

Pós-graduação: Especialização em *Data Science* e Inteligência Artificial  
Disciplina: Linguagem de Programação Python  
Turma: 2025/A  
Alunos:  
Celso Rodrigues de Freitas Júnior  
Dayane Cristina Santos Ferreira  
Francisco Acioli Gollo



## Definição do Problema

### Objetivo da análise:

Investigar como os hábitos e características dos estudantes influenciam seu desempenho acadêmico.

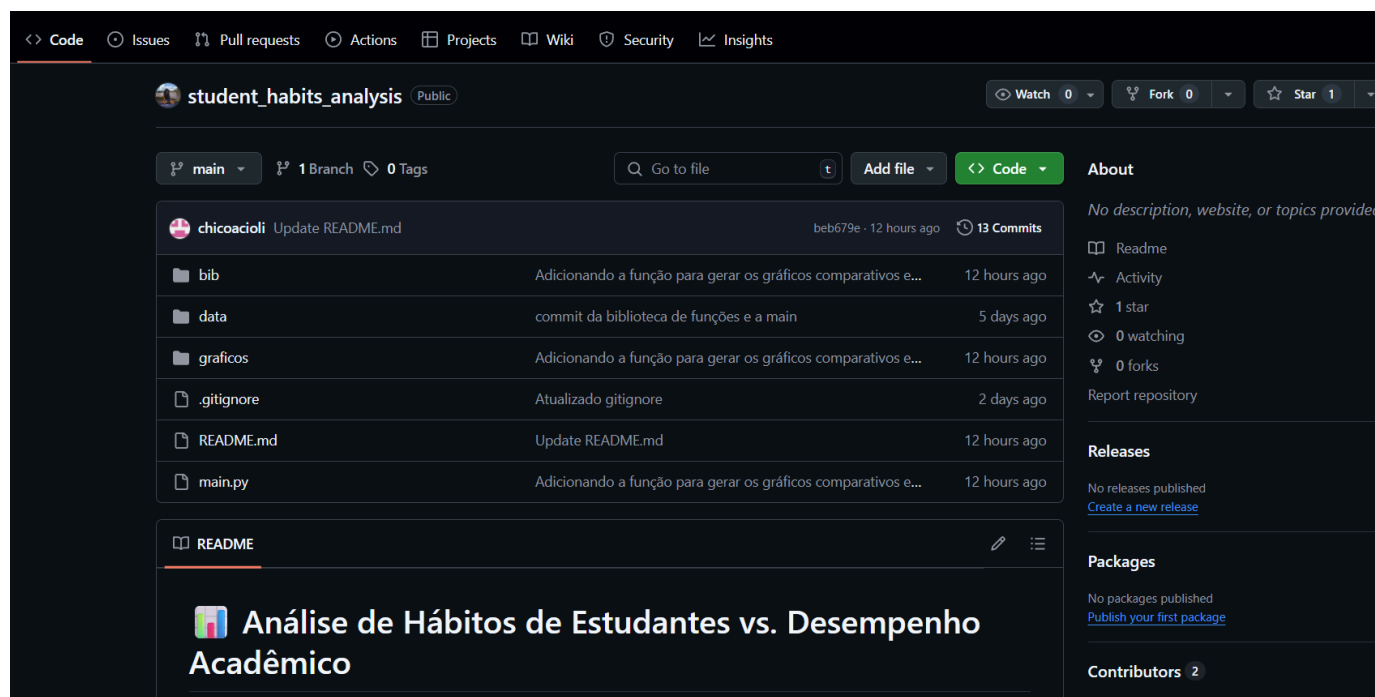
### Perguntas a responder:

Quais hábitos estão mais correlacionados com alto desempenho acadêmico?  
Existem diferenças significativas nos hábitos entre estudantes com alto e baixo desempenho?  
Como fatores como sono, uso de tecnologia e hábitos alimentares afetam as notas?  
Quais variáveis são os melhores preditores do desempenho acadêmico?

### Definir métricas de sucesso:

Identificar pelo menos 3 hábitos significativamente correlacionados com desempenho  
Produzir visualizações claras que comuniquem as relações encontradas

Repositório do Projeto: github: [https://github.com/ladyenay/student\\_habits\\_analysis#](https://github.com/ladyenay/student_habits_analysis#)



## Coleta de Dados

Fonte de dados: <https://www.kaggle.com/datasets/jayaantanaath/student-habits-vs-academic-performance/data>

É um conjunto de dados simulados para explorar os hábitos de vida que afetam o desempenho acadêmico dos alunos. Foi criado com base em padrões realistas para práticas educacionais. Este conjunto simula os hábitos diários de 1000 alunos e os compara com as notas das provas finais.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	student_id	age	gender	study_hours_per_day	social_media_hours	netflix_hours	part_time_job	attendance_percentage	sleep_hours	diet_quality	exercise_frequency	parental_education_level	internet_quality	mental_health_rating	extracurricular_participation	exam_score
2	S1000	23	Female	0.0	1.2	1.1	No	85.0	8.0	Fair	6	Master	Average	8	Yes	56.2
3	S1001	20	Female	6.9	2.8	2.3	No	97.3	4.6	Good	6	High School	Average	8	No	100.0
4	S1002	21	Male	1.4	3.1	1.3	No	94.8	8.0	Poor	1	High School	Poor	1	No	34.3
5	S1003	23	Female	1.0	3.9	1.0	No	71.0	9.2	Poor	4	Master	Good	1	Yes	26.8
6	S1004	19	Female	5.0	4.4	0.5	No	90.9	4.9	Fair	3	Master	Good	1	No	66.4
7	S1005	24	Male	7.2	1.3	0.0	No	82.9	7.4	Fair	1	Master	Average	4	No	100.0
8	S1006	21	Female	5.6	1.5	1.4	Yes	85.8	6.5	Good	2	Master	Poor	4	No	89.8
9	S1007	21	Female	4.3	1.0	2.0	Yes	77.7	4.6	Fair	0	Bachelor	Average	8	No	72.6
10	S1008	23	Female	4.4	2.2	1.7	No	100.0	7.1	Good	3	Bachelor	Good	1	No	78.9
11	S1009	18	Female	4.8	3.1	1.3	No	95.4	7.5	Good	5	Bachelor	Good	10	Yes	100.0
12	S1010	19	Female	4.6	3.7	0.8	No	77.6	5.8	Fair	1	None	Good	3	No	63.3
13	S1011	23	Male	3.9	2.4	2.5	No	71.7	7.9	Fair	2	Bachelor	Average	1	No	74.4
14	S1012	19	Female	3.7	2.1	0.4	Yes	81.1	4.5	Fair	1	Bachelor	Good	9	No	76.9
15	S1013	19	Female	3.4	2.7	2.7	No	89.3	4.7	Fair	4	Bachelor	Good	10	No	75.8
16	S1014	24	Male	2.4	1.5	0.7	No	87.4	6.7	Poor	6	Bachelor	Average	9	No	78.9
17	S1015	21	Male	3.1	5.0	1.0	No	97.5	6.5	Good	6	High School	Average	7	No	74.0
18	S1016	20	Male	1.0	0.6	0.2	No	92.9	5.6	Poor	3	High School	Poor	8	Yes	55.2
19	S1017	24	Female	3.4	2.7	1.2	No	94.7	7.5	Poor	0	High School	Average	1	Yes	70.8
20	S1018	24	Other	2.0	4.9	2.9	Yes	88.3	7.1	Good	2	High School	Good	5	No	43.9
21	S1019	19	Male	1.8	2.5	2.4	No	71.1	7.5	Fair	1	High School	Average	2	No	45.3
22	S1020	22	Female	3.8	2.3	2.9	No	83.0	6.4	Good	3	Master	Good	1	No	58.5
23	S1021	21	Male	5.6	2.1	2.4	No	95.6	7.2	Fair	1	High School	Good	3	No	82.5
24	S1022	18	Other	4.9	2.3	0.6	No	84.5	6.0	Fair	3	High School	Average	7	No	98.7
25	S1023	24	Female	1.1	4.1	1.4	Yes	90.0	9.0	Fair	6	Bachelor	Good	1	No	43.7
26	S1024	20	Female	2.0	0.0	0.9	Yes	81.8	5.5	Fair	4	Bachelor	Average	2	No	54.9
27	S1025	22	Male	4.9	4.3	3.3	No	74.7	9.0	Fair	1	High School	Average	2	Yes	69.9
28	S1026	22	Male	2.0	0.8	0.5	No	83.8	6.5	Good	4	Bachelor	Poor	4	No	73.5
29	S1027	18	Male	3.2	2.2	2.8	Yes	88.1	4.8	Fair	5	Bachelor	Average	3	No	71.1
30	S1028	24	Male	4.3	2.0	2.8	Yes	78.4	8.1	Good	5	High School	Poor	3	Yes	82.8
31	S1029	20	Female	2.0	3.2	3.8	Yes	82.6	6.7	Poor	6	High School	Poor	10	Yes	75.7
32	S1030	21	Female	3.7	0.6	1.3	No	75.6	7.5	Fair	2	Master	Average	5	No	70.6
33	S1031	17	Other	1.5	3.1	2.6	Yes	96.2	8.0	Fair	4	Bachelor	Average	3	Yes	51.3
34	S1032	20	Other	2.6	4.9	4.3	No	82.1	6.9	Good	3	Bachelor	Good	7	No	52.1
35	S1033	18	Female	4.0	0.9	1.1	No	99.5	5.8	Poor	2	None	Average	5	No	70.7
36	S1034	22	Female	1.1	2.7	0.1	No	100.0	7.7	Fair	1	High School	Good	4	No	51.2
37	S1035	21	Female	4.2	1.7	0.0	Yes	84.2	6.5	Good	1	Bachelor	Average	3	No	72.6
38	S1036	20	Female	3.5	3.0	2.6	No	74.9	7.3	Fair	3	High School	Good	4	No	70.7

