

# Consêquências do consumo de álcool em estudantes

Luisa Tavares e João Vitor

# Conjunto de Dados

## “Student Alcohol Consumption”

- Informações sobre 662 estudantes de Portugal
- Criado com o objetivo de verificar a relação entre o consumo de álcool e desempenho escolar
- Relevante pois problema de contexto real
- Atributos numéricos e categóricos.

# Análise inicial

*Raio-x*

## Análise estrutural dos dados brutos

- Entender a estrutura dos dados para detectar problemas e **a necessidade de modificações no dataset**
- Identificar tipos inválidos e inconsistências nos dados
- Verificação de dados faltantes, outliers e necessidades de discretizações
- Revelar problemas

---	-----	-----	-----	-----
0	school	649	non-null	object
1	sex	649	non-null	object
2	age	649	non-null	int64
3	address	649	non-null	object
4	famsize	649	non-null	object
5	Pstatus	649	non-null	object
6	Medu	649	non-null	int64
7	Fedu	649	non-null	int64
8	Mjob	649	non-null	object
9	Fjob	649	non-null	object
10	reason	649	non-null	object
11	guardian	649	non-null	object
12	traveltime	649	non-null	int64
13	studytime	649	non-null	int64
14	failures	649	non-null	int64
15	schoolsup	649	non-null	object
16	famsup	649	non-null	object
17	paid	649	non-null	object
18	activities	649	non-null	object
19	nursery	649	non-null	object
20	higher	649	non-null	object
21	internet	649	non-null	object
22	romantic	649	non-null	object

- Dataset bem estruturado, sem ruídos ou dados faltantes, bem tipado
- Necessidade de união entre dois datasets e de discretização e binarização de alguns atributos



# Atributos - explicação

## s Demográficos:

school: Escola do estudante (GP - Gabriel Pereira ou MS - Mousinho da Silva)

sex: Gênero ('F' - Feminino, 'M' - Masculino).

age: Idade (de 15 a 22 anos).

address: Tipo de endereço ('U' - Urbano, 'R' - Rural).

famsize: Tamanho da família ('LE3' - Menor ou igual a 3, 'GT3' - Maior que 3)

Pstatus: Status de coabitação dos pais ('T' - Moram juntos, 'A' - Separados)

## s Familiares e de Suporte:

Medu, Fedu: Nível de educação da mãe e do pai (0 a 4, onde 4 é ensino superior)

Mjob, Fjob: Trabalho da mãe e do pai.

guardian: Guardião legal do estudante.

famsup: Suporte educacional da família (sim/não).

schoolsup: Suporte educacional extra (reforço) (sim/não).

paid: Aulas particulares pagas na disciplina (sim/não).

famrel: Qualidade do relacionamento familiar (1 - muito ruim a 5 - excelente)

## bitos e Estilo de Vida:

- studytime: Tempo de estudo semanal (1 a 4, representando faixas de tempo)

- traveltime: Tempo de viagem de casa para a escola (1 a 4).

- goout: Frequência de saídas com amigos (1 - muito baixa a 5 - muito alta)

- Dalc, Walc: Consumo de álcool durante a semana e no fim de semana

- health: Estado de saúde atual (1 - muito ruim a 5 - muito bom).

- activities: Participa de atividades extracurriculares (sim/não).

- internet: Possui acesso à internet em casa (sim/não).

- romantic: Está em um relacionamento amoroso (sim/não).

## dos Acadêmicos:

- failures: Número de reprovações em classes anteriores.

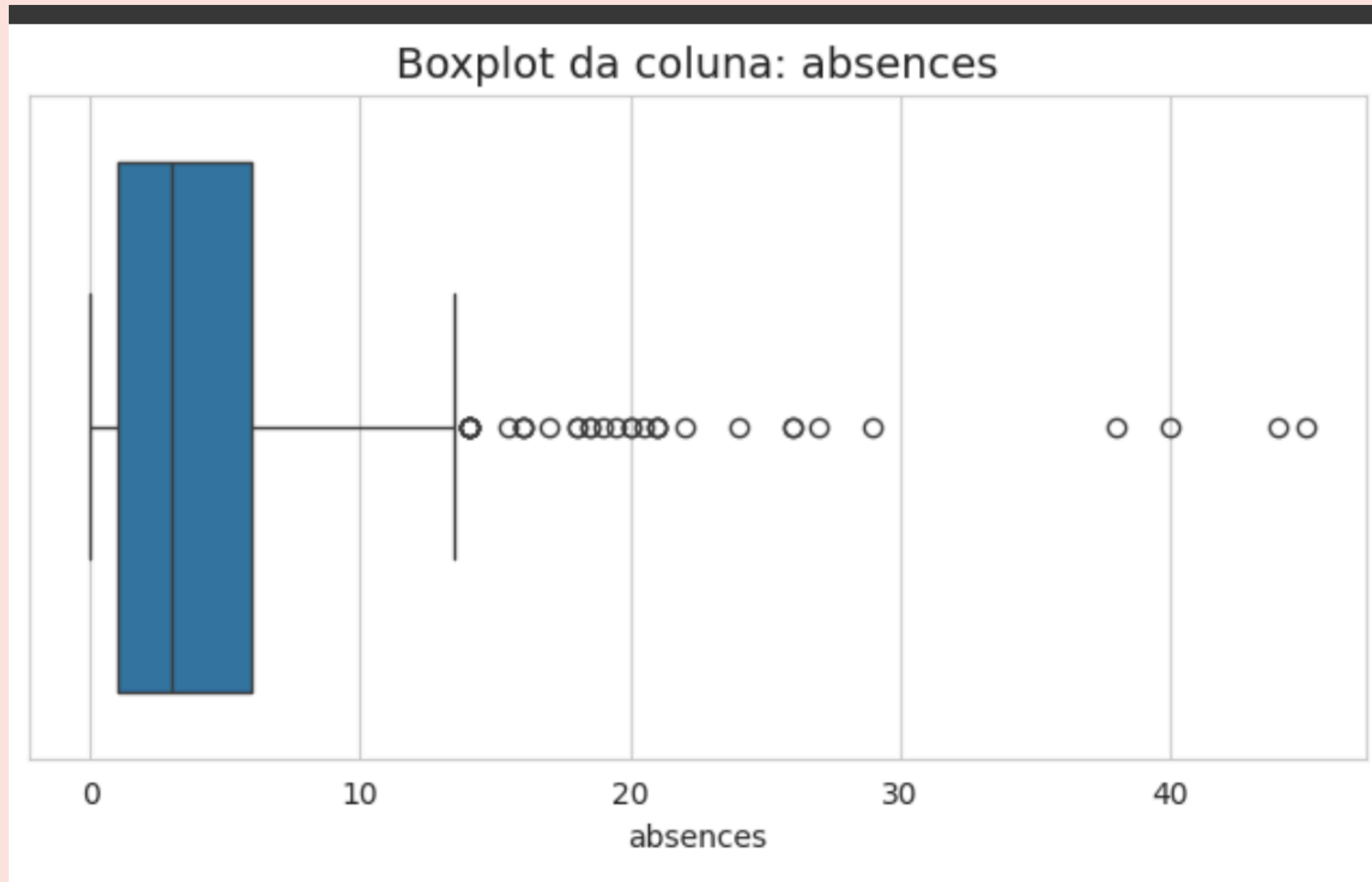
- absences: Número de faltas escolares.

