

Taxa de Desistência de Alunos em 2024 Fatores e Tendências

Alexandre Stalin Silva Alves^{1*}; Cassia Renata Pinheiro²

¹Instituto Monitor. Analista de Sistemas II. Av. Deputado Cantídio Sampaio, 5901 – Vila Souza; 02860-001 São Paulo, SP, Brasil

²Nome da Empresa ou Instituição (opcional). Titulação ou função ou departamento. Endereço completo (pessoal ou profissional) – Bairro; 00000-000 Cidade, Estado, País

*autor correspondente: nome@email.com

Taxa de Desistência de Alunos em 2024 Fatores e Tendências

Resumo

Este estudo procura investigar a evasão escolar em uma instituição de ensino no ano de 2024, por meio da análise estatística de dados institucionais. A pesquisa busca identificar correlações entre a variável ativo (0 para inativo ou 1 para ativo) e variáveis como faixa etária, curso, sexo e unidade. Com o apoio de ferramentas como Python e SQL, a análise pretende ir além do diagnóstico, oferecendo subsídios práticos para a formulação de estratégias institucionais voltadas à redução da evasão e ao fortalecimento da permanência estudantil.

Palavras-chave: evasão escolar; permanência estudantil; análise institucional; fatores de desistência.

Introdução

A desistência no sistema educacional é um fenômeno recorrente em diversas instituições e que tem gerado grande preocupação. A decisão de abandonar o curso pode ser relacionado a diversos fatores, como dificuldades financeiras, problemas acadêmicos, questões pessoais, entre outros fatores. Embora a evasão seja um tema bastante estudado, é crucial compreender como ela se manifesta em contextos específicos de cada instituição, seja de nível básico ao ensino superior.

De acordo com os indicadores estatísticos do censo em 2023, foi observado que houve uma queda nas matrículas no ensino básico em comparação a 2022 e que as taxas de aprovação também diminuiriam tanto no ensino fundamental como no ensino médio. Esse declínio pode estar relacionado a diversos fatores sendo socioeconômicos, falta de apoio educacional, além das sequelas que a pandemia deixou. No ensino superior, dados apontam que apenas 40% dos alunos concluem o curso, o que deixa evidente não apenas a dificuldade de acesso, mas também as barreiras que muitas pessoas enfrentam durante a trajetória acadêmica, como a pressão financeira, a adaptação ao ambiente universitário e a falta de apoio institucional. Esses dados revelam a urgência de implementar ações eficazes para combater a evasão escolar e garantir a permanência dos alunos em todos os níveis da educação.

Este estudo busca investigar a taxa de desistência dos alunos no ano de 2024. A partir desta análise, pretende-se identificar padrões e tendências que possam ajudar a entender melhor as causas desse fenômeno, além de contribuir de forma estratégica para que a instituição possa combater a evasão escolar.

Para realização do estudo, serão utilizados os dados de uma instituição específica, o que limitará a análise aos dados coletados por esta instituição o que pode evidenciar que, embora a evasão escolar seja amplamente discutida, ela se manifesta de diversas formas em cada instituição.

Considerando que a evasão escolar tem implicações significativas para as instituições, tanto em termos de planejamento acadêmico quanto financeiro, compreender os fatores que influenciam este evento pode ser fundamental para a criação de estratégias mais eficazes no contexto de cada instituição, minimizando o impacto causado pela evasão.

Metodologia ou Material e Métodos

Para realização do estudo, será utilizado a ferramenta Spyder, dentro do ambiente de desenvolvimento de software Anaconda, com a linguagem de programação Python. A coleta e o tratamento dos dados serão feitos a partir de um banco de dados SQL Server e exportados para um documento com a extensão .csv. A análise dos dados também será realizada no Python, executarei algumas análises exploratórias e estatísticas entre elas as tabelas de contingência aplicando o Teste Qui-Quadrado(X^2) para verificar se há correlação entre a variável ativo e as demais como unidade, curso, sexo e faixa etária. Além disso, pretendo utilizar mapas perceptuais com a técnica de Análise de correspondência múltipla (MCA), para representar graficamente as principais variáveis e padrões que contribuem para a condição de atividade dos alunos, permitindo uma análise visual mais eficiente.

