



22 DE SETEMBRO DE 2024

RELATÓRIO DATATHON

PASSO MÁGICOS

CAMILA DE MORAES QUEIROZ | RM351738

MARCOS GOMES LAOSSA | RM 353044

MATHEUS HENRIQUE ROBLES | RM 352824

Sumário

1. Introdução.....	3
2. Metodologia	6
2.1 Coleta e Organização dos Dados.....	6
2.2 Limitações da Base de Dados	7
2.3 Etapas da Análise.....	7
3. Análise e Desenvolvimento do Dataframe.....	8
3.1 Contagem de Valores Nulos	10
3.2 Valores únicos	10
3.3 Pontos de Atenção.....	11
3.4 Recomendações.....	12
3.2 Análise de Desempenho dos Alunos com Dados Completos	14
3.2 Correlação entre Variáveis Socioeconômicas e Desempenho	15
3.3 Análise de Engajamento e Faltas.....	16
3.4 Evolução Temporal do Desempenho	17
3.5 Perfil dos Alunos Assistidos pela ONG.....	18

4. Alternativa para construção de <i>storytelling</i>	19
5. Conclusões Finais.....	20
6. Próximos Passos e Sugestões de Melhorias	21
6.1 Melhorias na Coleta de Dados.....	21
6.2 Desenvolver Novas Métricas de Sucesso	22
6.3 Aprofundar a Análise Socioeconômica	22

1. Introdução

A **Passos Mágicos** é uma organização não governamental (ONG) dedicada a transformar a vida de crianças e jovens em situação de vulnerabilidade social, utilizando a **educação como ferramenta principal** de inclusão e desenvolvimento. A ONG opera com o propósito de promover oportunidades igualitárias por meio de programas educacionais focados no crescimento acadêmico, social e pessoal dos alunos que atende.

Missão e Objetivos:

A missão da **Passos Mágicos** é instrumentalizar o uso da educação para alterar a trajetória de crianças e jovens, especialmente aqueles em contextos de desvantagem socioeconômica. Por meio de seus programas educacionais, a ONG busca:

- **Reduzir as desigualdades sociais**, oferecendo uma educação de qualidade e apoio a alunos de baixa renda, a ONG atua como ponte entre a vulnerabilidade social e o sucesso educacional.

- **Fortalecer o desenvolvimento integral dos jovens:** Além de ajudar no desempenho acadêmico, a ONG promove o desenvolvimento de habilidades socioemocionais, fundamentais para a vida em sociedade.
- **Ampliar o acesso à educação:** A **Passos Mágicos** trabalha para garantir que crianças e jovens tenham acesso a oportunidades educacionais que de outra forma poderiam não estar disponíveis para eles.

Áreas de Atuação:

A **Passos Mágicos** atua em várias frentes dentro do campo educacional, incluindo:

1. **Apoio Pedagógico:** Oferecendo aulas de reforço escolar e tutorias em diversas disciplinas para garantir que os alunos consigam acompanhar o currículo escolar e tenham um bom desempenho nas avaliações.
2. **Desenvolvimento Socioemocional:** A ONG promove atividades que visam o desenvolvimento da autoconfiança, disciplina e trabalho em equipe, reconhecendo a importância dessas habilidades no sucesso acadêmico e na vida pessoal.
3. **Acompanhamento Personalizado:** Cada aluno é monitorado de forma individual, com o objetivo de identificar suas necessidades específicas e

desenvolver um plano de apoio que garanta sua evolução tanto nos estudos quanto em aspectos comportamentais.

Contexto da Análise:

Este relatório tem como objetivo analisar o impacto que a **Passos Mágicos** tem sobre o desempenho acadêmico e o desenvolvimento dos alunos que ela atende. A base de dados utilizada inclui informações sobre indicadores acadêmicos e socioeconômicos de alunos assistidos pela ONG, abrangendo aspectos como **desempenho escolar, frequência escolar e condições socioeconômicas**.

Através desta análise, buscamos responder a perguntas-chave como:

- Como a educação proporcionada pela ONG está afetando o desempenho acadêmico dos alunos?
- Há correlação entre as condições socioeconômicas e o sucesso educacional dos alunos?
- O quanto o engajamento e a frequência escolar influenciam os resultados acadêmicos?