Conhecendo os 16 campos deste dataset:

Nº	Campo	Descrição	Valores
1	student_id	Chave única de um estudante	S1000 à S1999
2	age	Idade do aluno	17 a 24
3	gender	Gênero do aluno	<ul style="list-style-type: none"><li>Male (Masculino)</li><li>Female (Feminino)</li><li>Other (Outro)</li></ul>
4	study_hours_per_day	Refere-se às horas estudadas por dia	0 a 8,3 horas
5	social_media_hours	Refere-se às horas em redes sociais por dia	0 a 7,2 horas
6	netflix_hours	Quantidade de horas de uso do Netflix por dia	0 a 5,4 horas
7	part_time_job	Apresenta se aluno possui trabalho de meio período	<ul style="list-style-type: none"><li>Yes (Sim)</li><li>No (Não)</li></ul>
8	attendance_percentage	Frequência do aluno no período antes do exame	56% a 100%
9	sleep_hours	Quantidade de horas dormidas por dia	3,2 a 10 horas
10	diet_quality	Qualidade das refeições no período	<ul style="list-style-type: none"><li>Fair (Justa)</li><li>Good (Boa)</li><li>Poor (Pobre)</li></ul>
11	exercise_frequency	Frequência de exercícios por semana	0 a 6 dias
12	parental_education_level	Nível de escolaridade dos pais	<ul style="list-style-type: none"><li>Bachelor (Bacharelado)</li><li>High School (Ensino Médio)</li><li>Master (Mestrado)</li><li>None (Desconhecido)</li></ul>
13	internet_quality	Qualidade da internet do aluno	<ul style="list-style-type: none"><li>Average (média)</li><li>Good (Boa)</li><li>Poor (Ruim)</li></ul>
14	mental_health_rating	Classificação da saúde mental do aluno	1 a 10
15	extracurricular_participation	Participação em atividades extracurriculares	<ul style="list-style-type: none"><li>Yes (Sim)</li><li>No (Não)</li></ul>
16	exam_score	Nota do exame	18,4 a 100 pontos

Limpeza e Pré-processamento (Data Cleaning & Wrangling)

VISÃO GERAL DOS DADOS:

Antes	Depois
<pre>&lt;class 'pandas.core.frame.DataFrame'&gt; RangeIndex: 1000 entries, 0 to 999 Data columns (total 16 columns): #   Column                                Non-Null Count  Dtype ---  --- 0   student_id                           1000 non-null   object 1   age                                   1000 non-null   int64 2   gender                               1000 non-null   object 3   study_hours_per_day                  1000 non-null   float64 4   social_media_hours                   1000 non-null   float64 5   netflix_hours                       1000 non-null   float64 6   part_time_job                       1000 non-null   object 7   attendance_percentage                1000 non-null   float64 8   sleep_hours                         1000 non-null   float64 9   diet_quality                        1000 non-null   object 10  exercise_frequency                  1000 non-null   int64 11  parental_education_level            909 non-null    object 12  internet_quality                    1000 non-null   object 13  mental_health_rating                1000 non-null   int64 14  extracurricular_participation       1000 non-null   object 15  exam_score                          1000 non-null   float64 dtypes: float64(6), int64(3), object(7) memory usage: 125.1+ KB None</pre>	<pre>VISÃO GERAL DOS DADOS: &lt;class 'pandas.core.frame.DataFrame'&gt; RangeIndex: 1000 entries, 0 to 999 Data columns (total 15 columns): #   Column                                Non-Null Count  Dtype ---  --- 0   age                                   1000 non-null   int64 1   gender                               1000 non-null   int64 2   study_hours_per_day                 1000 non-null   float64 3   social_media_hours                  1000 non-null   float64 4   netflix_hours                      1000 non-null   float64 5   part_time_job                      1000 non-null   int64 6   attendance_percentage                1000 non-null   float64 7   sleep_hours                        1000 non-null   float64 8   diet_quality                        1000 non-null   int64 9   exercise_frequency                  1000 non-null   int64 10  parental_education_level            1000 non-null   int64 11  internet_quality                    1000 non-null   int64 12  mental_health_rating                1000 non-null   int64 13  extracurricular_participation       1000 non-null   int64 14  exam_score                          1000 non-null   float64 dtypes: float64(6), int64(9) memory usage: 117.3 KB None</pre>

VISUALIZANDO OS 5 PRIMEIROS REGISTROS:

Antes	Depois
<pre>student_id age gender study_hours_per_day social_media_hours ... parental_education_level internet_quality mental_health_rating extracurricular_participation exam_score 0   S1000 23 Female 0.0 1.2 ... Master Average 8 Yes 56.2 1   S1001 20 Female 6.9 2.8 ... High School Average 8 No 100.0 2   S1002 21 Male 1.4 3.1 ... High School Poor 1 No 34.3 3   S1003 23 Female 1.0 3.9 ... Master Good 1 Yes 26.8 4   S1004 19 Female 5.0 4.4 ... Master Good 1 No 66.4  [5 rows x 16 columns]</pre>	<pre>age gender study_hours_per_day social_media_hours netflix_hours ... parental_education_level internet_quality mental_health_rating extracurricular_participation exam_score 0 23 0 0.0 1.2 1.1 ... 2 0 8 1 56.2 1 20 0 6.9 2.8 2.3 ... 1 0 8 0 100.0 2 21 1 1.4 3.1 1.3 ... 1 2 1 0 34.3 3 23 0 1.0 3.9 1.0 ... 2 1 1 1 26.8 4 19 0 5.0 4.4 0.5 ... 2 1 1 0 66.4  [5 rows x 15 columns]</pre>

VISUALIZANDO OS 5 ÚLTIMOS REGISTROS:

Antes	Depois
<pre>student_id age gender study_hours_per_day social_media_hours ... parental_education_level internet_quality mental_health_rating extracurricular_participation exam_score 995 S1995 21 Female 2.6 0.5 ... High School Good 6 Yes 76.1 996 S1996 17 Female 2.9 1.0 ... High School Average 6 Yes 65.9 997 S1997 20 Male 3.0 2.6 ... Bachelor Good 9 Yes 64.4 998 S1998 24 Male 5.4 4.1 ... Bachelor Average 1 No 69.7 999 S1999 19 Female 4.3 2.9 ... Bachelor Average 8 No 74.9  [5 rows x 16 columns]</pre>	<pre>age gender study_hours_per_day social_media_hours netflix_hours ... parental_education_level internet_quality mental_health_rating extracurricular_participation exam_score 995 21 0 2.6 0.5 1.6 ... 1 1 6 1 996 17 0 2.9 1.0 2.4 ... 1 0 6 1 997 20 1 3.0 2.6 1.3 ... 0 1 9 1 998 24 1 5.4 4.1 1.1 ... 0 0 1 0 999 19 0 4.3 2.9 1.9 ... 0 0 8 0  [5 rows x 15 columns]</pre>

