

Análise de Dados com Python

Leonardo Manzato

GitHub: github.com/leomzto

GitHub: [linkedin.com/in/leomzto](https://www.linkedin.com/in/leomzto)

4 de junho de 2025

Introdução

Neste case, meu objetivo é analisar como diferentes hábitos e características pessoais dos estudantes se relacionam com a performance acadêmica, a partir de um conjunto de dados que inclui informações como: idade, gênero, tempo médio de estudo por dia, uso de redes sociais, sono, qualidade da alimentação, entre outros fatores.

Utilizando dados estruturados — como porcentagem de frequência às aulas, horas diárias de estudo, de sono, uso de redes sociais, tempo em plataformas de streaming e se o estudante tem ou não um emprego de meio período — pretendo investigar perguntas como:

- Há uma relação entre o tempo de estudo diário e a frequência às aulas?
- Estudantes que dormem mais ou têm melhor alimentação têm maior frequência ou hábitos mais saudáveis de estudo?
- O uso de redes sociais ou streaming impacta negativamente a frequência?
- Quais variáveis parecem mais influenciar uma alta ou baixa *attendance-percentage*?

A ideia é explorar essas relações com visualizações, análises estatísticas básicas (como correlação) e interpretações que possam indicar quais hábitos são mais comuns entre estudantes com boa participação nas aulas — e possivelmente, melhor desempenho.

1 Análise Exploratória

1.1 Carregamento e limpeza dos dados

Os dados foram carregados do arquivo `student_habits_performance.csv`. A coluna `student_id` foi removida e as linhas com valores faltantes foram excluídas.

```
tabela = tabela.drop(columns="student_id")
tabela = tabela.dropna()
```

1.2 Relação entre tempo de estudo diário e frequência às aulas

Gráfico de barras da relação entre `study_hours_per_day` e `attendance_percentage` sugere que estudantes com rotinas de estudo tendem a ter mais frequência.

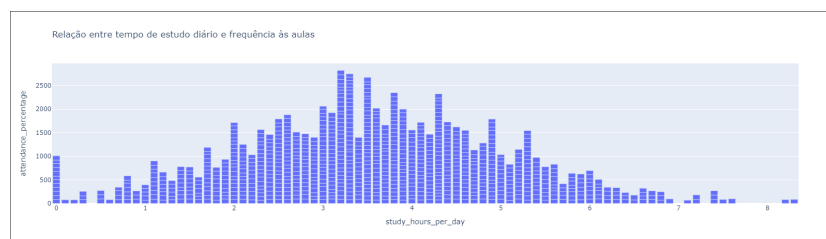


Figura 1: (Relação entre tempo de estudo e frequência)

1.3 Sono, alimentação e frequência

Gráficos indicam que estudantes com uma boa média de sono e melhor qualidade da dieta possuem maior frequência às aulas.

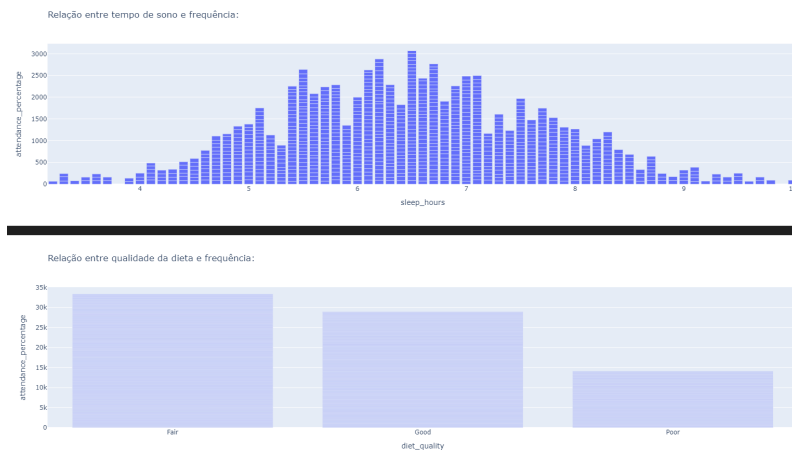


Figura 2: (Sono, alimentação e frequência)

1.4 Uso de redes sociais e streaming e impacto na frequência

Análise sugere que estudantes que passam mais de 4 horas diárias em redes sociais ou streaming tendem a ter menor frequência escolar.

