



**UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
CURSO DE ENGENHARIA DE COMPUTAÇÃO**

MAURÍCIO THIAGO FERREIRA DE LIMA

**Uma análise baseada em dados sobre o desempenho e a evasão escolar do
estudante da rede pública estadual do RN**

**NATAL-RN
2023**

Maurício Thiago Ferreira de Lima

Uma análise baseada em dados sobre o desempenho e a evasão escolar do estudante da rede pública estadual do RN

Trabalho de Conclusão de Curso na modalidade Monografia, submetido como parte dos requisitos necessários para conclusão do curso de Engenharia de Computação pela Universidade Federal do Rio Grande do Norte

Orientador: Prof. Dr. Luiz Affonso Guedes

Natal-RN
2023

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Lima, Maurício Thiago Ferreira de.

Uma análise baseada em dados sobre o desempenho e a evasão escolar do estudante da rede pública estadual do RN / Maurício Thiago Ferreira de Lima. - 2023.

49 f.: il.

Monografia (graduação) - Universidade Federal do Rio Grande do Norte, Centro de Tecnologia, Curso de Engenharia de Computação. Natal, RN, 2023.

Orientador: Prof. Dr. Luiz Affonso Guedes.

1. Ciência de Dados - Monografia. 2. Evasão Escolar - Monografia. 3. Dados Educacionais - Monografia. I. Guedes, Luiz Affonso. II. Título.

RN/UF/BCZM

CDU 004.6

Maurício Thiago Ferreira de Lima

Uma análise baseada em dados sobre o desempenho e a evasão escolar do estudante da rede pública estadual do RN

Trabalho de Conclusão de Curso na modalidade Monografia, submetido como parte dos requisitos necessários para conclusão do curso de Engenharia de Computação pela Universidade Federal do Rio Grande do Norte

Aprovado em 10 de julho de 2023:

Prof. Dr. Luiz Affonso Guedes

UFRN

Prof(a). Dr(a). Isabel Dillmann Nunes

UFRN

Janaína Silmara Silva Ramos

SEEC/RN

Natal-RN

2023

AGRADECIMENTOS

Agradeço à toda minha família em nome dos meus pais, Mara e Carlos, por toda educação e ensinamentos, dedicação e investimento de recursos materiais e tempo para que eu pudesse chegar até aqui. Agradeço pela paciência e compreensão, e pelo acolhimento, em momentos fáceis e difíceis, que só a família pode dar.

A todos os meus colegas e amigos que pude conhecer através da graduação e que me ajudaram de alguma forma a finalizar o curso, em nome de Amadeu, por toda amizade, companhia e ajuda durante a trajetória da graduação, meu muito obrigado.

Aos colegas de trabalho e amigos que dentre todas as tentativas de alcançar esse objetivo, se colocaram a disposição, seja emprestando livros, dividindo ideias, compartilhando os momentos difíceis, flexibilizando os horários de trabalho, em nome de Janaína, que me impulsionou e encorajou a chegar aqui, eu agradeço.

Agradeço à UFRN, instituição que me acolheu durante esses anos, e a todos os professores com quem pude aprender algo, especialmente ao professor Luiz Affonso, pela disponibilidade na orientação deste trabalho, pelo tempo que dedicou a mim, por sempre acreditar em mim e me dar segundas chances.

Agradeço em especial a Deus pela oportunidade de concluir esse ciclo, por me permitir conhecer todos esses, e por conduzir o meu caminho.

RESUMO

O Rio Grande do Norte possui historicamente, principalmente na Rede Estadual de Ensino, altos índices de evasão e reprovação, o que retrata um problema tanto educacional quanto social, que persiste na realidade do País, sobretudo, nos sistemas públicos de ensino. Em 2019 a taxa de evasão era de 9,5% e 17% de reprovação em todo o Estado. Tais dados são divulgados anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira possibilitando análises a respeito do desempenho dos estudantes. Considerando que tais dados podem auxiliar o poder público a planejar e executar políticas públicas e estratégias para assegurar a permanência do estudante na escola, este trabalho se propõe a utilizar técnicas de ciência de dados para analisar o desempenho acadêmico de alunos do Ensino Médio da rede pública de ensino do Estado do Rio Grande do Norte (RN), relacionando rendimento acadêmico, características socioeconômicas e evasão escolar. De posse dessas análises baseadas em dados, a Secretaria de Educação, da Cultura, do Esporte e do Lazer poderá articular ações integradas na reintegração do estudante evadido. A evasão difere do abandono. Enquanto que na primeira, o aluno não conclui a série escolar no ano corrente, a segunda indica que o estudante não concluiu todos os níveis para sua formação básica escolar. Para o desenvolvimento deste trabalho, foram analisados dados de estudantes do Ensino Médio nos anos de 2017 a 2019 e utilizando-se predominantemente as bibliotecas de manipulação e visualização de dados da linguagem Python. Os resultados apontam para uma forte influência da composição socioeconômica e rendimento acadêmico dos estudantes na taxa de evasão escolar.

Palavras-chave: Ciência de Dados, Evasão Escolar; Dados Educacionais.

ABSTRACT

Historically, especially in the State Education Network, Rio Grande do Norte has high rates of evasion and failure, which portrays both an educational and social problem that persists in the country's reality, especially in public education systems. In 2019, the dropout rate was 9.5% and 17% failure rate across the state. Such data are publicly disclosed by the National Institute of Educational Studies and Research Anísio Teixeira analyzing the performance of students. Considering that such data can help the public power to plan and execute public policies and strategies to guarantee the student's permanence in school, this work proposes to use data science techniques to analyze the academic performance of high school students from the public network of teaching in the State of Rio Grande do Norte (RN), relating academic performance, socioeconomic characteristics and school dropout. With these analyzes analyzed in data, the Secretariat of Education, Culture, Sport and Leisure will be able to articulate integrated actions in the reintegration of the dropped out student. Evasion is different from abandonment. While in the first, the student does not complete the school grade in the current year, the second indicates that the student has not completed all the levels for his/her basic school education. For the development of this work, data from high school students from 2017 to 2019 were analyzed, predominantly using Python language data manipulation and visualization libraries. The results point to a strong influence of the socioeconomic composition and academic performance of students on the school dropout rate.

Keywords: Data Science; School Dropout ; Educational Data.

LISTA DE ILUSTRAÇÕES

FIGURA		PÁGINA
1	Taxa de evasão de alunos da Rede Pública Estadual do Rio Grande do Norte (2007-2019)	21
2	Taxa de evasão de alunos do Ensino Médio da Rede Pública Estadual do Rio Grande do Norte (2007-2019)	22
3	Taxa de reprovação de alunos da Rede Pública Estadual do Rio Grande do Norte (2007-2019)	23
4	Taxa de reprovação de alunos do Ensino Médio da Rede Pública Estadual do Rio Grande do Norte (2007-2019).	24
5	Relatório com percentual de evasão por município do Rio Grande do Norte	27
6	Ciclo de atividades em ciência de dados	28
7	Portal de Gestão Escolar e Portal do Professor	31
8	Questionário socioeconômico	32
9	Método describe()	33
10	Método info()	34
11	Contando valores vazios com isnull().sum()	34
12	Método head()	35
13	Distribuição dos alunos por situação final	36
14	Total de alunos evadidos por ano	37
15	Relação de notas e frequência por série em 2017	38
16	Relação de notas e frequência por série em 2018	38
17	Relação de notas e frequência por série em 2019	39
18	Distribuição de estudantes evadidos por série e gênero em 2017, 2018 e 2019	40
19	Distribuição de notas por série para alunos evadidos em 2017, 2018 e 2019	41
20	Gráfico de violino mostrando a distribuição dos estudantes evadidos por idade ao longo dos anos 2017, 2018 e 2019 respectivamente.	42
21	Gráfico de violino mostrando a distribuição dos estudantes evadidos por idade ao longo dos anos 2017, 2018 e 2019 respectivamente.	43

22	Gráfico de pizza mostrando a contagem de estudante evadidos por condição socioeconômica	43
23	Gráfico de coordenadas paralelas relacionando notas e frequência para alunos evadidos	44
24	Utilização do método groupby()	45
25	Gráfico de bolhas interativo com cruzamento de variáveis	45

SUMÁRIO

CAPÍTULO		PÁGINA
1	INTRODUÇÃO	11
2	CIÊNCIA DE DADOS	13
2.1	Pandas	15
2.2	Plotly e Seaborn	16
2.3	Matplotlib	16
3	ORGANIZAÇÃO DOS SISTEMAS DE ENSINO: Normatizações para a garantia da permanência dos estudantes na escola	18
3.1	Indicadores Educacionais da Rede Estadual do Rio Grande do Norte(RN): Análise das taxas de rendimento dos alunos da Rede Pública Estadual do RN	20
3.2	Evasão e abandono escolar: conceitos e principais diferenças	24
4	O SISTEMA DE GERENCIAMENTO INTEGRADO DA EDUCAÇÃO – SIGEDUC	26
4.1	A evasão representada no SIGEDUC	26
5	ANÁLISE DE DESEMPENHO ESCOLAR BASEADA EM DADOS	28
5.1	Entendimento do problema e escolha da base de dados	28
5.2	Coleta dos dados	31
5.3	Tratamento dos dados	33
5.4	Análise exploratória dos dados	36
6	CONCLUSÃO	46
7	REFERÊNCIAS	47

1. INTRODUÇÃO

Assumindo-se que toda qualificação e aprendizado deve ser útil para o desenvolvimento da sociedade, e considerando a importância da educação básica na formação da sociedade como um todo, o tema proposto foi escolhido a fim de contribuir com a melhoria na qualidade da gestão da educação pública. Além disso, reforçar o conhecimento na área de ciência de dados, utilizando os recursos disponíveis na obtenção do objetivo principal.

Deste modo, o principal objetivo deste trabalho de conclusão de curso é utilizar técnicas de visualização científica de dados, eficaz na apresentação de informações quantitativas e categóricas de forma facilitada e dinâmica, para auxiliar na análise do desempenho escolar dos alunos da rede pública e contribuir para identificação dos alunos propensos a evadir, assim como contribuir para a permanência dos estudantes que precisam de um acompanhamento mais intensivo.

A análise do desempenho escolar dos estudantes é de fundamental importância para o desenvolvimento e aprimoramento do sistema educacional. O desempenho escolar pode ser analisado a partir de indicadores como nota e frequência. Compreender como os alunos estão progredindo durante a trajetória escolar permite não apenas identificar possíveis deficiências e lacunas no aprendizado, mas também possibilita a implementação de estratégias eficazes para melhorar a qualidade da educação. Nesse contexto, a ciência de dados surge como uma abordagem eficaz para o monitoramento e análise do tema em questão. Ao aplicar técnicas de coleta, processamento e análise de grandes volumes de dados, é possível extrair informações valiosas que auxiliam na tomada de decisões embasadas e orientadas por evidências. A ciência de dados possibilita a identificação de padrões, tendências e a compreensão das principais variáveis que influenciam o desempenho dos estudantes, permitindo assim a criação de estratégias personalizadas de ensino e intervenção dentro do espaço escolar ou dos órgãos de gestão (Barros, 2020). Para esta análise, foram observadas variáveis relacionadas ao contexto acadêmico como notas e frequência geral e nos componentes de Português e Matemática, bem como características do contexto socioeconômico como renda mensal familiar, adesão ao programa federal de distribuição de renda Bolsa Família, composição da membresia familiar e ocupação do estudante em atividade remunerada paralelamente à rotina escolar.

Para se atingir o objetivo principal, serão utilizados recursos de gráficos estatísticos dos pacotes Seaborn e Plotly da linguagem Python, além da biblioteca Pandas. Os dados analisados foram extraídos da base de dados do Sistema de Gerenciamento da Educação do

RN (SIGEduc) e resumidamente podem ser agrupados por abrangência de aspectos da vida do estudante, como por exemplo, dados socioeconômicos, dados pessoais, e dados de desempenho escolar como notas e frequências nos componentes curriculares. Todos os alunos desse conjunto de dados pertencem à etapa Ensino Médio, que contempla 1ª, 2ª e 3ª séries, e o período analisado está contido entre os anos de 2017 e 2019. A faixa histórica observada neste estudo é determinada, levado em consideração uma aprovação automática para todos os alunos da rede pública estadual em 2020 e 2021 dentro de um contexto específico advindo do isolamento social em decorrência da pandemia da Covid-19, e considerando também que a partir de 2021 o novo ensino médio trouxe uma gama de novas etapas dividindo a composição dos estudantes desse nível de ensino.

O restante deste trabalho está dividido em mais 5 capítulos. No capítulo 2, propõe-se primeiramente detalhar os conceitos fundamentais da ciência de dados, afinal é a principal ferramenta explorada neste trabalho, e de acordo com Rautemberg e Carmo (2019), é a área que atua diretamente no processo de transformação dos dados em informações e posteriormente em conhecimento. É proposto, uma breve apresentação dos principais recursos disponíveis através da linguagem de programação Python, para ciência de dados em todas as suas etapas. Nos capítulos 3 e 4 são apresentados, respectivamente, as definições e especificidades do sistema educacional brasileiro para que seja possível o entendimento do conceito principal deste trabalho, e o Sistema de Gerenciamento da Educação do Estado, bem como a forma de tratar a evasão escolar na rede, que permeia todos os capítulos. O capítulo 5 traz a análise baseada em dados do desempenho escolar da população estudada, e os passos utilizados para chegar nesse fim, apresentando o corpo e comportamento dos atributos, suas relações e resultados, com foco em uma apresentação detalhada dos conceitos que implicam no desenvolvimento escolar do estudante e dos fatores que influenciam na sua evasão. Por fim, no capítulo 7 são apresentadas as principais conclusões obtidas com a realização deste trabalho.

2. CIÊNCIA DE DADOS

Para que seja possível a melhor compreensão do que é a ciência de dados, entende-se o que é um dado. Segundo Amaral (2016), define que dados são fatos coletados e normalmente armazenados. E que a informação é o dado analisado e com algum significado. O conhecimento é a informação interpretada, entendida e aplicada para um fim.

Dados existem de vários formatos e através de origens diversas. Por exemplo, tabelas ou planilhas eletrônicas, arquivos de imagem, textos eletrônicos, são dados eletrônicos digitais, e assim como em qualquer outra ciência, a ciência de dados tem como principal objetivo, obter conhecimento e informação, de forma sistemática, desses dados. (Amaral, 2016)

A ciência de dados, embora semelhante em algum nível à estatística, possui características bem distintas. Enquanto a primeira se preocupa com o dado em todo o seu ciclo de vida, obtendo dados de fontes diversas, formalizando o tratamento dos dados, e explorando-os a fim de produzir informação relevante no processo de tomada de decisão, a outra é uma ferramenta que pode ser utilizada em alguma etapa desse processo (Rautemberg; Carmo, 2019). Para entender melhor esse conceito, será necessário compreender o ciclo de vida do projeto baseado em dados.