DESCRIÇÃO ESTATÍSTICA DOS DADOS:

Antes	<pre>count  age  study_hours_per_day  social_media_hours  netflix_hours  attendance_percentage  sleep_hours  exercise_frequency  mental_health_rating  exam_score mean  20.4980  3.55010  2.505500  1.819700  84.131700  6.470100  3.042000  5.438000  69.601500 std    2.3081  1.46889  1.172422  1.075118  9.399246  1.226377  2.025423  2.847501  16.888564 min    17.0000  0.00000  0.000000  0.000000  56.000000  3.200000  0.000000  1.000000  18.400000 25%    18.7500  2.60000  1.700000  1.000000  78.000000  5.600000  1.000000  3.000000  58.475000 50%    20.0000  3.50000  2.500000  1.800000  84.400000  6.500000  3.000000  5.000000  70.500000 75%    23.0000  4.50000  3.300000  2.525000  91.025000  7.300000  5.000000  8.000000  81.325000 max    24.0000  8.30000  7.200000  5.400000  100.000000  10.000000  6.000000  10.000000  100.000000</pre>
Depois	<pre>count  age  gender  study_hours_per_day  social_media_hours  ...  internet_quality  mental_health_rating  extracurricular_participation  exam_score mean  20.4980  0.561000  3.55010  2.505500  ...  0.771000  5.438000  0.318000  69.601500 std    2.3081  0.574987  1.46889  1.172422  ...  0.707856  2.847501  0.465932  16.888564 min    17.0000  0.000000  0.00000  0.000000  ...  0.000000  1.000000  0.000000  18.400000 25%    18.7500  0.000000  2.60000  1.700000  ...  0.000000  3.000000  0.000000  58.475000 50%    20.0000  1.000000  3.50000  2.500000  ...  1.000000  5.000000  0.000000  70.500000 75%    23.0000  1.000000  4.50000  3.300000  ...  1.000000  8.000000  1.000000  81.325000 max    24.0000  2.000000  8.30000  7.200000  ...  2.000000  10.000000  1.000000  100.000000</pre>

TOTAL DE DADOS DUPLICADOS:

Antes	Depois
TOTAL DE DADOS DUPLICADOS: 0	TOTAL DE DADOS DUPLICADOS: 0

LINHA x COLUNA:

Antes	Depois
LINHA x COLUNA: (1000, 16)	LINHA x COLUNA: (1000, 15)

TOTAL DE VALORES NULOS:

Preenchendo valores da coluna 12 (parental\_education\_level) de None para Untold

Antes	Depois
<pre>TOTAL DE VALORES NULOS: student_id      0 age             0 gender          0 study_hours_per_day  0 social_media_hours  0 netflix_hours   0 part_time_job   0 attendance_percentage  0 sleep_hours     0 diet_quality    0 exercise_frequency  0 parental_education_level  91 internet_quality  0 mental_health_rating  0 extracurricular_participation  0 exam_score      0 dtype: int64</pre>	<pre>TOTAL DE VALORES NULOS: age             0 gender          0 study_hours_per_day  0 social_media_hours  0 netflix_hours   0 part_time_job   0 attendance_percentage  0 sleep_hours     0 diet_quality    0 exercise_frequency  0 parental_education_level  0 internet_quality  0 mental_health_rating  0 extracurricular_participation  0 exam_score      0 dtype: int64</pre>

TIPOS DE DADOS:

Antes	Depois
<pre>TIPOS DE DADOS: student_id      object age             int64 gender          object study_hours_per_day float64 social_media_hours float64 netflix_hours   float64 part_time_job   object attendance_percentage float64 sleep_hours     float64 diet_quality     object exercise_frequency int64 parental_education_level object internet_quality object mental_health_rating int64 extracurricular_participation object exam_score      float64 dtype: object</pre>	<pre>TIPOS DE DADOS: age             int64 gender          int64 study_hours_per_day float64 social_media_hours float64 netflix_hours   float64 part_time_job   int64 attendance_percentage float64 sleep_hours     float64 diet_quality     int64 exercise_frequency int64 parental_education_level int64 internet_quality int64 mental_health_rating int64 extracurricular_participation int64 exam_score      float64 dtype: object PS C:\student_habits_analysis&gt; █</pre>

VALORES DAS COLUNAS CATEGÓRICAS:

Coluna	Valores	Mapeamento
GÊNERO	['Female' 'Male' 'Other']	Female → 0 Male → 1 Other → 2
TRABALHA	['No' 'Yes']	No → 0 Yes → 1
DIETA	['Fair' 'Good' 'Poor']	Fair → 0 Good → 1 Poor → 2
NÍVEL DE EDUCAÇÃO PARENTAL	['Master' 'High School' 'Bachelor' nan]	Bachelor → 0 High School → 1 Master → 2
QUALIDADE DA INTERNET	['Average' 'Poor' 'Good']	Average → 0 Good → 1 Poor → 2
PARTICIPAÇÃO EXTRACURRICULAR	['No' 'Yes']	No → 0 Yes → 1

