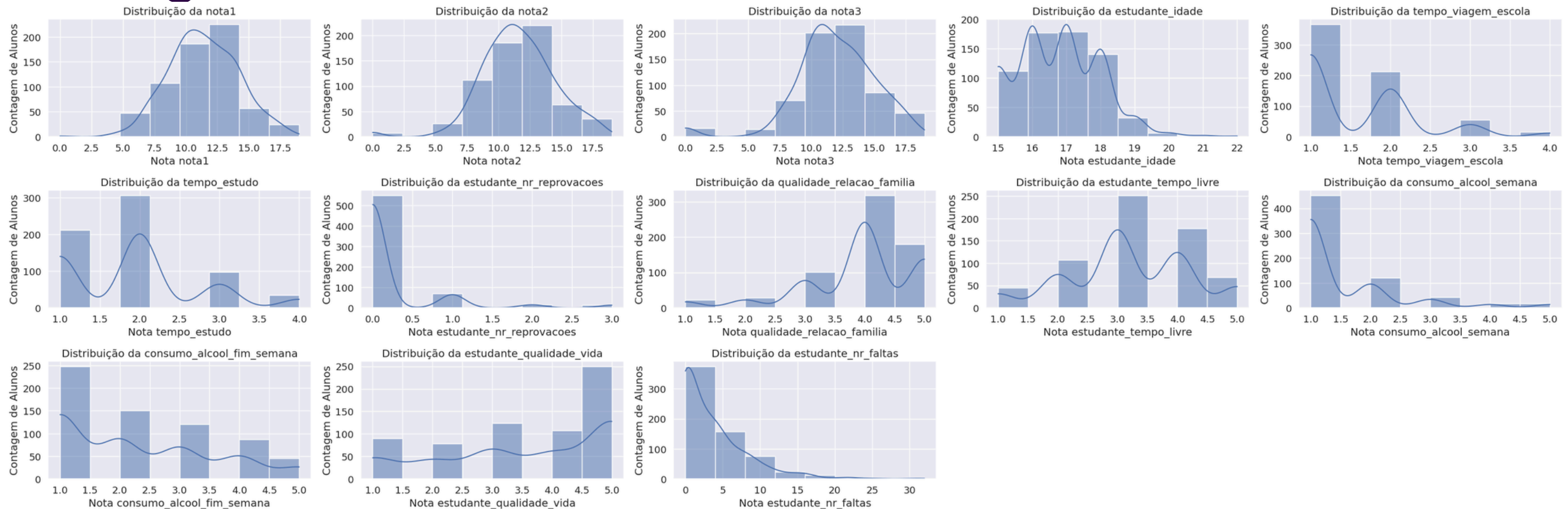


Análise da Distribuição das Variáveis Numéricas – Matemática



Após a análise exploratória dos dados, foram identificados os intervalos das principais variáveis numéricas da base. As notas (nota1, nota2 e nota3) variam de 0 a 20. A idade dos alunos está entre 15 e 22 anos. O tempo de viagem até a escola e o tempo de estudo variam de 1 a 4, representando categorias de duração. O número de reprovações anteriores varia de 0 a 3. Já as variáveis qualidade da relação familiar, tempo livre, consumo de álcool (durante a semana e no fim de semana) e qualidade de vida variam de 1 a 5. Por fim, o número de faltas apresenta uma ampla variação, indo de 0 a 70. Esses intervalos ajudam a compreender a distribuição e possíveis extremos nos dados analisados.

Análise da Distribuição das Variáveis Numéricas – Português

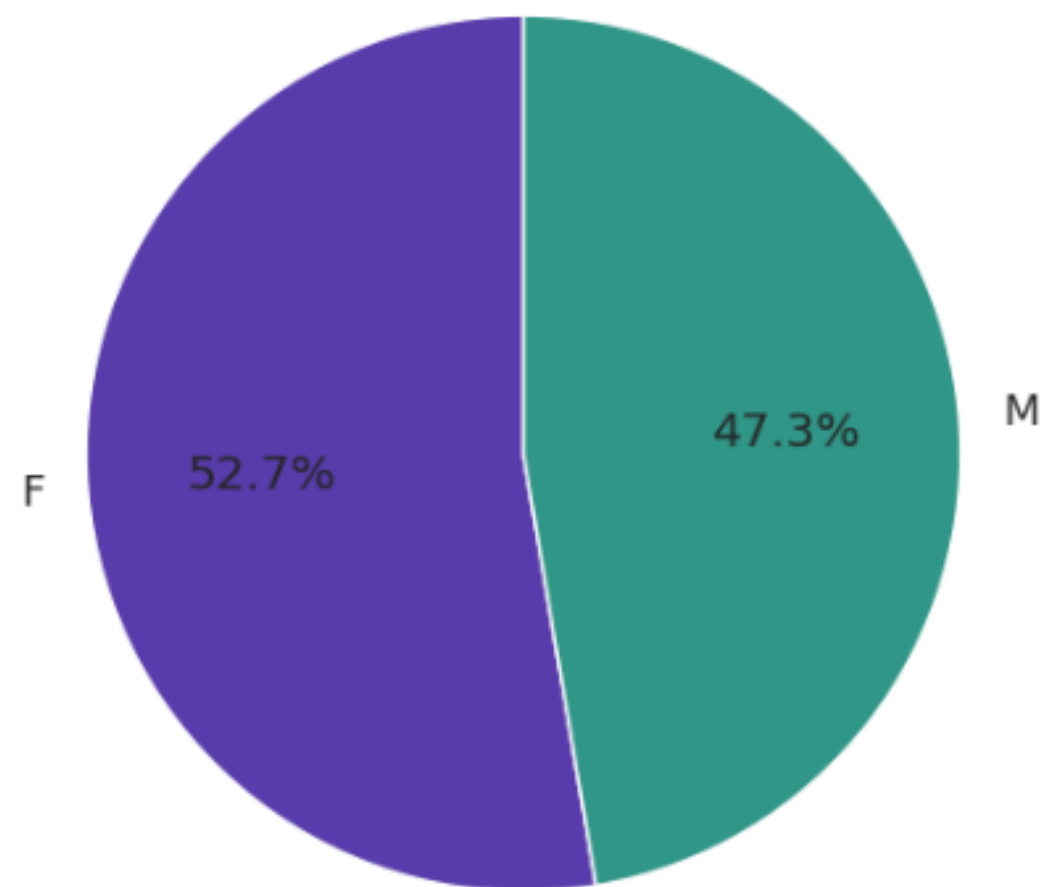


A análise exploratória revelou que a maioria das notas dos alunos (nota1, nota2 e nota3) está entre 10 e 15, indicando um desempenho médio. A idade varia de 15 a 22 anos, com predominância entre 16 e 18. A maior parte dos estudantes mora perto da escola (tempo de viagem 1) e estuda pouco (tempo de estudo 1 ou 2). A maioria nunca foi reprovada e relata ter boas relações familiares (valores 4 e 5). O tempo livre é mais comum nos níveis 3 e 4. O consumo de álcool durante a semana é baixo (valor 1), e no fim de semana, mais distribuído, mas ainda com predominância no nível 1. A qualidade de vida é avaliada positivamente por grande parte dos alunos (valor 5). Quanto às faltas, a maioria tem entre 0 e 10, indicando boa frequência escolar.

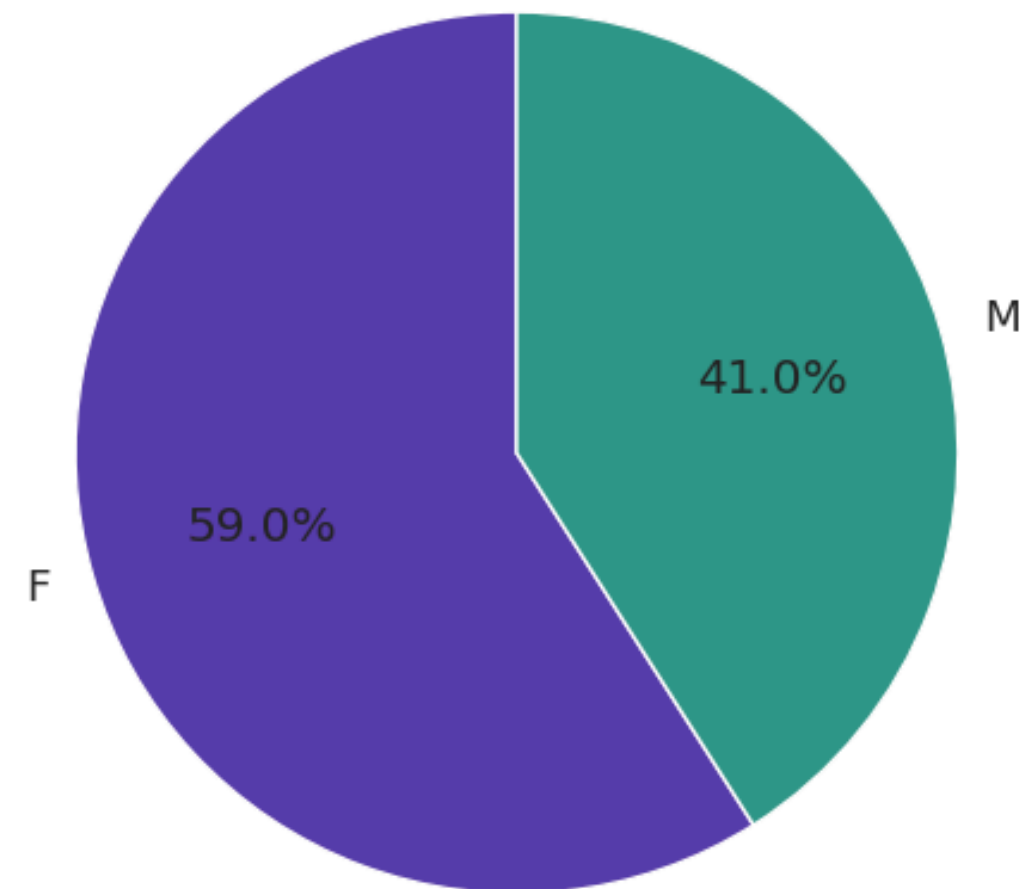
Distribuição de Estudantes por Gênero em Português e Matemática 🎓 🎓

A análise mostra uma predominância de alunas em ambas as bases de dados: são 383 alunas e 266 alunos em Língua Portuguesa, e 208 alunas e 187 alunos em Matemática.