De forma geral, de acordo com Tavares (2022) as principais etapas envolvidas no processo da ciência dos dados são resumidamente:

- Entendimento do problema: Identificar de forma clara o problema que precisa ser resolvido e entender os objetivos e requisitos do projeto.
- Coleta de dados: Obter os dados relevantes para o problema em questão, seja por meio de fontes públicas, bancos de dados internos ou sensores.
- Limpeza e tratamento dos dados: Avaliar a qualidade dos dados, tratar valores ausentes, eliminar ruídos, e transformar os dados em um formato adequado para análise.
- Exploração e análise de dados: Utilizar técnicas estatísticas e visualizações para explorar os dados, identificar padrões, tendências e relações entre as variáveis.
- Modelagem e algoritmos: Desenvolver modelos preditivos ou descritivos usando técnicas como aprendizado de máquina, mineração de dados ou estatística inferencial, dependendo do problema em questão.

De forma a contribuir com o perfeito entendimento da conclusão do processo de estudo dos dados, adiciona-se a etapa de comunicação e visualização dos resultados que tem como

objetivo comunicar os resultados de forma clara e eficaz, geralmente por meio de visualizações ou relatórios, para ajudar na tomada de decisões.

A análise das etapas envolvidas no processo de obtenção e uso dos dados, sejam eles quais forem, é essencial para compreender e delimitar de forma eficiente todos os atores envolvidos, a fim de proporcionar equilíbrio entre eles e obter uma máxima otimização de seu uso. Para Sant'ana (2016), o acesso a dados vem transformando todas as áreas de atuação humana, sobretudo nos últimos anos em função do aumento significativo de alternativas e meios de coleta, armazenamento e recuperação desses dados.

Sant'ana (2016) ressalta ainda a importância do planejamento na obtenção dos dados na etapa de coleta, visando atender uma necessidade específica sobre um contexto previamente determinado. Tal planejamento é requerido para algumas perguntas fundamentais que devem ser respondidas: qual o escopo da necessidade informacional? Quais os dados necessários e quais suas fontes? Existe risco de privacidade para os indivíduos envolvidos? Existe permissão para sua obtenção e utilização? Entre outras.

Nessa fase, é primordial a definição bem estruturada do planejamento para coleta dos dados, verificando cuidadosamente se a privacidade dos envolvidos está garantida, evitando o corrompimento das próximas etapas do ciclo de vida do projeto. Cabe aqui, um adendo referente à Lei Geral de Proteção de Dados (LGPD - Lei n. 13.709, de 14 de agosto de 2018), que de forma resumida tem por fundamentos a garantia do respeito à privacidade e da inviolabilidade da intimidade, da honra e da imagem, promovida pela instituição privada ou pública, garantindo o não livre acesso aos dados considerados sensíveis das pessoas naturais. Dados sensíveis de acordo com (MORAES, 2008), são aqueles associados às opções e características basilares da persona e, portanto, aptos a gerar situações de discriminação e desigualdade. É nesse momento também que devem ser preservados os direitos autorais vinculados aos dados que serão obtidos e utilizados.

Após obter os dados é necessário realizar o tratamento dos mesmos que consistem em garantir a qualidade dos dados, essa etapa segundo Teixeira (2003) pode ser resumida em ordenação, classificação e análise propriamente dita. Em seguida vem a etapa de análise de dados, de forma exploratória, apresentando visualmente os dados de maneira que seja possível obter informações úteis para os interessados, de acordo com o objetivo principal, e nesse processo se utiliza-se dos recursos da estatística. Silva (2021) define as últimas etapas desse ciclo de vida, como o processamento dos dados, onde são utilizadas técnicas de machine learning e estudo das principais variáveis e distribuições do sistema para a construção de um modelo sólido de aprendizagem de máquina, obtendo um resultado com a melhor precisão

possível. “Conceitualmente, o Aprendizado de Máquina é uma subárea da Inteligência Artificial que investiga a captura automatizada de modelos de abstração de informação a partir de registros contidos em (volumosas) bases de dados (Blum; Hopcroft; Kannan, 2018).

Sant’ana (2016) e Amaral (2016) refletem acerca da implementação de uma fase de descarte como final para o ciclo de vida dos dados, pois prevê um cenário futuro onde a capacidade de armazenamento será insuficiente diante da grande escala de coleta de dados.

Pode-se concluir que, no geral, ciência de dados é um campo interdisciplinar que combina conhecimentos de estatística, matemática, programação e domínio do negócio para obter ideias e conhecimentos úteis a partir de conjuntos de dados. O objetivo principal dessa ciência é resolver problemas complexos e tomar decisões baseadas em evidências, utilizando métodos analíticos e algoritmos avançados, extraíndo informação a partir de imensas bases de dados, complexas, dinâmicas, heterogêneas e distribuídas. (Bugnion; Manivannan; Nicolas, 2017)

É importante ressaltar que a ética desempenha um papel crucial na ciência de dados. A manipulação e o uso indevido de dados podem levar a consequências negativas, portanto, os cientistas de dados devem seguir princípios éticos, respeitar a privacidade e proteger as informações pessoais dos indivíduos.

2.1 Pandas

A biblioteca Pandas da linguagem Python possui grande responsabilidade no que diz respeito à grande utilização de Python para análise de dados. Oferece estruturas de dados de alto nível com funções que simplificam e agilizam o trabalho com dados estruturados ou tabulares (McKinney, 2018). McKinney ainda afirma que

“[...]o pandas combina as ideias de processamento de alto desempenho de arrays da NumPy com os recursos flexíveis de manipulação de dados das planilhas e dos bancos de dados relacionais (como o SQL). Ele disponibiliza uma funcionalidade sofisticada de indexação para facilitar a reformatação, a manipulação, as agregações e a seleção de subconjuntos de dados. Como a manipulação, a preparação e a limpeza de dados são habilidades importantes na análise de dados.”
(2018, p.20)

Seu principal objeto é o DataFrame, que é algo como uma matriz com colunas nomeadas por um cabeçalho e células que podem ser preenchidas por elementos de diferentes

tipos. Trazendo para o contexto da ciência de dados, o DataFrame pode ser considerado como a visualização dos dados em tabela onde cada linha representa um registro na base de dados.

Em resumo, “Pandas é uma biblioteca licenciada com código aberto que oferece estruturas de dados de alto desempenho e de fácil utilização voltado à análise de dados para a linguagem de programação Python” (Coelho, 2017).

2.2 Plotly e Seaborn

O Plotly e o Seaborn são duas bibliotecas amplamente utilizadas para visualização de dados em Python, embora tenham abordagens um pouco diferentes. Ambas as bibliotecas fornecem recursos poderosos para criar gráficos estatísticos e explorar conjuntos de dados, inclusive com possibilidade de integração entre diversas linguagens de programação no caso do Plotly.

O Plotly é uma biblioteca de visualização de dados interativa que permite criar gráficos dinâmicos, muito úteis para visualizações em dashboards e aplicativos web. Ele oferece suporte a uma ampla variedade de tipos de gráficos, como gráficos de dispersão, de barras, de linhas e de superfície, e permite a interação do usuário com eles por meio de zoom, rotação e seleção de pontos. (MOURA, 2022)

Por outro lado, o Seaborn é uma biblioteca de visualização estatística construída sobre o Matplotlib. Seu objetivo é fornecer uma interface simples para criar gráficos estatísticos atraentes e informativos. O Seaborn oferece estilos pré definidos que facilitam a criação de gráficos com uma aparência visualmente agradável e funcional. Além disso, também fornece recursos avançados para visualizações estatísticas, como plots de distribuição, box plots, gráficos de regressão e mapas de calor. (WASKON, 2021)

Em resumo, enquanto o Plotly é mais adequado para visualizações interativas e voltadas para a web, o Seaborn é mais focado em gráficos estatísticos esteticamente agradáveis e análise exploratória de dados. Ambas as bibliotecas são extremamente úteis para diferentes finalidades e podem ser combinadas em projetos de visualização de dados, dependendo dos requisitos específicos de cada projeto.

2.3 Matplotlib

O Matplotlib é uma biblioteca de visualização de dados amplamente utilizada na linguagem de programação Python, é referência quando se trata de plotagem de gráficos e

visualizações de dados em formato bidimensional. Embora existam outras, esta é a mais popular e amplamente utilizada ao criar objetos de apresentação estáticos, animados e interativos. (McKinney, 2018).

O Matplotlib oferece suporte a vários tipos de gráficos, como gráficos de linhas, gráficos de dispersão, histogramas, gráficos de barras, gráficos de pizza, gráficos de contorno, entre outros. Além disso, é possível criar gráficos em vários subplots ou adicionar elementos de texto, anotações e legendas para melhorar a legibilidade e a compreensão dos gráficos. (Robert Johansson, 2021, p155)

3. ORGANIZAÇÃO DOS SISTEMAS DE ENSINO: Normatizações para a garantia da permanência dos estudantes na escola

A Lei de Diretrizes e Bases da Educação Nacional (LDB), Lei n. 9.394, de 20 de dezembro de 1996, estabelece que é dever do Estado brasileiro garantir educação básica e gratuita para todos, tendo vista assegurar “[...] o pleno desenvolvimento do educando, seu preparo para o exercício da cidadania e sua qualificação para o trabalho” (BRASIL, 1996).

A referida Lei traz alguns aspectos que são importantes para esta análise. A legislação define a educação básica em três níveis: pré-escola, ensino fundamental e ensino médio, garantindo sua execução mesmo para alunos fora de faixa do ensino regular, com características e modalidades adequadas à realidade dos educandos, inclusive com relação aos turnos das aulas, permitindo que aqueles que trabalhem, não sejam prejudicados quanto ao ensino.

A respeito da organização dos sistemas educacionais, a Constituição Federal de 1988 e a atual Lei de Diretrizes e Bases estabelecem atribuições e responsabilidades aos entes federados (poderes administrativos) para manter os sistemas de ensino. Entre as atribuições da União, cabe a este elaborar o Plano Nacional de Educação (PNE), em colaboração com os Estados, Distrito Federal e Municípios; assegurar assistência técnica e financeira a todos os entes federados para manter os sistemas de ensino; e coletar, analisar e disseminar informações sobre educação no país. Entre as atribuições dos Estados, cabe definir com os Municípios formas de colaboração na oferta do ensino fundamental, garantindo sua execução, e priorizar a oferta do ensino médio de acordo com a demanda. Os Municípios, por sua vez, além de outras responsabilidades, deverão ofertar o ensino infantil (pré-escolar), e priorizar a oferta do ensino fundamental de acordo com a necessidade, garantindo sua execução.

O Governo Estadual do RN, através da Secretaria Estadual de Educação, da Cultura, do Esporte e Lazer (SEEC), detém sob sua responsabilidade as escolas da rede estadual que ofertam o ensino fundamental e o ensino médio. É também responsabilidade desta secretaria garantir a elaboração e execução de políticas e planos educacionais, em consonância com as diretrizes e planos nacionais de educação, e o Plano Estadual de Educação (2016) integrando e coordenando as suas ações e as dos seus Municípios, a fim de promover o êxito no sistema educacional desta localidade.

Na Lei n. 9.394/1996, no art. 3, são definidos princípios que devem orientar o ensino, dentre os quais destaca-se o inciso I, que expressa “igualdade de condições para o acesso e permanência na escola”. Com base nesse princípio, a lei prevê que todo estudante deve ter

assegurado as condições de acesso à escola, de modo que o poder público deve propor ações para o ingresso e permanência do estudante.

Com a implementação da Lei Nº 12.013, de 6 de agosto de 2009 que altera o art. 12 da Lei no 9.394, de 20 de dezembro de 1996, determinando às instituições de ensino obrigatoriedade no envio de informações escolares aos pais sobre a frequência e rendimento dos alunos, destaque também para a Lei nº 13.803, de 10 de janeiro de 2019, que estabelece a obrigatoriedade da unidade escolar “VIII – notificar ao Conselho Tutelar do Município a relação dos alunos que apresentem quantidade de faltas acima de 30% (trinta por cento) do percentual permitido em lei;”. Provocando a escola e secretarias para uma articulação necessária para acompanhamento dos estudantes evadidos.

A frequência dos estudantes, segundo o art. 23 disposto na Lei 9.394/96, define-se, no inciso VI, que “o controle de frequência fica a cargo da escola, conforme o disposto no seu regimento e nas normas do respectivo sistema de ensino, exigida a frequência mínima de setenta e cinco por cento do total de horas letivas para aprovação” (BRASIL, 1996). Não somente a Constituição Federal de 1988 e a Lei 9.394/96 trata sobre a universalização da educação,

Em novembro de 2009 foi promulgada a emenda constitucional nº 59, que determinava a ampliação da obrigatoriedade e gratuidade do ensino a todas as etapas da educação básica no Brasil, dos quatro aos dezessete anos de idade, prevendo sua implantação progressiva até o ano de 2016. Desse modo, a matrícula escolar obrigatória no Brasil, que até então ia dos 6 aos 14 anos de idade – o que corresponde ao Ensino Fundamental –, foi estendida também para as faixas etárias que correspondem à Pré-Escola (4-5 anos) e ao Ensino Médio (15-17 anos). O Plano Nacional de Educação (PNE), por sua vez, em consonância com a emenda constitucional nº 59, entre suas 20 metas previa, justamente, universalizar, até 2016, o atendimento escolar da população de 4 a 5 (primeira meta) e de 15 a 17 anos (terceira meta), além de garantir a matrícula escolar de toda a população de 6 a 14 anos (segunda meta).
(SALATA, 2019, p.1)

É com a inserção do Ensino Médio enquanto etapa de ensino obrigatória que a legislação reforça de que a universalização é um direito a ser garantido, e que devem ser dadas as condições a este alunado, especificamente na faixa de 15 a 17 anos, para sua permanência na via escolar até a conclusão da educação básica. No entanto, um dos problemas enfrentados nesta etapa de ensino é a evasão/abandono dos alunos.

A definição de abandono e evasão é multifacetada, e compreendida de forma distinta no campo educacional. Conforme Silva Filho e Araújo (2017) as várias formas de

interpretação não permitem chegar a uma definição precisa do que seria “evasão e abandono escolar”, uma vez que esta requer uma compreensão da complexidade dos motivos de ingresso e a trajetória dos permanentes, dos desistentes e egressos desse público. Nessa tentativa de definir o conceito de abandono e evasão fica mais desafiador quando entendemos que esse movimento de saída da escola é multifatorial, pois deve considerar o contexto sócio-histórico, cultural e político que o aluno está inserido, devemos considerar os

Fatores internos e externos, como drogas, tempo na escola, sucessivas reprovações, falta de incentivo da família e da escola, necessidade de trabalhar, excesso de conteúdo escolar, alcoolismo, localização da escola, vandalismo, falta de formação de valores e preparo para o mundo do trabalho, podem ser considerados decisivos no momento de ficar ou sair da escola, engrossando a fila do desemprego. O censo escolar de 2007(Inep/MEC) afirma que a evasão escolar entre jovens é alarmante. O Brasil tem a terceira maior taxa de abandono escolar entre os 100 países com maior IDH e no PNUD e a menor média de anos de estudo entre os países da América do Sul. (SILVA FILHO e ARAÚJO, 2017, p. 35)

Devido a complexidade do tema, é difícil definir as motivações para que o aluno abandone a escola. Conforme Salata (2019) a evasão deve ser compreendida como a finalização de um processo de desengajamento por parte do estudante em relação à escola, tanto no que se refere às suas taxas de rendimento – que se expressa no acúmulo de faltas e maior repetência – quanto social – que expõe a fragilidade dos laços estabelecidos naquele espaço em relação aos contextos nos quais os estudantes estão inseridos.

