

Análise exploratória do perfil dos alunos e casos de insucessos nas disciplinas do curso de Ciência da Computação na UFERSA – Mossoró – RN - Brasil

Exploratory analysis of student profiles and cases of failure in subjects of the Computer Science course at UFERSA – Mossoró – Brazil

DOI: 10.55905/rcssv12n7-011

Received on: October 13th, 2023

Accepted on: November 17th, 2023

João Victor Leite dos Santos

Graduado em Ciência e Tecnologia

Instituição: Universidade Federal Rural do Semi-Árido

Endereço: Av. Francisco Mota, 572, Mossoró-RN, CEP: 59625-900

E-mail: joao.santos67982@alunos.ufersa.edu.br

Angélica Félix de Castro

Doutora em Geodinâmica

Instituição: Universidade Federal Rural do Semi-Árido

Endereço: Av. Francisco Mota, 572, Mossoró-RN, CEP: 59625-900

E-mail: angelica@ufersa.edu.br

Amanda Gondim de Oliveira

Doutora em Engenharia Elétrica

Instituição: Universidade Federal Rural do Semi-Árido

Endereço: Av. Francisco Mota, 572, Mossoró-RN, CEP: 59625-900

E-mail: amandagondim@ufersa.edu.br

RESUMO

O trabalho a seguir realiza uma análise exploratória de dados com a finalidade de compreender o perfil do curso e seus índices de insucessos nas disciplinas. Este trabalho tem como objetivo auxiliar os membros do conselho, bem como a coordenação do curso, para realizar o planejamento visando a tomada de decisão, utilizando a linguagem Python, por meio das bibliotecas Pandas, Matplotlib e Seaborn, foi realizado a extração de informações das bases de dados fornecidas. Assim, foi possível compreender as tendências dentro de um curso de graduação, ou seja, conhecer quantos estudantes são do sexo masculino, quais disciplinas mais reprovam e qual a porcentagem de alunos formados ao longo dos anos, é de extrema importância para entender o quadro geral do curso.

Palavras-chave: perfil dos estudantes, retenção, evasão, análise exploratória.

ABSTRACT

The following work carries out an exploratory data analysis with the purpose of understanding the profile of the course and its failure rates in the subjects. This work aims to assist board members, as well as course coordination, to carry out planning aimed at decision making, using the Python language, through the Pandas, Matplotlib and Seaborn libraries, extracting information from the databases provided. Thus, it was possible to

understand the trends within an undergraduate course, that is, knowing how many students are male, which subjects they fail the most and what percentage of students graduated over the years, is extremely important to understand the general picture. of the course.

Keywords: student profile, retention, evasion, exploratory analysis.

1 INTRODUÇÃO

Nos últimos anos estamos produzindo cada vez mais dados em maior volume e velocidade, esses dados possuem formas e características, sendo produzidos por diversos setores. Atualmente existem mais de 2,7 zettabytes de dados disponíveis digitalmente, e deverá crescer para 180 zettabytes até 2025 (CETAX, 2022). Esse crescimento até o momento não indica haver limites, em termos de armazenamento não parece haver sinais de desgaste, nem quanto ao processamento desses dados. A partir disso, podemos concluir que não chegamos ao pico de produção e armazenamento, ao contrário, existindo perspectivas de crescimento para os próximos anos, tendo em vista as condições na qual nos encontramos.

Dados isolados podem não ter nenhuma informação agregada, portanto, transformar dados em informações pode ser bastante útil para tomadores(as) de decisões, que diariamente determinam o destino de diversas organizações. A área da educação é um nicho bastante amplo para se explorar, com dados que podem ir do ensino básico até o ensino superior. No Brasil, 55,5% dos estudantes que iniciaram a graduação em 2017 haviam desistido em 2021 (CAFARDO, 2023). Segundo dados do Mapa do Ensino Superior, no Brasil em 2023, a evasão nos cursos da área de Tecnologia da Informação (TI) supera a média das demais áreas do conhecimento, sendo de 38,5% (EDUCA + BRASIL, 2023).

