

AUTOR: FELIPE CIDADE SOARES

**ANÁLISE EXPLORATÓRIA DE DADOS APLICADA À CONCLUSÃO DE CURSOS
ONLINE**

Janeiro de 2026

SUMÁRIO

1. INTRODUÇÃO	1
2. OBJETIVOS DE ANÁLISE.....	1
2.1. Objetivo Geral	1
2.2. Objetivos Específicos	1
3. FONTE E ESTRUTURA DOS DADOS.....	1
4. EXTRAÇÃO E PREPARAÇÃO DOS DADOS	2
5. ANÁLISES REALIZADAS	3
Atividades perdidas vs conclusão	3
Dias desde o último acesso	3
Tempo total dedicado ao curso.....	4
Taxa de conclusão de vídeos.....	4
Esforço versus desempenho	5
Correlação entre variáveis.....	5
6. CONCLUSÃO	6

1. INTRODUÇÃO

O ensino online teve um aumento significativo, principalmente durante a pandemia de COVID-19. Nesse contexto, as instituições de ensino precisaram se adaptar e implementar novas tecnologias, o que acabou proporcionando uma maior integração de novas tecnologias no ensino. Além disso, a educação online facilitou o acesso a cursos para estudantes em áreas mais remotas, contribuindo para a democratização do acesso à educação.

O trabalho em questão visa realizar uma análise exploratória dos dados para entender as características dos estudantes e dos cursos, identificar os fatores que influenciam a conclusão de cursos, por meio dos padrões de engajamento e gerar insights acionáveis para acompanhamento.

2. OBJETIVOS DE ANÁLISE

2.1. Objetivo Geral

O objetivo geral deste trabalho é realizar uma análise dos fatores que levam os estudantes a concluir ou não cursos remotos.

2.2. Objetivos Específicos

Os objetivos específicos deste trabalho são:

- Realizar uma análise exploratória dos dados para entender as características dos estudantes e dos cursos.
- Identificar os fatores que influenciam a conclusão de cursos, por meio dos padrões de engajamento.
- Gerar insights acionáveis para acompanhamento.

3. FONTE E ESTRUTURA DOS DADOS

A origem dos dados é de um dataset que pode ser encontrado na plataforma Kaggle, no link <https://www.kaggle.com/datasets/nisargpatel344/student-course-completion-prediction-dataset>. O dataset contém informações sobre os estudantes, como idade, gênero, tempo de gasto na plataforma, formação, entre outros, e também as informações sobre os cursos, como o nome do curso, a duração, o preço, entre outros. A variável de interesse está na coluna "Completed", que indica se o estudante concluiu o curso ou não.

4. EXTRAÇÃO E PREPARAÇÃO DOS DADOS

O dataset foi extraído do Kaggle e salvo no arquivo "student_course_completion.csv". O arquivo foi carregado através da biblioteca pandas, da linguagem Python e os dados foram explorados para entender as características dos estudantes e dos cursos. O código fonte está disponível no repositório do GitHub: <https://github.com/cidade-felipe/Student-Course-Completion-Prediction-Dataset>. As variáveis foram selecionadas para a análise estão categorizadas, listadas e descritas na tabela abaixo.

Categoria	Coluna	Descrição
Variável alvo	Completed	Indica se o curso foi concluído
Variável derivada	Completed_Label	Rótulo categórico da conclusão do curso
Engajamento	Login_Frequency	Frequência de acessos à plataforma
Engajamento	Time_Spent_Hours	Total de horas dedicadas ao curso
Engajamento	Days_Since_Last_Login	Dias desde o último acesso
Engajamento	Video_Completion_Rate	Percentual de vídeos concluídos
Avaliação acadêmica	Quiz_Score_Avg	Média das notas dos quizzes
Avaliação acadêmica	Assignments_Submitted	Quantidade de atividades entregues
Avaliação acadêmica	Assignments_Missed	Quantidade de atividades não entregues