A variável `total_screen_time` (soma de horas em redes sociais e streaming) reforça essa tendência.

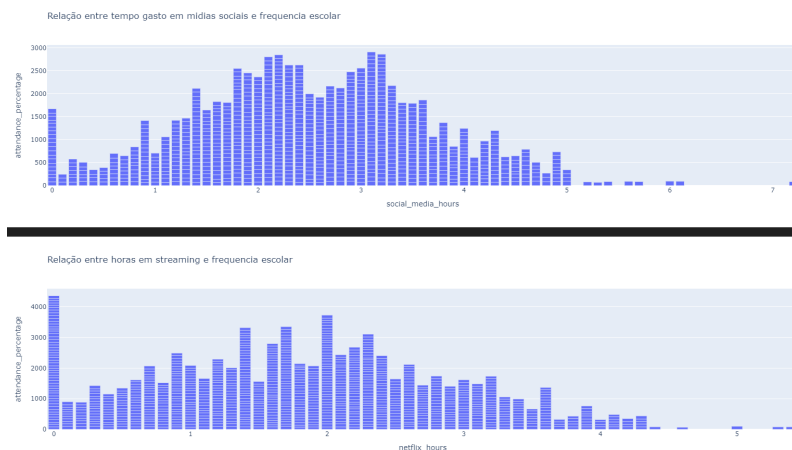


Figura 3: (Uso de redes sociais e streaming e impacto na frequência)

1.5 Saúde mental e frequência

A análise inicial sugere que não há impacto claro da saúde mental na frequência escolar.

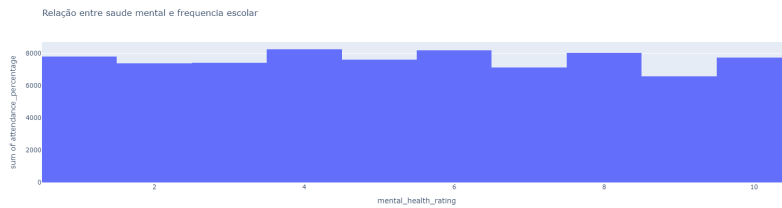


Figura 4: (Saúde mental e frequência)

1.6 Alimentação e notas

Estudantes com má alimentação tendem a possuir notas menores.

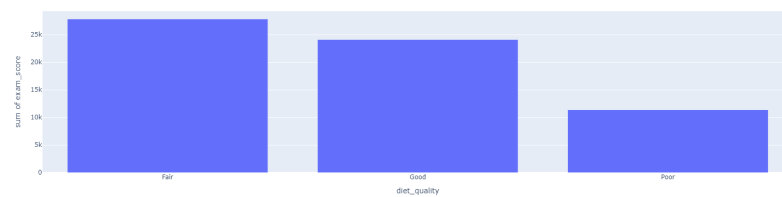


Figura 5: (Alimentação e notas)

1.7 Frequência e notas

Estudantes com baixa frequência possuem menores notas em exames.