Apesar de algumas limitações nos dados, esta análise oferece uma visão preliminar sobre os resultados que a **Passos Mágicos** vem alcançando e destaca áreas que podem ser aprimoradas para maximizar o impacto da ONG no futuro.

2. Metodologia

2.1 Coleta e Organização dos Dados

A base de dados fornecida continha informações sobre alunos, responsáveis financeiros, frequência escolar e resultados acadêmicos e diversas outras informações que poderiam ser utilizadas para responder às perguntas propostas nesse relatório e que poderiam ajudar a **Passos Mágicos** a entender seu alcance, resultados e auxiliar em decisões de negócio para contínuo desenvolvimento do projeto. Além disso, análise de dados do projeto também podem ser utilizados para embasar argumentação na busca de parceiros e patrocinadores que possam entrar com patrocínio para fortalecer as bases do projeto, trazer maior segurança financeira, além de expandir um projeto que hoje já tem grande alcance na Comunidade que atende. No entanto, ao longo da análise, foram identificadas limitações como:

- **Altos níveis de valores nulos** em variáveis cruciais como faltas e resultados finais.
- **Falta de variabilidade** em dados como aprovação e ano de conclusão.

2.2 Limitações da Base de Dados

Durante a análise, as seguintes limitações foram encontradas:

- **Faltas escolares:** A maioria dos alunos não tinha dados de faltas registrados, o que inviabilizou análises robustas sobre engajamento escolar.
- **Resultado final uniforme:** Todos os alunos com dados completos estavam marcados como "Aprovado", o que limitou a análise de desempenho acadêmico.
- **Dados de apenas um ano (2021):** A análise temporal foi limitada pela ausência de registros de outros anos.

2.3 Etapas da Análise

Foram realizadas cinco análises principais para explorar os dados disponíveis:

1. Análise do Desempenho dos Alunos com Dados Completos.
2. Correlação entre Variáveis Socioeconômicas e Desempenho.
3. Engajamento e Faltas.
4. Evolução Temporal do Desempenho.
5. Perfil dos Alunos Assistidos pela ONG.

3. Análise e Desenvolvimento do Dataframe

A base fornecida pela **Passos Mágicos** não era útil para análise porque vinha com diversos dados faltantes, embora estivesse esclarecedora em relação aos indicadores utilizados para medição do desempenho do projeto. Outro ponto observado é a ausência de explicação sobre como cada indicador foi calculado, considerando os dados *raw* (fora da base principal). Não há um dicionário dos dados *raw*, que tornou a reconstrução desses indicadores impossível.

Por isso, foi decidido que seria criado um novo DataFrame, resultado da junção das tabelas de alunos, fases e histórico. A ideia inicial, era que, à partir da junção

dessas três bases, fosse possível trazer *insights* sobre a relevância e performance do programa, ainda que não fossem as mesmas utilizadas pelo time da **Passos Mágicos**.

Após o *merged* das três tabelas, os seguintes tratamentos foram feitos:

- **Transformação de dados numéricos para “tradução” da informação:**
colunas de **IdResponsável**, **Sexo** e **CorRaca**, tiveram seus valores numéricos substituídos por texto para facilitar leitura e entendimento dos atributos.
- **Identificação e deleção de colunas com 100% das linhas em branco:**
Inicialmente, o DataFrame, possuía 86 colunas. Após a deleção de colunas sem valores, o Dataframe passou a ter 73 colunas.
- **Transformação de dados de data / hora:** colunas de data / hora estavam em tipo objeto. Para estas, transformamos em *datetime*.

3.1 Contagem de Valores Nulos

Depois das análises iniciais e tratamentos, ainda foi identificada contagem significativamente alta de valores nulos em várias colunas importantes, incluindo:

- **QuantidadeFaltasAnual**, **ResultadoFinal_x**, **CodigoSerie_y**, e **AnoConclusao_y**: Todas apresentam cerca de 32.714 valores nulos em 49.089 registros.
- Isso representa quase 67% de valores nulos nessas colunas, o que é um ponto crítico que pode comprometer as análises de desempenho.