5. ANÁLISES REALIZADAS

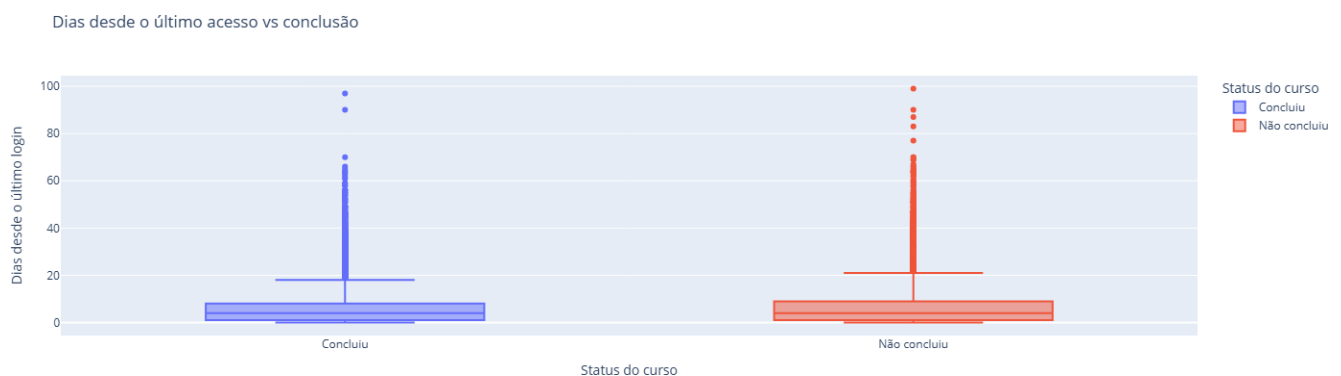
Foram realizadas análises exploratórias dos dados para entender os fatores que influenciam a conclusão de cursos, por meio dos padrões de engajamento.

Atividades perdidas vs conclusão



Alunos que concluíram o curso têm uma mediana menor de atividades perdidas e uma distribuição mais concentrada. Já os que não concluíram apresentam maior dispersão e valores máximos mais altos. Isso sugere que perder atividades não é um evento isolado, mas um comportamento recorrente associado à evasão. Um aluno não abandona um curso porque perdeu uma atividade, mas sim pelo desengajamento progressivo gerado pela sequência de atividades perdidas.

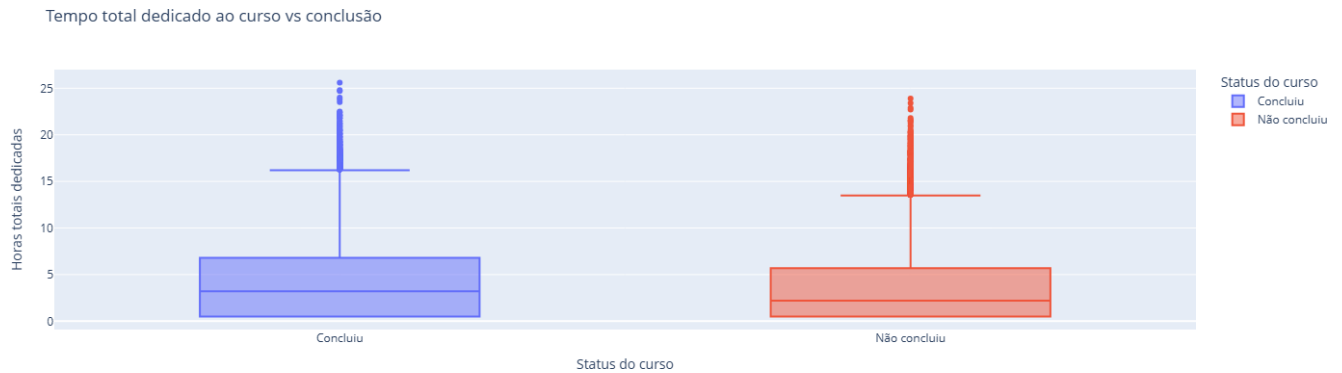
Dias desde o último acesso



Quem concluiu o curso tende a ter poucos dias desde o último login, com a maior parte dos valores concentrados perto de zero. Já os alunos que não concluíram apresentam caudas longas, com muitos casos de longos períodos sem acesso. Quanto maior o tempo sem

acessar a plataforma, menor a chance de conclusão.