Um termômetro para identificarmos os níveis de evasão é a partir dos dados educacionais oficiais, construídos a partir de análise dos dados de Censo Escolar¹ em que se analisam dentre outros dados, as taxas de rendimentos dos alunos, neste trabalho será dado enfoque a discussão nos dados referentes à evasão e reprovação.

3.1 Indicadores Educacionais da Rede Estadual do Rio Grande do Norte(RN): Análise das taxas de rendimento dos alunos da Rede Pública Estadual do RN

O Censo Escolar é realizado anualmente e abrange todos os níveis de ensino (Educação Infantil, Ensino Fundamental e Ensino Médio) e modalidades (Ensino Regular, Educação Especial e Educação de Jovens e Adultos). Ferreira (2016), explica que são

¹ O Censo Escolar é o principal instrumento de coleta de informações da educação básica e a mais importante pesquisa estatística educacional brasileira. É coordenado pelo Inep e realizado em regime de colaboração entre as secretarias estaduais e municipais de educação e com a participação de todas as escolas públicas e privadas do país (SOUZA, 2019)

coletados dados tanto a respeito da infra-estrutura da escola, como sobre o corpo docente, matrículas, jornada escolar, rendimento dentre outros.

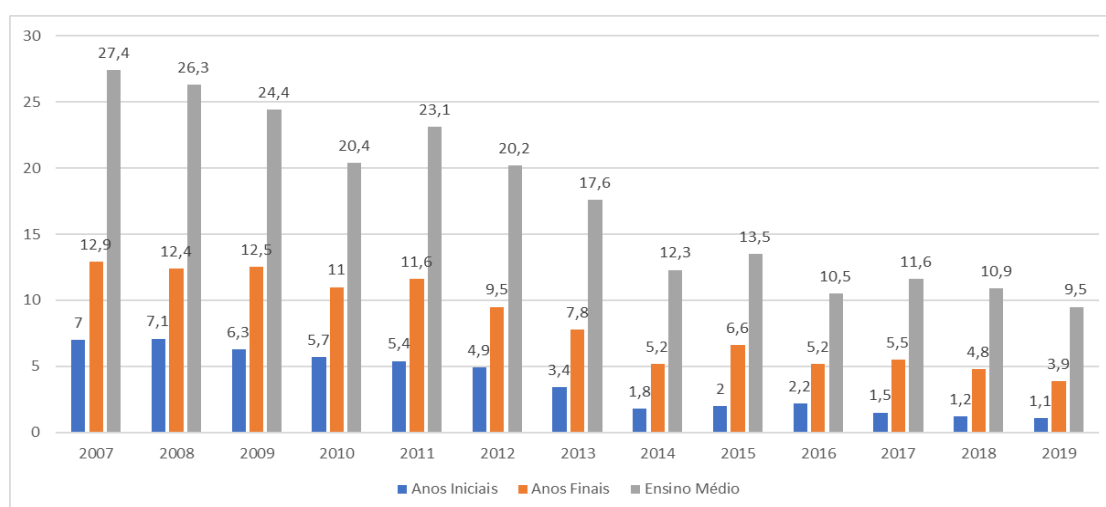
Assim como em outras áreas, o fortalecimento do sistema de coleta de dados educacionais através do censo escolar, promove um avanço no desenvolvimento de políticas públicas relacionadas ao ensino no Brasil.

De acordo com Diniz (1999) o Censo é uma ferramenta de formulação, planejamento, monitoramento, e acompanhamento das políticas públicas, auxiliando no processo de tomada de decisões, contribuindo inclusive com a transparência no que diz respeito às ações realizadas no contexto educacional no país.

Todo o processo é executado pelo INEP/MEC, em parceria com as Secretarias de Educação Estaduais e Municipais, e as escolas da rede pública, particular conveniada e privada do Brasil. Essa pesquisa tem sido aprimorada a cada ano, por meio do Sistema Educacenso, possibilitando a oferta de informações de qualidade para a formulação de políticas públicas e execução de programas na área da educação.

O Estado do Rio Grande do Norte, referente às escolas estaduais da rede pública, possui altas taxas de evasão desde a implementação da publicização das taxas de rendimentos pelo INEP em 2007, chegando a mais de 27% no Ensino Médio, conforme pode ser visualizado na Figura 1.

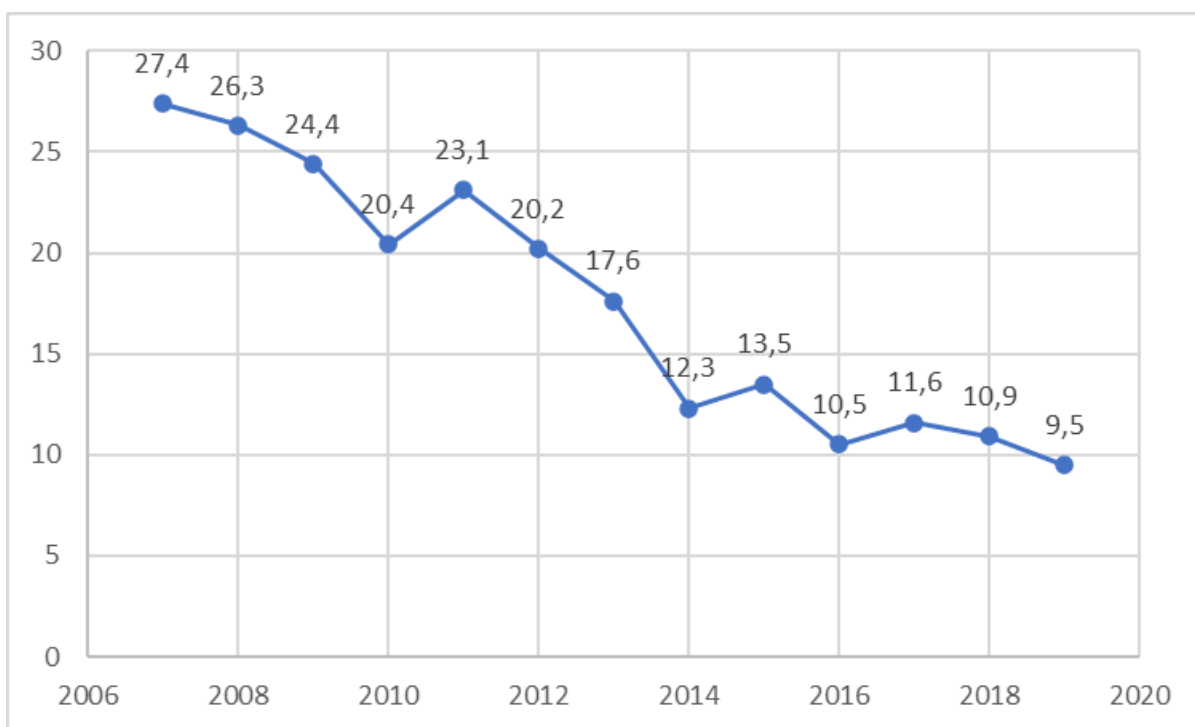
Figura 1: Taxa de evasão de alunos da Rede Pública Estadual do Rio Grande do Norte (2007-2019)



Fonte: Inep (2021)

Na série histórica, nos últimos doze anos vemos uma redução da evasão dos alunos da Rede Pública do Rio Grande do Norte. No Ensino Médio destacamos a redução de 27% em 2007 para 9,5% em 2019. Nos anos iniciais do ensino fundamental, a etapa com os menores índices de evasão, vemos uma redução de 7% para 1,1 %, o que nos mostra que estamos próximo de garantir a universalização desta etapa de ensino. Nos anos finais do ensino fundamental, em 2007, a evasão está em 12,9 pontos percentuais, e diminuiu para 3,9% em 2019, também apresentando uma melhoria no referido índice. No entanto, reforçamos que as taxas de evasão da etapa do Ensino Médio, mesmo com a diminuição dos índices de evasão ao longo de doze anos, são valores preocupantes, e nos alertam sobre a quantidade considerável de alunos que deixam de frequentar a escola. O que é possível apreender a partir dos dados, é que, comparado com as etapas de ensino ofertadas pela rede pública estadual, o Ensino Médio é uma das etapas que apresenta uma alta taxa de evasão de estudantes.

Figura 2: Taxa de evasão de alunos do Ensino Médio da Rede Pública Estadual do Rio Grande do Norte (2007-2019)



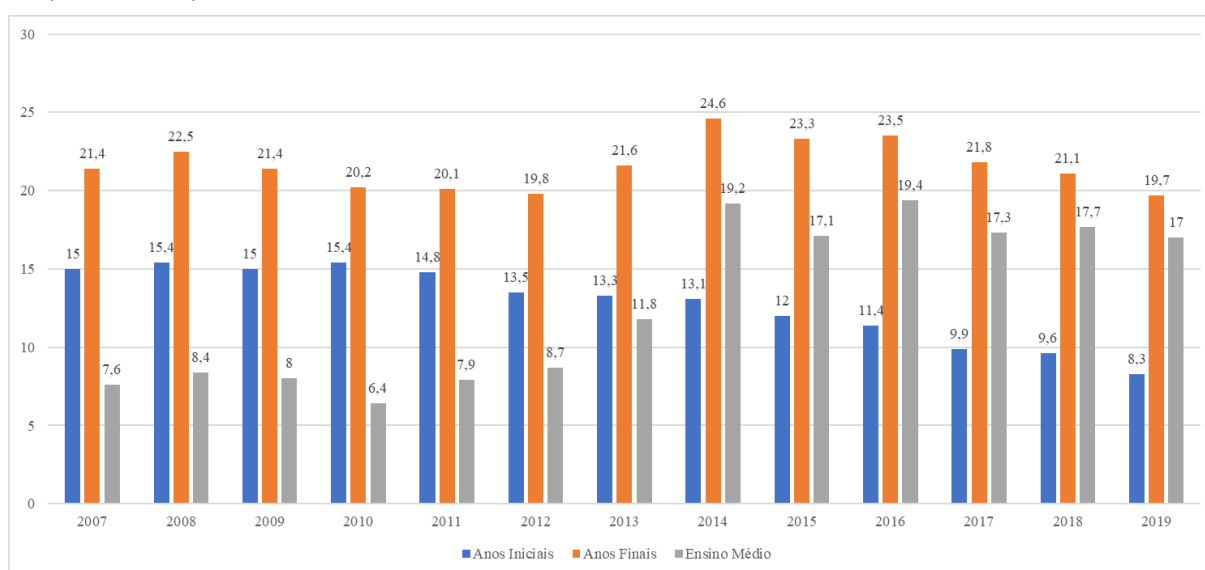
Fonte: Inep (2021)

Não é possível precisar acerca do aumento da taxa de abandono referente a 2011, mas pode-se considerar que esse aumento seja decorrente da fragilidade ou ausência de políticas educacionais que assegurem a permanência do estudante no ambiente escolar. Outros fatores

também podem ser considerados pertinentes a este aumento do indicador no ano de 2011, como as condições socioeconômicas dos estudantes.

Outro indicador importante para definir a taxa de rendimento dos alunos é a Taxa de Reprovação dos estudantes, que ao contrário das taxas de abandono, este indicador possui elevação dos indicadores, sendo os anos finais do ensino fundamental a etapa que apresenta número mais preocupantes.

Figura 3 - Taxa de reprovação de alunos da Rede Pública Estadual do Rio Grande do Norte (2007-2019)

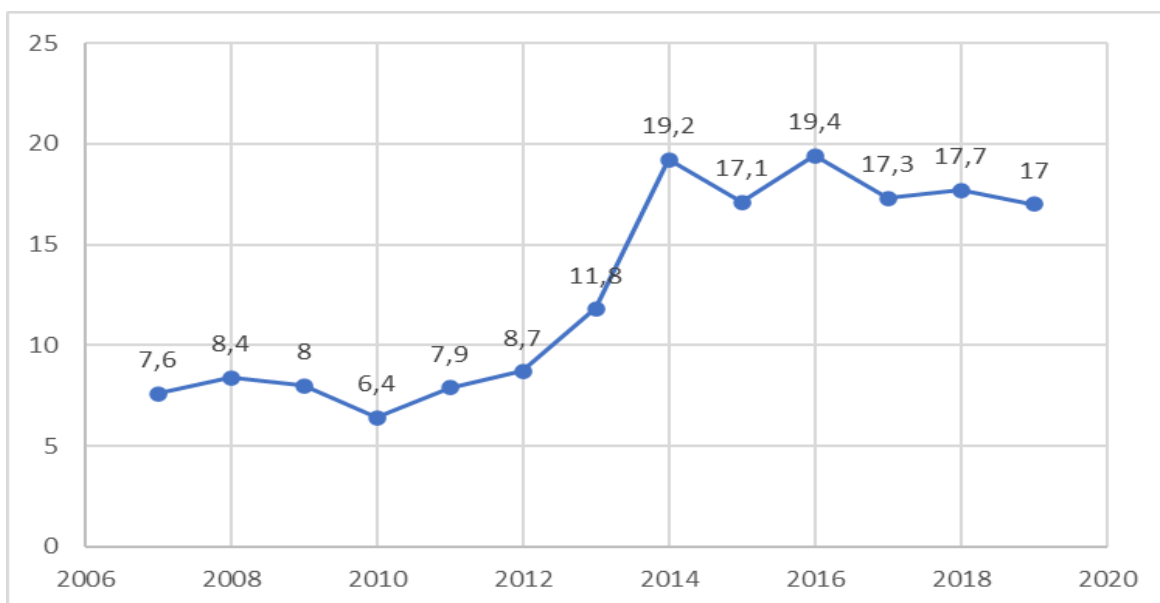


Fonte: Inep (2021)

Conforme a Figura 3, percebe-se uma queda da reprovação nos anos iniciais de maneira gradativa, em 2007 o índice é de 15% e chega a 8,3% em 2019. Já nos anos finais o índice se mantém praticamente o mesmo, em 2007 fica em 21,4%, no ano de 2019 chega a 19,7%, chegando a 24,6% em 2014, onde atinge seu maior índice de reprovação. No Ensino Médio identificamos que o índice de reprovação praticamente dobra ao longo dos anos, saindo de 7,6% em 2007 e chegando a 17% em 2019, sendo o maior nível de reprovação nesta etapa no ano de 2016 com 19,4% de reprovação.