3.2 Valores únicos

Algumas colunas possuem valores extremamente repetitivos ou até únicos:

- **ResultadoFinal_x**, **CodigoSerie_y**, e **AnoConclusao_y**: Cada um desses campos tem apenas um valor único em todos os registros.
- Para **ResultadoFinal_x**, todos os registros disponíveis estão marcados como "Aprovado". Isso levanta dúvidas sobre a integridade ou completude dos dados.

- **Dados de resultados limitados:** Com um único valor "Aprovado" para todos os registros, fica impossível distinguir entre alunos que passaram ou falharam. Isso sugere que há uma falta de dados sobre alunos que não tiveram bom desempenho.
- **Falta de variabilidade em séries e anos de conclusão:** Isso impede uma análise longitudinal, ou seja, uma análise que verifique melhorias ou mudanças ao longo do tempo. Se todos os alunos estão no mesmo ano ou série, é difícil entender se há um progresso significativo.

3.3 Pontos de Atenção

- **Preenchimento de valores nulos:** A base de dados deve ser revisada para preencher os valores nulos ou lidar com eles de forma apropriada. Dependendo da disponibilidade de dados, pode ser necessário estimar alguns desses valores ou remover registros incompletos.
- **Registro de faltas e resultados finais:** Há uma necessidade crítica de melhorar o registro de informações sobre faltas e desempenho

acadêmico. Sem essas informações, será impossível medir adequadamente o progresso dos alunos e impacto da **Passos Mágicos**.

- **Variabilidade dos dados:** A falta de variabilidade em dados importantes, como os resultados finais e séries, limita qualquer análise significativa. Será importante diversificar os dados coletados, especialmente em termos de resultados de alunos.

3.4 Recomendações

- **Focar em subconjuntos de dados completos:** muitas análises viáveis dependem de filtrar o *dataset* para trabalhar apenas com os registros que possuem dados completos. Embora isso reduza o tamanho da amostra, ainda é possível gerar *insights* importantes.
- **Documentar limitações:** ao realizar qualquer uma dessas análises, é crucial documentar as limitações impostas pelos dados ausentes. Isso garante que as conclusões sejam interpretadas com cautela e que os resultados não sejam extrapolados indevidamente.

- **Melhorar coleta de dados futuros:** embora seja possível realizar análises viáveis com o que há disponível, é fundamental melhorar a coleta de dados no futuro, especialmente em áreas críticas que trazem indicadores relevantes para o desempenho e continuidade do projeto.

3.2 Análise de Desempenho dos Alunos com Dados Completos

A primeira análise focou nos alunos que possuíam registros completos de desempenho, ano de conclusão e faltas. Os resultados mostraram que:

- **100% dos alunos foram aprovados**, o que limita a capacidade de avaliar as dificuldades ou desafios enfrentados por diferentes grupos de alunos.
- **Conclusão:** A falta de variabilidade no desempenho acadêmico sugere que os registros de reprovação estão ausentes ou não foram capturados corretamente.