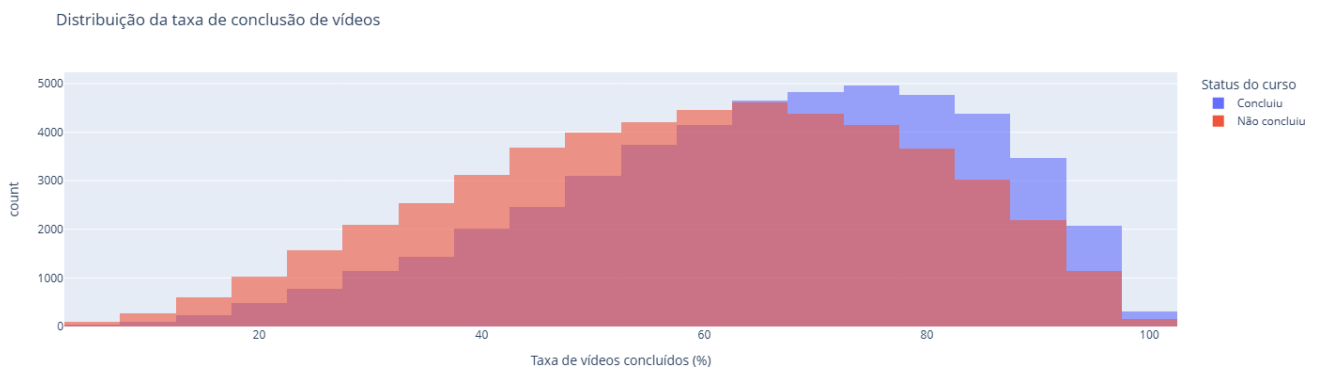
Tempo total dedicado ao curso



Os alunos que concluíram apresentam uma mediana de horas dedicadas claramente maior, além de uma distribuição deslocada para valores mais altos. Já os que não concluíram concentram-se em poucas horas de dedicação, com alguns outliers que indicam tentativas pontuais, mas sem consistência.

Aqui aparece a diferença entre esforço contínuo e esforço esporádico. Dedicar mais tempo não garante conclusão, mas dedicar pouco praticamente garante o oposto.

Taxa de conclusão de vídeos



A distribuição dos alunos que concluíram está visivelmente deslocada para taxas mais altas de consumo de vídeo, enquanto os que não concluíram concentram-se em valores médios e baixos. Isso indica que assistir ao conteúdo até o fim é um bom proxy de engajamento real. Não é só entrar na plataforma, é consumir o material de forma consistente.

Esforço versus desempenho

Relação entre esforço e desempenho

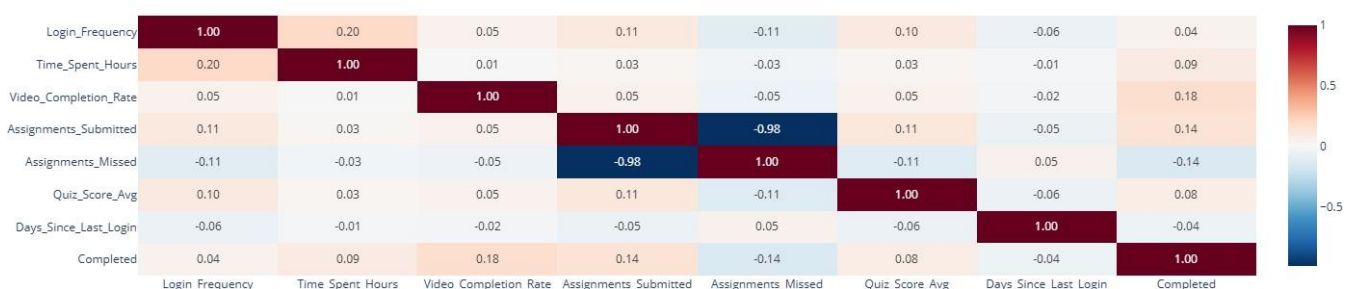


Existe uma tendência geral de que mais horas dedicadas estejam associadas a melhores médias nos quizzes, mas o mais importante é a separação visual dos grupos. Alunos que concluem aparecem com mais frequência na região de maior esforço e melhor desempenho, enquanto os que não concluem se espalham mais nos níveis baixos e intermediários.

Esse gráfico é ótimo para discussão, porque mostra que esforço ajuda, mas não explica tudo. Existem alunos que estudam bastante e mesmo assim não concluem, o que abre espaço para fatores como organização, dificuldades conceituais ou contexto pessoal.

Correlação entre variáveis

Mapa de correlação entre engajamento, desempenho e conclusão



As correlações com a variável Completed são moderadas, não exageradas, o que é esperado em dados comportamentais reais. Destacam-se a correlação positiva com Video_Completion_Rate e Assignments_Submitted, e a correlação negativa com Assignments_Missed. O destaque mais forte do mapa é a correlação negativa quase perfeita entre Assignments_Submitted e Assignments_Missed, o que é coerente e ajuda a validar a consistência do dataset.

6. CONCLUSÃO

A análise exploratória dos dados revelou que os fatores que influenciam a conclusão de cursos são multivariados, combinando engajamento, esforço e desempenho acadêmico. Nota-se que nenhuma variável isolada explica completamente a conclusão. Assim, é possível concluir que a conclusão de cursos remotos é um fenômeno multivariado, combinando engajamento, esforço e desempenho acadêmico.