- G1, G2: Notas do primeiro e segundo período (0 a 20).

- G3: Nota final (0 a 20) - nossa principal variável de resultado.

# Estudo e preparação

- **Unificação Inteligente:** Juntamos os dados de Matemática e Português, agregando a média das notas dos alunos duplicados para não perder informação.
- **Engenharia de Features:** Criamos colunas-chave: `alcool_total` e `passed`(Sim, Não).
- **Limpeza:** Verificamos a qualidade dos dados para verificar realmente a não existência de ruídos
- **Binarização de atributos booleanos:** Transformamos em 1 e 0 os atributos de sim e não



- **Não removemos esses outliers**, pois eles representam um comportamento real e importante de um grupo de alunos.

# Discretização:

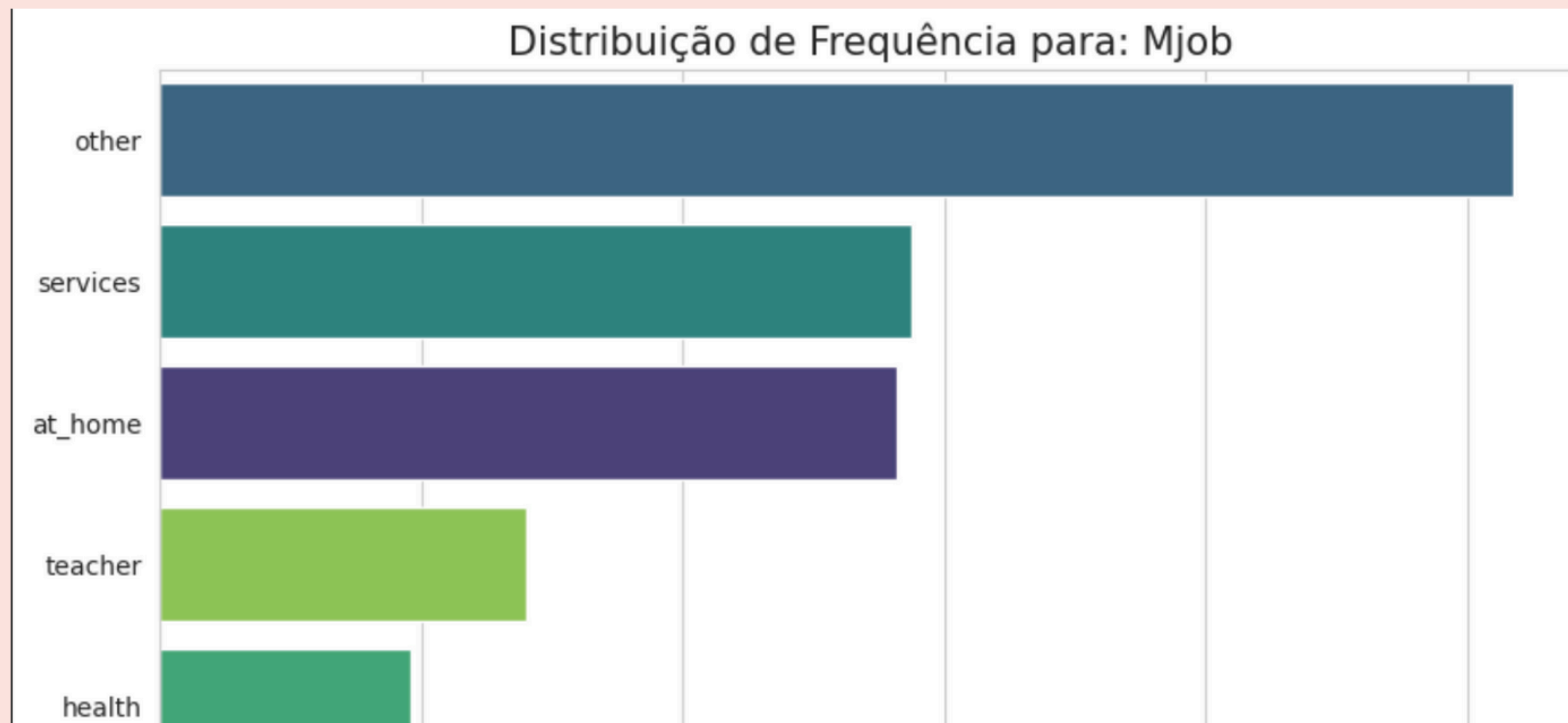
Exemplos das novas colunas

	age	faixa_etaria	absences	faixa_faltas	G3	desempenho_final
213	18	17-18 anos	7.0	Faltas Moderadas (5-9)	13.0	Regular (10-13)
402	18	17-18 anos	7.0	Faltas Moderadas (5-9)	12.0	Regular (10-13)
447	15	15-16 anos	0.0	Nenhuma Falta	9.0	Insuficiente (0-9)
185	18	17-18 anos	2.0	Poucas Faltas (1-4)	11.0	Regular (10-13)
541	18	17-18 anos	0.0	Nenhuma Falta	15.0	Bom (14-17)
222	18	17-18 anos	5.0	Faltas Moderadas (5-9)	11.0	Regular (10-13)
247	15	15-16 anos	6.0	Faltas Moderadas (5-9)	7.0	Insuficiente (0-9)
55	16	15-16 anos	3.0	Poucas Faltas (1-4)	13.5	Bom (14-17)
430	19	19+ anos	8.0	Faltas Moderadas (5-9)	12.5	Regular (10-13)
617	17	17-18 anos	2.0	Poucas Faltas (1-4)	9.0	Insuficiente (0-9)