Extração e consumo dos dados

Com livre acesso ao banco de dados da instituição foi realizado inicialmente uma consulta que retorna as variáveis que serão estudadas, foi utilizado a linguagem SQL no SQL Server Management Studio para realizar a filtragem das tabelas processo importante na etapa de ETL (Extract, Transform, Load) neste caso como já temos acesso podemos transformar os dados antes de extrair o que facilita o processo neste caso. Foram escolhidas 5 variáveis para o estudo sendo elas a unidade em quem o aluno foi matriculado sendo elas São Paulo, Minas Gerais e Paraná, faixa etária com intervalo de idade dos 14 anos aos 17 anos, dos 18 anos aos 29 anos e dos 30 anos ou mais, sexo indicando o sexo do aluno entre masculino e feminino, curso e ativo indicando se o aluno está ativo ou não sendo 1 como ativo e 0 como inativo, como os dados são de 2014 temos diversos alunos formados então considere o status formado como 1 ao invés de 0 indicando que ele permaneceu ativo até o fim do curso que normalmente dura cerca de 6 meses e pode ser postergado.

| | unidade | faixa_etaria | sexo | curso | ativo | data_matricula |
|----|-----------|--------------|-----------|------------------------------------|-------|----------------|
| 1 | São Paulo | 18-29 | MASCULINO | TÉCNICO EM LOGÍSTICA | 1 | 2024/01/01 |
| 2 | Paraná | 30+ | MASCULINO | TÉCNICO EM ELETRÔNICA | 1 | 2024/01/01 |
| 3 | São Paulo | 30+ | FEMININO | TÉCNICO EM ADMINISTRAÇÃO | 1 | 2024/01/01 |
| 4 | São Paulo | 18-29 | MASCULINO | TÉCNICO EM MECATRÔNICA | 1 | 2024/01/01 |
| 5 | São Paulo | 30+ | MASCULINO | TÉCNICO EM TRANSAÇÕES IMOBILIÁRIAS | 1 | 2024/01/01 |
| 6 | São Paulo | 30+ | FEMININO | TÉCNICO EM SEGURANÇA DO TRABALHO | 0 | 2024/01/02 |
| 7 | São Paulo | 18-29 | MASCULINO | TÉCNICO EM ADMINISTRAÇÃO | 0 | 2024/01/02 |
| 8 | São Paulo | 18-29 | FEMININO | TÉCNICO EM TRANSAÇÕES IMOBILIÁRIAS | 1 | 2024/01/02 |
| 9 | São Paulo | 30+ | FEMININO | TÉCNICO EM TRANSAÇÕES IMOBILIÁRIAS | 0 | 2024/01/02 |
| 10 | São Paulo | 30+ | FEMININO | TÉCNICO EM TRANSAÇÕES IMOBILIÁRIAS | 1 | 2024/01/02 |
| 11 | São Paulo | 30+ | FEMININO | TÉCNICO EM AUTOMAÇÃO INDUSTRIAL | 1 | 2024/01/02 |
| 12 | São Paulo | 18-29 | MASCULINO | TÉCNICO EM LOGÍSTICA | 1 | 2024/01/02 |
| 13 | São Paulo | 30+ | MASCULINO | TÉCNICO EM TRANSAÇÕES IMOBILIÁRIAS | 1 | 2024/01/02 |
| 14 | São Paulo | 30+ | FEMININO | TÉCNICO EM ADMINISTRAÇÃO | 1 | 2024/01/02 |
| 15 | São Paulo | 30+ | MASCULINO | TÉCNICO EM SEGURANÇA DO TRABALHO | 1 | 2024/01/02 |

Tabela 1. Parte da amostra de dados coletados.

Fonte: Dados originais da pesquisa

Exploração dos dados

Para a exploração e análise estatística dos dados coletados, foi utilizada a linguagem Python, por meio do ambiente de desenvolvimento Spyder. O primeiro passo consiste em carregar para o dataset os dados contidos no arquivo .csv, previamente gerado a partir da linguagem SQL e depois extraído.