Distribuição de Alunos por Sexo - Mat



Distribuição de Alunos por Sexo - Port



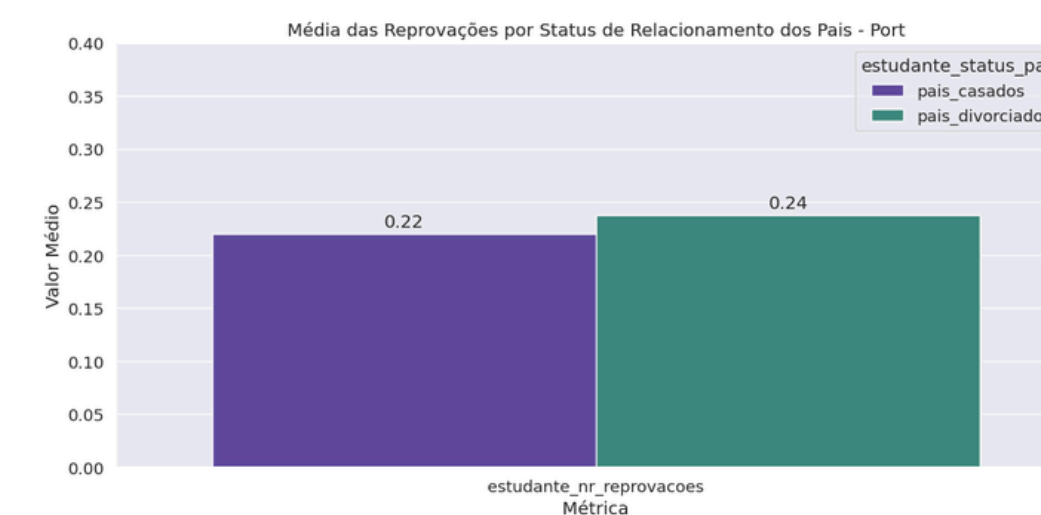
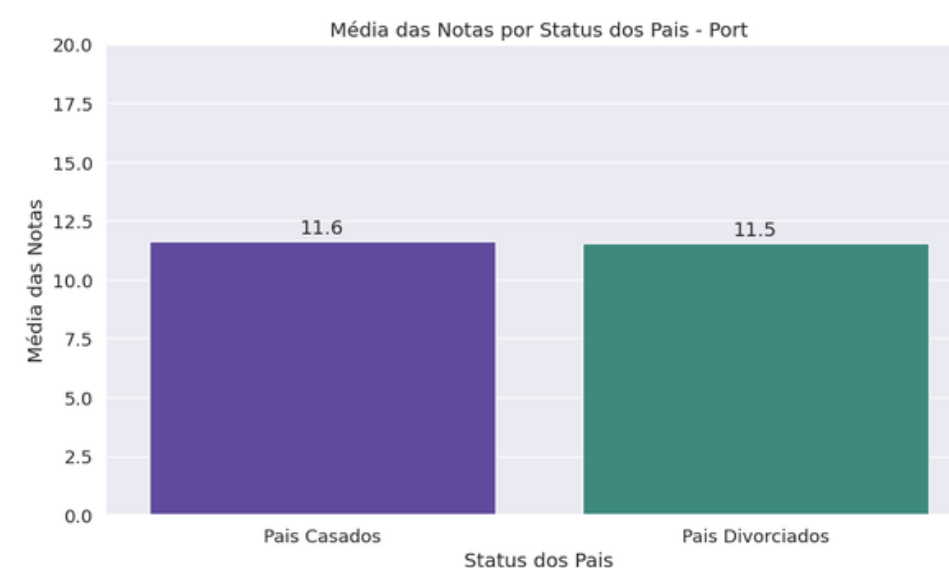
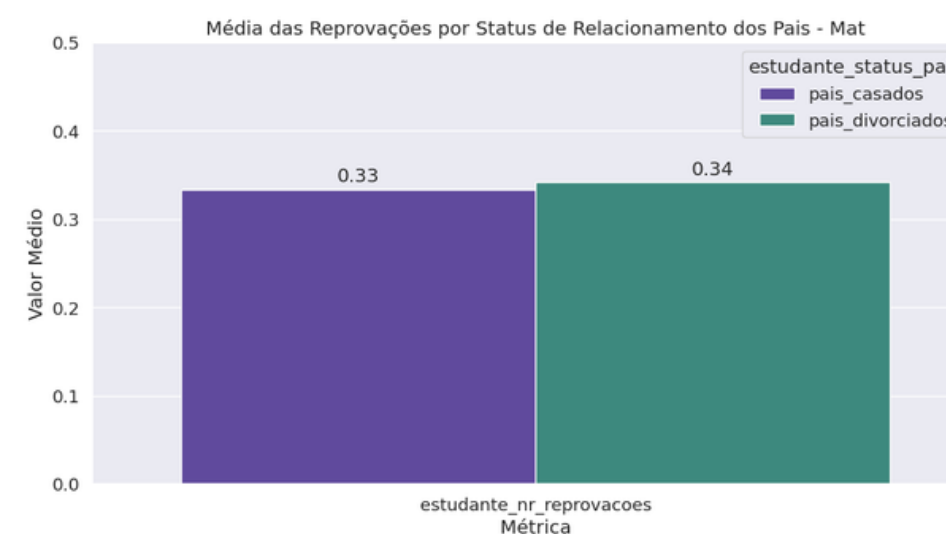
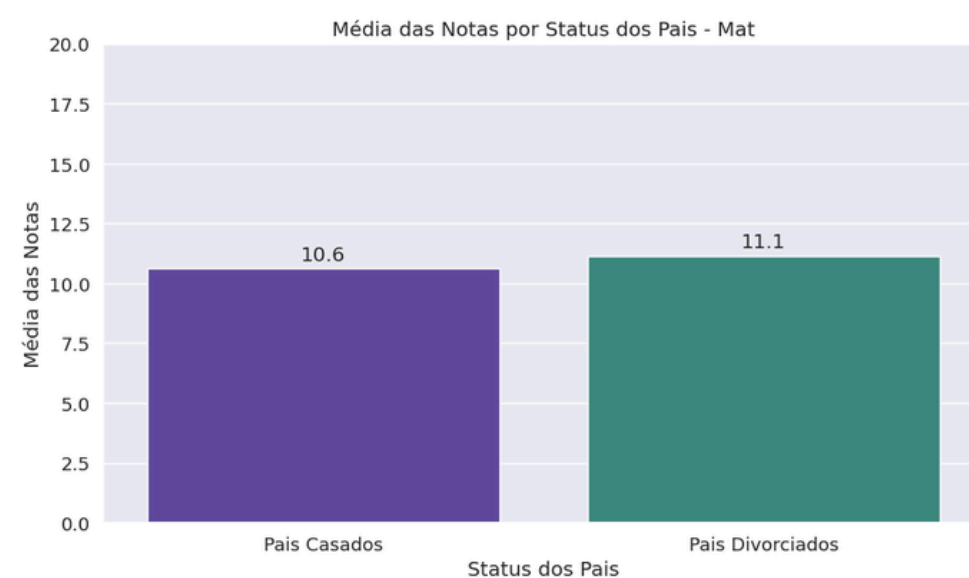
Status dos Pais

Na disciplina de Matemática, alunos com pais casados tiveram média de 10,6, enquanto os com pais divorciados tiveram média de 11,1. A média de reprovações foi praticamente igual: 0,33 para pais casados e 0,34 para divorciados.

Em Português, a média das notas foi de 11,2 para alunos com pais casados e 11,4 para os com pais divorciados. A média de reprovações também foi muito próxima: 0,26 e 0,27, respectivamente.

Conclusão:

As diferenças entre os grupos são pequenas tanto nas notas quanto nas reprovações. Portanto, o status civil dos pais não parece influenciar significativamente o desempenho escolar dos alunos neste conjunto de dados.