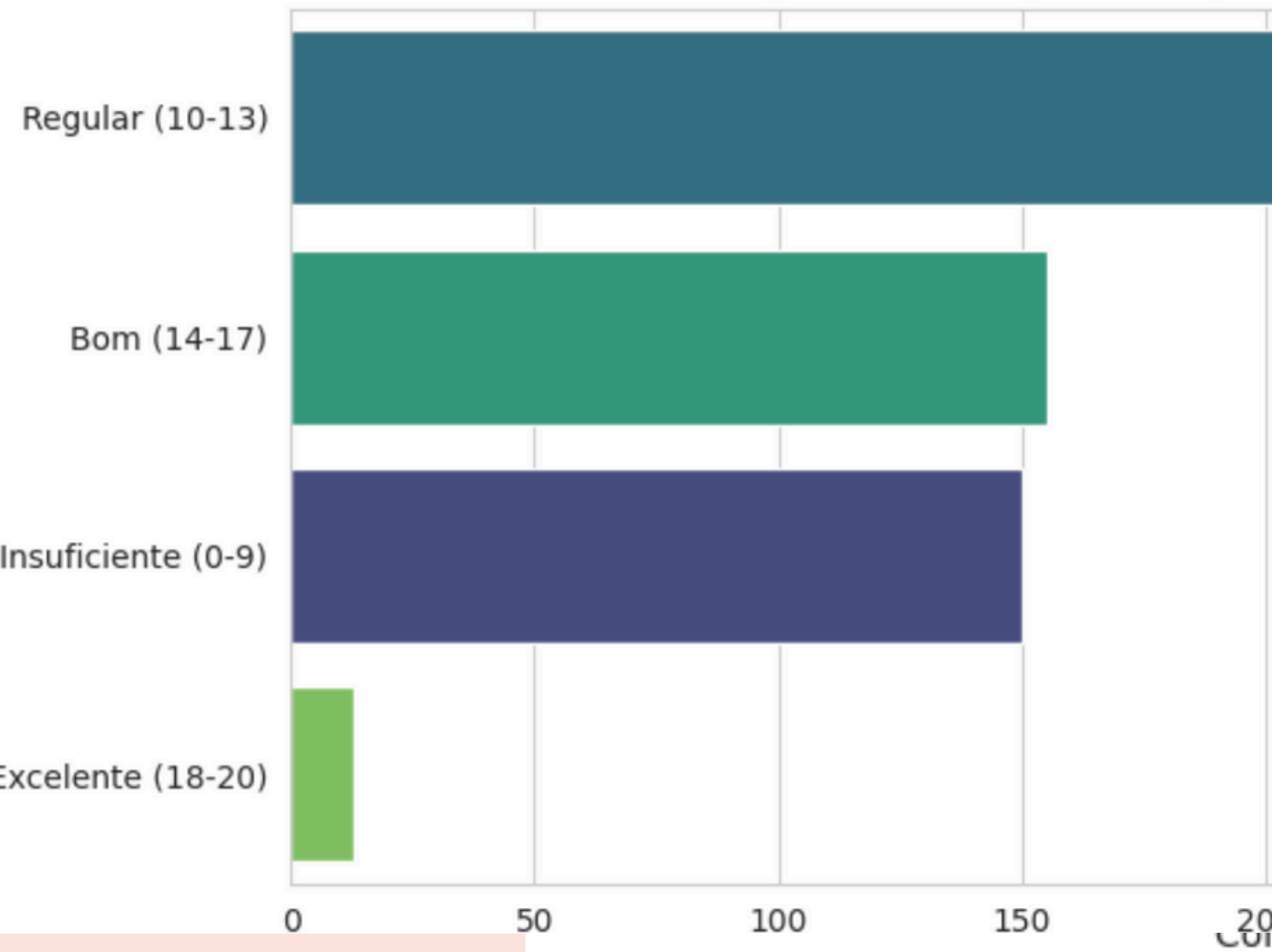


# Análise Univariada

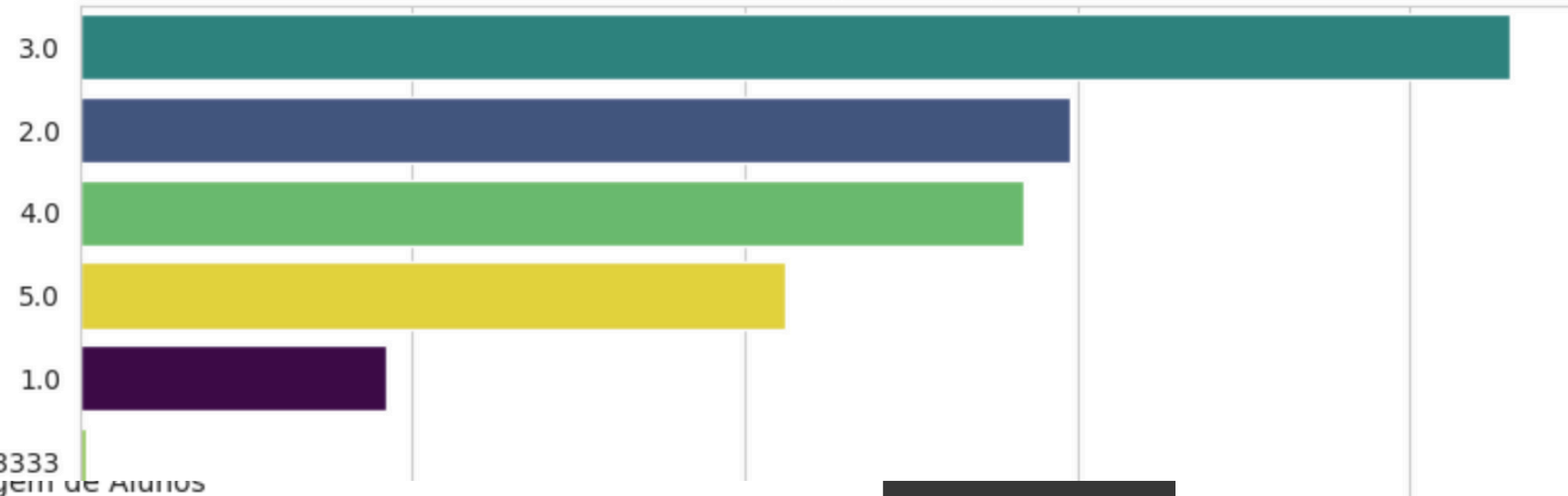
- **Utilização de gráficos:** verificar distribuição dos dados, valores mais comuns, valores possíveis em cada atributo, para analisar como se portam unitariamente os atributos:



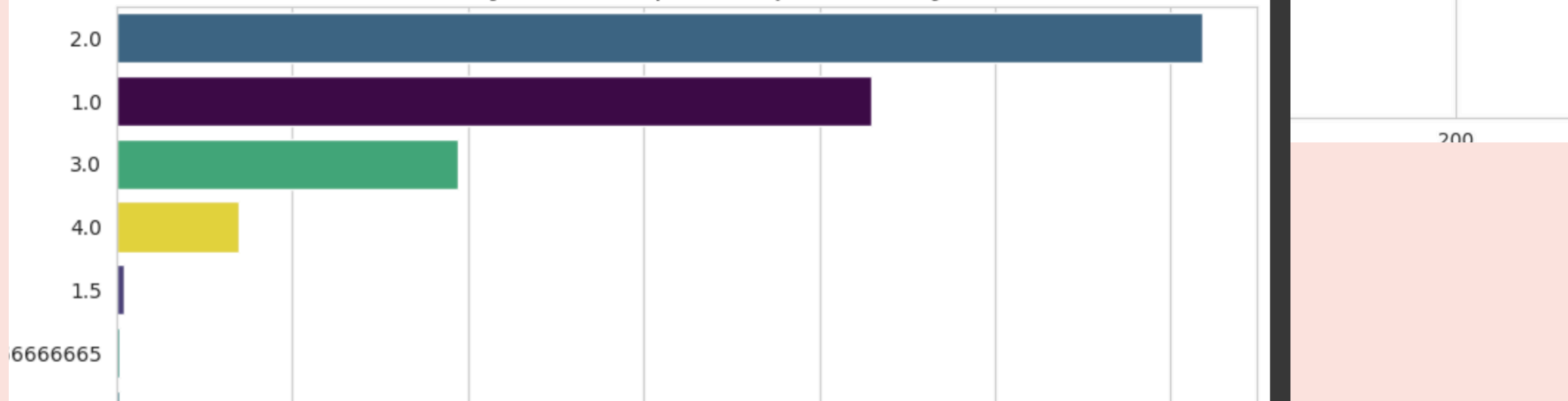
Distribuição de Frequência para: desempenho\_final



Distribuição de Frequência para: goout



Distribuição de Frequência para: studytime



----- Análise da Coluna: 'higher' -----

Contagem Percentual (%)

higher

1	591	89.27
---	-----	-------

0	71	10.73
---	----	-------

----- Análise da Coluna: 'romantic' -----

Contagem Percentual (%)

romantic

0	414	62.54
---	-----	-------

1	248	37.46
---	-----	-------

Análise da Coluna: 'sex'