Antes de iniciar a análise de correlação, foi necessário realizar uma etapa exploratória, com objetivo de compreender melhor os valores de cada variável selecionada. Para isso, foi utilizado a função .info() do Python, cujo retorno é apresentado na Figura 1. Essa função trás informações relevantes como o número de colunas, os nomes das colunas, a quantidade de registros não nulos e o tipo de cada variável.

As figuras 2 a 6 apresentam a distribuição de frequência das variáveis analisadas, foi obtido utilizando a função .value_counts() do Python. Por meio dessas representações foi possível identificar padrões significativos no conjunto de dados. Por exemplo, foi observado que a maior concentração de alunos fica na unidade de São Paulo, com uma diferença expressiva em relação as demais unidades, enquanto a faixa etária observasse que grande parte dos alunos possuem 30 anos ou mais, na variável curso os cursos que mais possuem alunos tanto ativos quanto inativos são o Técnico em Transações Imobiliárias e o EJA (Educação de Jovens e Adultos), na variável sexo notou-se que existem mais alunos do sexo

masculino do que feminino porém não é tão expressivo mas ainda assim chama a atenção ao observar diferente da variável ativos onde a quantidade de alunos ativos é muito expressiva em comparação aos alunos inativos, sugerindo um alto índice de permanência.

As figuras 7 a 10 apresentam as tabelas de contingência relacionadas a variável ativo, geradas por meio da função `pd.crosstab()`. Essas tabelas permitem observar a discrepância entre alunos ativos e inativos em relação a outras variáveis. No caso da faixa etária os dados reforçam que os alunos com 30 anos ou mais representam o principal público e consequentemente também correspondem ao grupo com maior evasão.

As figuras 11 a 14 apresentam tabelas que analisam a significância estatística entre a variável ativo e as outras variáveis, por meio da função `chi2_contingency()`. Esses testes permitem verificar se existe associação entre as variáveis, auxiliando na seleção de variáveis relevantes para análises posteriores. Observa-se associação entre as variáveis ativo x unidade (p-valor = 0,0; Qui-Quadrado = 74,21) e ativo x curso (p-valor = 0,0; Qui-Quadrado = 162,75). Em contraste as associações entre ativo x faixa_etaria (p-valor = 0,9674; Qui-Quadrado = 0,07) e ativo x sexo (p-valor = 0,4266; Qui-quadrado = 0,63) não foram estatisticamente significativas, indicando independência entre essas variáveis.

Mesmo que as variáveis `faixa_etaria` e `sexo` não apresentem associação estatisticamente significativa com a variável ativo, conforme indicado pelos valores elevados do p-value nos testes Qui-Quadrado(X^2), a aplicação dos mapas perceptuais continua sendo uma ferramenta valiosa. Isso se deve ao fato de que os mapas perceptuais proporcionam uma representação gráfica multidimensional das relações entre variáveis, possibilitando a identificação de agrupamentos e tendências que podem não ser capturados por testes estatísticos tradicionais. Assim, esses mapas atuam como uma ferramenta complementar à análise, enriquecendo a interpretação dos dados ao revelar aspectos estruturais e padrões sutis que contribuem para a compreensão do comportamento dos alunos.

```
Data columns (total 7 columns):
```

| # | Column | Non-Null Count | Dtype |
|---|----------------|----------------|---------|
| 0 | UNIDADE | 6835 non-null | object |
| 1 | FAIXA_ETARIA | 6835 non-null | object |
| 2 | SEXO | 6835 non-null | object |
| 3 | CURSO | 6835 non-null | object |
| 4 | DATA_MATRICULA | 6835 non-null | object |
| 5 | STATUS | 6835 non-null | object |
| 6 | ATIVO | 6835 non-null | float64 |

Figura 1. Resumo das variáveis do dataset utilizado

Fonte: Dados originais da pesquisa

| Frequência da variável: Unidade | | Frequência da variável: Faixa Etária | |
|---------------------------------|------------|--------------------------------------|------------|
| Unidade | Frequência | Faixa Etária | Frequência |
| São Paulo | 5630 | 30+ | 4203 |
| Paraná | 835 | 18-29 | 2592 |
| Minas Gerais | 365 | 14-17 | 35 |