Devido a isso, este trabalho dedica-se a realizar um estudo com foco no perfil dos alunos (2006-2023) e nos índices de insucessos (2017-2022), ou seja: reprovações, cancelamentos e trancamentos nas disciplinas do Curso de Ciência da Computação da Universidade Federal Rural do Semi-Árido, Campus Mossoró. Primeiro foi realizado um mapeamento do perfil dos estudantes, com dados de 2006 até 2023, que representam todos os alunos que participaram do curso desde sua criação. Em seguida iremos realizar o mapeamento dos casos de insucessos dos estudantes, utilizando dados de 2017 até 2022. Este trabalho possui a finalidade de servir de apoio à coordenação de curso e aos membros

dos conselhos, com o objetivo de auxiliar na tomada de decisão permitindo a esses agentes conhecerem o perfil exato de seus alunos, bem como o perfil de insucessos nas disciplinas. Nas próximas seções, serão apresentados a fundamentação teórica, materiais e métodos, resultados obtidos, e finalmente a conclusão, onde se fará um diagnóstico a respeito do que foi desenvolvido durante o trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Serão apresentados nessa seção alguns tópicos de natureza importante para a compreensão da realização do trabalho proposto. Os seguintes tópicos serão: i) Análise exploratória; ii) Estatística descritiva; iii) Ciência de dados.

2.1 ANÁLISE EXPLORATÓRIA DE DADOS (AED)

É um conceito onde o analista anteriormente à aplicação de estatística, examina os dados disponíveis, explorando e testando as eventuais relações entre variáveis, obtendo um entendimento básico dos dados.

Passando desta etapa, seguindo para a análise descritiva, na qual é bastante importante pois o pesquisador consegue obter uma compreensão geral dos aspectos quantitativos e qualitativos dos dados, realizando sínteses para se extrair informações, podendo responder perguntas relacionadas a esse conjunto de dados.

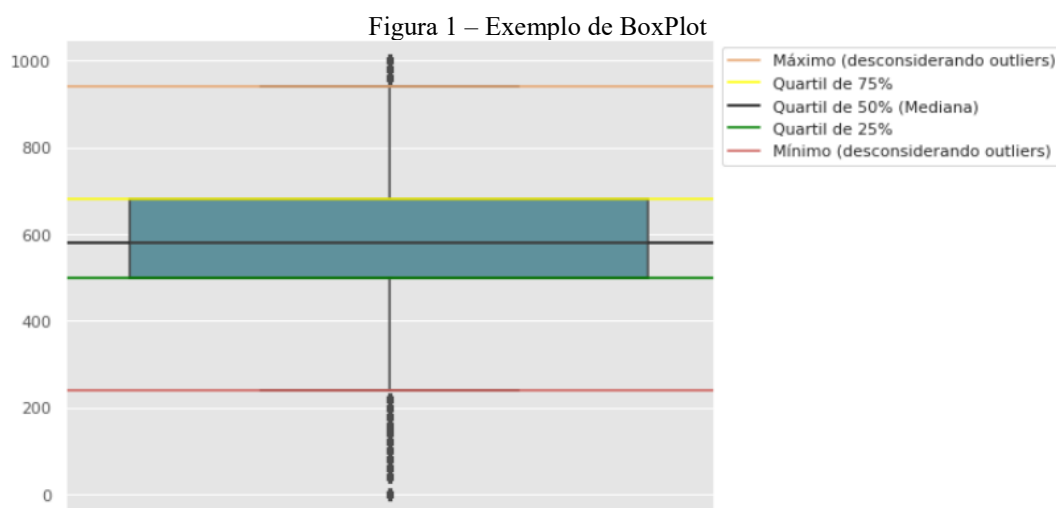
Em resumo, na análise exploratória se extrai informações de uma base de dados, porém sem as suposições de modelos probabilísticos (Machine Learning) (EACH USP, 2011). Na análise exploratória de dados, é comum a apresentação dos resultados por meio de gráficos e tabelas, sendo assim, na próxima seção serão apresentados os principais tipos de gráficos utilizados nesse tipo de trabalho.

2.1.1 Gráficos

São representações de informações obtidas de um conjunto de dados, sendo representadas por figuras geométricas, com o objetivo de tornar fácil a interpretação. Utilizando gráficos é possível encontrar padrões, realizar comparações e visualizar resultados, sendo utilizados em diversas áreas do conhecimento (SILVA, 2023).

- Gráfico de linha: Utilizado para representar as variáveis dependentes do tempo, é usado quando necessário haver uma representação das tendências em um determinado período (TIBCO, 2023a).