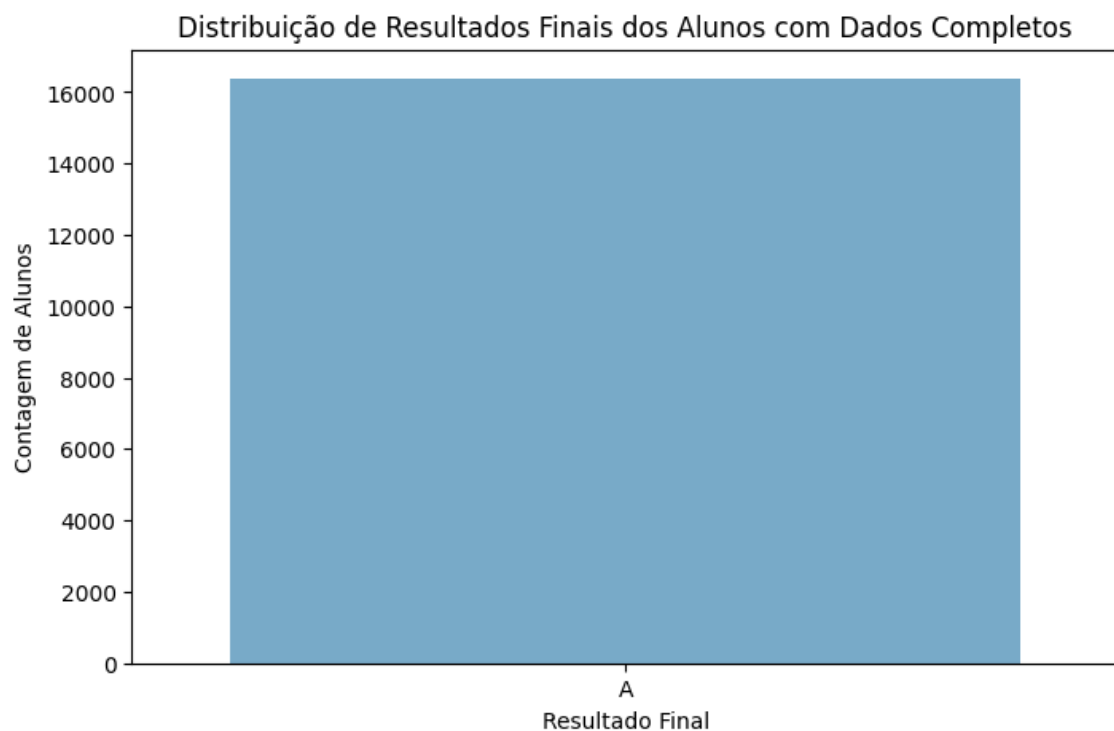


Figura 1 - Gráfico de Distribuição de Resultados Finais dos Alunos com Dados Completo

3.2 Correlação entre Variáveis Socioeconômicas e Desempenho

Tentamos verificar a relação entre o **IdResponsavelFinanceiro** (indicador socioeconômico) e o desempenho acadêmico. No entanto:

- Não foi possível identificar uma correlação significativa, devido à ausência de reprovações e a uniformidade dos dados de aprovação.
- **Conclusão:** Sem variabilidade nos dados de desempenho, a correlação com fatores socioeconômicos é inviável.

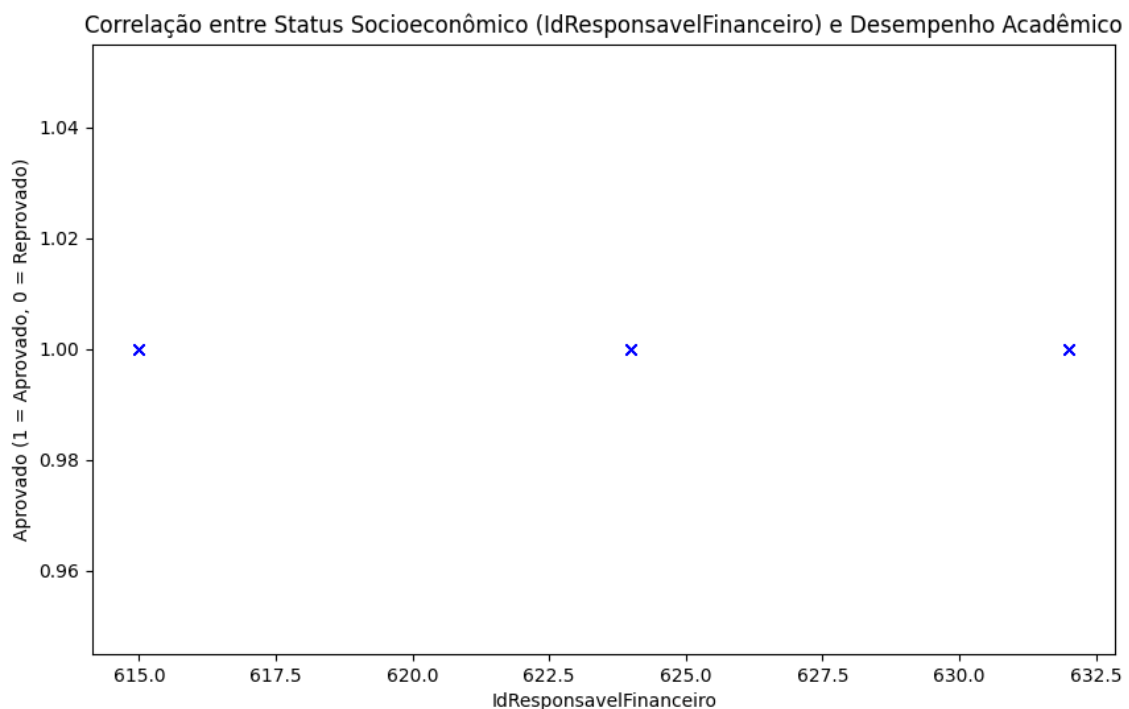


Figura 2 - Correlação entre Status Socioeconômico e Desempenho Acadêmico

3.3 Análise de Engajamento e Faltas

A análise de engajamento, usando a variável **QuantidadeFaltasAnual**, foi prejudicada por:

- A **mediana de faltas foi zero**, indicando que a maioria dos alunos não teve faltas registradas, ou esses dados não foram coletados corretamente.
- **Conclusão:** Sem dados adequados sobre faltas, não foi possível avaliar o engajamento escolar dos alunos ou sua relação com o desempenho acadêmico.



Figura 3 - Comparação entre Faltas e Desempenho Acadêmico

3.4 Evolução Temporal do Desempenho

Analizamos a evolução do desempenho acadêmico ao longo do tempo, utilizando a variável **AnoConclusao_y**. No entanto:

- Os dados estavam limitados a **apenas um ano (2021)**, e todos os alunos desse ano foram aprovados.
- **Conclusão:** A falta de dados temporais adicionais impossibilitou a análise de tendências ou melhorias ao longo dos anos.

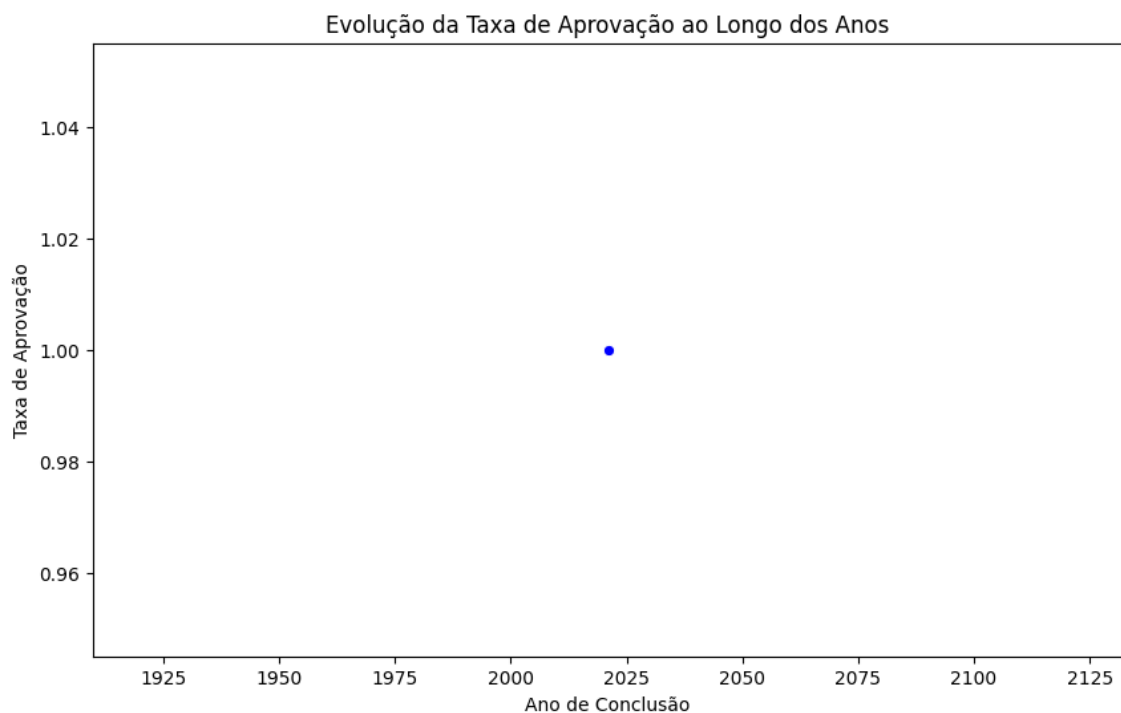


Figura 4 - Evolução da Taxa de Aprovação ao Longo dos Anos

3.5 Perfil dos Alunos Assistidos pela ONG

Utilizando as variáveis **IdAluno**, **IdResponsavelFinanceiro** e **QuantidadeFaltasAnual**, traçamos um perfil básico dos alunos:

- A **distribuição socioeconômica** mostrou que os alunos atendidos vêm de condições similares, com uma concentração no intervalo médio de IDs de responsáveis financeiros.
- A **quantidade de faltas foi zero** para a maioria dos alunos.