Vemos na série histórica que as taxas de reprovação e de abandono seguem caminhos inversamente proporcionais, existe uma diminuição do abandono dos alunos, embora seja ainda um indicador consideravelmente alto, considerando que a rede estadual de ensino, conforme os dados na plataforma SIGEDUC, possui pouco mais de 200 mil estudantes matriculados.

Figura 4 - Taxa de reprovação de alunos do Ensino Médio da Rede Pública Estadual do Rio Grande do Norte (2007-2019).



Fonte: Inep (2021)

Verificamos na Figura 4 que há um aumento exponencial e preocupante, a partir de 2011, nos percentuais de reprovação, e alcançando os maiores níveis em 2014 e 2016, sendo 19,2 % e 19,4% respectivamente. Diante do alto índice de reprovação, que pode se tornar uma das razões para a evasão do aluno, é necessário que através de um sistema de monitoramento dos alunos, administrado pelo Órgão Central (SEEC), Diretorias Regionais e Escolas Estaduais colaborem com a criação de estratégias para diminuir estes índices. Conforme o Decreto Estadual 15543 de 18 de Julho de 2001 a estrutura organizacional da Secretaria Estadual de Educação, de Cultura, do Esporte e do Lazer é composta pelo órgão central (SEEC), diretorias regionais de educação e cultura e escolas estaduais

3.2 Evasão e abandono escolar: conceitos e principais diferenças

Evasão escolar é um termo utilizado para descrever a situação em que um estudante deixa de frequentar a escola sem concluir o curso ou obter uma qualificação adequada. É um fenômeno complexo que envolve múltiplos fatores, como questões socioeconômicas, falta de motivação, problemas familiares, dificuldades de aprendizagem, entre outros. A evasão escolar é vista como um problema social e educacional significativo, pois priva os indivíduos

de oportunidades de desenvolvimento pessoal e profissional, além de afetar negativamente a sociedade como um todo.

De acordo com a UNESCO (Organização das Nações Unidas para a Educação, a Ciência e a Cultura), a evasão escolar é um problema global que afeta milhões de crianças e jovens. Segundo relatório recente da UNESCO intitulado "Leave No One Behind: Building a Pathway to Universal Education", a evasão escolar continua sendo um obstáculo importante para a conquista dos Objetivos de Desenvolvimento Sustentável relacionados à educação.

Pesquisas têm identificado uma série de fatores que contribuem para a evasão escolar. Dentre eles, destacam-se as questões socioeconômicas, como pobreza, desigualdades sociais e falta de recursos financeiros, que podem levar os alunos a abandonarem a escola devido à necessidade de trabalhar para ajudar suas famílias. Além disso, problemas familiares, como violência doméstica, abuso, negligência e divórcio, também podem ter um impacto significativo na evasão escolar (UNESCO, 2021).

Outros fatores relacionados à evasão escolar incluem falta de motivação, desinteresse pelo conteúdo ensinado, dificuldades de aprendizagem não resolvidas, baixa qualidade da educação, bullying e discriminação, entre outros fatores que podem ser identificados a partir do baixo rendimento acadêmico dos estudantes e frequência ao espaço escolar (OECD, 2020).

O abandono escolar, por sua vez, está relacionado ao ato de interromper definitivamente os estudos sem a intenção de retornar à escola. É uma forma extrema de evasão escolar, em que o estudante desiste de prosseguir sua trajetória educacional e não busca outras alternativas de formação ou qualificação. O abandono escolar é frequentemente associado a problemas mais profundos, como desmotivação extrema, falta de perspectivas futuras, insatisfação com o sistema educacional, entre outros.

Em resumo, a evasão escolar refere-se à saída não concluída de um estudante da escola, enquanto o abandono escolar representa a interrupção definitiva dos estudos sem a intenção de retornar. Ambos os conceitos estão relacionados à falta de continuidade educacional e trazem consigo implicações importantes para o desenvolvimento individual e social.

Para combater a evasão escolar, políticas e programas de intervenção têm sido implementados em diferentes países. Essas estratégias incluem o fortalecimento do apoio socioeconômico às famílias, o desenvolvimento de medidas preventivas, como tutorias e acompanhamento dos alunos em risco, e a criação de programas de reengajamento escolar para recuperar aqueles que já abandonaram a escola (OECD, 2020).

4. O SISTEMA DE GERENCIAMENTO INTEGRADO DA EDUCAÇÃO – SIGEDUC

O SIGEduc é um sistema de gerenciamento específico para a educação, desenvolvido pela SIG Software e Consultoria em Tecnologia da Informação. Barbosa, Silva e Sant’ana (2015) o descrevem como um sistema utilizado para fins de controle escolar mais diretamente relacionado com a Educação básica, tendo como proposta informatizar toda a dimensão acadêmica e administrativa da rede pública de ensino estadual do RN. O mesmo possui diversos módulos e portais que permitem a utilização para fins pedagógicos além de administrativos.

O Sistema, implantado e em funcionamento desde 2012 na rede estadual do RN, foi fundamental para a continuidade da educação básica da rede pública do Estado, durante o enfrentamento do isolamento social provocado pela pandemia do Covid-19. Novos módulos e funcionalidades como a Escola Virtual (ambiente onde aluno e professor se comunicam diretamente através de materiais e atividades), foram acrescentados ao SIGEduc.(Flor, 2020).

A implantação do SIGEduc foi mais uma etapa no processo de avanço tecnológico enfrentado por todas as esferas da sociedade, promovendo a inserção cada vez mais ativa das Tecnologias de Informação e Comunicação (TICs) no cotidiano das instituições públicas e privadas, em quaisquer áreas, como saúde, segurança e inclusive educação. Araújo (2019) afirma que o desenvolvimento desses sistemas computacionais promovem uma melhoria no processo de ensino aprendizagem na educação.

A comunidade acadêmica da rede de ensino pública estadual é composta pelos estudantes e seus familiares, docentes e servidores nas mais diversas funções. Todos esses, participam ativamente como usuários do SIGEduc. Assim como qualquer outro sistema, o SIGEduc possui um banco de dados onde armazena informações coletadas pelos seus mais diversos usuários e às mantém protegidas, utilizando-as para fins de melhor gerenciamento da educação, através de ferramentas como atualização cadastral anual dos servidores e questionário socioeconômico para os estudantes no momento da matrícula online. Todos os dados recolhidos através do SIGEDUC são migrados para a plataforma do governo federal Educacenso, para atualização anual do censo brasileiro da educação.

4.1 A evasão representada no SIGEDUC

Através do SIGEDUC, a comunidade acadêmica, composta pelos órgãos gestores, professores, alunos e familiares, consegue obter informações de desempenho escolar dos

estudantes, desde as notas nos componentes da estrutura curricular, até relatórios e gráficos informativos a respeito das aprovações e reprovações. Porém, não existe algo especificamente útil para o monitoramento da evasão na rede dentro do sistema. É possível obter relatórios contendo o percentual de evasão por município, DIREC, escola, e etapa de ensino, como mostra a figura 5.

Figura 5 – Relatório com percentual de evasão por município do Rio Grande do Norte

The screenshot displays the SEEC - SIGEduc system interface. At the top, the header shows the user name MAURICIO THIAGO F. DE LIMA and the session time 01:00. Below the header, there are navigation tabs for 'Turma', 'Diá...', 'Educação Espec...', 'Enem', and 'Educação Hospitalar e Domicil...'. The main menu area is divided into several sections:

- Avanço de Estudos**
 - Avaliar Avanço de Estudos (14)
 - Relatório de Solicitação de Avanço de Estudos
 - Gerenciar Séries que Permitem Avanço de Estudo
- Retenções**
 - Retenções por Componente
 - Retenções por DIREC/Município
 - Retenções por Município - Sintético
 - Retenções por Escola
- Documentos**
 - Emitir Boletim do Estudante
 - Ata de Resultados Finais
 - Histórico Escolar
 - Acompanhamento da Digitação da Ata de Coleta de Resultados
 - Relatório Sintético de Lançamento da Ata
- Deixou de Frequentar** (highlighted with a red box)
 - Percentual de Alunos que Deixaram de Frequentar por DIREC/Município
 - Percentual de Alunos que Deixaram de Frequentar por Município - Analítico
 - Percentual de Alunos que Deixaram de Frequentar por Escola
 - Percentual de Alunos que Deixaram de Frequentar por Etapa de Ensino/Ano Escolar/Série
 - Relatório de Previsão de Abandono Escolar
- Frequência**
 - Relatório de Taxa de Frequência/Nota por Componente
 - Relatório de Taxa de Frequência/Nota por Etapa de Ensino
 - Relatório de Taxa de Frequência/Nota por DIREC/Município
 - Relatório de Taxa de Frequência/Nota por Município - Analítico
 - Relatório de Taxa de Frequência/Nota por Escola
- Histórico**
 - Implantar Histórico

At the bottom of the interface, there is a 'Menu Principal' button and a footer with system information: SIGEduc | GPD/SEEC - (84) 3232-1316 -seec-app4.imd.ufrn.br.srv4inst1 - v20230622_1100.

Fonte: Sistema Integrado de Gestão da Educação

Existe também um relatório de previsão de abandono organizado por município, DIREC, Escola, turmas que determina a chance em percentual de um determinado estudante abandonar a escola.

5. ANÁLISE DE DESEMPENHO ESCOLAR BASEADA EM DADOS

Neste trabalho, as atividades de planejamento, obtenção e análise dos dados serão realizadas nas seguintes etapas: identificação e escolha da base de dados, coleta, estruturação da base de dados e seleção de variáveis, tratamento dos dados, análise exploratória e visualização dos dados. O ciclo de atividades com suas respectivas etapas pode ser visualizado na Figura 6.

Figura 6 – Ciclo de atividades em ciência de dados



Fonte: Elaborado pelo autor

5.1 Entendimento do problema e escolha da base de dados

Este trabalho adota uma metodologia baseada em ciência de dados para analisar o desempenho escolar dos alunos da rede pública de educação do Estado do RN que cursaram o Ensino Médio entre os anos de 2017 e 2019. Pretende-se verificar variáveis de maior impacto em relação à evasão escolar na amostra analisada, bem como identificar a relação entre as variáveis e sua distribuição, como por exemplo a idade do estudante, gênero, condição socioeconômica, notas e frequência. Segundo Sampieri, Collado e Lucio (2013), estudos sobre correlação entre os atributos têm por finalidade conhecer a relação ou grau de associação que existe entre dois ou mais conceitos, categorias ou variáveis em um contexto específico.

As variáveis escolhidas para a análise são divididas em quatro categorias e estão descritas na Tabela 1.

Tabela 1 – Descrição das variáveis de análise

Detalhamento das variáveis		
Dados pessoais	ESTUDANTE	Nome do estudante
	CPF	CPF do estudante
	GENERO	Gênero do estudante (masculino ou feminino)
	IDADE	Idade do estudante ao cursar a série
	ENDERECO	Rua, número, bairro e CEP da residência do estudante
Dados de matrícula	ANO	Ano de matrícula (2017 a 2019)
	DIREC	DIREC a qual o município pertence (1ª a 16ª)
	CIDADE	Cidade onde a escola está localizada
	ESCOLA	Escola de matrícula
	TURMA	Turma do estudante no ano letivo
	ETAPA	Etapa de ensino (Ensino Médio)
	SERIE	Série de matrícula (1ª a 3ª)
	TURNOS	Turno de matrícula (Matutino, Vespertino ou Integral)
	MATRICULA	Número de matrícula do estudante
Rendimento escolar	SITUACAO_FINAL	Situação final no ano letivo (Aprovado, Reprovado, Deixou de frequentar)
	QUANTIDADE_REPROVACAO	Quantidade de vezes que o estudante reprovou na mesma série em toda a vida escola
	QUANTIDADE_EVASAO	Quantidade de vezes que o estudante evadiu na mesma série em toda a vida escola
	FREQUENCIA	Média anual de frequência
	NOTAS	Média anual de notas
	NOTA_PORTUGUES	Nota final de português

	FREQUENCIA_PORTUGUES	Frequência final de português
	NOTA_MATEMATICA	Nota final de matemática
	FREQUENCIA_MATEMATICA	Frequência final de matemática
Contexto socioeconômico	RENDA_MENSAL	Renda mensal familiar (Até 1 salário mínimo, Mais de 1 até 2 salários mínimos, Mais de 2 até 5 salários mínimos, Mais de 5 até 10 salários mínimos, Mais de 10 até 20 salários mínimos, Mais de 20 salários mínimos)
	MEMBROS_FAMILIA	Quantas pessoas moram com o estudante (Um, Dois, Três, Quatro a seis, Sete ou Mais, Não mora com familiares)
	BOLSA_FAMILIA	Participante do bolsa família (sim ou não)
	TRABALHA	Se o estudante trabalha (sim ou não)

Fonte: Elaborada pelo autor

A frequência total anual e a nota anual são calculados a partir da soma dos resultados em cada componente dividida pelo total deles. Estão sendo consideradas apenas duas casas decimais de precisão.