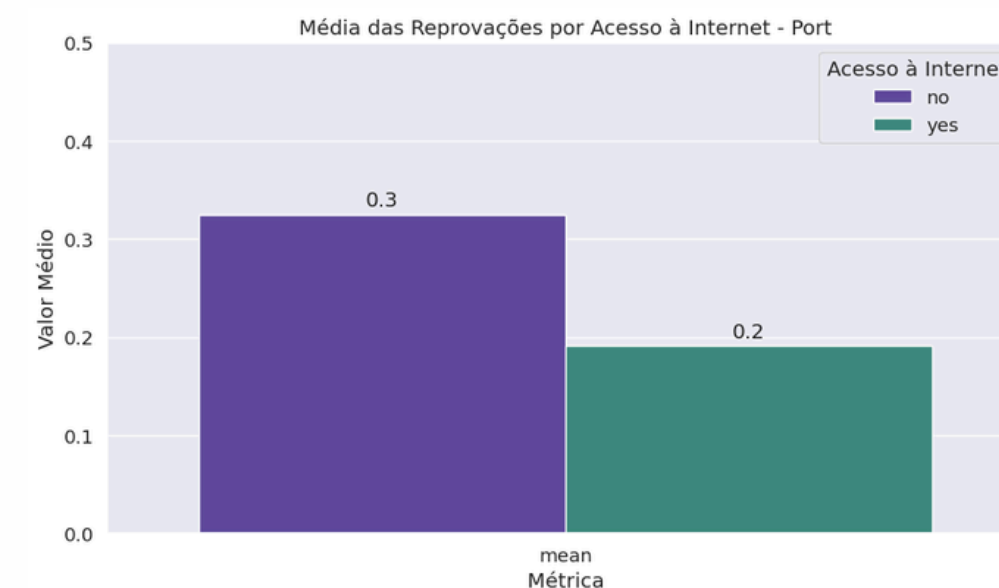
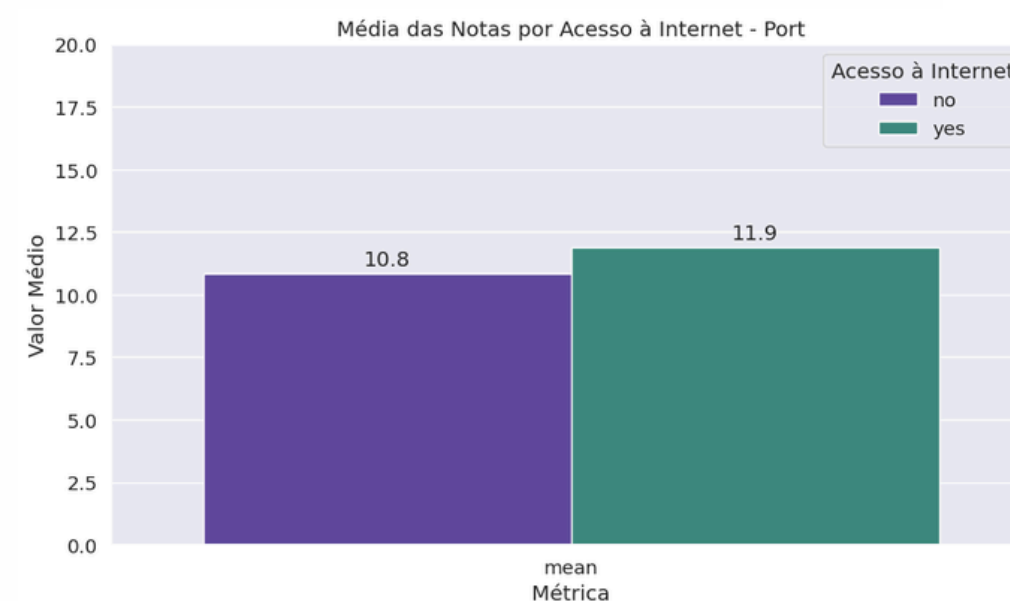
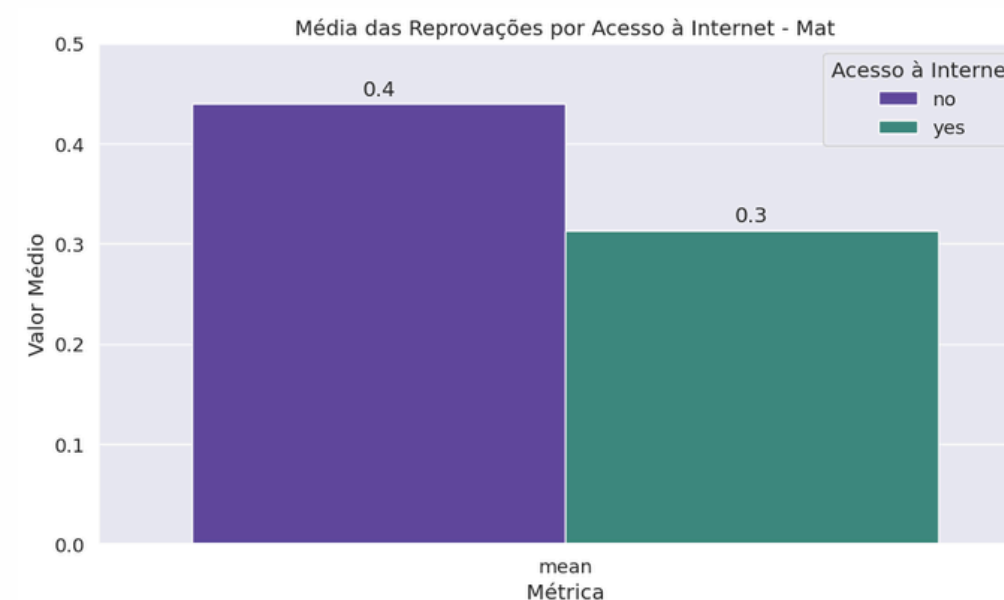
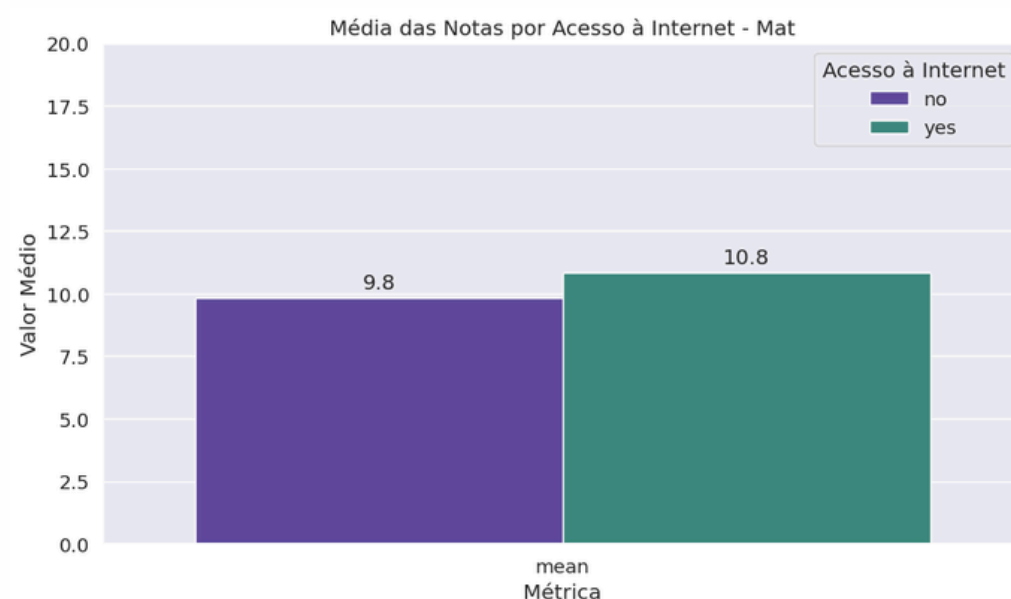
Acesso à internet

Em Matemática, alunos com acesso à internet tiveram média de 10,8, enquanto os que não têm acesso ficaram com 9,8. A média de reprovações foi de 0,3 para quem tem internet e 0,4 para quem não tem.

Em Português, a média foi de 11,5 com acesso à internet e 10,5 sem acesso. A média de reprovações foi de 0,2 com internet e 0,4 sem.

Conclusão:

O acesso à internet está associado a melhores notas e menos reprovações nas duas disciplinas. Isso mostra a importância da inclusão digital como apoio ao desempenho escolar dos alunos.