GRÁFICO DE BARRA

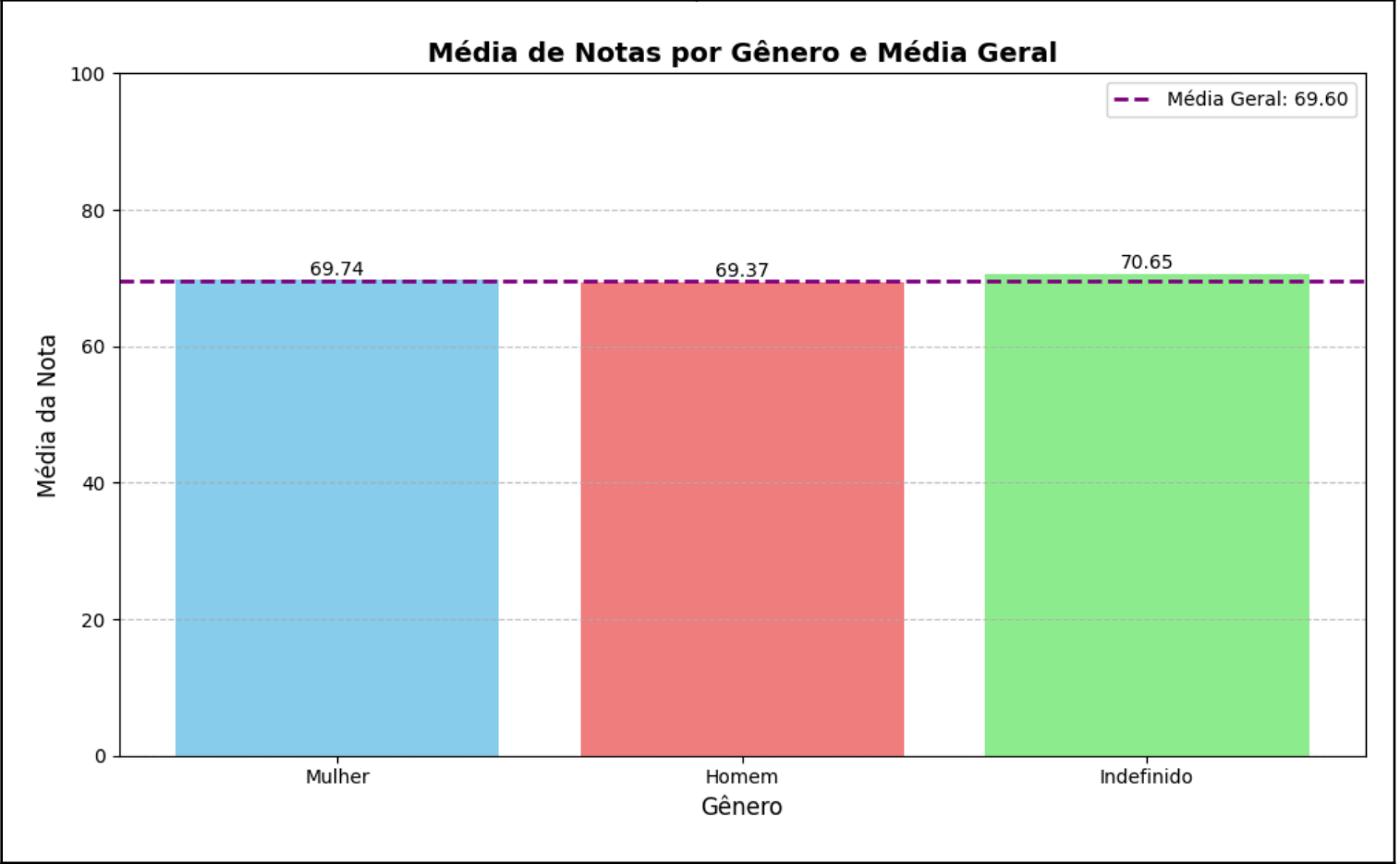
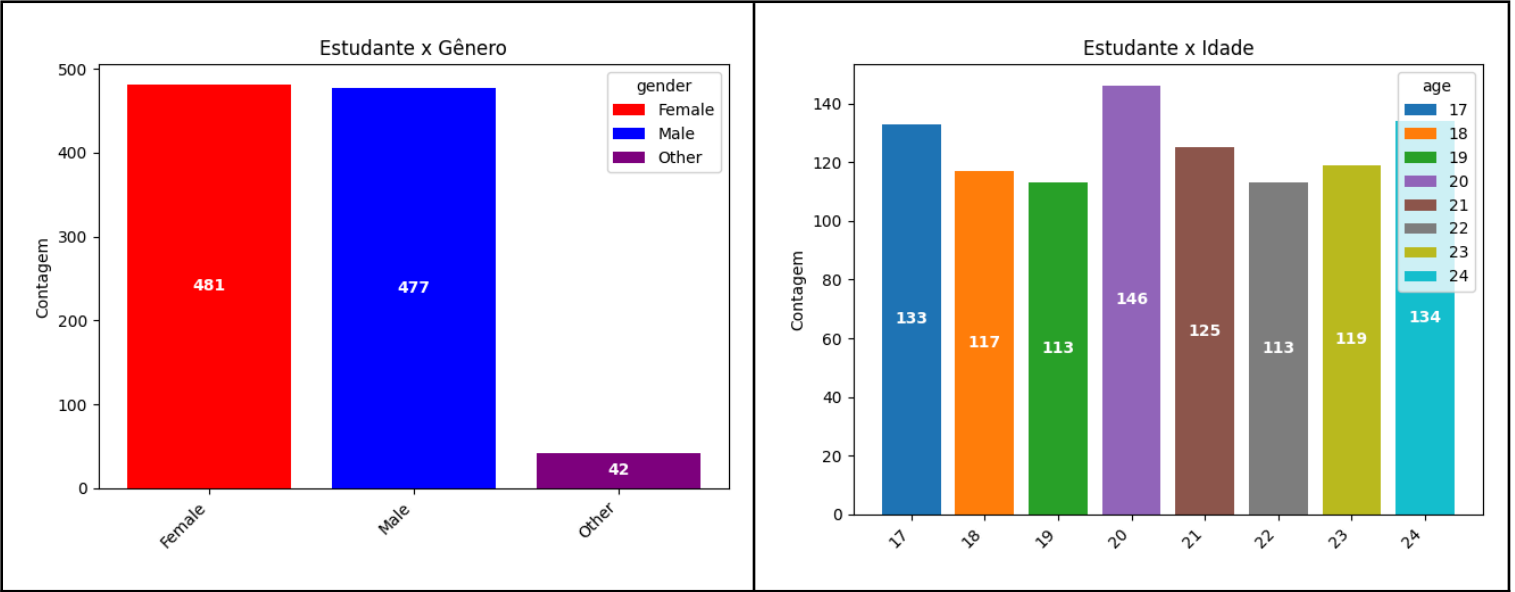