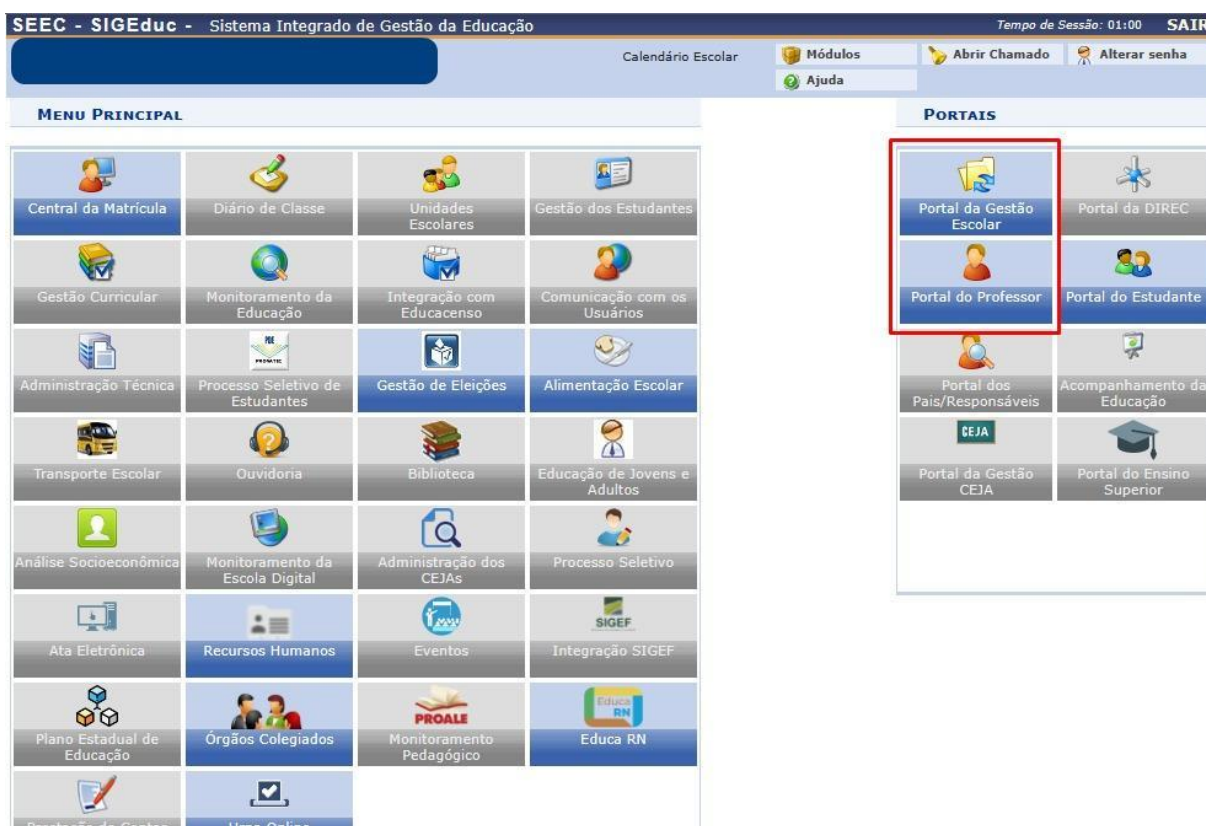
Deseja-se analisar os dados dentro de uma série histórica que abrange os anos de 2017 a 2019, para cada município e microrregião delimitada pelas Diretorias de Ensino DIRECs, observando o comportamento e desenvolvimento escolar do estudante de acordo com as variáveis estabelecidas na Tabela 1. Embora existam variações na oferta da etapa de ensino, esse conjunto de dados considera apenas estudantes pertencentes a etapa Ensino Médio regular, dos turnos matutino, vespertino e integral.

5.2 Coleta dos Dados

Os dados utilizados nesta pesquisa são originados da base de dados do SIGEduc. São dados coletados e atualizados pelos usuários periodicamente. Os dados relacionados ao desempenho escolar, frequência e notas dos estudantes, são coletados durante todo o período de aulas e atualizados através da consolidação ao final do ano letivo. Os dados de matrícula e os dados relacionados ao contexto socioeconômico dos estudantes são obtidos durante o processo de matrículas através do portal público do SIGEDUC, onde a comunidade acadêmica externa consegue realizar as solicitações de matrícula no período correspondente, ou são obtidos no decorrer do ano letivo onde de acordo com a disponibilidade da escola, são feitas novas matrículas.

Dentre os diversos módulos do Sistema, existem os portais da gestão escolar e do professor, cuja interface de acesso é mostrada na Figura 7. Esses são os ambientes para matrículas durante o ano letivo e registros dos dados acadêmicos, respectivamente.

Figura 7 – Portal de Gestão Escolar e Portal do Professor



Fonte: Sistema Integrado de Gestão da Educação

Quando aberto o período de solicitação de matrículas para o novo ano letivo, é público que os alunos, pais, familiares e responsáveis podem se inscrever inserindo dados pessoais, da matrícula desejada e o questionário socioeconômico. A Figura 8 mostra um exemplo desse questionário.

Figura 8 – Questionário socioeconômico

MATRÍCULA > QUESTIONÁRIO SOCIOECONÔMICO				
CONFIRMAR NOVO QUESTIONÁRIO				
Título: Questionário Sócio Econômico				
PERGUNTAS				
Pergunta	Alternativas		Tipo	Obrigatória
Qual sua religião?	Nenhuma Católica Anglicana Protestante Espírita Judáica Umbanda Candomblé Outras		ÚNICA ESCOLHA	Sim
Tipo de Residencia em que você mora:	Casa de Alvenaria Casa de barro/taipa Apartamento Orfanato, República, casa de estudantes ou pensionato Outro		ÚNICA ESCOLHA	Sim
Situação de moradia:	Mora em casa própria dos pais Mora em casa própria dos parentes Mora em casa alugada pelos pais Mora em casa alugada pelos parentes Mora em sua própria casa/Mora em casa alugada por você Mora em Orfanato, república, casa de estudantes ou pensionato Mora em casa de amigos Mora em casa situada em área de invasão		ÚNICA ESCOLHA	Sim
Local de Moradia			ÚNICA ESCOLHA	Sim
Indique as pessoas que residem com você:			MULTIPLA ESCOLHA	Sim

Fonte: Sistema Integrado de Gestão da Educação

Os dados para esta pesquisa foram solicitados por meio de processo administrativo de nº 00410002003131/2022-84 à SEEC a disponibilização dos dados de interesse para este estudo. Embora não sejam utilizados dados considerados sensíveis, foi necessário seguir o protocolo de solicitação de cessão de acordo com a LGPD. Após a deliberação por parte do órgão possuidor, os dados foram cedidos em formato .csv como solicitado.

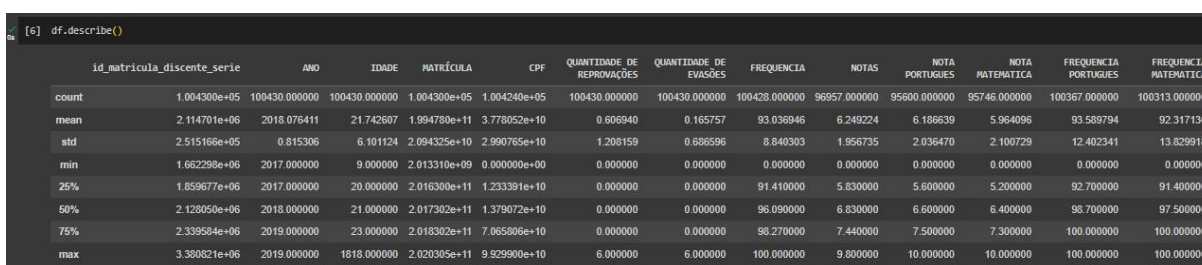
5.3 Tratamento dos Dados

A parte que corresponde ao tratamento dos dados é muito importante para o bom desenvolvimento das próximas etapas do projeto. É nesse momento em que os dados são organizados de acordo com o objetivo da pesquisa e das análises. Os dados foram manuseados através de importação no ambiente de desenvolvimento virtual Google Collaboratory através da biblioteca Pandas e convertidos em DataFrames.. A importação é feita através do comando `pd.read_csv()`.

Após isso, utilizando a função `.info()` é possível observar um panorama abrangente do DataFrame e suas características, colunas ou variáveis, quantidade de registros ou linhas, tipos de dados etc. Isso ajuda a refinar ainda mais o tratamento dos dados, deixando o conjunto mais adequado para o trabalho. Uma informação importante obtida através desta função é a quantidade de campos vazios, que quando visualizamos o DataFrame, é representado pela sigla NaN, do inglês Not a Number. Saber isso é importante para quando se estiver trabalhando com dados numéricos escolher a melhor estratégia de tratamento: preenchimento, substituição, ou exclusão desses campos.

O método `describe()` no Pandas é utilizado para gerar um resumo estatístico dos dados contidos em um DataFrame. Esse comando fornece informações essenciais, como a contagem de observações, a média, o desvio padrão, os valores mínimo e máximo, e os quartis (25%, 50% e 75%) para cada coluna numérica do DataFrame, bastante útil para o conjunto de dados analisado neste trabalho, por se tratar de notas e valores de medição de frequência por exemplo.

Figura 9 – Método `describe()`



```
[6] df.describe()
```

	id_matricula_discente_serie	ANO	IDADE	MATRÍCULA	CPF	QUANTIDADE DE REPROVAÇÕES	QUANTIDADE DE EVASÕES	FREQUENCIA	NOTAS	NOTA PORTUGUES	NOTA MATEMATICA	FREQUENCIA PORTUGUES	FREQUENCIA MATEMATICA
count	1.004300e+05	100430.000000	100430.000000	1.004300e+05	1.004240e+05	100430.000000	100430.000000	100428.000000	95957.000000	95600.000000	95746.000000	100367.000000	100313.000000
mean	2.114701e+06	2018.076411	21.742607	1.994780e+11	3.778052e+10	0.606940	0.165757	93.036946	6.249224	6.186639	5.964096	93.589794	92.317136
std	2.515166e+05	0.815306	6.101124	2.094325e+10	2.990765e+10	1.208159	0.686596	8.840303	1.956735	2.036470	2.100729	12.402341	13.829918
min	1.662298e+06	2017.000000	9.000000	2.013310e+09	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.859677e+06	2017.000000	20.000000	2.016300e+11	1.233391e+10	0.000000	0.000000	91.410000	5.830000	5.600000	5.200000	92.700000	91.400000
50%	2.128050e+06	2018.000000	21.000000	2.017302e+11	1.379072e+10	0.000000	0.000000	96.090000	6.830000	6.800000	6.400000	98.700000	97.500000
75%	2.339584e+06	2019.000000	23.000000	2.018302e+11	7.065806e+10	0.000000	0.000000	98.270000	7.440000	7.500000	7.300000	100.000000	100.000000
max	3.380821e+06	2019.000000	1818.000000	2.020305e+11	9.929900e+10	6.000000	6.000000	100.000000	9.800000	10.000000	10.000000	100.000000	100.000000

Fonte: Elaborada pelo Autor

Figura 10 – Método info()

```
[7] df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100430 entries, 0 to 100429
Data columns (total 30 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   id_matricula_discente_serie              100430 non-null  int64
1   ANO                                       100430 non-null  int64
2   DIREC                                    100430 non-null  object
3   CIDADE                                   100430 non-null  object
4   ESCOLA                                   100430 non-null  object
5   TURMA                                    100430 non-null  object
6   ESTUDANTE                                100430 non-null  object
7   GÊNERO                                   100430 non-null  object
8   IDADE                                    100430 non-null  float64
9   ENDEREÇO                                  100239 non-null  object
10  BAIRRO                                    99274 non-null  object
11  CEP                                       99982 non-null  object
12  RENDA MENSAL                             100426 non-null  object
13  MEMBROS DA FAMÍLIA                       100422 non-null  object
14  BOLSA FAMÍLIA                            100430 non-null  object
15  TRABALHA                                  100413 non-null  object
16  MATRÍCULA                                100430 non-null  int64
17  CPF                                       100424 non-null  float64
18  ETAPA                                    100430 non-null  object
19  SÉRIE                                    100430 non-null  object
20  TURNO                                    100430 non-null  object
21  SITUAÇÃO FINAL                          100430 non-null  object
22  QUANTIDADE DE REPROVAÇÕES                100430 non-null  int64
23  QUANTIDADE DE EVASÕES                   100430 non-null  int64
24  FREQUENCIA                               100428 non-null  float64
25  NOTAS                                    96957 non-null  float64
26  NOTA PORTUGUES                           95600 non-null  float64
27  NOTA MATEMATICA                          95746 non-null  float64
28  FREQUENCIA PORTUGUES                     100367 non-null  float64
29  FREQUENCIA MATEMATICA                    100313 non-null  float64
dtypes: float64(8), int64(5), object(17)
memory usage: 23.0+ MB
```

Fonte: Elaborada pelo Autor

A Figura 10 possibilita visualizar a aplicação do método info() no DataFrame indicando o tipo de cada variável, e se existe a presença de valores nulos. Para ver essa informação mais detalhadamente utiliza-se o método isnull().sum() como mostrado na Figura 11, que retorna o total de campos nulos por coluna do DataFrame (DF).

Figura 11 – Contando valores vazios com isnull().sum()

```
df.isnull().sum().sort_values(ascending=False)[:10]
NOTA PORTUGUES      4830
NOTA MATEMATICA     4684
NOTAS                3473
BAIRRO              1156
CEP                  448
ENDEREÇO             191
FREQUENCIA MATEMATICA  117
FREQUENCIA PORTUGUES  63
TRABALHA             17
MEMBROS DA FAMÍLIA   8
dtype: int64
```

Fonte: Elaborada pelo Autor

A partir das informações obtidas com os métodos detalhados anteriormente, decidiu-se remover as colunas 'id_matricula_discente_serie', 'ENDEREÇO', 'BAIRRO', 'CEP', 'MATRÍCULA', 'ETAPA', 'TURNO', 'CPF' que possuíam variáveis não úteis para o foco dessa análise, utilizando o método `drop()`, bem como decidiu-se também a exclusão das linhas que possuíam dados faltantes, através do método `dropna()`. Como deseja-se verificar o desempenho escolar dos estudantes, sobretudo dos alunos evadidos, a partir de informações como a nota e frequência, não é prudente preencher esses campos nulos com nenhum outro valor aproximado. O método `shape` do Pandas mostra a quantidade de linhas e colunas do DF, nesse caso originalmente possuindo 100430 linhas e 30 colunas. Após os ajustes o DF passou a ter 96755 linhas e 22 colunas. É interessante observar que a ausência de muitos registros de notas e frequência, como percebido na Figura 11, indica que todos esses registros estão divididos entre os estudantes que não foram aprovados.

Foi necessário utilizar o método `replace()` para normalizar as informações da coluna 'SÉRIE' pois percebeu-se mais de uma nomenclatura se referindo à mesma série, dessa forma divergindo dados.

Através do método `head()` é possível visualizar uma quantidade de linhas do DF, como demonstrado na Figura 12, para verificar o formato dos dados contidos na base.