----- Análise da Coluna: 'paid' -----

Contagem Percentual (%)

paid

0	469	70.85
---	-----	-------

1	193	29.15
---	-----	-------

----- Análise da Coluna: 'famsize' -----

Contagem Percentual (%)

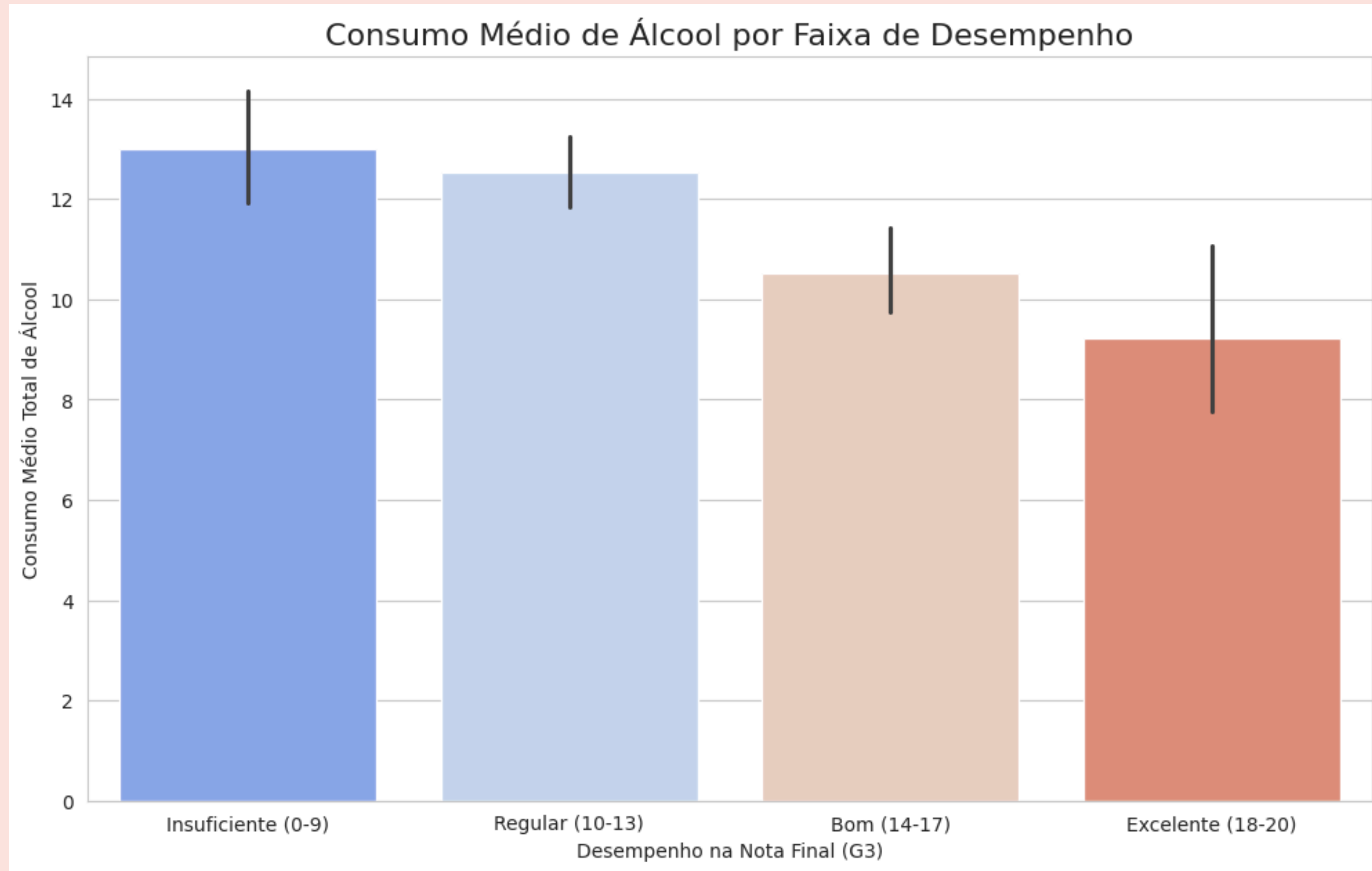
famsize

GT3	461	69.64
-----	-----	-------