| Frequência da variável: Sexo | | Frequência da variável: Ativo | |
|------------------------------|------------|-------------------------------|------------|
| Sexo | Frequência | Ativo | Frequência |
| MASCULINO | 3932 | 1.0 | 5560 |
| FEMININO | 2898 | 0.0 | 1270 |

| Frequência da variável: Curso | |
|------------------------------------|------------|
| Curso | Frequência |
| TÉCNICO EM TRANSAÇÕES IMOBILIÁRIAS | 4414 |
| EJA | 619 |
| TÉCNICO EM SEGURANÇA DO TRABALHO | 432 |
| TÉCNICO EM ELETRÔNICA | 325 |
| TÉCNICO EM ELETROTÉCNICA | 309 |
| TÉCNICO EM ADMINISTRAÇÃO | 199 |
| TÉCNICO EM MECATRÔNICA | 117 |
| TÉCNICO EM SECRETARIADO | 106 |
| TÉCNICO EM LOGÍSTICA | 92 |
| TÉCNICO EM AUTOMAÇÃO INDUSTRIAL | 79 |
| TÉCNICO EM CONTABILIDADE | 78 |
| TÉCNICO EM SECRETARIA ESCOLAR | 30 |
| TÉCNICO EM REDES DE COMPUTADORES | 16 |
| TÉCNICO EM ENFERMAGEM | 14 |

Figuras 2 a 6. Tabelas de frequência das variáveis analisadas

Fonte: Dados originais da pesquisa

| UNIDADE | Minas Gerais | Paraná | São Paulo | FAIXA_ETARIA | 14-17 | 18-29 | 30+ |
|---------|--------------|--------|-----------|--------------|-------|-------|------|
| ATIVO | | | | ATIVO | | | |
| 0.0 | 32 | 86 | 1152 | 0.0 | 7 | 484 | 779 |
| 1.0 | 333 | 749 | 4478 | 1.0 | 28 | 2108 | 3424 |

| SEXO | FEMININO | MASCULINO | CURSO | EJA | ... | TÉCNICO EM TRANSAÇÕES IMOBILIÁRIAS |
|-------|----------|-----------|-------|-----|-----|------------------------------------|
| ATIVO | | | ATIVO | | ... | |
| 0.0 | 552 | 718 | 0.0 | 212 | ... | 758 |
| 1.0 | 2346 | 3214 | 1.0 | 407 | ... | 3656 |

Figuras 7 a 10. Tabelas de contingência em relação à variável ATIVO

Fonte: Dados originais da pesquisa

| | |
|---|--|
| <pre>===== Teste de Independência: ATIVO x UNIDADE ----- Estatística χ^2 : 74.21 P-valor : 0.0 Graus de liberdade : 2 =====</pre> | <pre>===== Teste de Independência: ATIVO x FAIXA_ETARIA ----- Estatística χ^2 : 0.07 P-valor : 0.9674 Graus de liberdade : 2 =====</pre> |
| <pre>===== Teste de Independência: ATIVO x SEXO ----- Estatística χ^2 : 0.63 P-valor : 0.4266 Graus de liberdade : 1 =====</pre> | <pre>===== Teste de Independência: ATIVO x CURSO ----- Estatística χ^2 : 162.75 P-valor : 0.0 Graus de liberdade : 13 =====</pre> |

Figuras 11 a 14. Tabela indicando os resultados do p-value e o Qui-quadrado (X^2) de cada variável

Fonte: Dados originais da pesquisa

Resultados Preliminares

Para aprofundar a compreensão dos dados após a fase exploratória, foram desenvolvidas visualizações gráficas com o objetivo de identificar padrões e relações entre as variáveis. A técnica utilizada foi a Análise de Correspondência Múltipla (MCA), que permite representar graficamente as associações entre categorias de diferentes variáveis. Os gráficos gerados chamados de mapas perceptuais foram construídos em Python, com apoio da biblioteca Plotly para a geração das imagens. Nesses mapas, categorias que aparecem próximas indicam maior associação entre si, facilitando a interpretação visual da relação entre a variável ativo e as outras variáveis do conjunto de dados.

Na figura 15, que representa a variável unidade, observamos que a unidade São Paulo apresenta maior proximidade ao dado 1 da variável ativo o que representa grande associação entre os alunos que estão ativos ou formados com a unidade de São Paulo, por outro lado a unidade de Minas Gerais está posicionada bem distante indicando uma fraca associação com a variável ativo.