GRÁFICO DE PIZZA

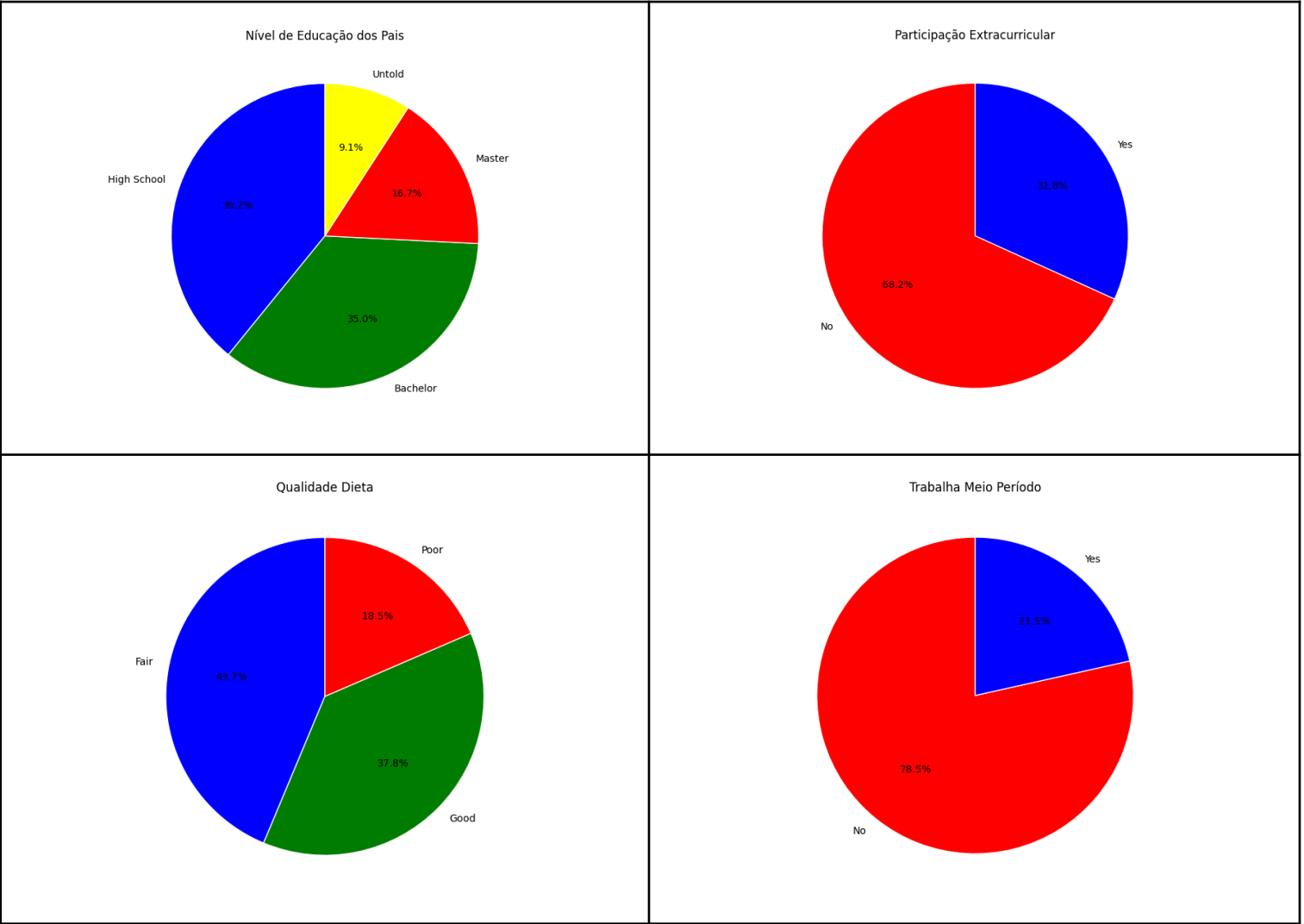
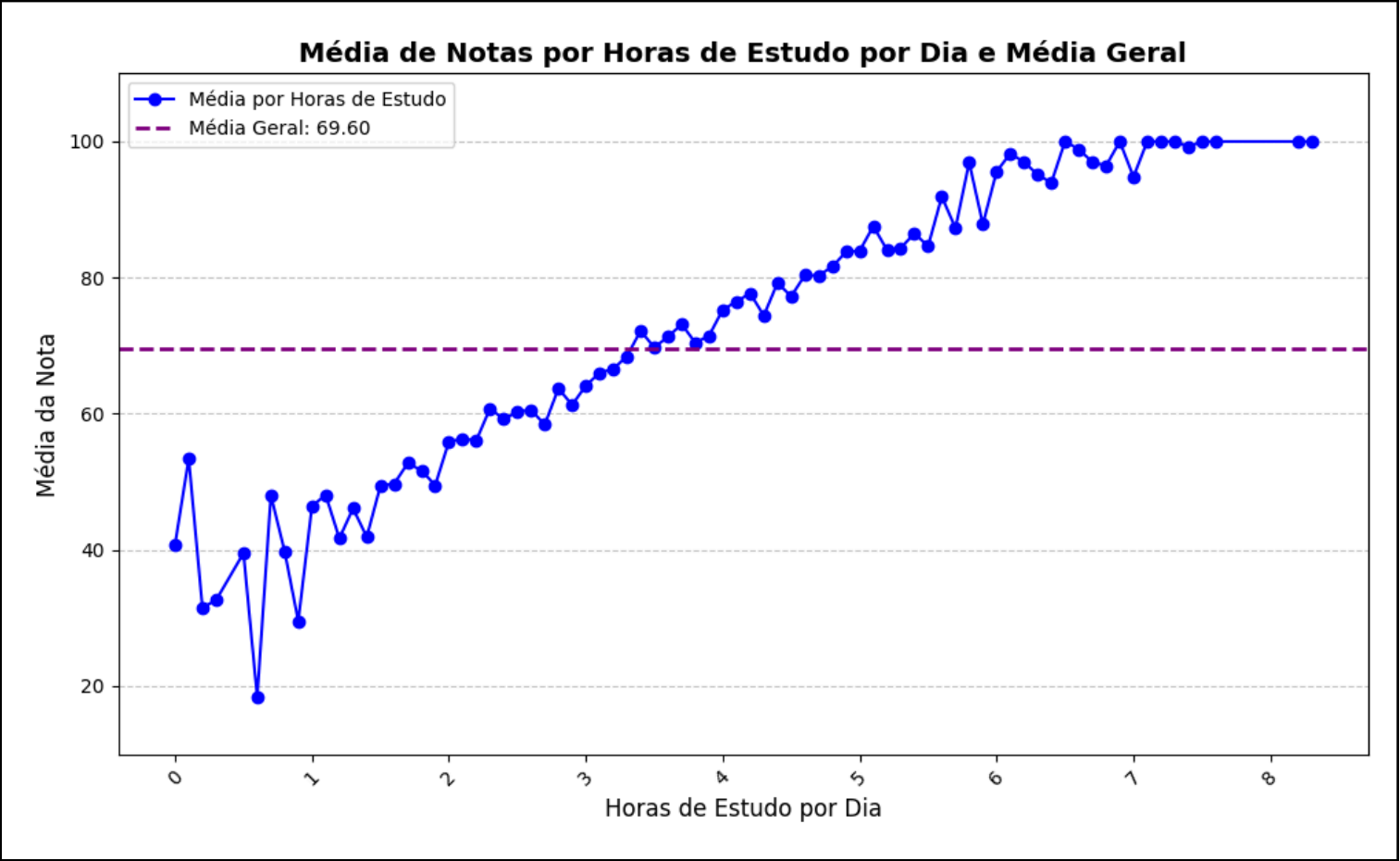
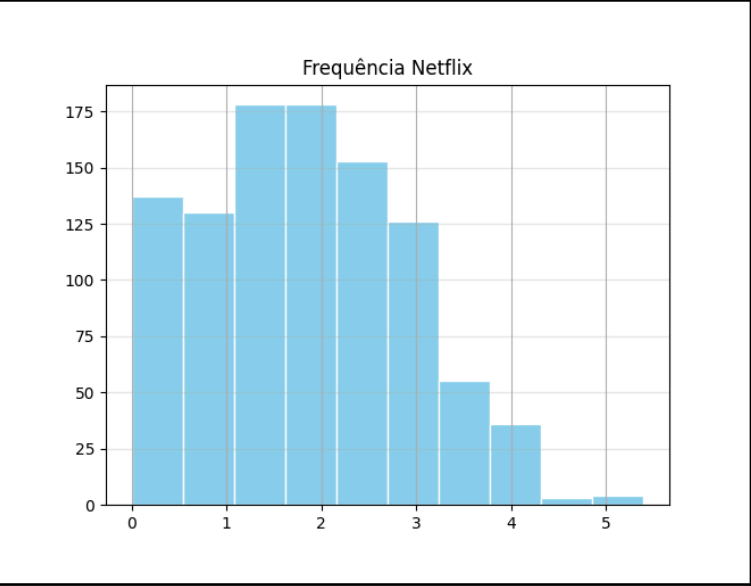
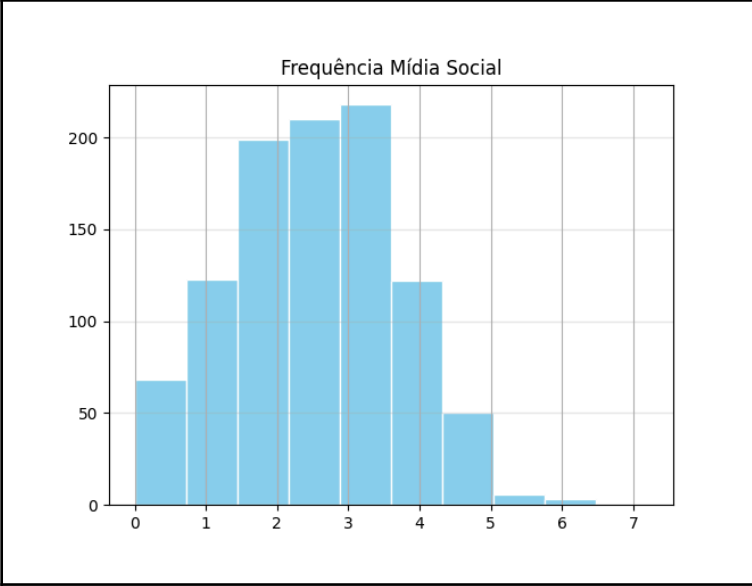