Figura 12 - Método head()

ANO	DIREC	CIDADE	ESCOLA	TURMA	ESTUDANTE	GÊNERO	IDADE	RENDA MENSAL	MEMBROS DA FAMÍLIA	SÉRIE	SITUAÇÃO FINAL	QUANTIDADE DE REPROVAÇÕES	QUANTIDADE DE EVASÕES	FREQUENCIA	NOTAS	NOTA PORTUGUES	NOTA MATEMATICA	FREQUENCIA PORTUGUES	FREQUENCIA MATEMATICA
3473	09ª DIREC - CURRAIS NOVOS	ACARI	ESCOLA ESTADUAL PROFESSORA RAQUELA BRANDÃO DE ...	PROEMM3ªSA	ALYSSANDRA VIANA FONSECA	FEMININO	22.0	Mais de 1 até 2 salários mínimos	Quatro a seis	3ª SÉRIE	APROVADO	0	0	99.08	9.80	10.0	10.0	100.0	100.0
3474	12ª DIREC - MOSSORÓ	AREIA BRANCA	ESCOLA ESTADUAL CONSELHEIRO BRITO GUERRA	EMM3A	JARDIANNY KELLIS DE SOUSA CARNEIRO	FEMININO	21.0	Mais de 1 até 2 salários mínimos	Quatro a seis	3ª SÉRIE	APROVADO	0	0	99.21	9.80	9.1	10.0	100.0	100.0
3475	09ª DIREC - CURRAIS NOVOS	CURRAIS NOVOS	ESCOLA ESTADUAL TRISTÃO DE BARROS	PIINT2ªB	FRANCISCO FERNANDO CAPE	MASCULINO	20.0	Mais de 1 até 2 salários mínimos	Quatro a seis	2ª SÉRIE	APROVADO	0	0	99.23	9.80	9.9	10.0	98.4	100.0
3476	09ª DIREC - CURRAIS NOVOS	ACARI	ESCOLA ESTADUAL DOUTOR JOSÉ GONÇALVES DE MEDEIROS	EMV2A	RIELLY CRISTINA ALVES GOMES	FEMININO	21.0	Mais de 1 até 2 salários mínimos	Três	2ª SÉRIE	APROVADO	0	0	94.12	9.74	9.4	10.0	98.3	93.3
3477	09ª DIREC - CURRAIS NOVOS	ACARI	ESCOLA ESTADUAL DOUTOR JOSÉ GONÇALVES DE MEDEIROS	EMV1A	POLIANA DE SOUSA MEDEIROS	FEMININO	19.0	Até 1 salário mínimo	Quatro a seis	1ª SÉRIE	APROVADO	0	0	99.24	9.71	9.7	10.0	98.8	98.8

5 rows x 22 columns

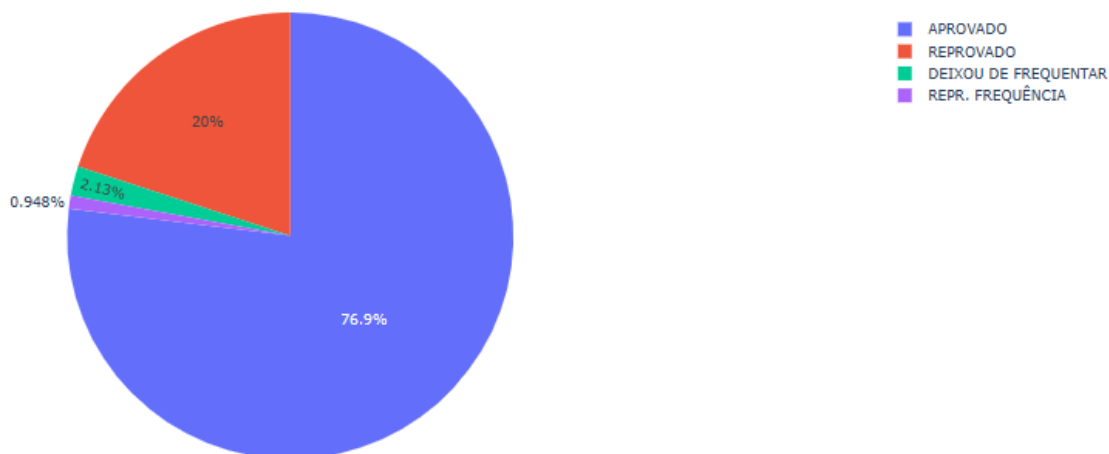
Fonte: Elaborada pelo autor

5.4 Análise Exploratória

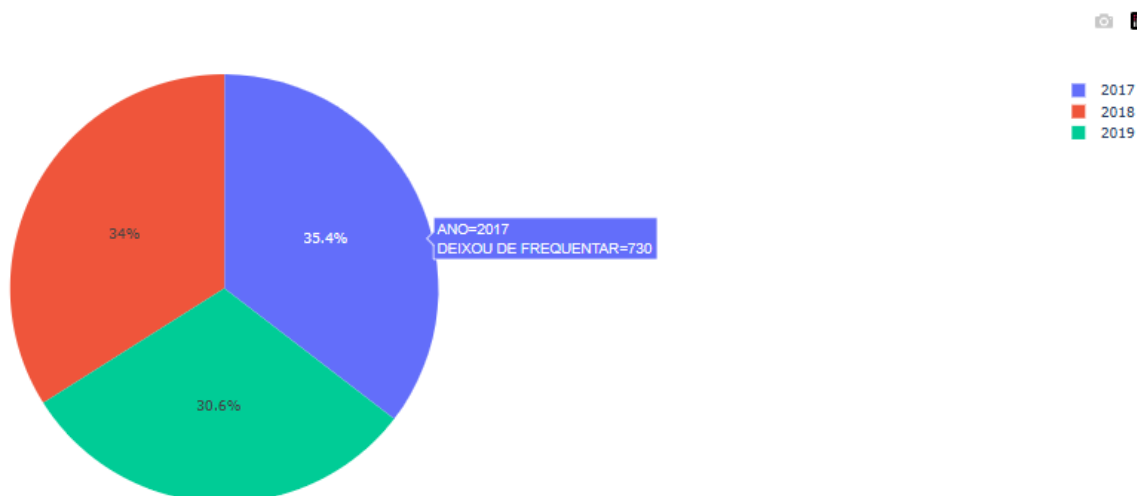
A análise exploratória dos dados é a etapa onde se observa os dados, procurando suas características principais para entender o que fazer a partir delas. É nesse momento que pode-se fazer perguntas sobre eles e buscar respondê-las através, geralmente, dos recursos de visualização dos dados estatísticos. Barros (2020) afirma que a análise exploratória de dados (em inglês, Exploratory Data Analysis - EDA) é uma maneira de fazer avaliações iniciais sobre a distribuição populacional das variáveis usando os dados das amostras observadas, e além disso maximizar a percepção de um conjunto de dados, descobrir estruturas subjacentes, extrair variáveis importantes, testar premissas, detectar outliers e anomalias.

A primeira análise a ser feita é a distribuição dos estudantes da amostra, por situação final. Isso possibilitará o entendimento da porcentagem de alunos evadidos, que é o foco deste trabalho. A Figura 13 mostra essa distribuição indicando que apenas 2,13% dos alunos foram registrados com status ‘Deixou de Frequentar’, e mais de 20% dos alunos foram registrados como ‘Reprovados’. A Figura 14 indica o total de alunos evadidos por ano.

Figura 13 - Distribuição dos alunos por situação final



Fonte: Elaborada pelo autor

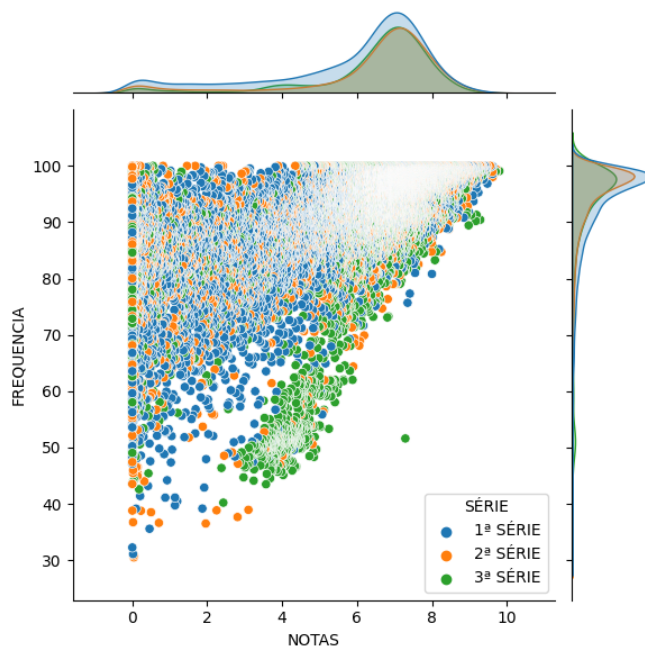
Figura 14 - Total de alunos evadidos por ano

Fonte: Elaborada pelo autor

Para a elaboração da Figura 14 foi necessário separar a coluna ‘Situação final’ em outras quatro colunas de acordo com a categoria. Isso foi feito através do método `pd.get_dummies()` que transforma o dado categórico em numérico e transforma cada valor distinto da coluna original em uma nova coluna que passa a ser preenchida com valores inteiros 1 ou 0 de acordo com a ocorrência daquela informação.

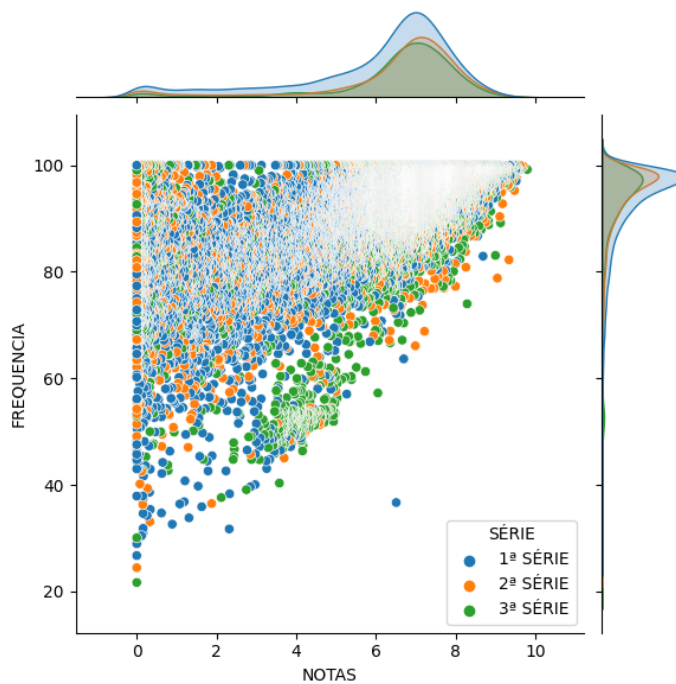
Como é possível verificar na Figura 14 em cada ano de análise a ocorrência de evasão se manteve equivalente, proporcionando uma divisão de dados equilibrada por ano. Diante disso, os alunos foram agrupados por ano, para uma análise mais profunda. As Figuras 15, 16 e 17 mostram a relação entre nota e frequência entre os alunos agrupados por série em cada ano. Em cada gráfico pode-se visualizar um agrupamento de estudantes com baixa nota e frequência pertencentes à 3ª série, enquanto os da 1ª série possuem as maiores notas e melhores frequências e os alunos da 2ª série são melhor distribuídos. Para verificar se há alguma relação entre esses resultados e a evasão escolar, será observada a distribuição de alunos evadidos por série em cada ano, na Figura 18.

Figura 15 - Relação de notas e frequência por série em 2017

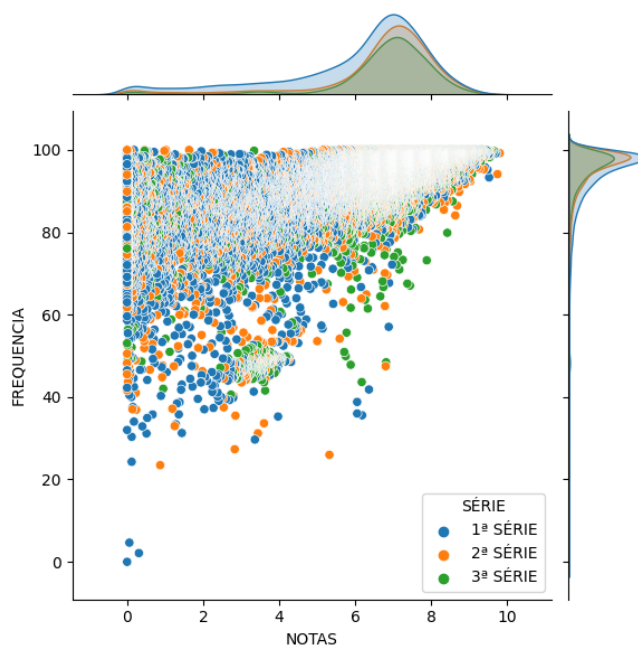


Fonte: Elaborada pelo autor

Figura 16 - Relação de notas e frequência por série em 2018



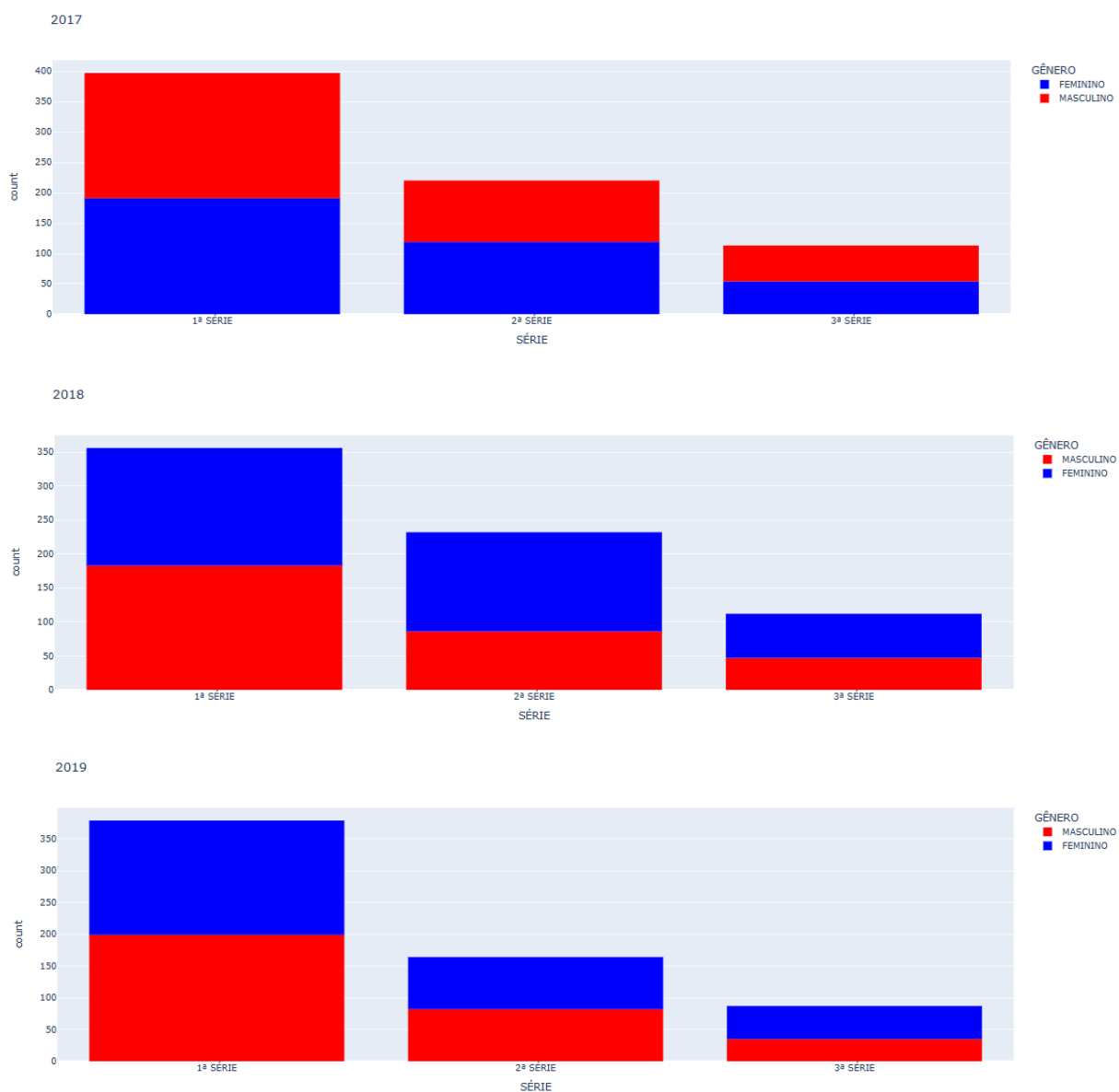
Fonte: Elaborada pelo autor

Figura 17 - Relação de notas e frequência por série em 2019

Fonte: Elaborada pelo autor

A Figura 18, não confirma a relação entre o número de evasão e a série com menores índices de nota e frequência. Algo a se analisar, é o percentual de estudantes por série. Dessa forma, se a comparação do total de estudantes por série e total de evadidos por série se mostrar proporcional, então a hipótese ganharia força. O gênero dos estudantes também se apresenta equilibrado ao variar a série e o ano. Isso confirma que o índice de nota e frequência podem indicar evasão, mas o gênero e a série parecem não ser muito relevantes.