LE3	201	30.36
-----	-----	-------

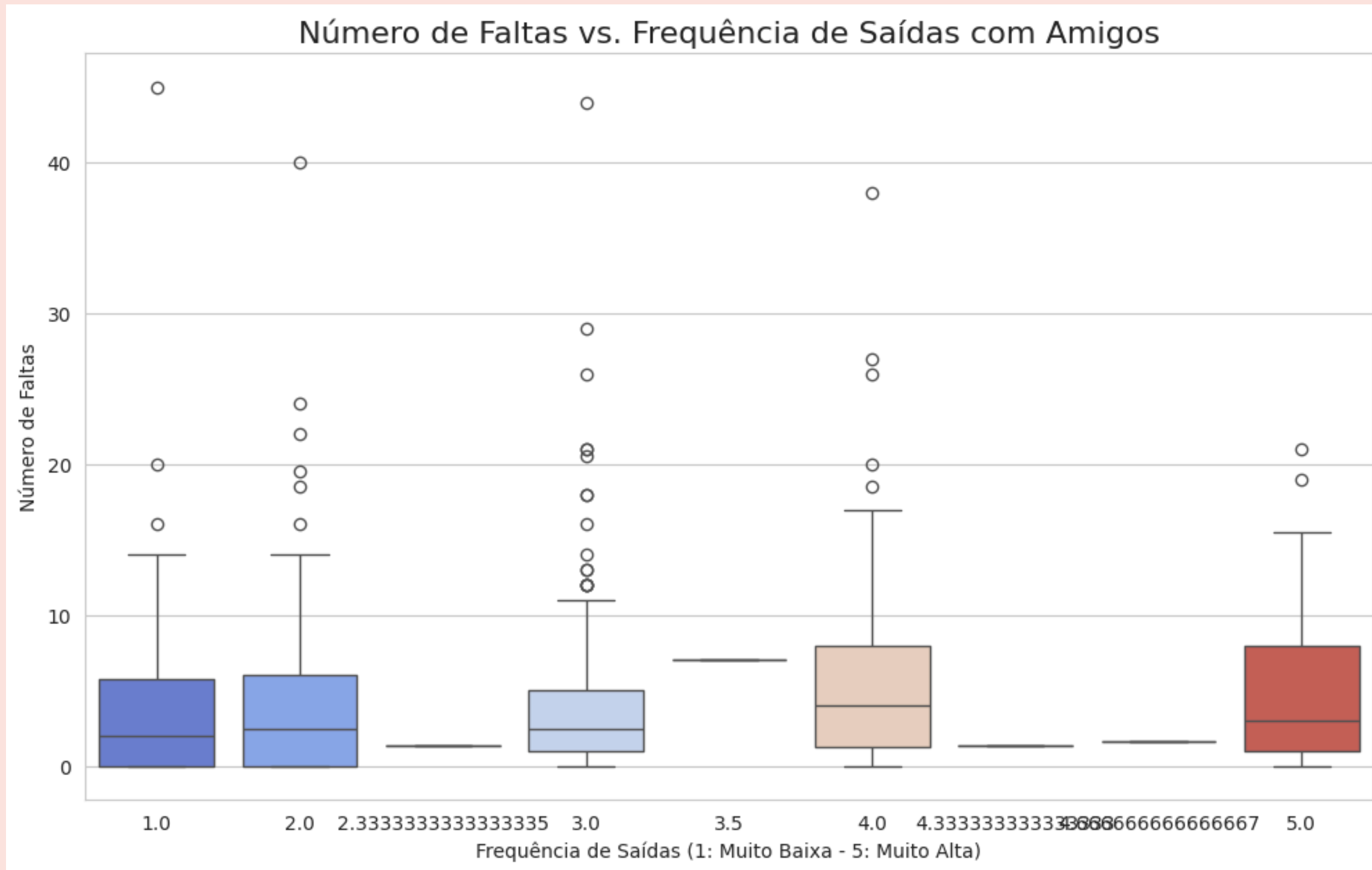
# Análise Bivariada

## 1 Consumo de Álcool vs. Desempenho - Afeta mesmo?



# Análise Bivariada

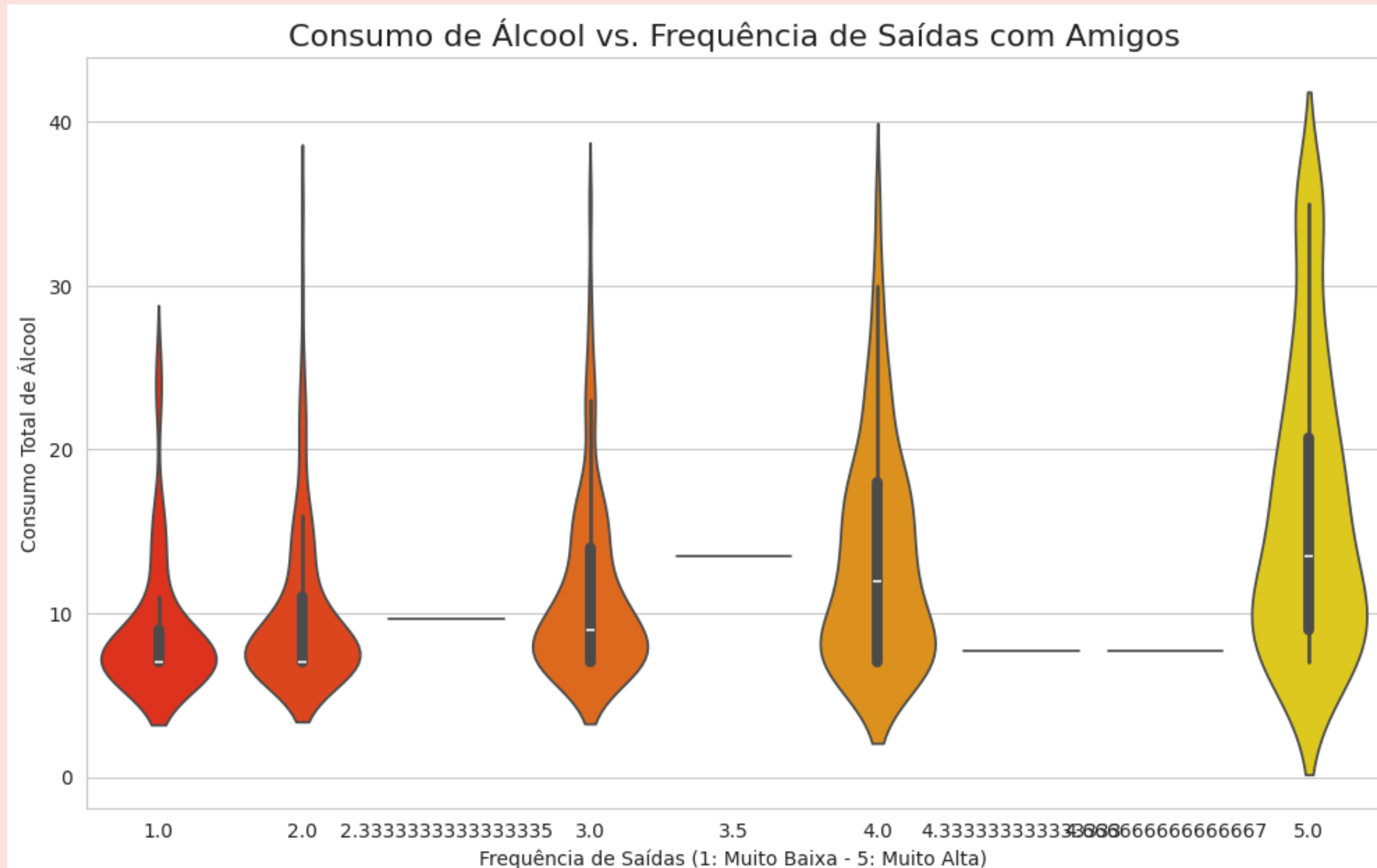
## 2 O Efeito em Cadeia: Vida Social e Faltas