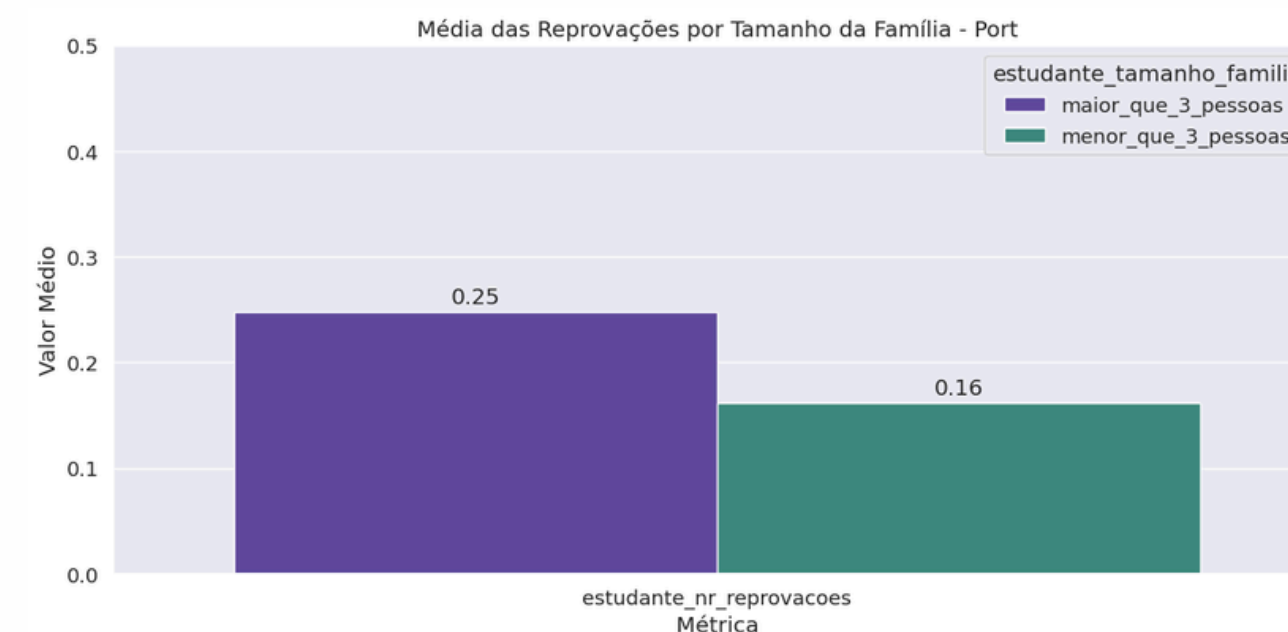
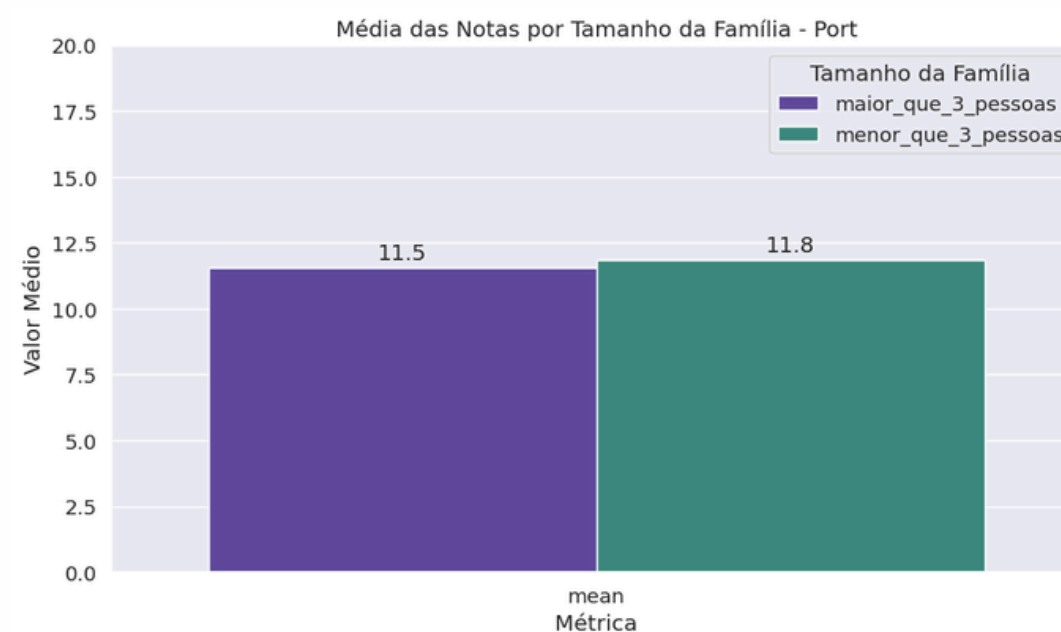
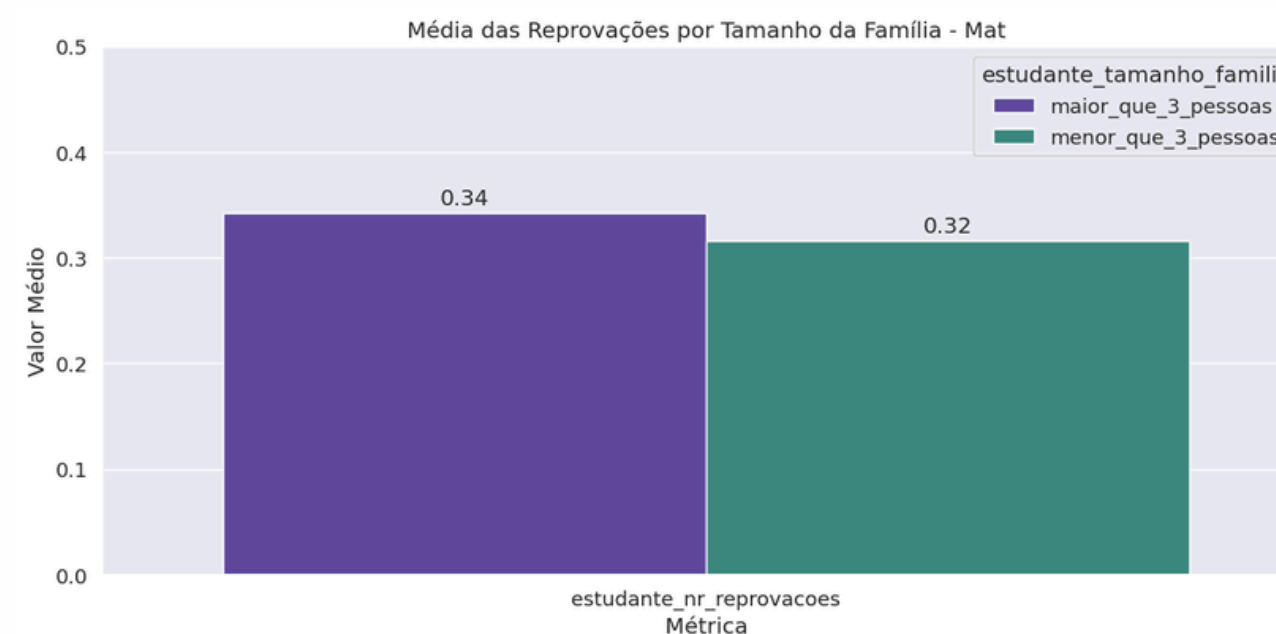
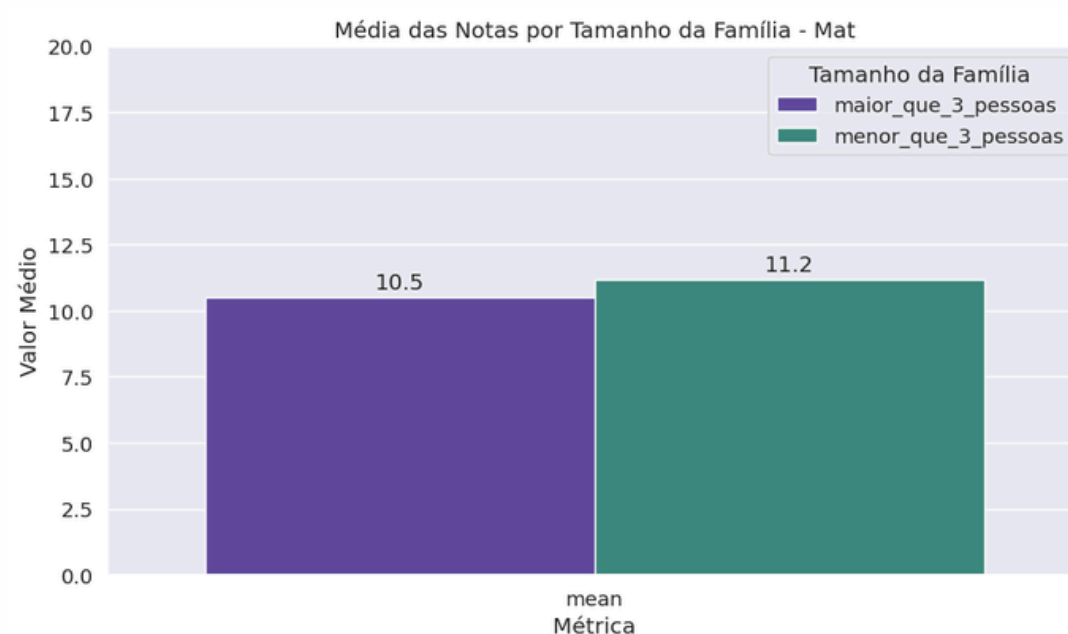




Tamanho da Família (famsize)

Em Matemática, alunos de famílias com até 3 pessoas tiveram média de 11,2 pontos, com 0,32 reprovações. Já os de famílias com mais de 3 pessoas apresentaram média de 0,34 reprovações.