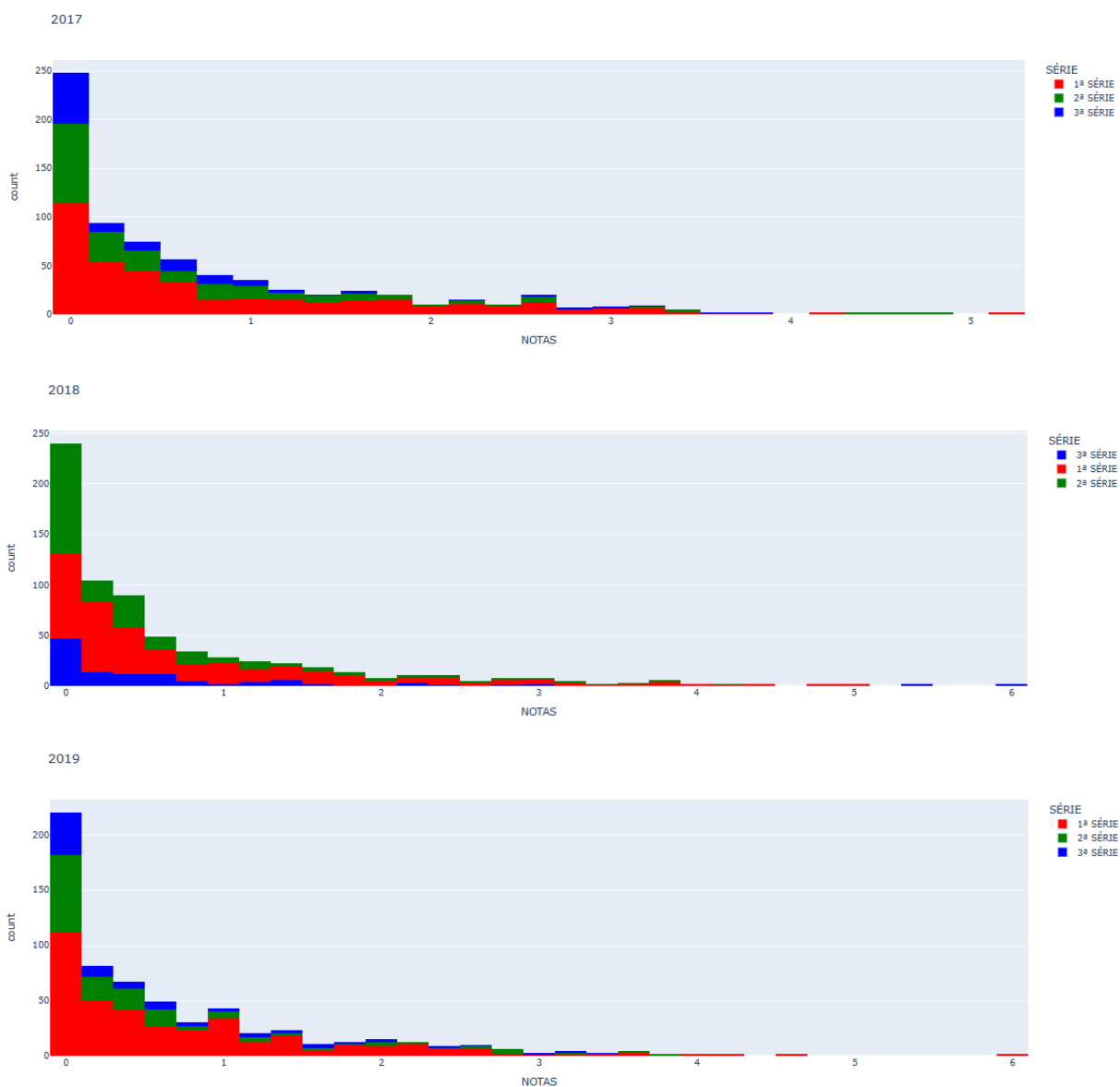
Figura 18 - Distribuição de estudantes evadidos por série e gênero em 2017, 2018 e 2019



Fonte: Elaborada pelo autor

A Figura 19 corrobora com a ideia apresentada nas Figuras 15, 16 e 17 reforçando que dentro do contexto de alunos evadidos, a grande maioria das notas zeradas é pertencente aos alunos da 3ª série.

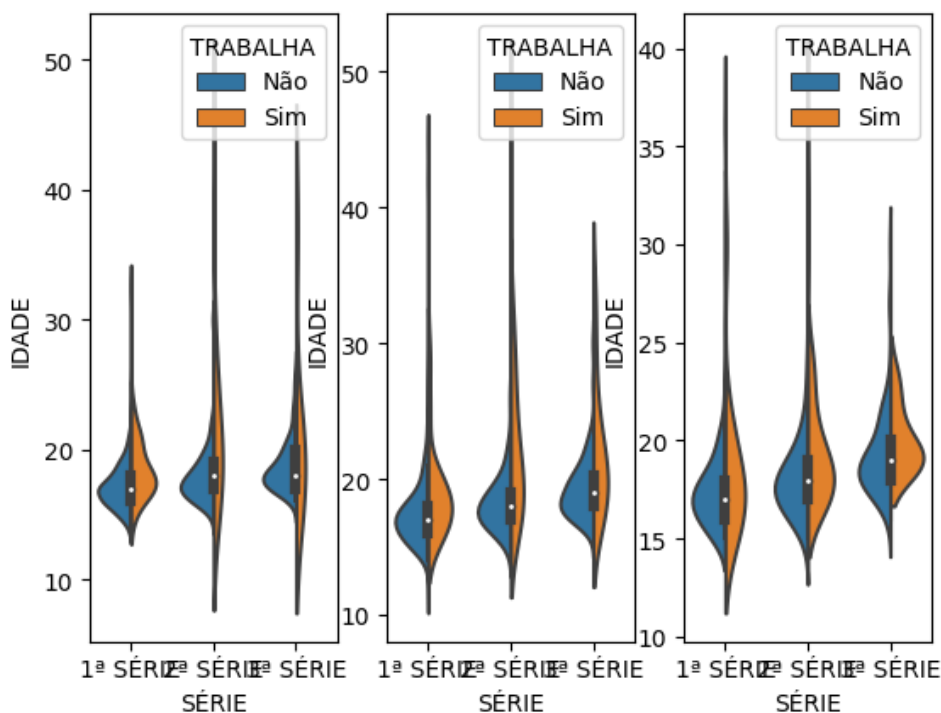
Figura 19 - Distribuição de notas por série para alunos evadidos em 2017, 2018 e 2019



Fonte: Elaborada pelo autor

Outra análise a ser feita se refere a idade do estudante evadido, durante os anos, agrupado por série. Essa análise, através da Figura 20 que apresenta um gráfico do tipo violino, indica que existem outliers no campo 'Idade' que podem ou não prejudicar a leitura correta das informações, a depender da aplicação do negócio.

Figura 20 - Gráfico de violino mostrando a distribuição dos estudantes evadidos por idade ao longo dos anos 2017, 2018 e 2019 respectivamente.

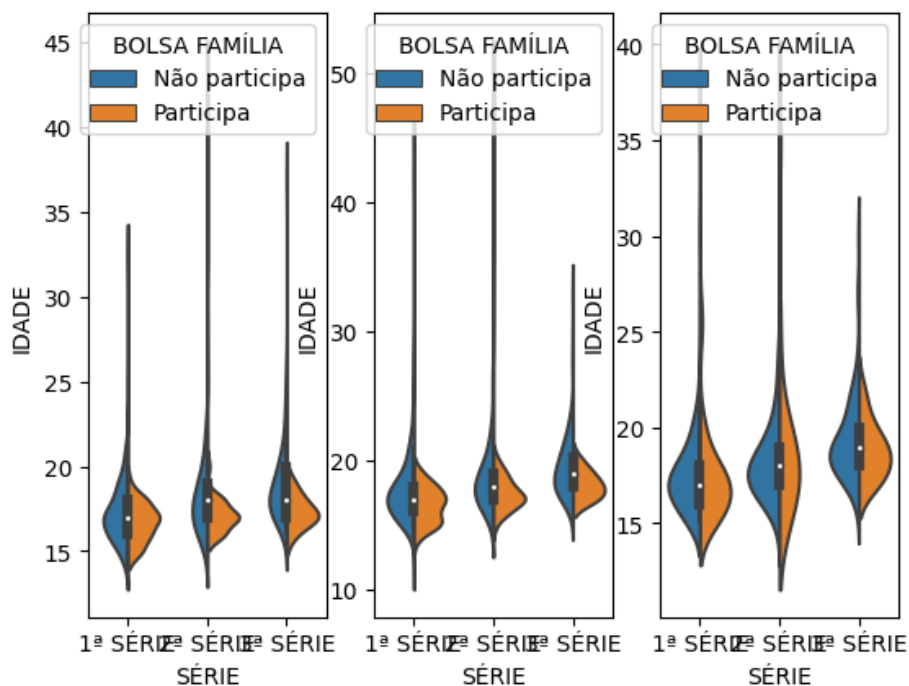


Fonte: Elaborada pelo autor

A Figura 20 mostra também a distribuição dos estudantes evadidos que trabalham em relação à idade. É possível observar que a maior parte dos evadidos não trabalhavam no ano em questão, porém os que trabalhavam e evadiram, possuíam idade mais avançada. O que pode significar uma possível relação entre o índice de evasão e o contexto socioeconômico em que o estudante está situado. Utilizando a mesma lógica, verificou-se a relação entre os estudantes que faziam parte do programa governamental de assistência social ‘Bolsa Família’, a idade e o fato de terem evadido. O resultado pode ser conferido na Figura 21.

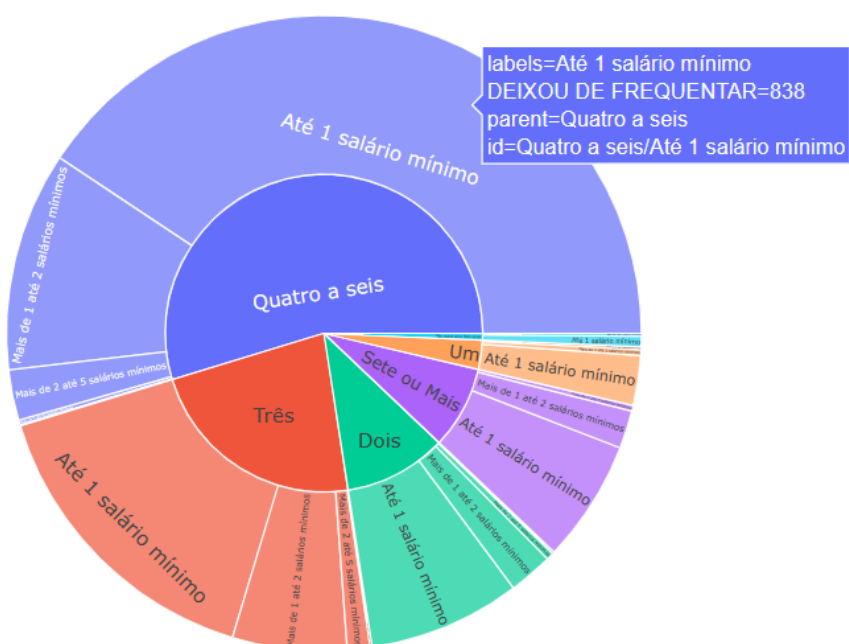
Existe uma menor incidência dos estudantes evadidos que participavam do Bolsa Família em comparação aos estudantes que não participavam. Isso pode indicar que existe relação das variáveis socioeconômicas com o resultado final do estudante evadido. Para investigar essa hipótese, irá se analisar a proporção de outros dados nesse contexto.

Figura 21 - Gráfico de violino mostrando a distribuição dos estudantes evadidos por idade ao longo dos anos 2017, 2018 e 2019 respectivamente.



Fonte: Elaborada pelo autor

Figura 22 - Gráfico de pizza mostrando a contagem de estudante evadidos por condição socioeconômica

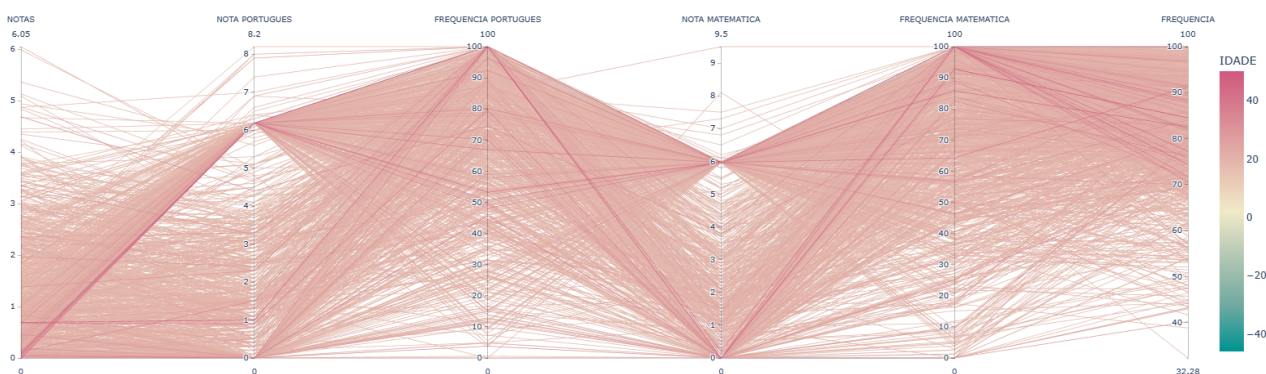


Fonte: Elaborada pelo autor

A Figura 22 mostra que a grande maioria dos estudantes evadidos possuíam uma até 1 salário mínimo como renda familiar mensal e entre 4 e 6 membros no contexto familiar, indicando que a condição de pobreza impacta diretamente no quantitativo de evasão.