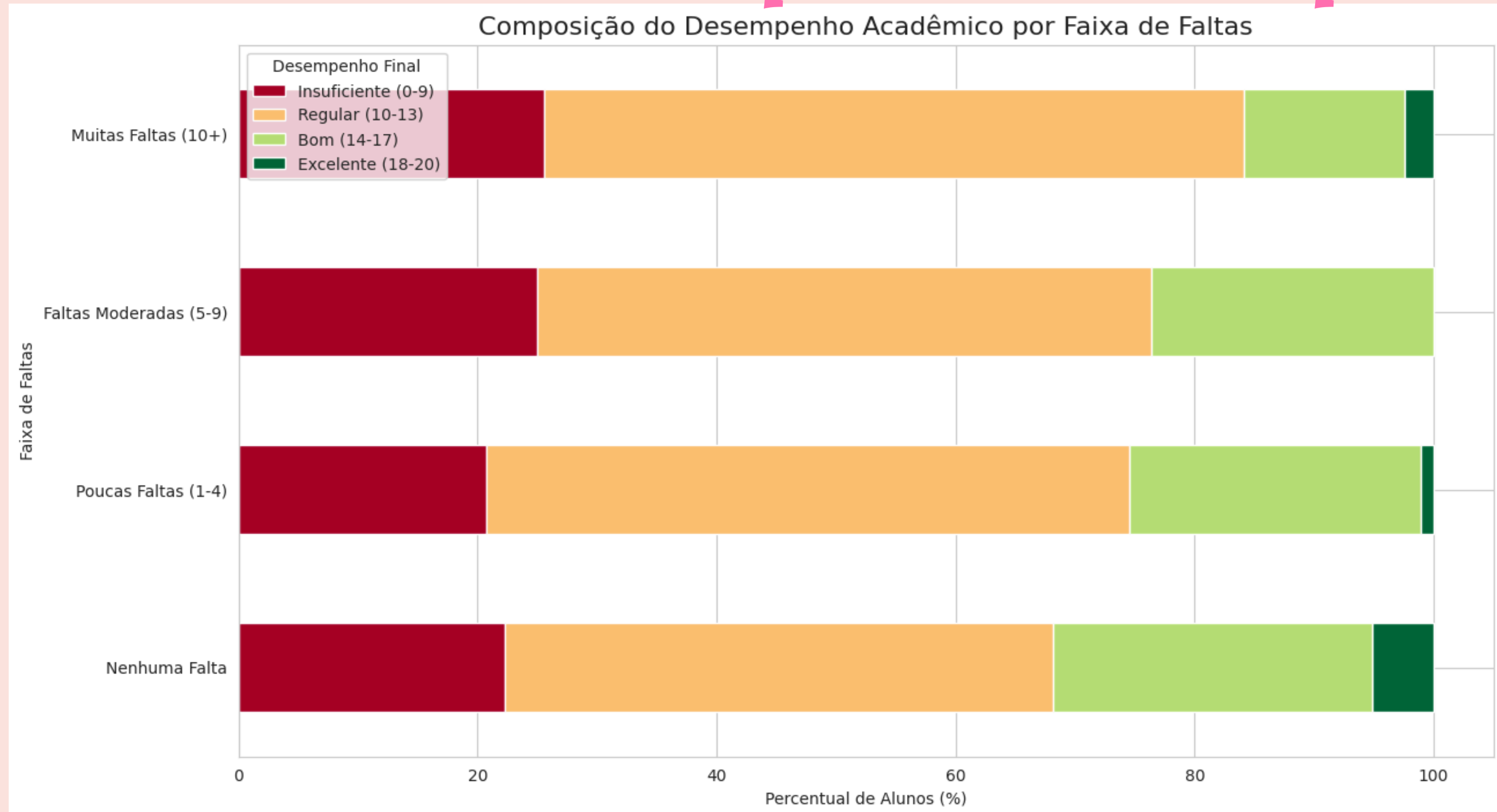
# Análise Bivariada

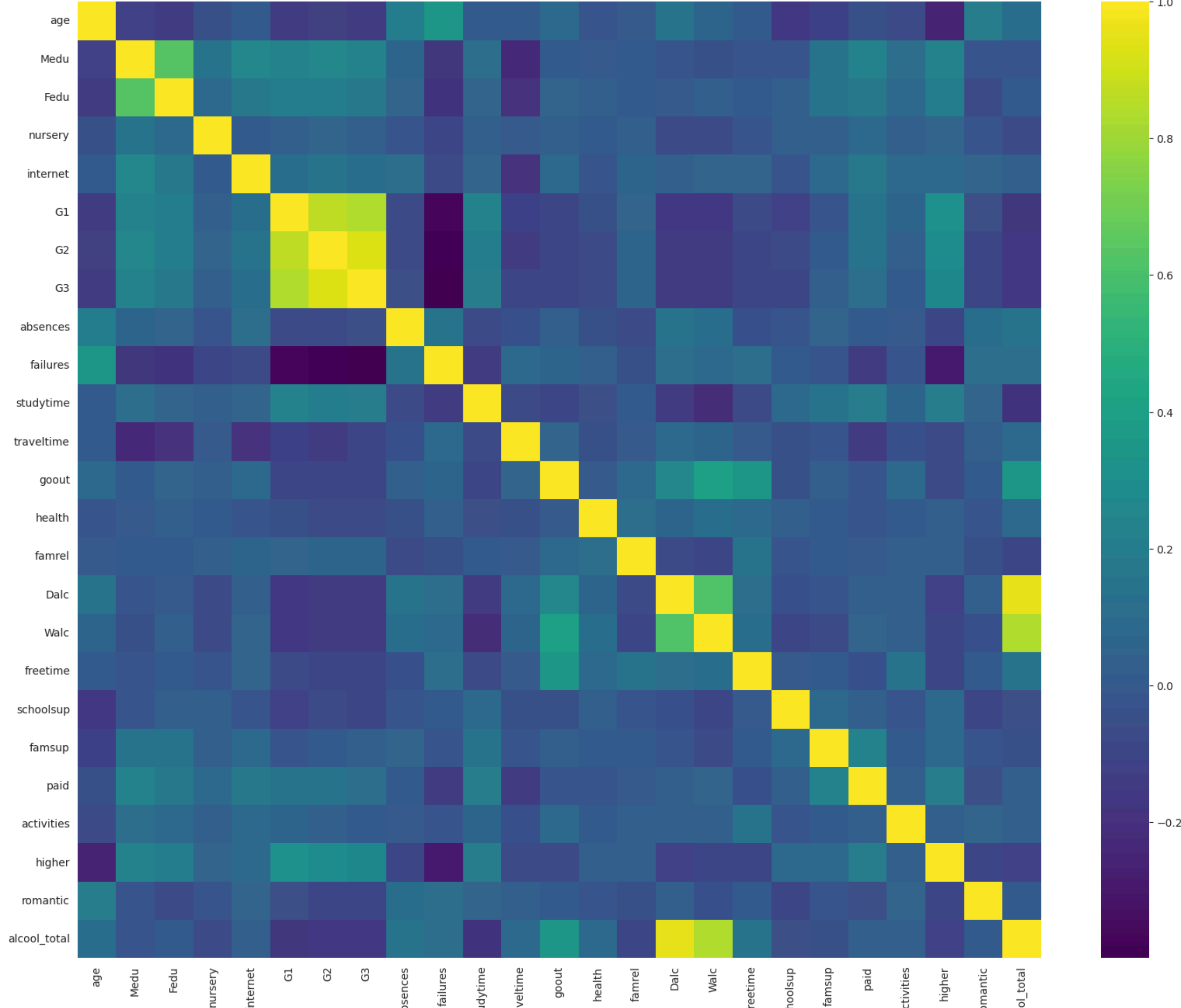
## 3 A vida social relacionada ao consumo de álcool?