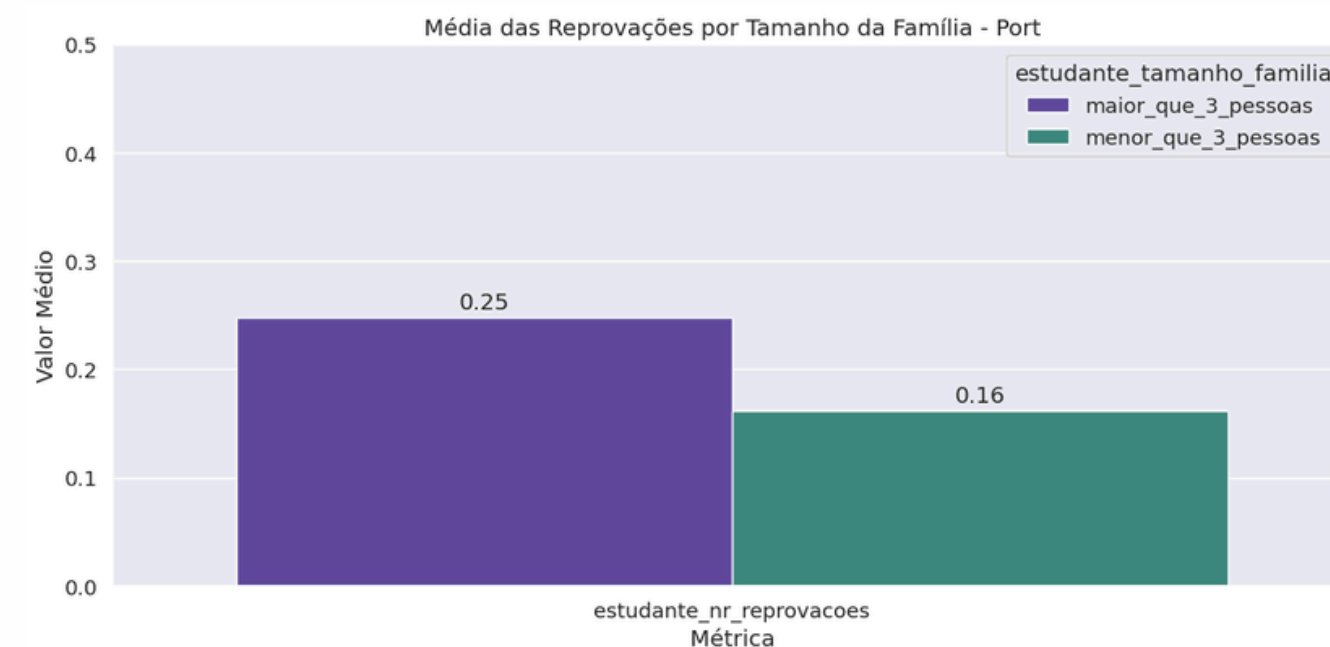
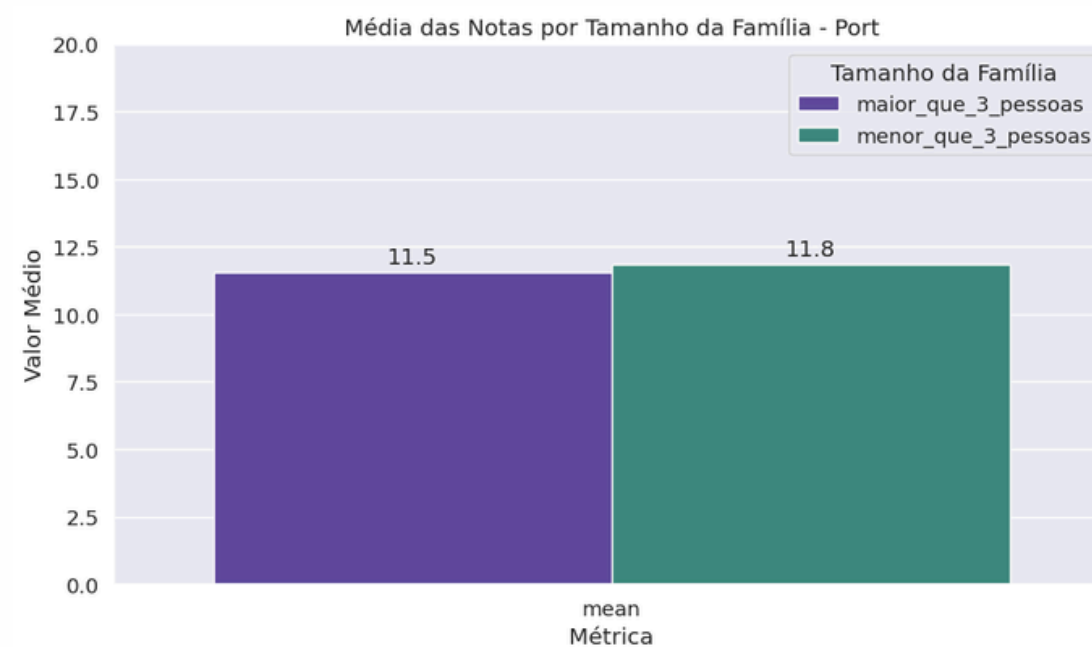
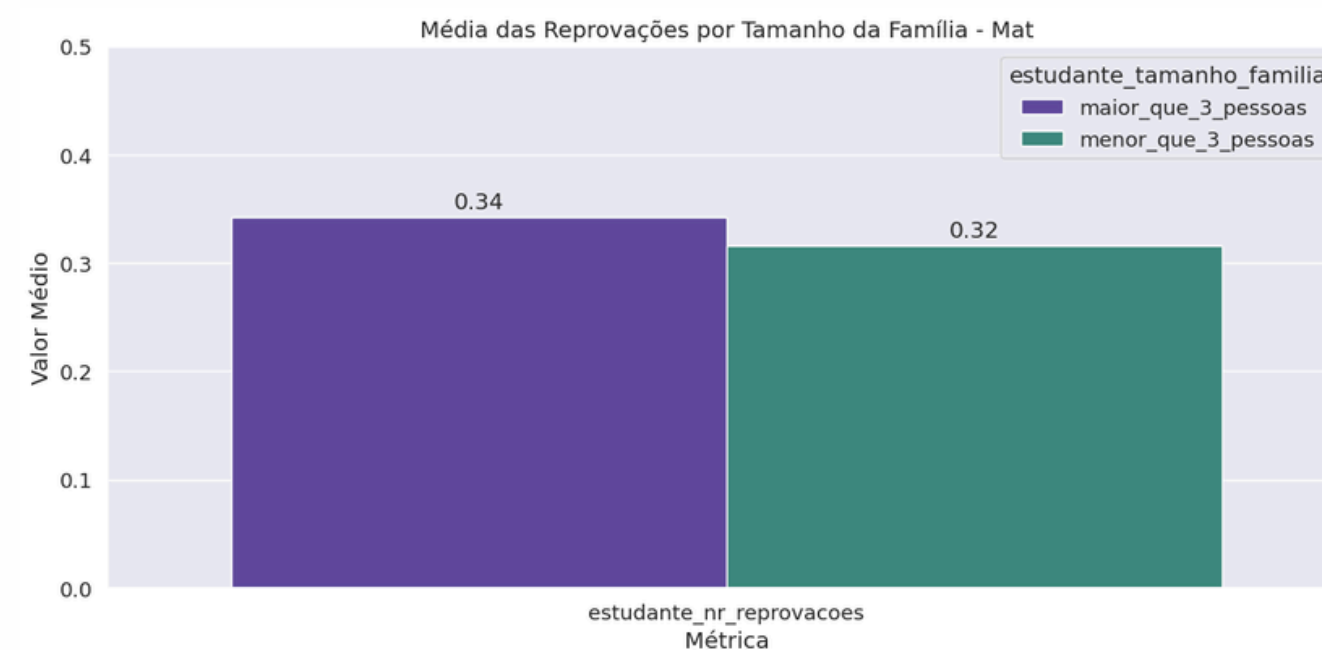
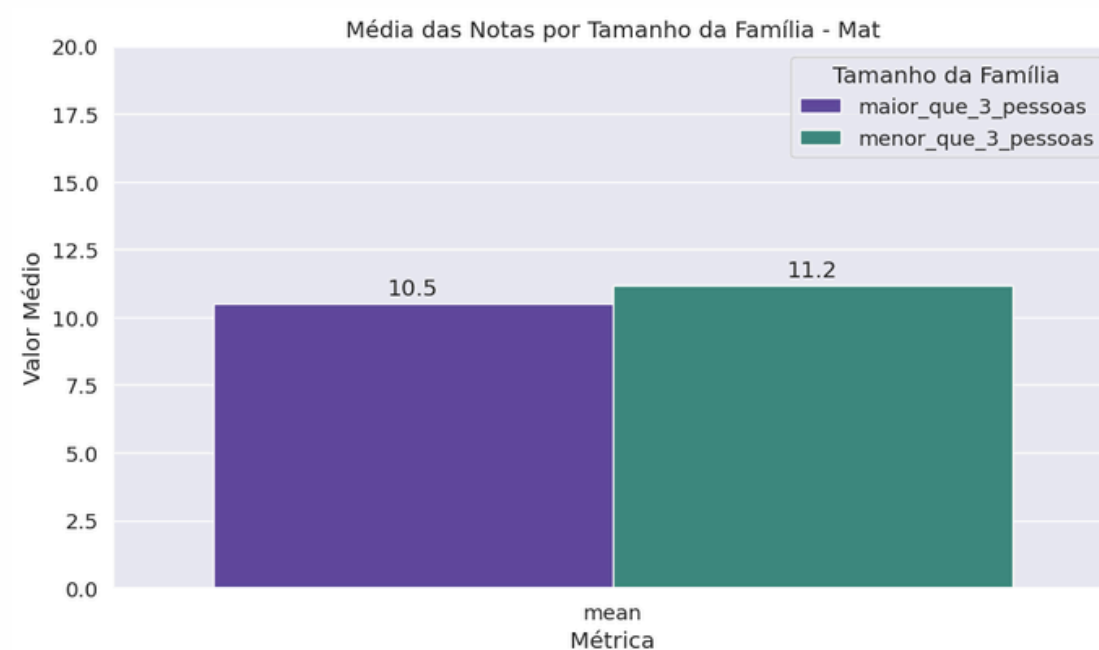
Em Português, alunos de famílias com até 3 pessoas tiveram média de 11,6 pontos e 0,16 reprovações. Alunos de famílias maiores tiveram média de 11,1 pontos e 0,25 reprovações