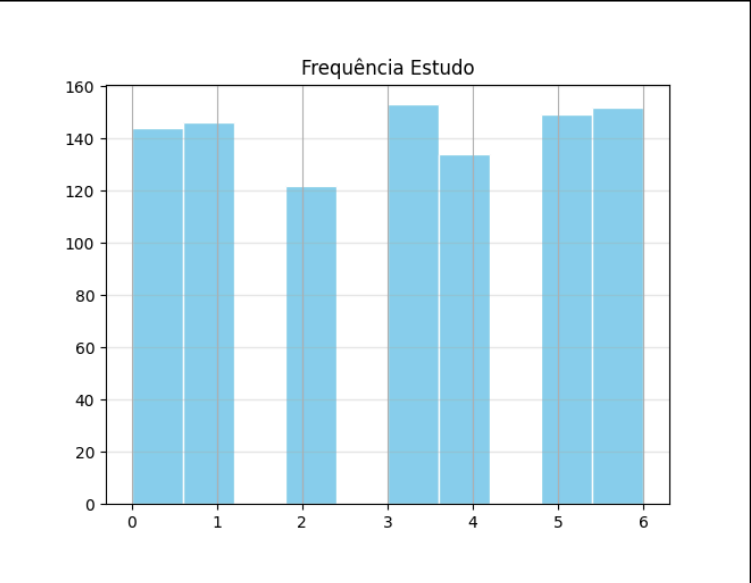
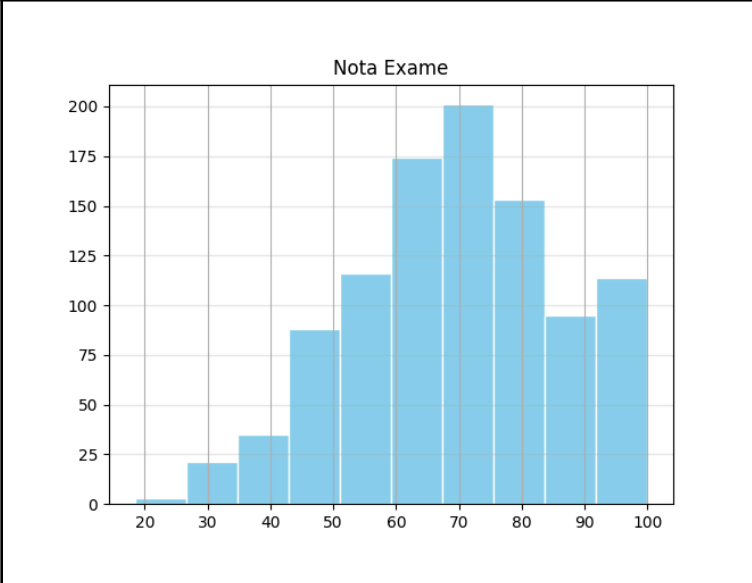
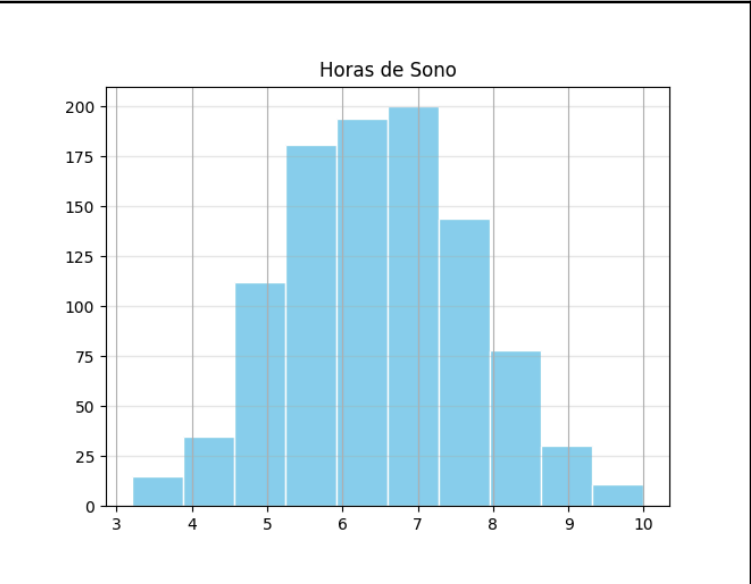
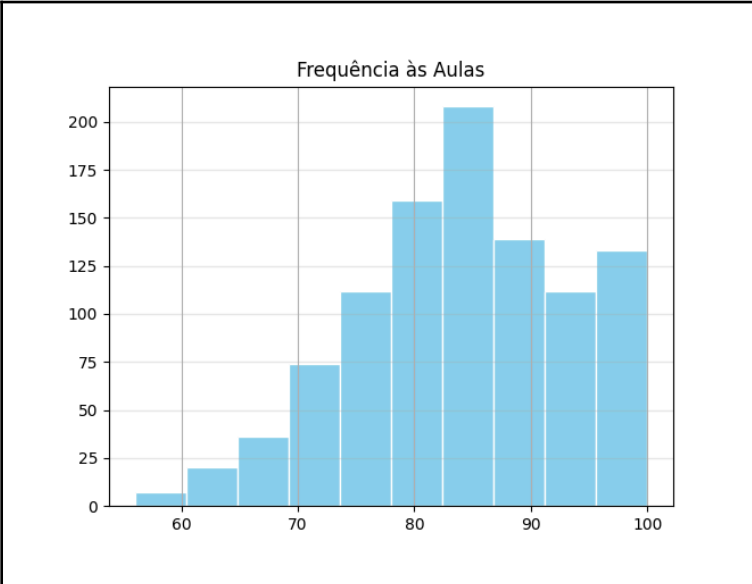