# Análise Bivariada

## 4 A assiduidade tem impacto no desempenho ?

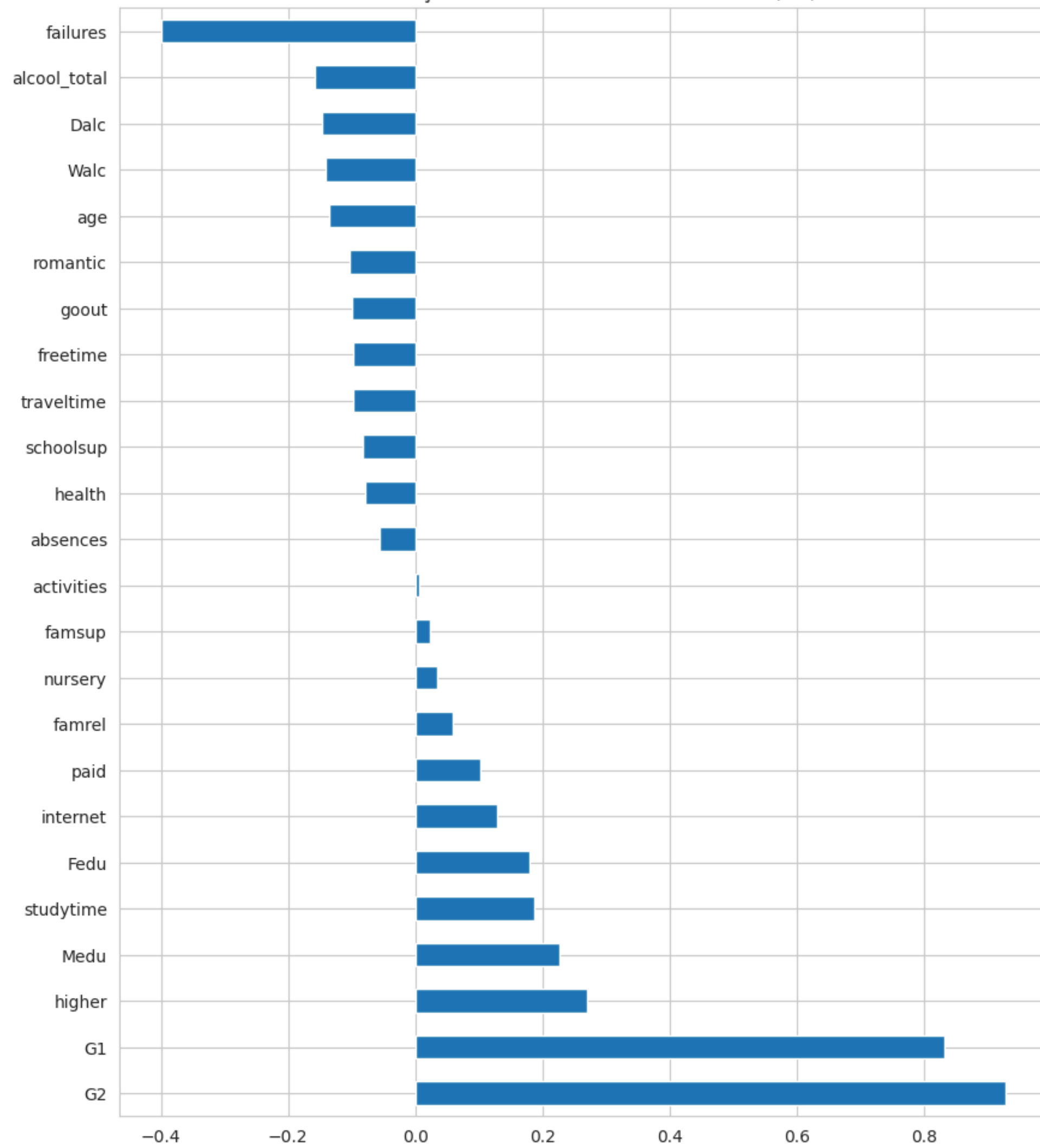




# HEATMAP

- Aspiração e Ambiente Familiar (higher, Medu, Fedu)
- Desempenho Anterior (G1-G3 + failures)
- Uso do tempo
  - freetime vs studytime vs goout

Correlação das Variáveis com a Nota Final (G3)



# Fatores de Sucesso e Risco

## Além do Álcool: O Que Realmente Pesa nas Notas?

As variáveis com maior impacto negativo no desempenho final (G3) são

- histórico de reprovações (failures)
- consumo de álcool (alcohol\_total) = -0.16

Por outro lado, os fatores com maior impacto positivo são:

- a educação dos pais (Medu e Fedu)
- o tempo de estudo (studytime) = 0.19
- e o desejo de cursar o ensino superior (higher)."

**CSF:** Suporte Educacional (schoolsup, -0.08), (famsup, 0.02), (paid, 0.10)

# Fatores de Sucesso e Risco

## Além do Álcool: O Que Realmente Pesa nas Notas?

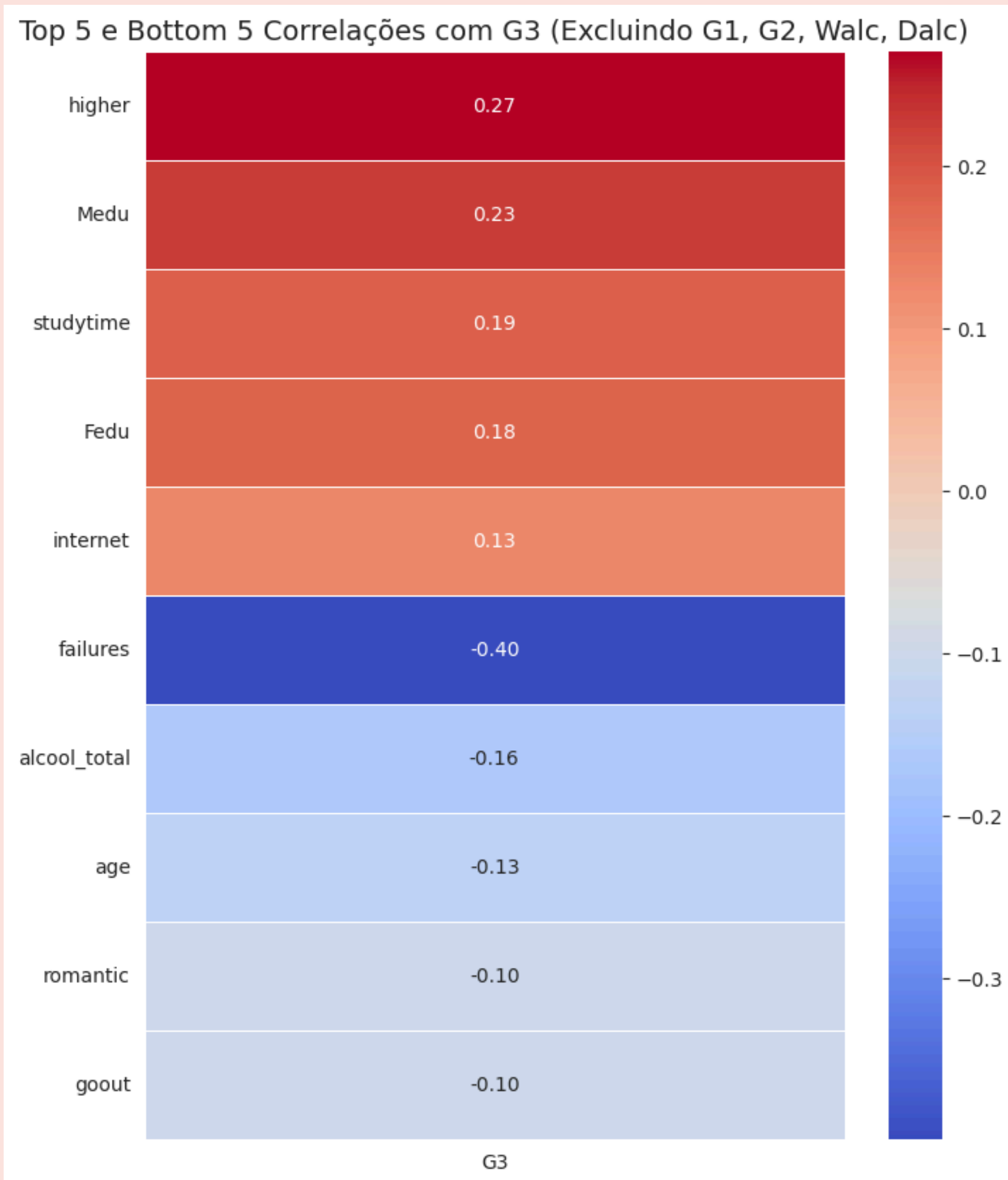
As variáveis com maior impacto negativo no desempenho final (G3) são

- histórico de reprovações (failures)
- consumo de álcool (alcohol\_total) = -0.16

Por outro lado, os fatores com maior impacto positivo são:

- a educação dos pais (Medu e Fedu)
- o tempo de estudo (studytime) = 0.19
- e o desejo de cursar o ensino superior (higher)."