- Gráfico de Colunas: Utilizado para apresentar dados com categorias representadas por um retângulo, onde pode-se realizar comparações entre outros elementos presentes no gráfico (TIBCO, 2023b).
- Gráfico de Setores: Utilizado para representar dados de natureza estática, consistindo em poucas variáveis de forma a facilitar seu entendimento, este gráfico possui suas informações distribuídas em um formato circular. Possuindo o objetivo de realizar, uma comparação da contribuição de cada uma das partes com números percentuais (TIBCO, 2023c).
- Histograma: É utilizado para representar a distribuição de frequência de dados de uma variável, sendo utilizado quando se possui um conjunto de dados que possui uma única variável independente (TIBCO, 2023d).
- Boxplot: É utilizado para representar a distribuição dos dados, mostrando relações entre uma variável numérica e um agrupamento de variável, sendo usado para exibir e demonstrar graficamente como grupos de dados numéricos estão localizados, espalhados ou distorcidos – mostrando o quão amplos os valores dos dados estão espalhados utilizando o resumo de cinco números, os quais são: mínimo, primeiro quartil (Q1), mediana, terceiro quartil (Q3), máximo (TIBCO, 2023e). A figura 1 abaixo mostra um exemplo do que seria o gráfico de BoxPlot:



Fonte: Adaptado de SOUZA *et al.* (2023).

2.2 ESTATÍSTICA DESCRITIVA

A Estatística Descritiva realiza aplicações de técnicas para realizar a descrição, organização, e resumir um determinado conjunto de dados. De acordo com o IME

(Instituto Militar de Engenharia), a partir dela, é realizada a etapa inicial da análise de dados, onde usamos suas técnicas visando a compreensão dos dados.

Devido a evolução da computação, essa área da estatística se beneficiou bastante devido a quantidade de dados que se poderia processar (PEREIRA, 2019). Dentro da Estatística Descritiva, possui-se um conjunto de métricas que auxilia na descrição do conjunto de dados, abaixo apresentam-se algumas delas:

- Média: É uma medida de tendência central que se descreve pela soma dos valores de um conjunto de dados, dividida pelo número total dos valores.
- Mediana: Assim como a média, é uma medida de tendência central. No geral, o termo Mediana se refere ao que está entre dois pontos e dois extremos. Mediana nos diz que metade (50%) dos valores do conjunto de dados está abaixo dela e a outra metade está acima dela.
- Quartis: São valores que dividem os dados ordenados em quatro partes iguais, a partir disso é possível compreender a dispersão e a tendência central de um conjunto de dados.
- Desvio Padrão: Grau de dispersão de um conjunto de dados em relação à média.

2.3 CIÊNCIA DE DADOS

A Ciência de Dados é um campo de estudo interdisciplinar na qual possui a finalidade de obter informações de bases de dados, passando desde as etapas de extração, transformação e carregamento (ETL) dos dados, até a análise final, onde se irá extrair os insights. A Ciência de Dados é constituída pelas disciplinas: Matemática e Estatística, Ciência da Computação e Conhecimento específico das áreas de atuação (CETAX, 2022).

Importante informar que Souza *et al.* (2023) utilizou técnicas estatísticas e da ciência de dados para identificar características relevantes dos participantes do Exame Nacional do Ensino Médio do ano de 2019 e possíveis relações dessas características com o desempenho deles.

3 MATERIAIS E MÉTODOS

Nesta seção serão apresentadas as ferramentas utilizadas para compor este trabalho, sendo descrito o processo de extração/coleta e tratamento dos dados.

3.1 EXTRAÇÃO/COLETA E TRATAMENTO DE DADOS

Para a construção da base de dados, foram utilizados dois relatórios em formato PDF, extraídos via Sigaa através da coordenação do curso. O primeiro relatório obtinha os dados a respeito dos insucessos (reprovação, cancelamento etc.) do curso de Ciência da Computação, dos anos de 2017 até 2022. O segundo relatório possui os dados a respeito do sexo dos estudantes, status (ativo, cancelado etc.) em relação ao ano de 2023, e ano de ingresso do estudante. Ambos os relatórios são referentes ao Campus Mossoró.

O tratamento desses dados foi realizado a partir da importação dos relatórios para o Microsoft Excel, organizando de forma manual as linhas e colunas para que fosse importado para o ambiente Jupyter. Ao final do tratamento, a base de dados de insucessos possuía 2936 registros e 7 colunas, enquanto a base que possuía os dados sobre sexo e status tinha 936 registros e 5 colunas. Ambas as bases de dados foram trabalhadas individualmente, para em seguida serem unificadas para a realização da última etapa da análise.