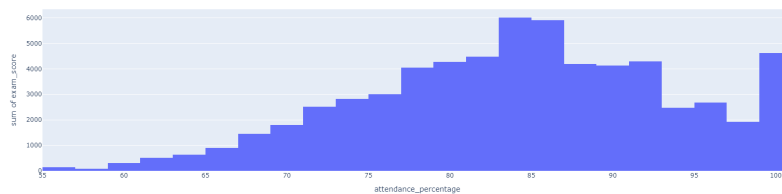


Figura 6: Descrição da imagem (Frequência e notas)

2 Análise de Correlação

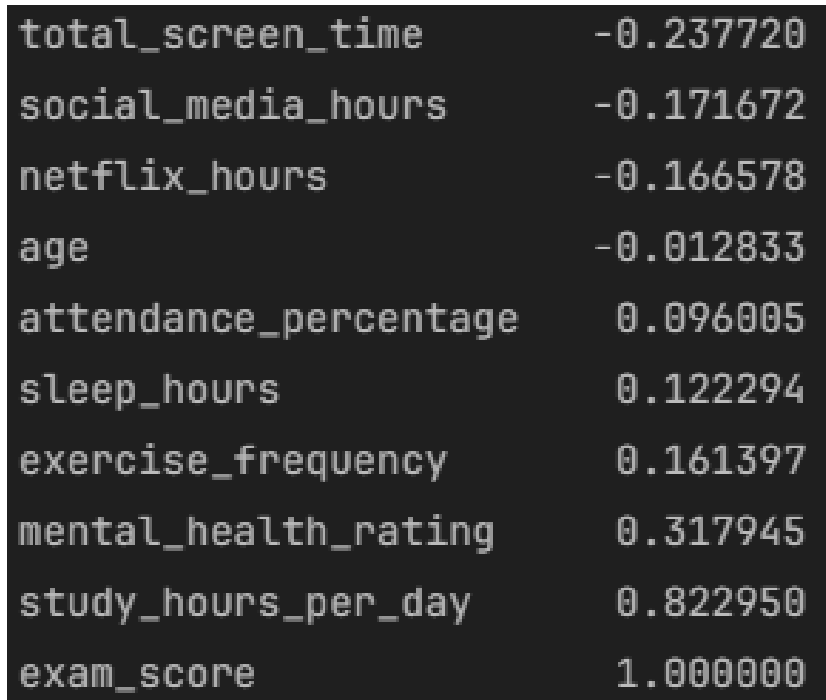
2.1 Correlação entre attendance_percentage e outras variáveis

A variável com maior correlação positiva com a frequência é exam_score.

age	-0.030203
mental_health_rating	-0.017096
netflix_hours	-0.015008
exercise_frequency	-0.010563
sleep_hours	0.001972
total_screen_time	0.026015
study_hours_per_day	0.028967
social_media_hours	0.049356
exam_score	0.096005
attendance_percentage	1.000000

Figura 7: (Correlação: frequência e outros dados)

2.2 Correlação entre exam_score e outras variáveis



total_screen_time	-0.237720
social_media_hours	-0.171672
netflix_hours	-0.166578
age	-0.012833
attendance_percentage	0.096005
sleep_hours	0.122294
exercise_frequency	0.161397
mental_health_rating	0.317945
study_hours_per_day	0.822950
exam_score	1.000000

Figura 8: (Correlação: Desempenho e outros dados)

As variáveis com maior correlação positiva com o desempenho são:

- `study_hours_per_day`: mais horas de estudo, melhores notas.
- `mental_health_rating`: melhor saúde mental, melhores notas.

Já `social_media_hours` apresenta correlação negativa: mais tempo em redes sociais, menor desempenho.

3 Conclusões

- Existe uma relação positiva, porém moderada, entre tempo de estudo diário e frequência às aulas.
- Sono e qualidade da alimentação podem influenciar a frequência, porém com correlações mais fracas.
- O uso excessivo de redes sociais e streaming impacta negativamente tanto a frequência quanto o desempenho.
- O desempenho acadêmico e hábitos relacionados a estudo e saúde mental parecem ser os principais indicadores de uma boa frequência.

Ferramentas Utilizadas

- Dataset: Student Habits vs Academic Performance (Kaggle)
- Ambiente de desenvolvimento: Visual Studio Code com Jupyter Notebook
- Linguagem: Python 3
- Bibliotecas Python: pandas, plotly