GRÁFICO DE LINHA

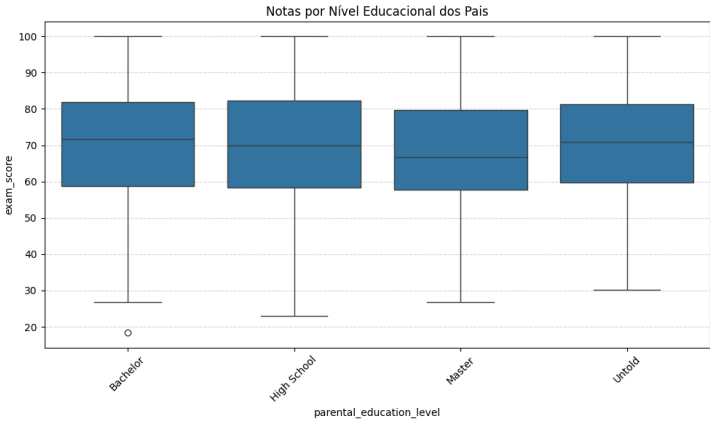


HISTOGRAMA

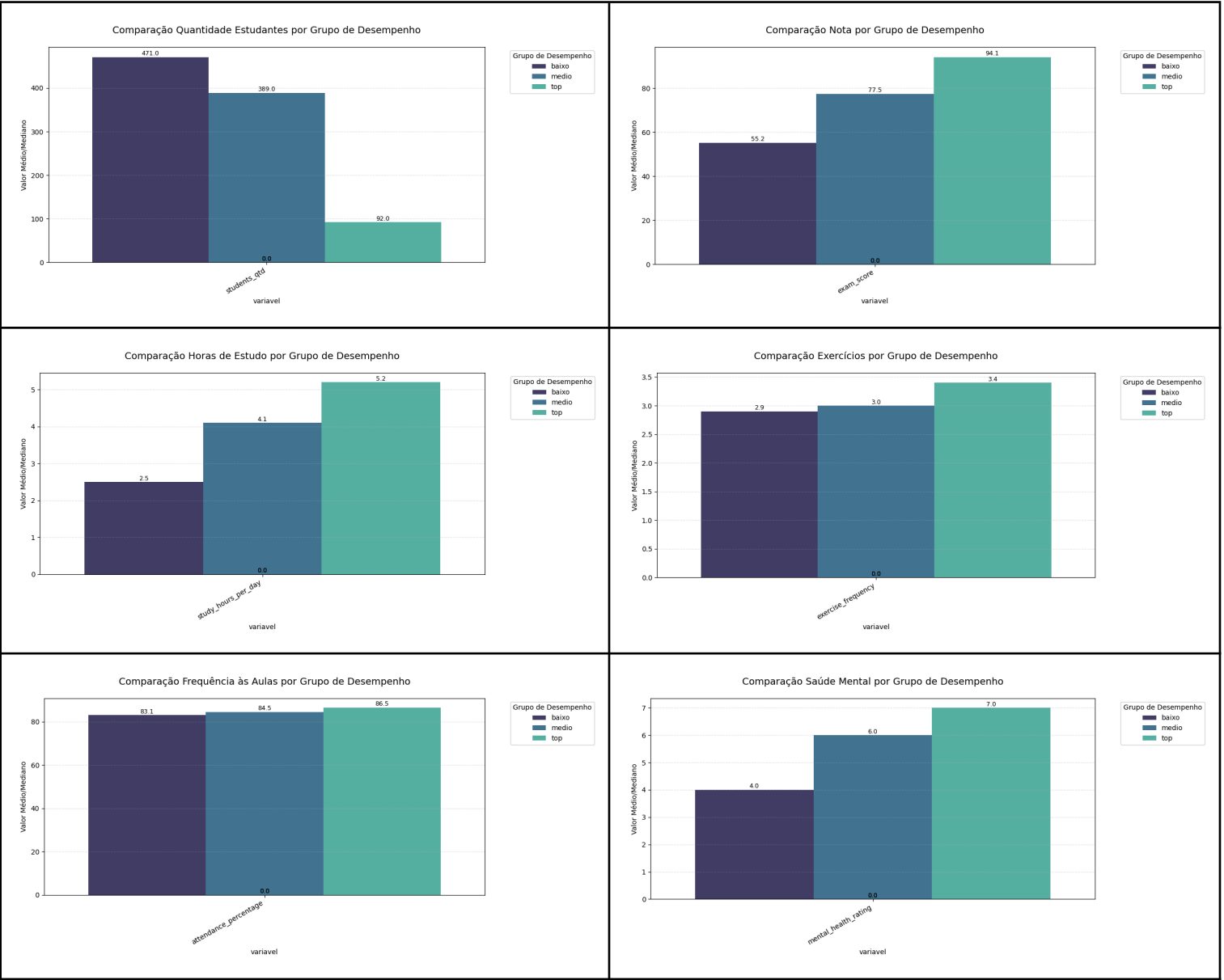


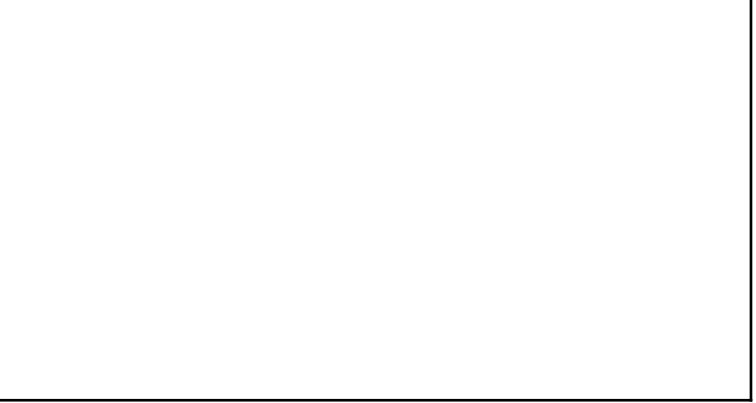
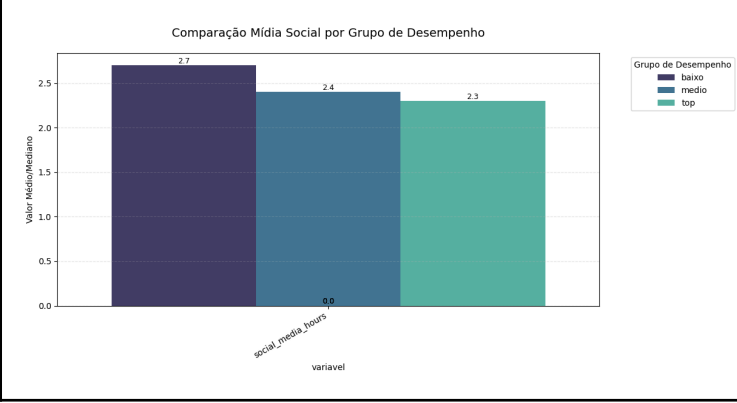
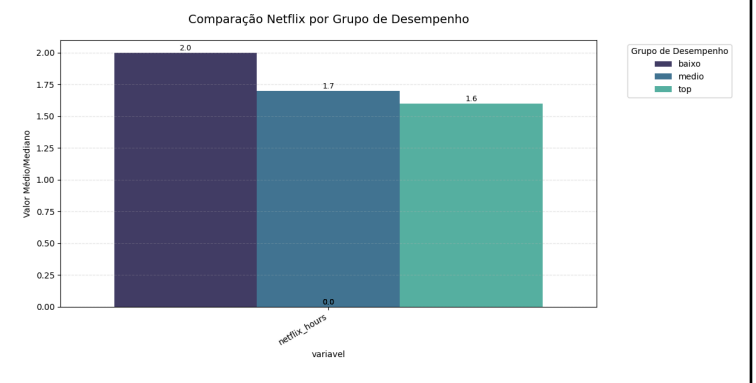
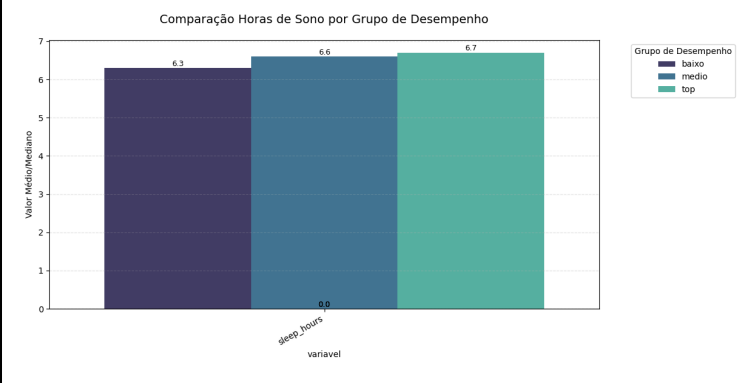