Na figura 16, relacionada a variável curso, nota-se que muitos cursos estão fortemente associados com a variável ativo, no entanto alguns cursos chamam a atenção por estarem bem afastados o que indica que existem pouca associação com ativos entre eles está o Técnico em Enfermagem, Técnico em Contabilidade, Técnico em Administração e Técnico em Secretaria Escolar, essa distância pode refletir também que há pouco engajamento nestes cursos, o que merece atenção em análises futuras.

A figura 17, por sua vez, analisa a relação entre faixa_etaria e a variável ativos, os dados revelam que existe maior associação entre os alunos com 30 anos ou mais e a variável ativos, os alunos de 18 aos 29 também tem forte associação, porém os alunos menores de idade têm fraca associação possivelmente por conta dos cursos serem incompatíveis com alunos menores de idade por se tratar de cursos técnicos, porém existem algumas exceções.

Por fim, a figura 18 mostra um mapa perceptual 2D, não foi utilizado o 3D devido às características das variáveis envolvidas. Nessa visualização, observa-se o comportamento da variável sexo em relação à variável ativo, a categoria masculino aparece mais próxima da variável ativo, o que sugere uma associação mais forte. Isso pode refletir um padrão de maior permanência entre alunos do sexo masculino no período analisado, enquanto o grupo feminino aparece mais afastado, indicando uma relação menos expressiva.

Essas representações visuais contribuem de forma significativa para identificação de padrões comportamentais dentro da base de dados, permitindo direcionar decisões futuras com base em evidências visuais e estatísticas.