**CSF:** Suporte Educacional (schoolsup, -0.08), (famsup, 0.02), (paid, 0.10)



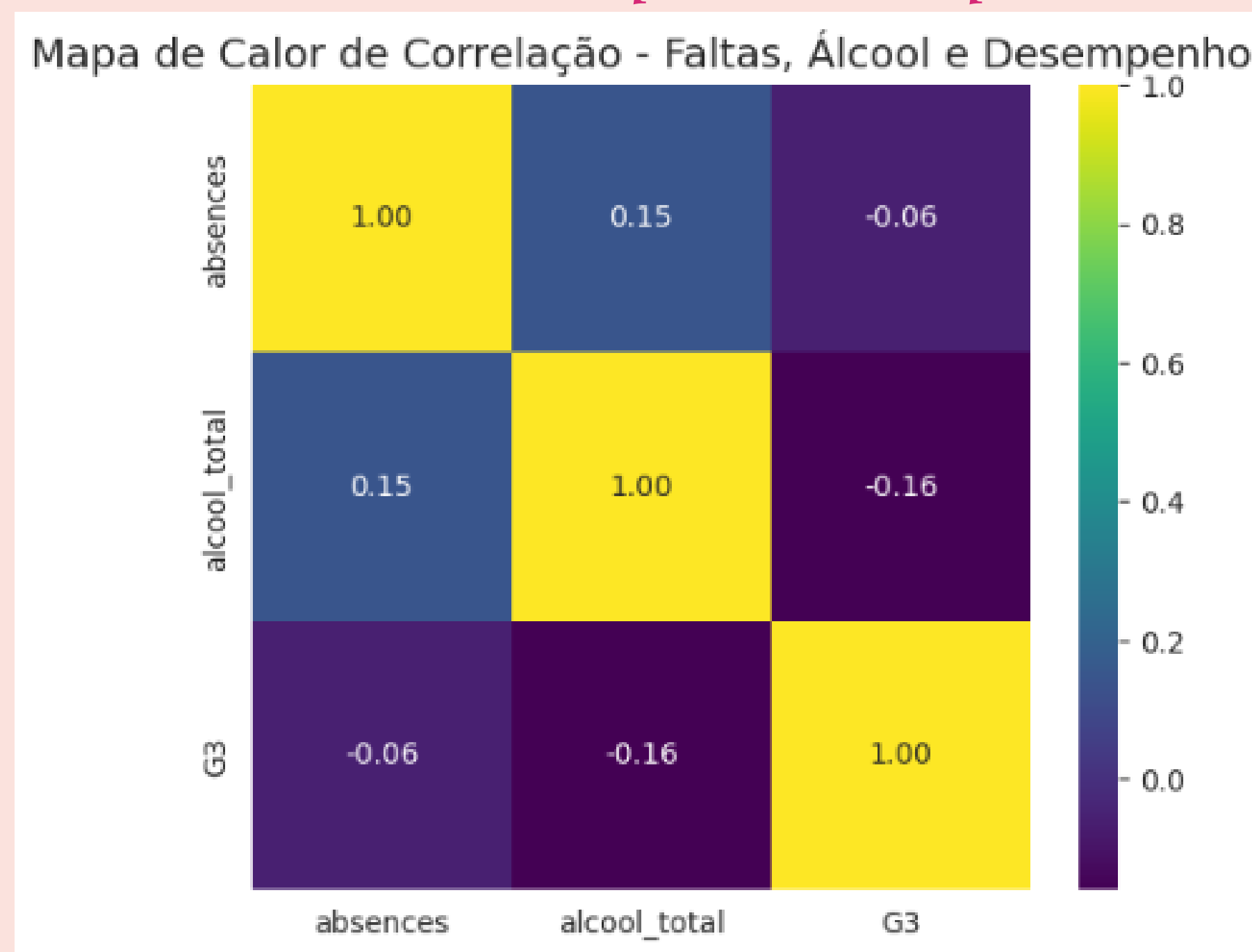


# Perguntas

1. Qual a relação entre o consumo de álcool e o desempenho acadêmico?
2. Como o estilo de vida social (saídas, faltas) se conecta com o desempenho?
3. Quais outros fatores, como o ambiente familiar, influenciam no sucesso?

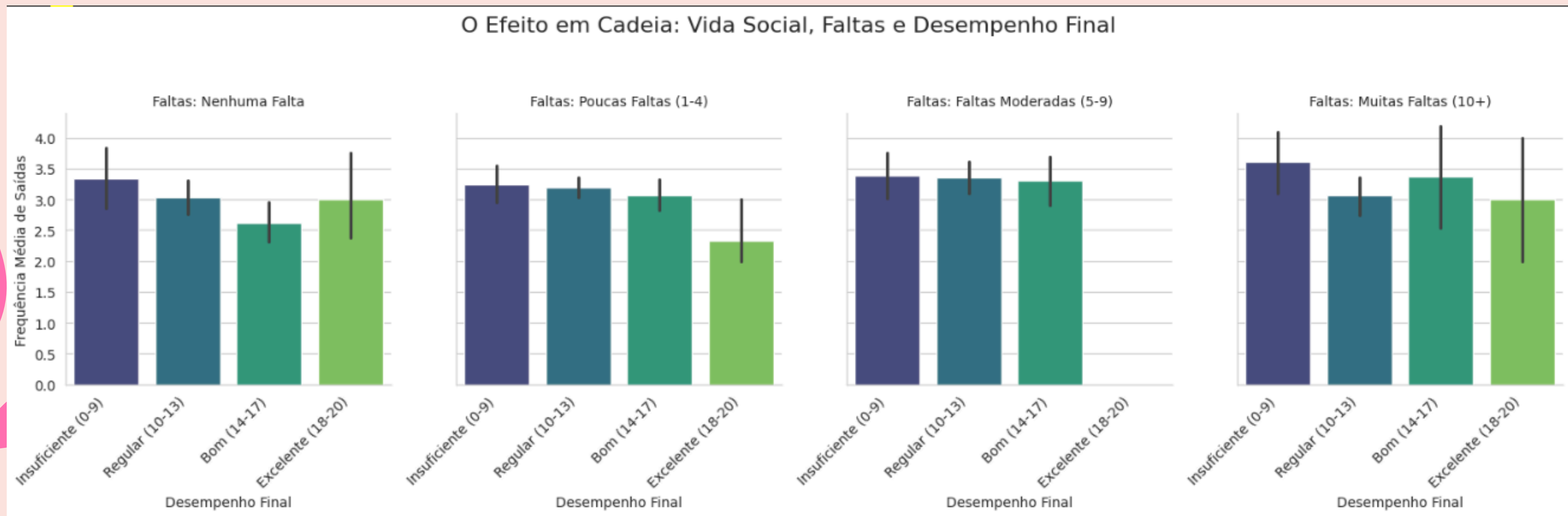
# Resultados

**Relação Negativa Confirmada:** Consumo de álcool e ocorrência de faltas estão associados a um pior desempenho acadêmico.



# Resultados

**A "Cadeia de Risco":** Uma vida social muito ativa se conecta com mais faltas e, por fim, com um desempenho mais baixo.



# Resultados

**Fatores de Sucesso:** Um bom histórico de notas e um ambiente familiar favorável são fatores cruciais para o sucesso.

**Descoberta Inesperada:** A relação das faltas com a aprovação é complexa, impactando mais a qualidade do desempenho (a chance de tirar notas altas).

# Perguntas



**Muito  
Obrigado!**