BOXPLOT

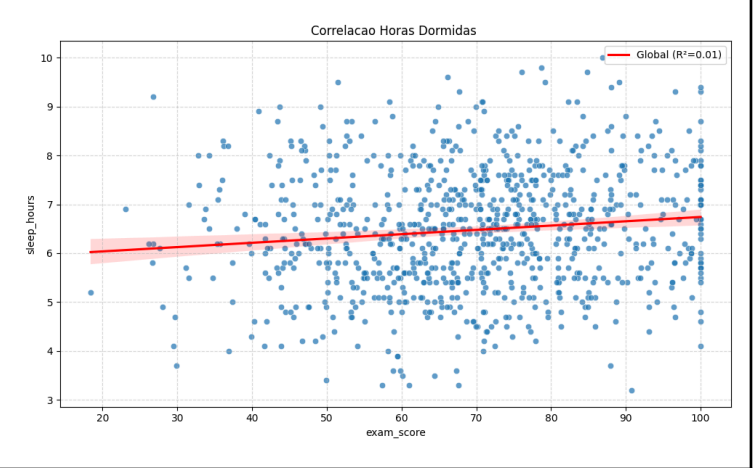
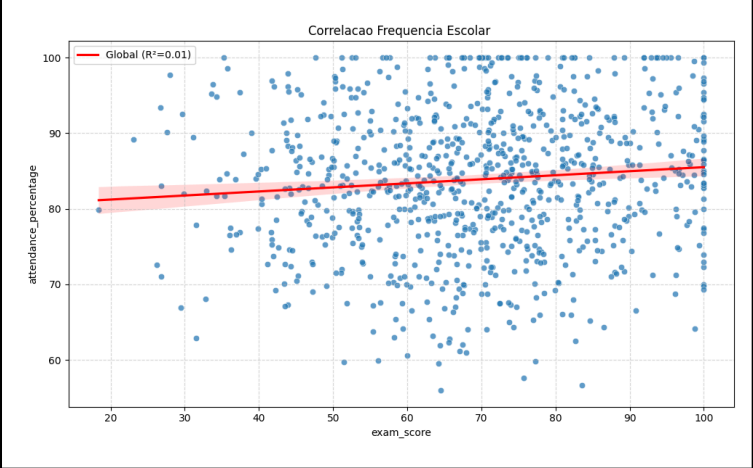
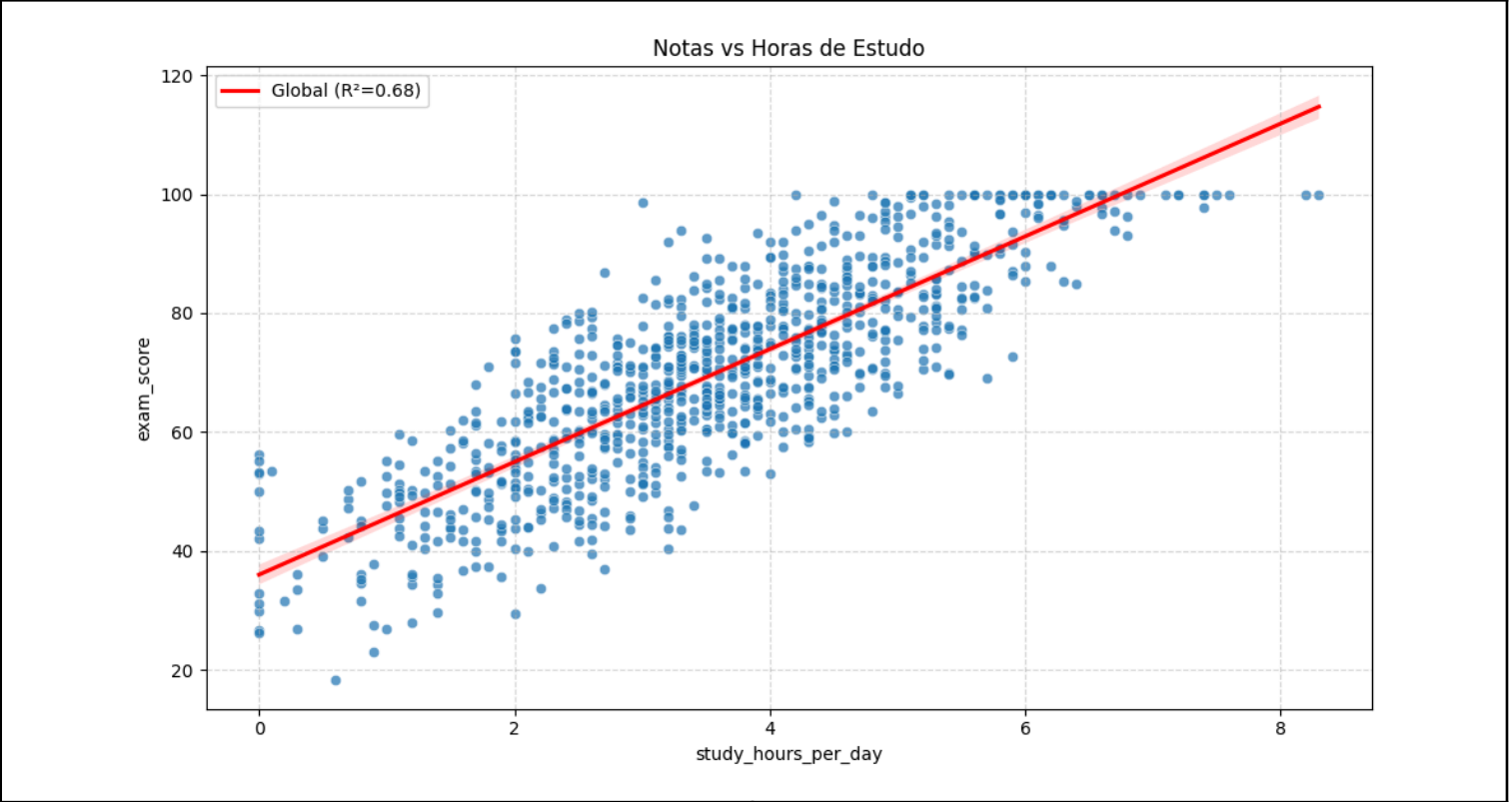


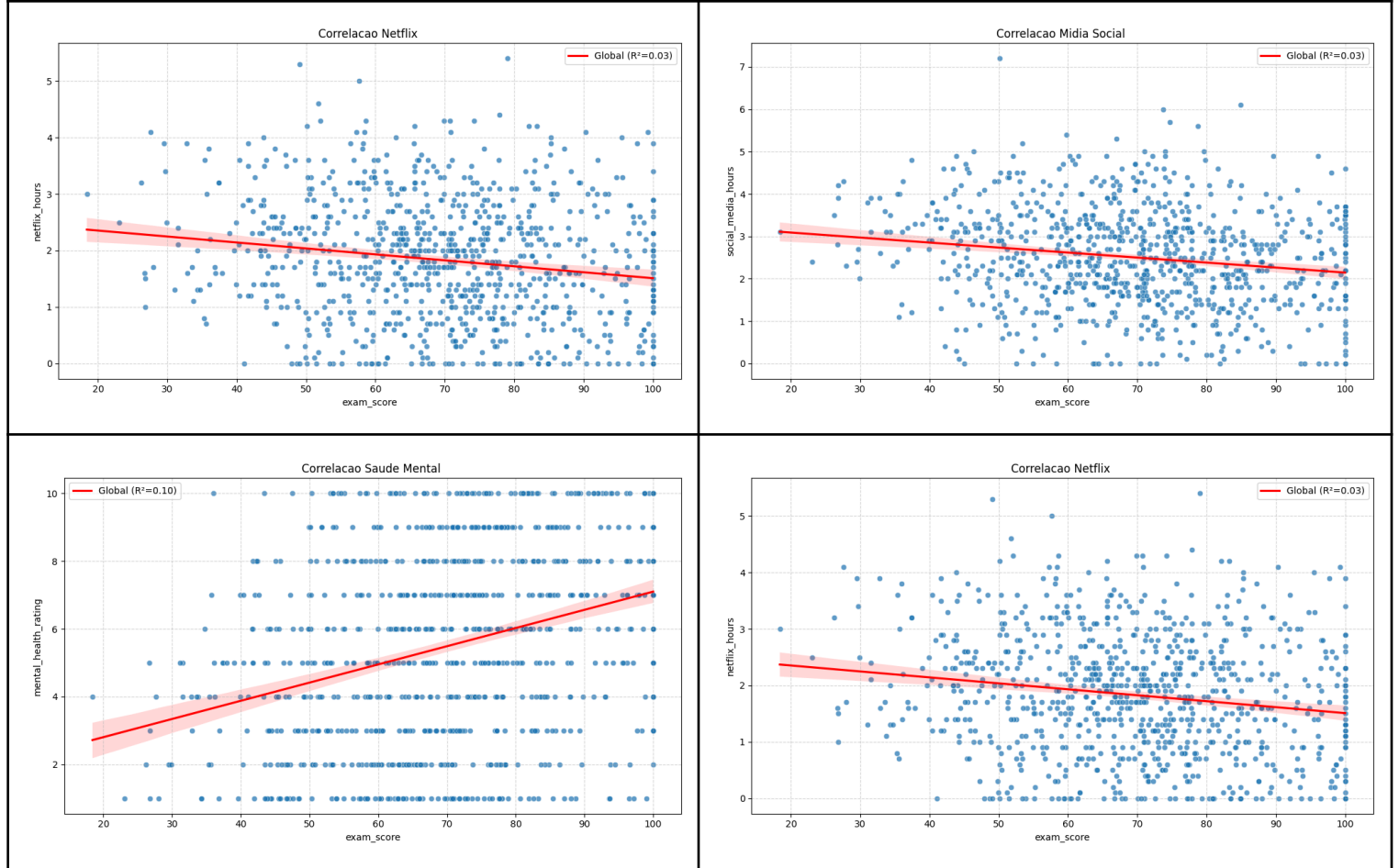
COMPARAÇÃO





REGRESSÃO LINEAR





## Interpretação dos Resultados

### 1. Hábitos mais correlacionados com alto desempenho:

Mais horas de estudos diários

Boa saúde mental

Maior frequência escolar

### 2. Diferenças entre alto e baixo desempenho:

A correlação entre algumas variáveis foi baixa: muitas horas dormidas e usar pouco mídias sociais possuem baixa correlação com o desempenho. Entretanto os alunos top dormem mais, usam menos redes sociais e se exercitam mais do que alunos de baixo desempenho (agregação de ganhos marginais).

O que vai de acordo com o livro Hábitos Atômicos (James Clear). No livro o autor relata como o acúmulo de pequenos detalhes é fundamental para um desempenho de excelência.

### 3. Melhores preditores de desempenho:

Horas de estudo (5.2h ideal)

Horas de sono (7h ideal)

Uso de redes sociais ( $\leq 1.6$ h)

Frequência escolar ( $\geq 86.5\%$ )

Exercícios (3.4x/semana)