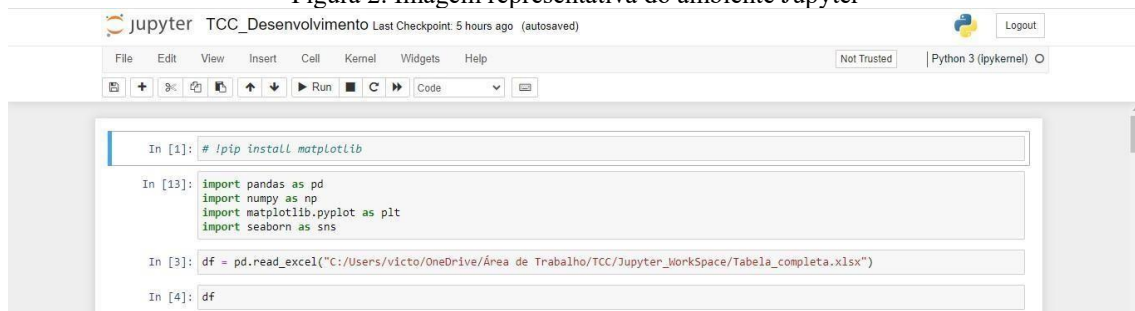
3.2 IMPORTAÇÃO DA BASE DE DADOS PARA O JUPYTER NOTEBOOK

Em seguida foi realizada a importação para o Jupyter Notebook. Este trabalho utilizou dessa ferramenta para realizar as consultas e plotagem dos gráficos, que serão apresentados posteriormente. O ambiente é descrito em seu próprio site como:

“[...] Um projeto de código aberto sem fins lucrativos, nascido do Projeto IPython em 2014, à medida que evoluiu para apoiar a ciência de dados interativa e a computação científica em todas as linguagens de programação. O Jupyter sempre será um software 100% de código aberto, gratuito para uso de todos e lançado sob os termos liberais da licença BSD modificada (JUPYTER ORG, 2022).”

Na figura 2 está o ambiente de trabalho na qual foram executadas as consultas e gráficos que serão apresentados posteriormente neste trabalho.

Figura 2. Imagem representativa do ambiente Jupyter



```

jupyter TCC_Desenvolvimento Last Checkpoint: 5 hours ago (autosaved)
File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)
In [1]: # !pip install matplotlib
In [13]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
In [3]: df = pd.read_excel("C:/Users/victo/OneDrive/Área de Trabalho/TCC/Jupyter_WorkSpace/Tabela_completa.xlsx")
In [4]: df
  
```

Fonte: Autoria própria.

3.3 LINGUAGEM DE PROGRAMAÇÃO

Neste trabalho optou-se em utilizar a linguagem Python, tendo em vista diversas vantagens que ela apresenta, tais como sua utilização massificada em projetos de análise de dados, a variedade de bibliotecas disponíveis, e por fim, por possuir uma comunidade ativa no mundo todo, na qual disponibiliza uma quantidade enorme de exemplos e resolução de problemas. As bibliotecas utilizadas para desenvolver o estudo foram (SOUZA et al., 2023):

- **Pandas:** é uma biblioteca Python utilizada para análise de dados. Com ela, foi feita a leitura, tratamento e processamento dos dados.
- **Matplotlib:** é uma biblioteca utilizada para criar gráficos diversos para tipos de dados variados. Grande parte dos gráficos apresentados neste trabalho foram feitos utilizando esta biblioteca.
- **Seaborn:** é mais uma biblioteca que auxilia na criação de gráficos. Costuma ter um layout mais apresentável que os gráficos criados pelo matplotlib.