Além do contexto socioeconômico, é possível obter informações importantes dos registros de frequência e nota dos estudantes. A Figura 23 retoma essa análise com uma abordagem mais direta. O gráfico de coordenadas paralelas disponibiliza uma visão geral da situação. As linhas, que indicam cada estudante evadido, se comportam de forma diferente quando afetado pelas variáveis, ‘Notas’ e ‘Frequência’.

Figura 23 - Gráfico de coordenadas paralelas relacionando notas e frequência para alunos evadidos



Fonte: Elaborado pelo autor

Existe uma particularidade no SIGEduc, em que os professores não registram as frequências e automaticamente, por definição do órgão gestor, todos os estudantes recebem frequência máxima, uma vez que a mesma implica em reprovação, caso o estudante atinja o inferior a 75% de presenças. O professor então deverá registrar os dias em que o estudante estiver ausente apenas. Como pode-se ver, a grande maioria dos registros indicam notas baixas, porém frequência máxima, o que nem sempre reflete a realidade.

Em uma última análise, deseja-se verificar o comportamento de um gráfico de bolhas no contexto de uma escola, relacionando notas, frequência, renda mensal da família, e quantos membros residem com o estudante. Para transformar a variável categórica ‘MEMBROS DA FAMÍLIA’ foi utilizada a abordagem Label Encoding, uma codificação de rótulo que atribui um número inteiro único a cada categoria presente na coluna categórica. Para escolher a

escola, foi-se verificando através do método `groupby()`, quais as maiores ocorrências de evasão dentre as DIRECs, Cidades e Escolas, como mostra o exemplo da Figura 24.

Figura 24 - Utilização do método `groupby()`

```

grupo = df_evadidos.groupby('DIREC')['DEIXOU DE FREQUENTAR'].count()
print(grupo)

```

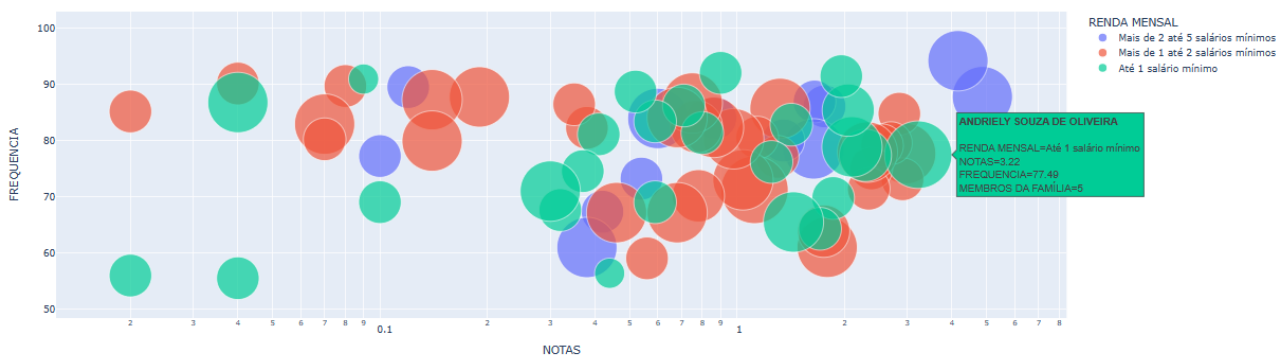
DIREC	DEIXOU DE FREQUENTAR
01ª DIREC - NATAL	497
02ª DIREC - PARNAMIRIM	399
03ª DIREC - NOVA CRUZ	103
04ª DIREC - SÃO PAULO DO POTENGI	32
05ª DIREC - CEARÁ MIRIM	211
06ª DIREC - MACAU	82
07ª DIREC - SANTA CRUZ	71
08ª DIREC - ANGICOS	15
09ª DIREC - CURRAIS NOVOS	41
10ª DIREC - CAICÓ	29
11ª DIREC - ASSU	102
12ª DIREC - MOSSORÓ	224
13ª DIREC - APODI	47
14ª DIREC - UMARIZAL	50
15ª DIREC - PAU DOS FERROS	125
16ª DIREC - JOÃO CÂMARA	32

Name: DEIXOU DE FREQUENTAR, dtype: int64

Fonte: Elaborado pelo autor

Chegou-se então a ESCOLA ESTADUAL PROFESSOR VARELA BARCA, da Cidade de Natal, 1ª DIREC. A Figura 25 mostra o gráfico de bolhas. O painel demonstrado na Figura 25 representa uma proposta de quadro de monitoramento escolar para acompanhamento dos estudantes a fim de identificar aqueles mais propensos a evadir de acordo com a análise feita neste trabalho. Apesar da frequência ser um dado problemático de acordo com as análises feitas nesse trabalho, propõe utilizá-la como parâmetro em outros contextos de gestão escolar, onde a medição da informação é feita de forma coerente com a realidade.

Figura 25 - Gráfico de bolhas interativo com cruzamento de variáveis



Fonte: Elaborado pelo autor

6. CONCLUSÃO

Como apresentado nos capítulos anteriores, a evasão escolar no ensino médio é um problema grave, silencioso e preocupante. Nesse sentido, o objetivo principal deste trabalho foi analisar dados e desenvolver recursos visuais eficazes para o monitoramento dos estudantes, como o painel obtido com o gráfico de bolhas disponibilizado pelo Plotly que de forma dinâmica apresenta os estudantes da Escola Estadual Professor Varela Barca, relacionando-os com variáveis de desempenho acadêmico e contexto socioeconômico, visando identificar precocemente os indícios de evasão e proporcionar possíveis intervenções no contexto da gestão escolar, baseando-se em dados de alunos do Ensino Médio da rede de educação pública do Estado do RN entre os anos de 2017 e 2019. Através de uma abordagem multidisciplinar e da utilização de recursos tecnológicos, sugere-se aprimorar as práticas de acompanhamento e gestão educacional, a fim de diminuir as taxas de evasão escolar e garantir que todos os alunos tenham acesso e permanência a uma educação de qualidade.

A visualização científica dos dados educacionais obtida através dos recursos de ciência de dados, revelou-se de extrema importância ao longo deste estudo, proporcionando resultados interessantes, como a predominância de estudantes de renda salarial familiar de até 1 salário mínimo e rendimento escolar baixo, despertando conhecimento sobre os padrões e tendências relacionados à evasão escolar. Através da coleta e interpretação dos dados, foi possível identificar variáveis mais e menos relevantes, que influenciam diretamente na evasão dos estudantes, como questões socioeconômicas, desempenho escolar, entre outros. Também foi possível compreender melhor o contexto de coleta desses dados e as fragilidades envolvidas no monitoramento da educação de modo geral, com rupturas e falhas por parte do sistema gerenciado pelo órgão gestor da educação no Estado, como por exemplo a interpretação ambígua da frequência escolar proporcionada pelo atual formato de registro de presenças e faltas por parte do professor.

Por fim, é fundamental ressaltar que este trabalho representa apenas um pontapé inicial para uma área de pesquisa em constante evolução, instigando que outros trabalhos possam continuar a explorar e expandir o conhecimento nessa área, importante e carente, sobretudo no contexto da educação pública estadual do RN. Sugere-se aprofundamento nessa linha de pesquisa, considerando dados mais profundos sobre o contexto socioeconômico, regional e cultural, por exemplo, evidenciando mais características desse público múltiplo que é o aluno da rede pública do Estado do RN.

7. REFERÊNCIAS

- AMARAL**, Fernando. Introdução à ciência de dados: mineração de dados e big data. Alta Books Editora, 2016.
- ARAÚJO**, Aliete Tuylla da Silva. Análise da usabilidade do Sistema Integrado de Gestão da Educação (SIGEduc) das escolas públicas do Rio Grande do Norte. 2019. 78 f. Monografia (Graduação) - Universidade Federal Rural do Semi-árido, 2019.
- BARBOSA**, Thiago Moura; **DA SILVA**, Iago Sinésio Ferris; **SANT'ANA**, Alex Sandro Coitinho. Os desafios do uso pedagógico do SIGEduc no contexto de escolas públicas dos municípios de Angicos/RN e Santana do Matos/RN. Educação & Tecnologia, v. 20, n. 3, 2019.
- BARROS**, Thiago Medeiros. Um processo orientado a dados para geração de modelo de predição de evasão escolar / Thiago Medeiros Barros. - Natal, 2020. 119 f.: il.
- BATISTA**, Gustavo Enrique de Almeida Prado Alves. Pré-processamento de dados em aprendizado de máquina supervisionado. 2003. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, University of São Paulo, São Carlos, 2003. doi:10.11606/T.55.2003.tde-06102003-160219. Acesso em: 2023-07-14.
- BRASIL**. Constituição da República Federativa do Brasil. Diário Oficial da União, República Federativa do Brasil, Brasília, 5 out. 1988.
- BRASIL**. Lei nº 9.394, de 20 de dezembro de 1996. Estabelece as Diretrizes e Bases da Educação Nacional. Diário Oficial da União, Brasília, DF, Brasília, DF, 23 dez. 1996.
- BRASIL**. Lei nº 12.013, de 06 de agosto de 2009. Altera o art. 12 da Lei no 9.394, de 20 de dezembro de 1996, determinando às instituições de ensino obrigatoriedade no envio de informações escolares aos pais, conviventes ou não com seus filhos. Diário Oficial da União, Brasília, DF, 07 ago. 2009.
- BLUM**, Avrim; **HOPCROFT**, John; **KANNAN**, Ravi. Foundations of Data Science (2018). URL: <https://www.cs.cornell.edu/jeh/book.pdf>, 2018.
- COELHO**, A. S. Introdução a análise de dados com python e pandas. Anais Eletrônicos ENUCOMP, p. 862-876, 2017.
- DINIZ**, Ednar. O censo escolar. Revista Brasileira de Estudos Pedagógicos, v. 80, n. 194, 1999.
- FERREIRA**, Gisele. Investigação acerca dos fatores determinantes para a conclusão do Ensino Fundamental utilizando Mineração de Dados Educacionais no Censo Escolar da

Educação Básica do INEP 2014. In: Anais dos Workshops do Congresso Brasileiro de Informática na Educação. 2015. p. 1034.

FLOR, Ana Paula dos Santos Oliveira. Sistema integrado de gestão da educação do Rio Grande do Norte - SIGEDUC e escola digital como espaço pedagógico. 2019. 180f. Tese (Doutorado em Educação) - Centro de Educação, Universidade Federal do Rio Grande do Norte, Natal 2019.

FREITAS, André Lage; **SANTANA JUNIOR**, O. V. Machine Learning: desafios para um Brasil competitivo. Revista da Sociedade Brasileira de Computação, v. 38, n. 01, 2019.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO

TEIXEIRA. Dados Abertos. Indicadores Educacionais. Taxas de Rendimento. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/indicadores-educacionais/taxas-de-rendimento>>. Acesso em: 01 ago. 2021.

MCKINNEY, Wes. Python para análise de dados: Tratamento de dados com Pandas, NumPy e IPython. Novatec Editora, 2018.

MORAES, Maria Celina Bodin de (Org.). Apresentação do autor e da obra. In: RODOTÀ, Stefano. A vida na sociedade de vigilância: A privacidade hoje. Rio de Janeiro: Renovar, 2008. p. 1-12. Tradução: Danilo Doneda e Luciana Cabral Doneda.

MOURA, Bruna Rego de. Análise de Dados sobre Crimes Violentos Letais Intencionais no Rio Grande do Norte / Bruna Rego de Moura. - 2022. 42 f.: il.

OECD. (2020). Understanding the Socio-economic Impact of Early School Leaving. Disponível em:

<https://www.oecd.org/education/understanding-the-socio-economic-impact-of-early-school-leaving-3>. Acesso em 20 de jun. de 2023.

RIO GRANDE DO NORTE. Lei nº 10.049, de 27 de janeiro de 2016. Aprova o Plano Estadual de Educação do Rio Grande do Norte (2015-2025) e dá outras providências. Diário Oficial do Estado do Rio Grande do Norte, Natal, 27 jan. 2016. Disponível em: http://www.diariooficial.rn.gov.br/dei/dorn3/docview.aspx?id_jor=00000001&data=20160128&id_doc=524349. Acesso em: 18 jan. 23.

RAUTENBERG, Sandro; **DO CARMO**, Paulo Ricardo Viviurka. Big data e ciência de dados: complementariedade conceitual no processo de tomada de decisão. Brazilian Journal of Information Science: research trends, v. 13, n. 1, p. 56-67, 2019.

SANT'ANA, Ricardo César Gonçalves. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. Informação & Informação, v. 21, n. 2, p. 116-142, 2016.

- SAMPIERI**, Collado; **COLLADO**, F. C. Lucio. Metodología de la Investigación, v. 4, 2013.
- SALATA**, André. Razões da evasão: abandono escolar entre jovens no Brasil. Interseções. Revista de Estudos Interdisciplinares, v. 21, n. 21-1, 2019.
- SILVA**, José António. Desenvolvimento de Modelos (de Machine Learning) Baseados em Dados de Monitorização Contínua de Barragens de Betão para Interpretação do Comportamento Estrutural Observado. 2022. 112f. Dissertação (Mestrado em Engenharia Civil) – Faculdade de Engenharia, Universidade do Porto, Porto, Portugal, 2021.
- SOUZA**, Elizeu Clementino de; **RAMOS**, Michael Daian Pacheco. Indicadores Educacionais do território do Piemonte da Diamantina-Bahia: apontamentos iniciais. R. Educ. Públ., Cuiabá, v. 28, n. 69, p. 677-699, set. 2019. Disponível em <http://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S2238-20972019000300677&lng=pt&nrm=iso>. acessos em 29 jan. 2023. Epub 15-Jun-2020. <https://doi.org/10.29286/rep.v28i69.8179>.
- SILVA**, Igor Rodrigues Sousa; **SILVA**, Rogério Oliveira da. Linguagem De Programação Python Python Programming Language. Revista Tecnologias em Projeção, v. 10, p. 55-71, 2019.
- TEIXEIRA**, Enise Barth. A análise de dados na pesquisa científica: importância e desafios em estudos organizacionais. Desenvolvimento em questão, v. 1, n. 2, p. 177-201, 2003.
- SILVA**, Filho, R. B., & **ARAÚJO**, R. M. de L. (2017). Evasão e abandono escolar na educação básica no Brasil: fatores, causas e possíveis consequências. Educação Por Escrito, 8(1), 35–48. <https://doi.org/10.15448/2179-8435.2017.1.24527>
- TAVARES**, Sayonara Alice Cirilo. Análise de Comportamento de Alunos na Plataforma Multiprova Baseada em Técnicas de Ciência de Dados / Sayonara Alice Cirilo Tavares. - 2022. 41f.: il.
- WASKON**, M. L., (2021). Seaborn: Visualização estatística de dados. Jornal de Software de Código Aberto, 6(60), 3021, <https://doi.org/10.21105/joss.03021>.