Sexo dos Estudantes

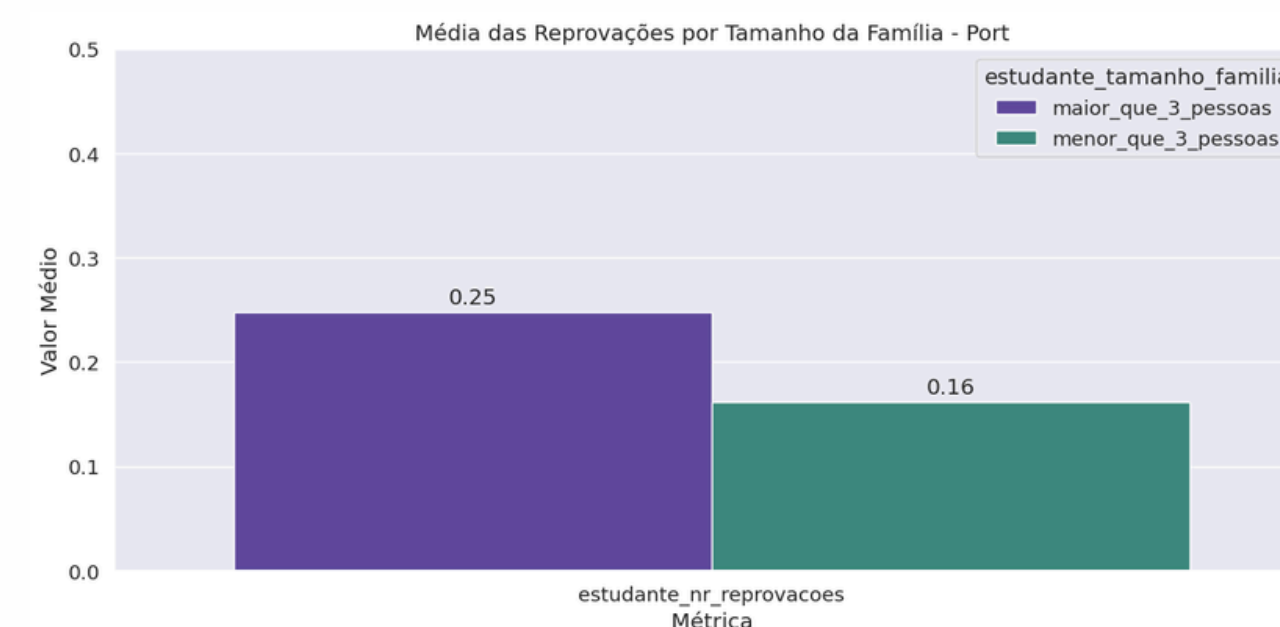
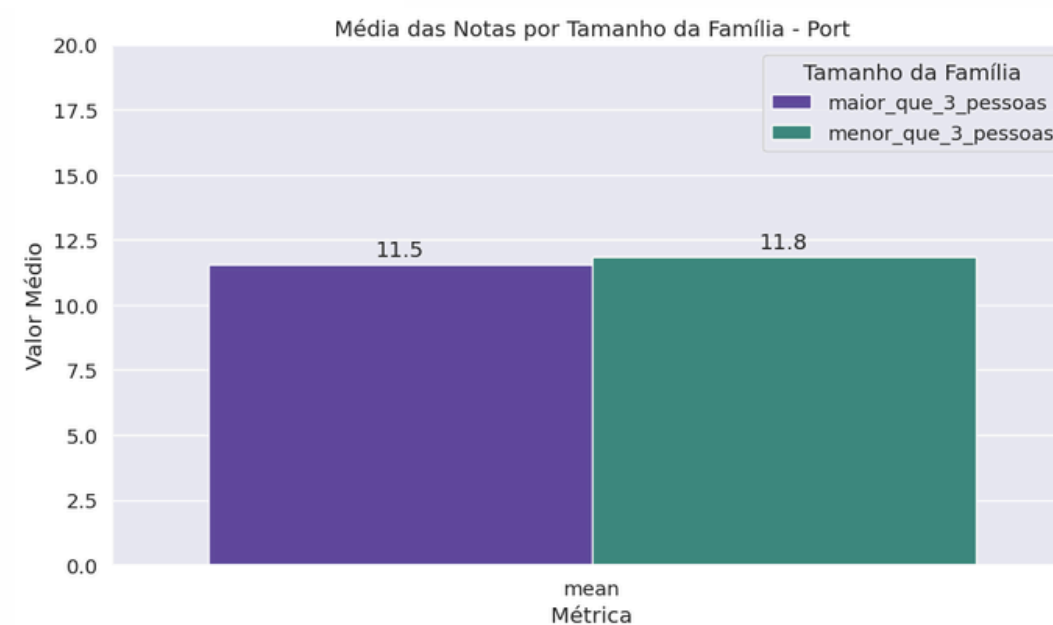
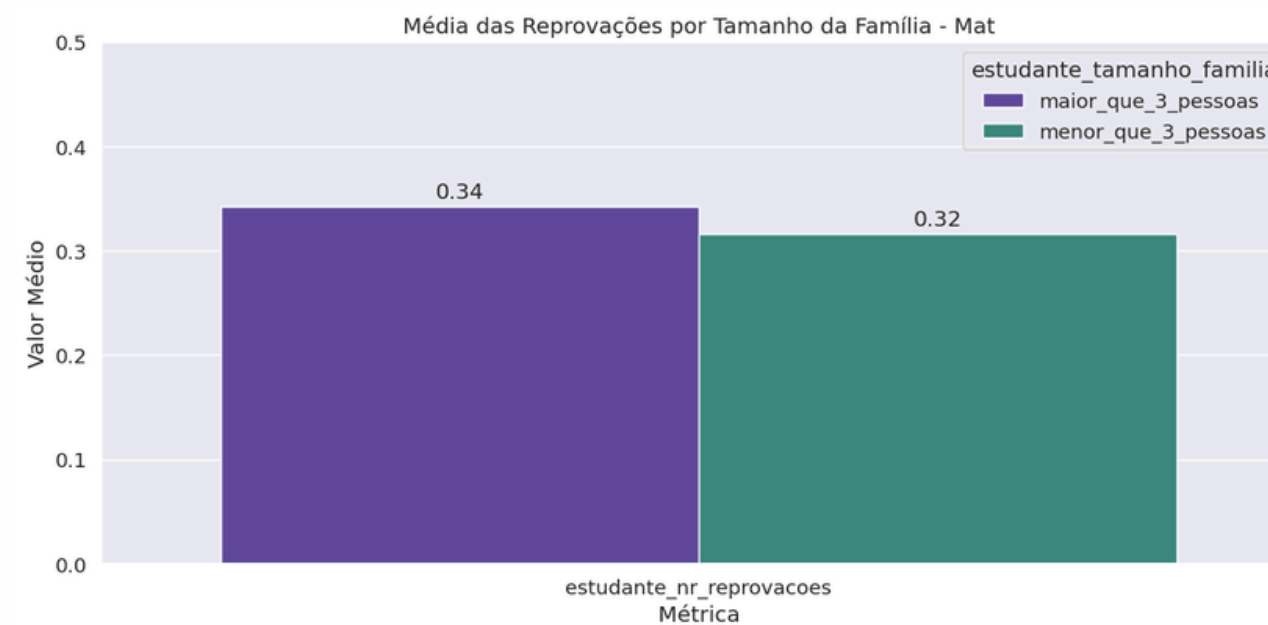
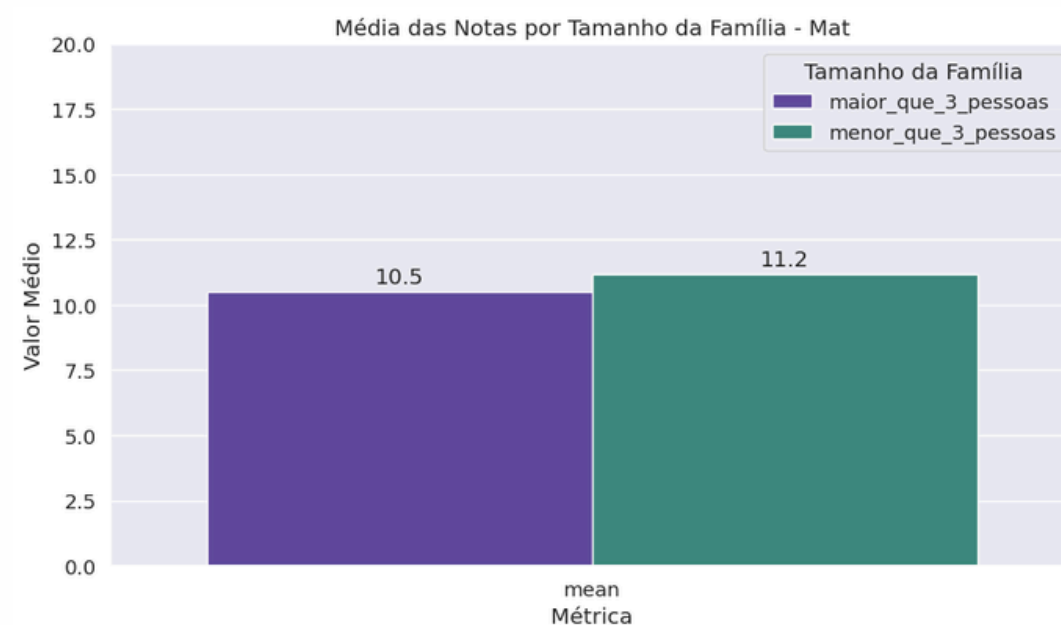
Em Matemática, meninos tiveram média de 11,1 pontos e 0,37 reprovações; meninas, média de 10,3 pontos e 0,30 reprovações. Em Português, meninas tiveram média de 11,3 pontos e 0,23 reprovações; meninos, 11,2 pontos e 0,31 reprovações. Meninos tiveram notas mais altas em Matemática, mas reprovaram mais. Meninas se saíram melhor em Português e tiveram menos reprovações no geral.





Relacionamento Amoroso

Alunos que não têm relacionamento amoroso apresentam notas médias mais altas e menos reprovações tanto em Matemática quanto em Português. Em Matemática, esses alunos tiveram média de 10,95 pontos e 0,29 reprovações, enquanto os que estão em relacionamento tiveram média de 10,14 pontos e 0,43 reprovações. Em Português, alunos sem relacionamento tiveram média de 11,48 pontos e 0,23 reprovações, já os que estão em relacionamento tiveram média de 10,88 pontos e 0,33 reprovações. Isso indica que, neste conjunto de dados, estar em um relacionamento amoroso pode estar associado a um desempenho escolar ligeiramente inferior.