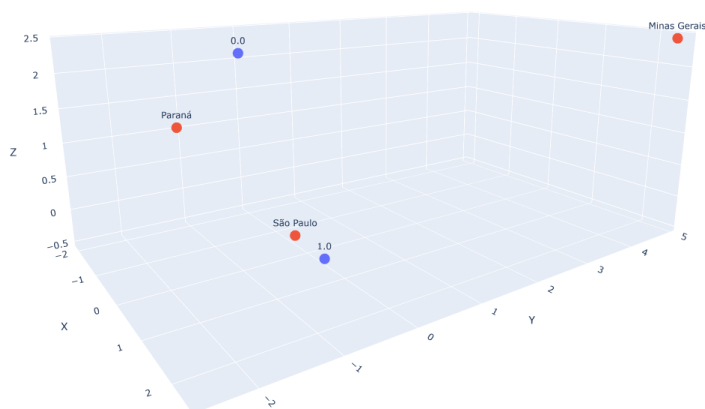


Figura 15. Mapa Perceptual das variáveis ATIVO x UNIDADE

Fonte: Dados originais da pesquisa

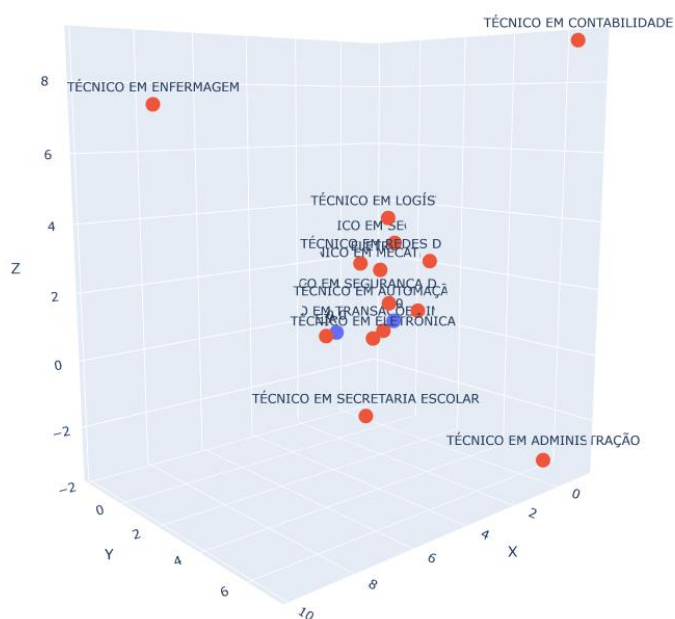


Figura 16. Mapa Perceptual das variáveis ATIVO x CURSO

Fonte: Dados originais da pesquisa

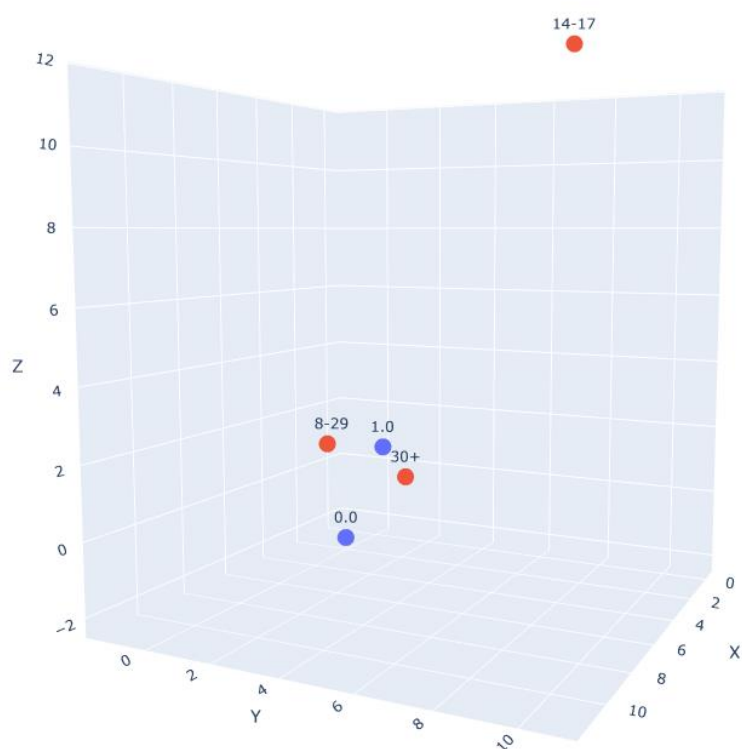


Figura 17. Mapa Perceptual das variáveis ATIVO x FAIXA_ETARIA

Fonte: Dados originais da pesquisa

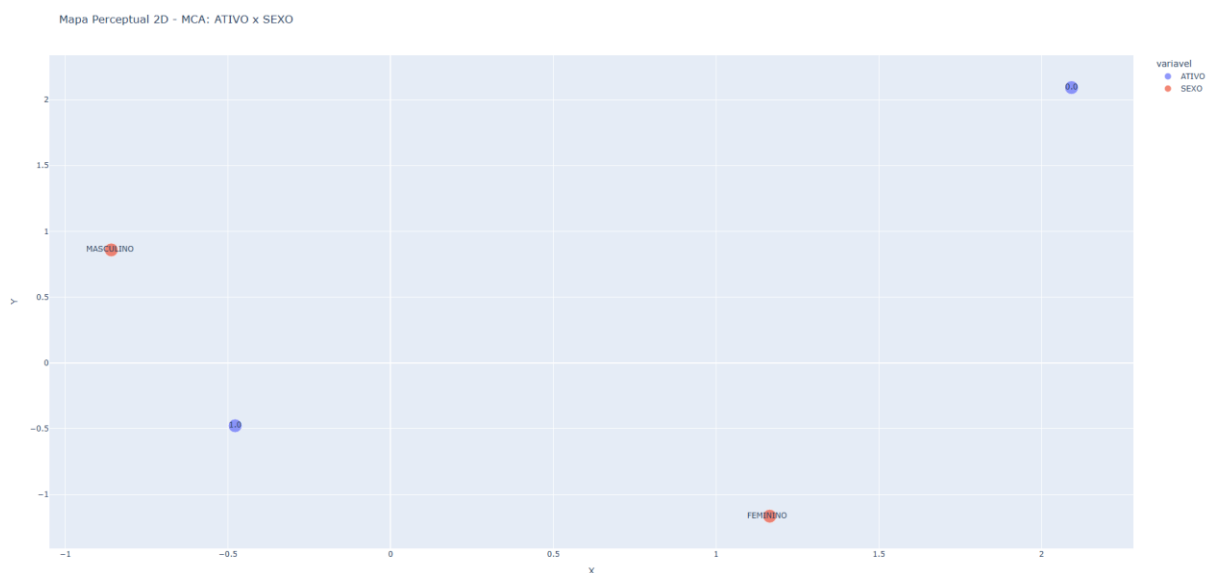


Figura 18. Mapa Perceptual das variáveis ATIVO x SEXO

Fonte: Dados originais da pesquisa

Conclusão

Levando em conta os aspectos discutidos, observa-se que embora algumas variáveis analisadas não apresentem forte correlação com o fato de o aluno estar ativo ou não, foi possível identificar tendências relevantes. Fatores como faixa etária, sexo e curso indicam que parte da evasão pode estar relacionada ao perfil do público atendido pela instituição. Em contrapartida, a variável curso e unidade apenas evidenciou uma maior concentração de alunos na unidade de São Paulo e quais cursos são mais evidentes, sem demonstrar um impacto significativo na evasão que, de modo geral apresentou índices muito baixos em todas as unidades.

Esse cenário sugere que as principais causas da evasão estejam ligadas a fatores externos à instituição. Entre eles, destacam-se questões como a necessidade de conciliar trabalho e estudo, a falta de tempo devido a responsabilidades familiares, dificuldades financeiras para arcar com custos indiretos de formação (transporte, alimentação e acesso à tecnologia), e até mesmo problemas de saúde física ou emocional. Além disso, a ausência de uma rede de apoio e a falta de motivação pessoal podem contribuir para o abandono do curso, especialmente entre estudantes que enfrentam uma rotina exaustiva ou vivem em contextos sociais vulneráveis.

Dessa forma, torna-se essencial que a instituição realize um trabalho mais aprofundado de escuta e análise da realidade dos alunos, por meio de instrumentos como entrevistas, questionários socioeconômicos e acompanhamento individualizado. A partir disso, é possível desenvolver estratégias de apoio pedagógico, psicológico e financeiro promovendo um ambiente mais acolhedor e inclusivo, que favoreça a permanência e o sucesso acadêmico.