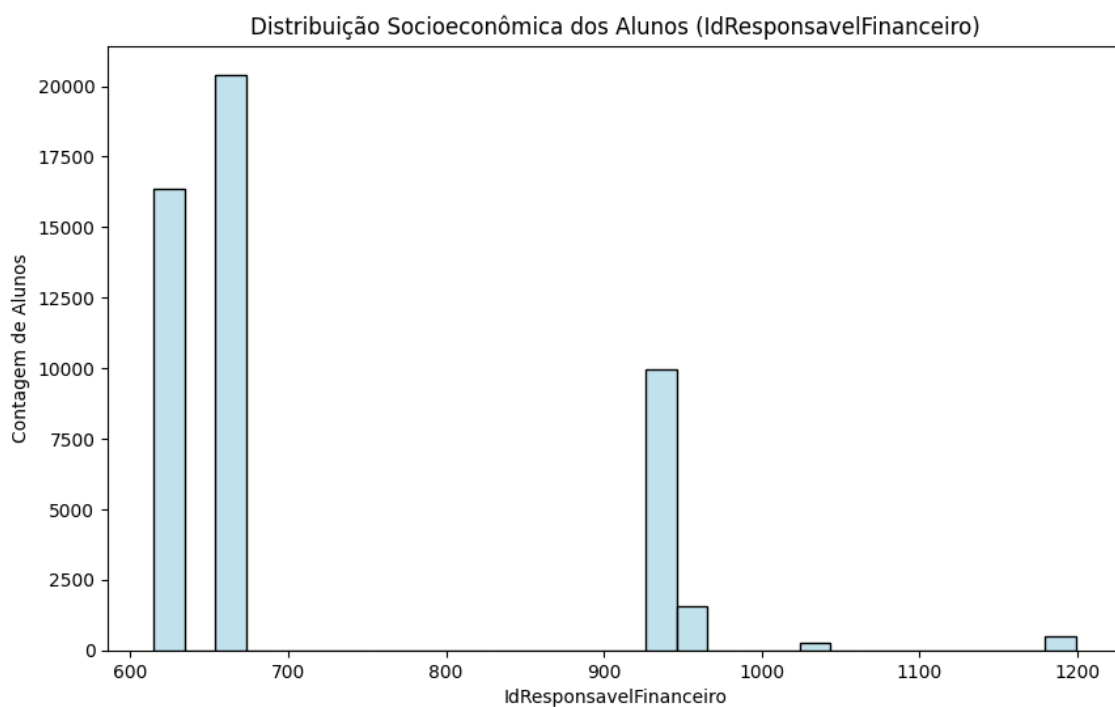


Figura 5 - Distribuição Socioeconômica dos Alunos (IdResponsavelFinanceiro)