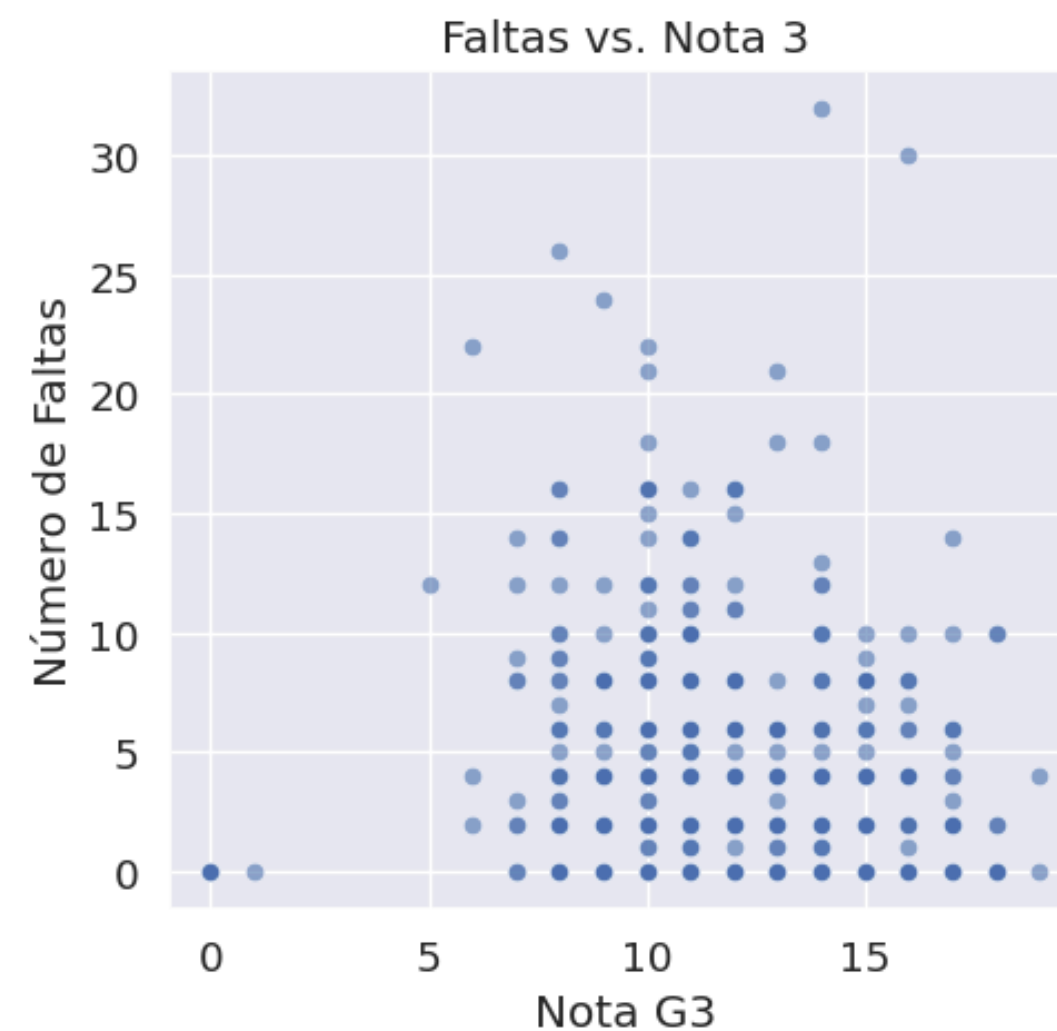
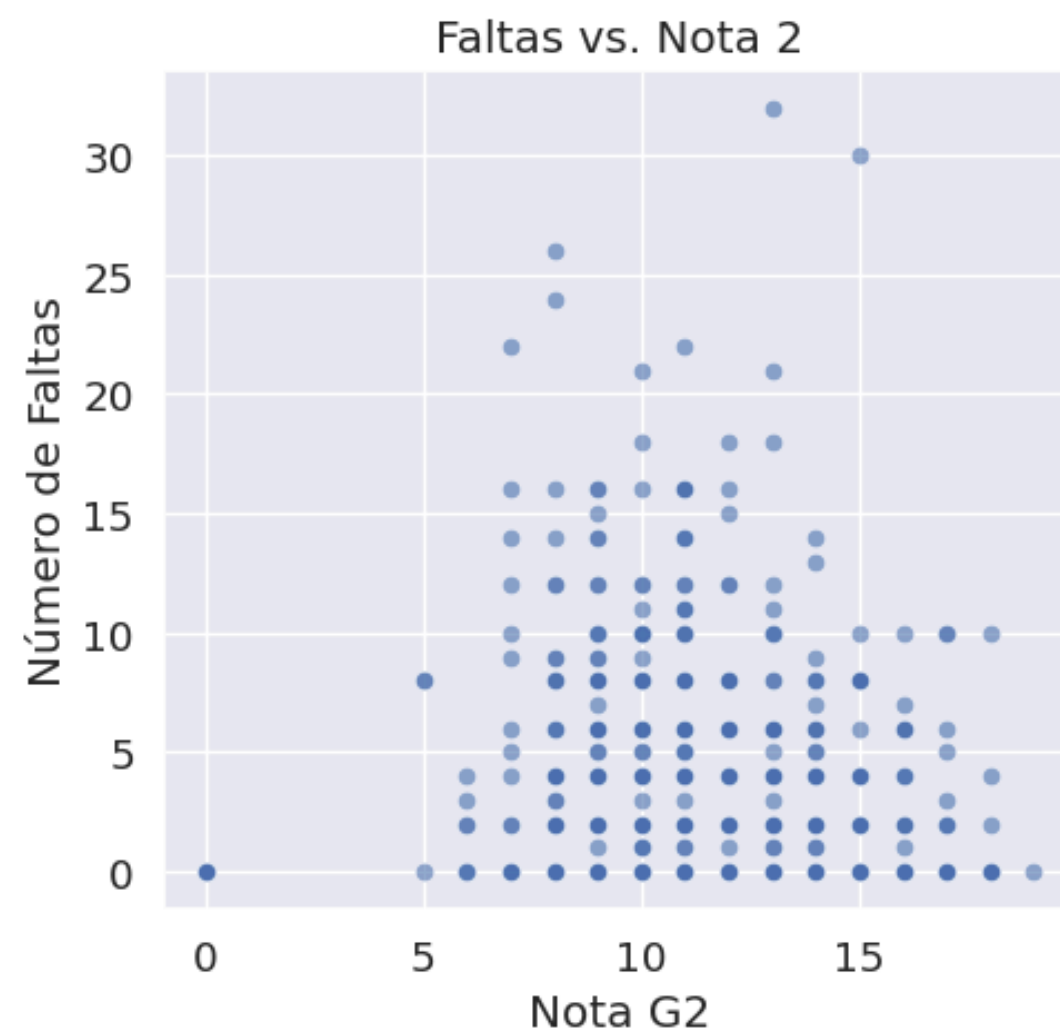
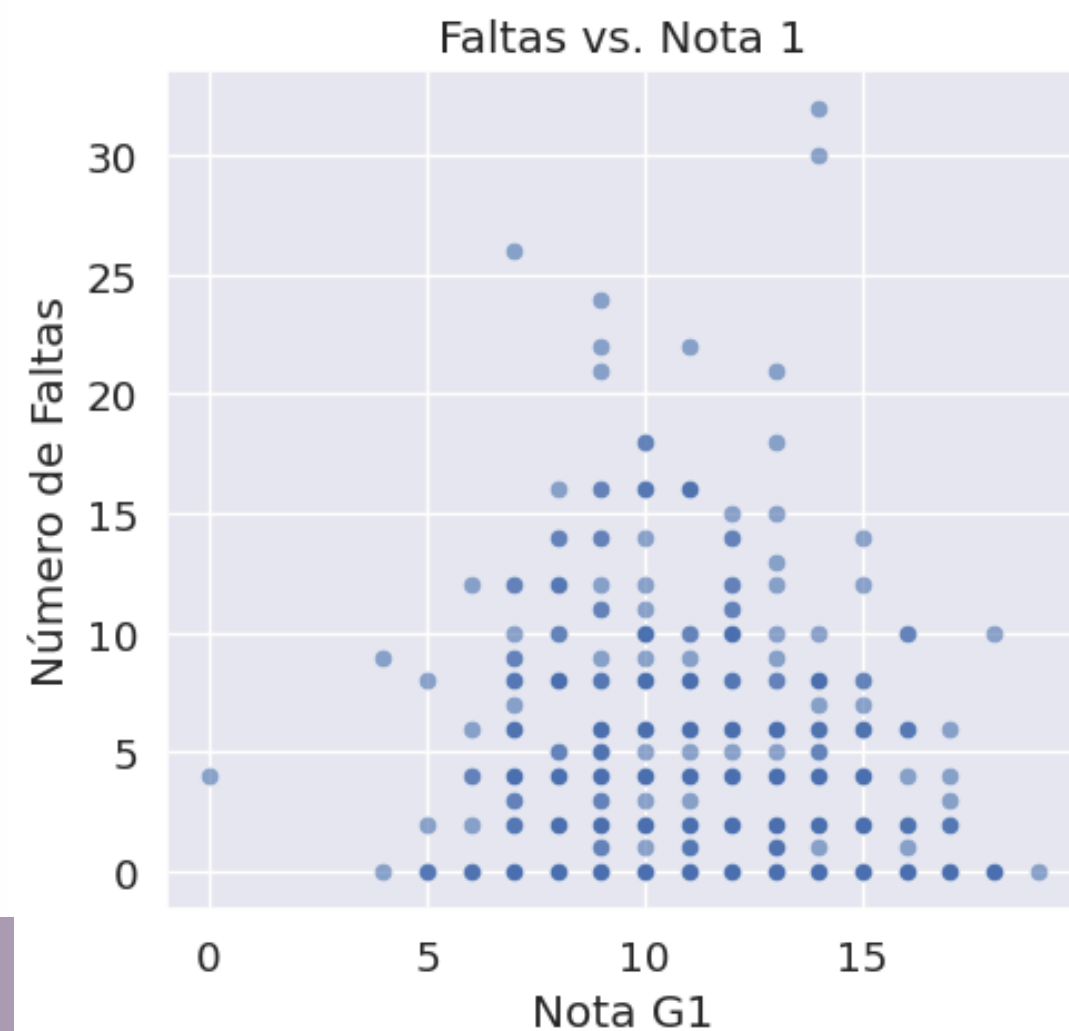




Visualizando a Relação entre Faltas e Notas

O gráfico mostra que, de forma geral, alunos com mais faltas tendem a ter notas menores nas três avaliações (nota1, nota2 e nota3).