Agradecimento

“Porque para mim tenho por certo que as aflições deste tempo presente não são para comparar com a glória que em nós há de ser revelada.”

Em tempos difíceis, palavras como essas não trazem respostas fáceis, mas despertam em nós a perspectiva de que o presente momento não define tudo, e que o futuro reserva valor e esperança.

Palavras assim não apenas preenchem o tempo — elas permanecem. Sou profundamente grato por toda a minha trajetória.

Referências

Censo Escolar, resultados, 2023. Disponível em:

<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar/resultados>

Acesso em: 24 de mar. 2025

Censo Escolar, Em dez anos, 40% dos que iniciaram um curso o concluíram, 2022.

Disponível em: <https://www.gov.br/inep/pt-br/assuntos/noticias/censo-da-educacao-superior/em-dez-anos-40-dos-que-iniciaram-um-curso-o-concluíram>

Acesso em: 24 de mar. 2025

Amanda Garcia, 55,5% dos alunos desistem antes de completar ensino superior, aponta relatório, 2023.

Disponível em: <https://www.cnnbrasil.com.br/nacional/555-dos-alunos-desistem-antes-de-completar-ensino-superior-aponta-relatorio/>

Acesso em: 24 de mar. 2025

Wallace Cardozo, Censo Escolar 2022 aponta aumento em taxas de reprovação e evasão escolar na rede pública, 2023.

Disponível em: <https://www.itausocial.org.br/noticias/censo-escolar-2022-aponta-aumento-em-taxas-de-reprovacao-e-evacao-escolar-na-rede-publica/>

Acesso em: 24 de mar. 2025

Jornal Nacional, Censo Escolar confirma impacto negativo da pandemia na educação básica, 2022. Disponível em: <https://g1.globo.com/jornal-nacional/noticia/2022/05/20/censo-escolar-confirma-impacto-negativo-da-pandemia-na-educacao-basica.ghtml>

Acesso em: 24 de mar. 2025

Fávero, Luiz Paulo; Belfiore, Patrícia. (2024). Manual de análise de dados: estatística e machine learning com Excel®, SPSS®, Stata®, R® e Python®. 2 ed. Rio de Janeiro: LTC.