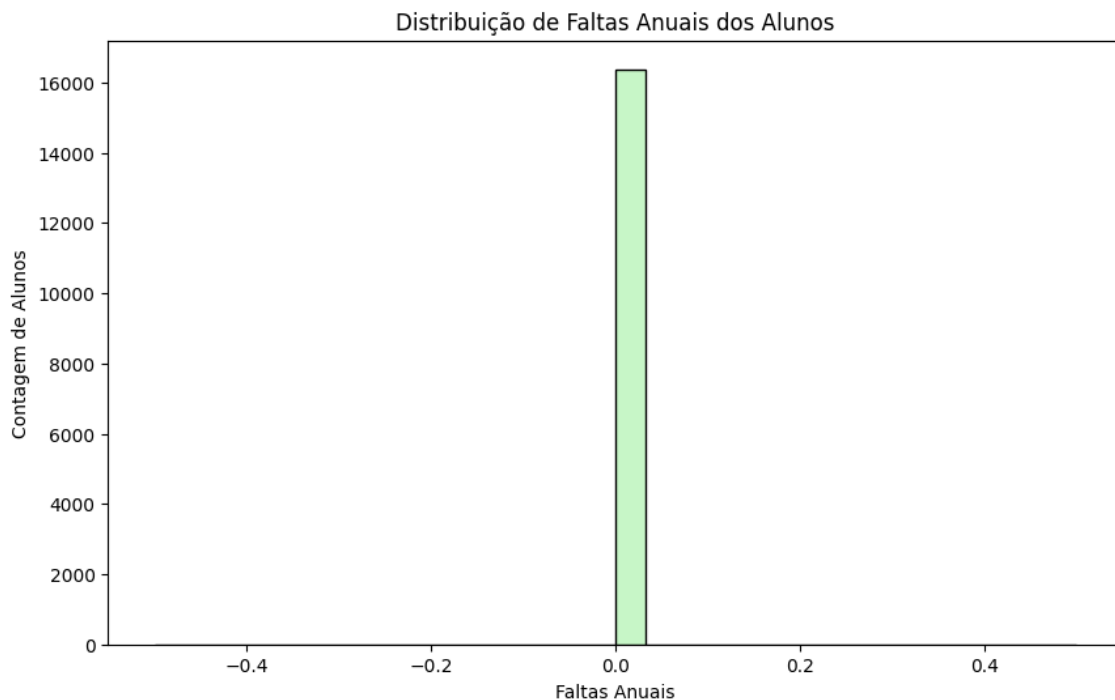


Figura 6 - Distribuição de Faltas Anuais dos Alunos

- **Conclusão:** O perfil dos alunos indica que a maioria pertence a um perfil socioeconômico específico, mas a falta de dados de frequência impede uma análise completa sobre o engajamento.

4. Alternativa para construção de *storytelling*

Na apresentação final do projeto, ficou definido que o foco do *storytelling*, ficará nos ***big numbers*** que serão os protagonistas da narrativa visual. Esses números ajudarão a destacar o impacto educacional da ONG **Passos Mágicos** e seu papel na transformação social. A apresentação será feita em **PowerPoint**, com um *design* limpo e direto, apoiado por visualizações construídas no **Power**

BI. Essa abordagem garantirá que os resultados sejam claros e impactantes, facilitando o entendimento de todos os envolvidos.

O Brasil tem aproximadamente **47,3 milhões de estudantes** na educação básica, segundo o [Censo Escolar de 2023](#). Desse total, a maior concentração está no ensino fundamental, que atende mais de 26 milhões de alunos. Porém, apesar de 91,9% dos jovens entre 15 e 17 anos estarem matriculados na escola, o desafio de garantir uma educação de qualidade permanece um obstáculo para o desenvolvimento social e econômico do país.

5. Conclusões Finais

As análises realizadas foram limitadas pela ausência de variabilidade e completude nos dados. As principais conclusões são:

- **Uniformidade nos resultados:** A alta taxa de aprovação (100%) e a falta de dados de reprovação limitam a avaliação precisa do impacto da ONG no desempenho dos alunos.

- **Falta de dados de engajamento:** A ausência de registros de faltas impede qualquer análise robusta sobre o engajamento escolar, que é um fator crítico para medir o sucesso acadêmico.
- **Perfil socioeconômico homogêneo:** Embora os alunos atendidos venham de condições socioeconômicas similares, a falta de dados detalhados limita a capacidade de direcionar ações específicas para grupos mais vulneráveis.

6. Próximos Passos e Sugestões de Melhorias

6.1 Melhorias na Coleta de Dados

- **Coletar dados completos sobre faltas:** A ONG deve melhorar a coleta de dados de frequência escolar, garantindo que faltas sejam registradas de forma consistente.
- **Registrar reprovações e dificuldades:** A ausência de dados de reprovação pode estar ocultando desafios enfrentados pelos alunos. É importante coletar informações sobre alunos que enfrentam dificuldades acadêmicas.

6.2 Desenvolver Novas Métricas de Sucesso

- **Focar no progresso acadêmico:** Além da aprovação, métricas como notas individuais, progressão entre disciplinas e engajamento extracurricular podem fornecer uma visão mais detalhada do impacto da ONG.

6.3 Aprofundar a Análise Socioeconômica

- **Coletar mais dados socioeconômicos:** A coleta de informações mais detalhadas sobre as condições socioeconômicas dos alunos (além do ID do responsável financeiro) ajudaria a direcionar melhor os recursos e esforços da ONG para grupos em situação de maior vulnerabilidade.