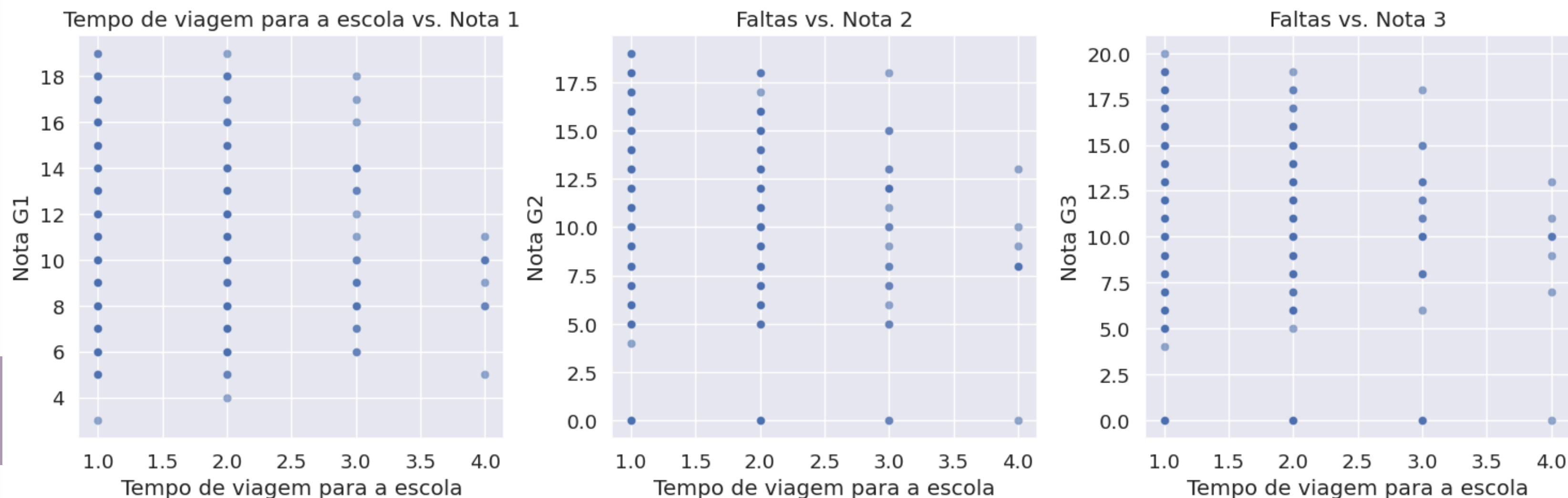
Há uma concentração maior de alunos com poucas faltas e notas intermediárias a altas. Alunos com muitas faltas apresentam uma maior variação, mas geralmente notas mais baixas.





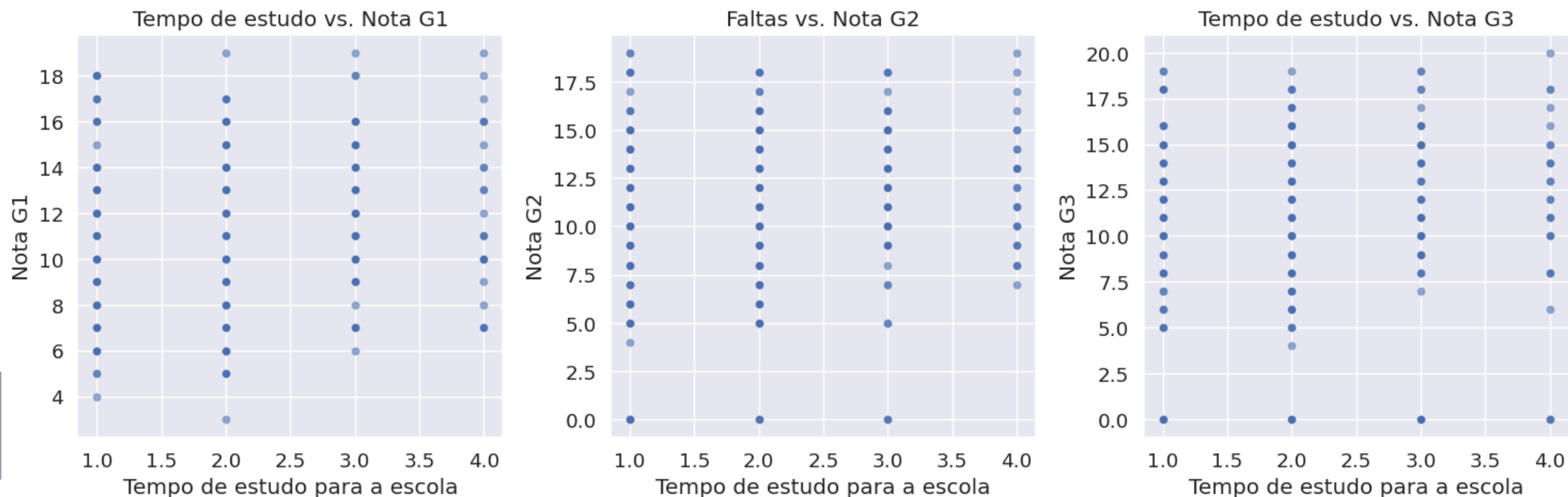
Visualizando a Relação entre Tempo de Viagem e Notas

O gráfico mostra a relação entre o tempo de viagem até a escola e as notas dos alunos em três períodos (G1, G2 e G3). De forma geral, observa-se que alunos com menor tempo de viagem (menos de 30 minutos) tendem a apresentar maior variação e, em alguns casos, melhores notas. Já alunos com maior tempo de deslocamento (mais de 1 hora) costumam ter menos variação e notas um pouco mais baixas. No entanto, a relação não é muito forte ou definitiva, indicando que o tempo de viagem pode influenciar o desempenho escolar, mas não é o único fator determinante.

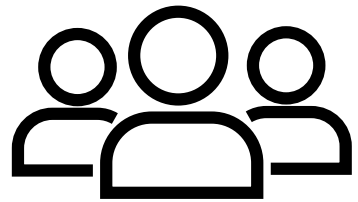


Visualizando a Relação entre Tempo de Estudo e Notas

O gráfico mostra que há, sim, uma relação entre o tempo de estudo semanal e as notas dos alunos. Alunos que estudam mais tendem a ter notas mais altas, especialmente nas avaliações finais (nota2 e nota3). Já os alunos que estudam menos apresentam uma concentração maior de notas baixas, incluindo muitas notas zero. Apesar de ainda haver variação, a tendência geral indica que quanto maior o tempo de estudo, melhor o desempenho escolar.

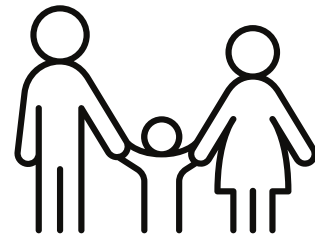


Conclusões



Distribuição Etária

- Nos DataFrames de Matemática e Português, a maioria dos alunos analisados está na faixa etária entre 15 e 22 anos, indicando que os dados se concentram principalmente em estudantes do ensino médio ou início da vida adulta.



Status dos Pais (Pstatus)

- O status civil dos pais teve pouca influência no desempenho dos alunos, com diferenças mínimas nas notas e nas médias de reprovação em Matemática e Português.



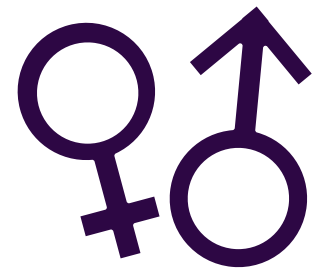
Acesso à Internet (internet)

- Alunos com acesso à internet apresentam notas mais altas e menores taxas de reprovação em Matemática e Português, indicando que a inclusão digital contribui positivamente para o desempenho escolar.

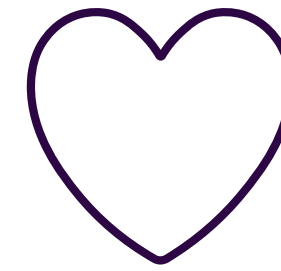
Conclusões



- Alunos de famílias menores (até 3 pessoas) tendem a ter médias mais altas e menos reprovações em Matemática e Português, indicando que o tamanho da família pode influenciar positivamente o desempenho escolar.

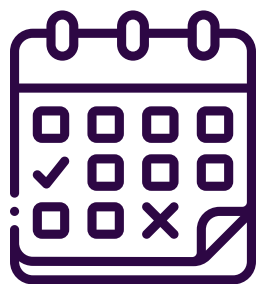


- Meninos tiveram melhores notas em Matemática, mas mais reprovações, enquanto meninas se saíram melhor em Português e apresentaram menos reprovações no geral.



- Alunos sem relacionamento amoroso apresentam notas melhores e menos reprovações, sugerindo que estar em um relacionamento pode impactar negativamente o desempenho escolar.

Conclusões



Faltas x notas (absence)

- Alunos com mais faltas tendem a ter notas mais baixas, enquanto aqueles com poucas faltas apresentam melhor desempenho geral.



Tempo de viagem x notas (traveltime)

- Alunos com menor tempo de viagem tendem a ter notas melhores, mas a influência do tempo de deslocamento no desempenho escolar é moderada e não determinante.



Tempo de estudo x notas (studytime)

- Alunos que estudam mais tendem a obter notas mais altas, especialmente nas avaliações finais, enquanto quem estuda menos apresenta mais notas baixas.