4 RESULTADOS E DISCUSSÃO

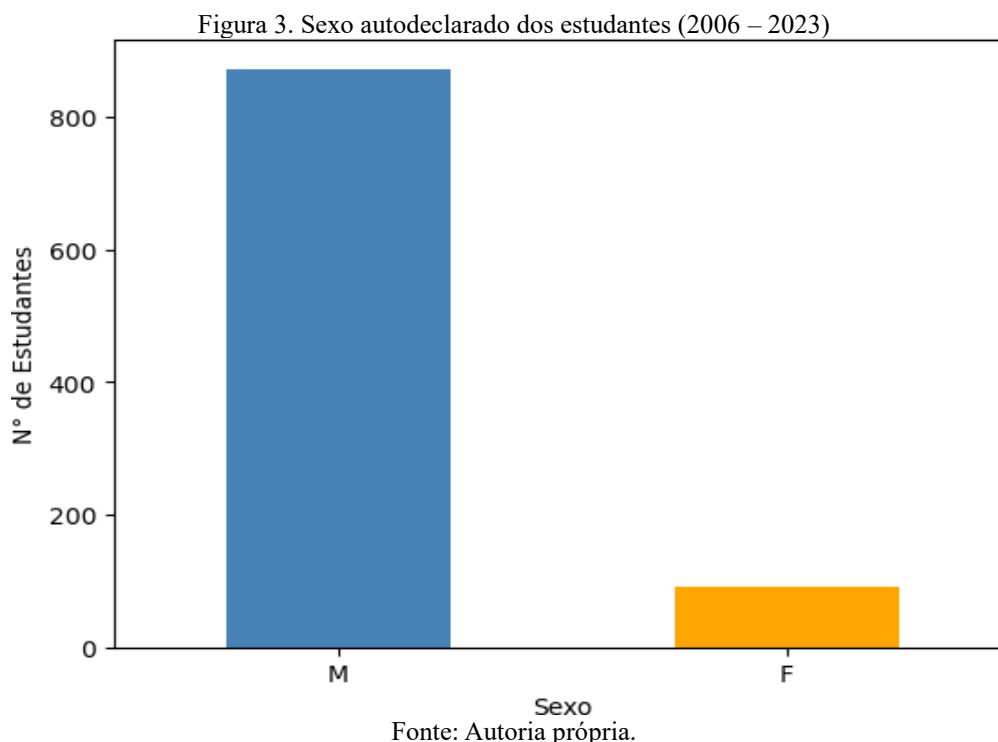
Esta seção irá apresentar os resultados que foram extraídos através dos gráficos plotados no ambiente Jupyter, serão apresentados os resultados obtidos da análise exploratória da base de dados, que está dividido em 3 partes, a primeira irá apresentar a análise do perfil dos alunos (2006-2023), em seguida iremos apresentar as análises dos casos de insucessos (2017-2022), onde primeiro não iremos considerar as variáveis: status, sexo, ano de ingresso. Na parte final, serão consideradas essas variáveis nos casos de insucessos para entendermos os comportamentos das variáveis analisadas.

4.1 ANÁLISE DO PERFIL DOS ESTUDANTES

Nesta seção, será realizada a análise do perfil dos estudantes do curso de Ciência da Computação, a base de dados contém informações de 2006 até 2023, possuindo 936 registros e 5 colunas.

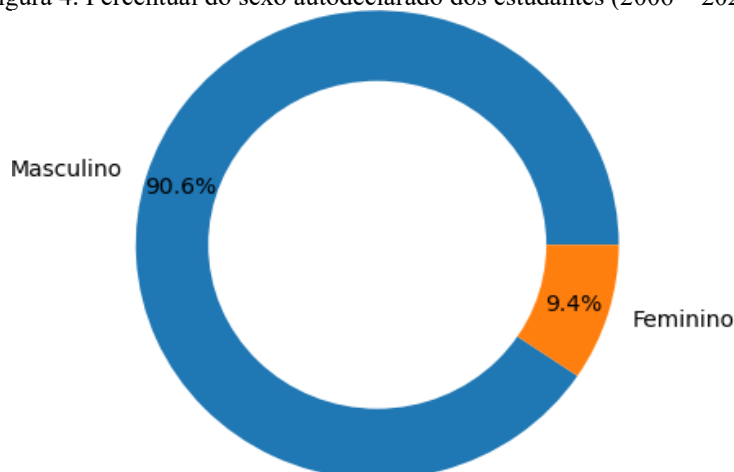
4.1.1 Em relação ao sexo dos estudantes

A Figura 3 apresenta a seguinte informação, dos 17 anos de existência do curso de Ciência da Computação da UFERSA - Mossoró, a quantidade de estudantes que passaram pelo curso foi de 963 alunos, dentre esses estudantes, o quantitativo de pessoas do sexo masculino são de 872 alunos, representando 90,6% dos estudantes. Enquanto o quantitativo referente as pessoas do sexo feminino, são de 91 alunas, representando 9,4% do sexo feminino.



A Figura 4 apresenta a distribuição dos estudantes por gênero em termos percentuais.

Figura 4. Percentual do sexo autodeclarado dos estudantes (2006 – 2023)



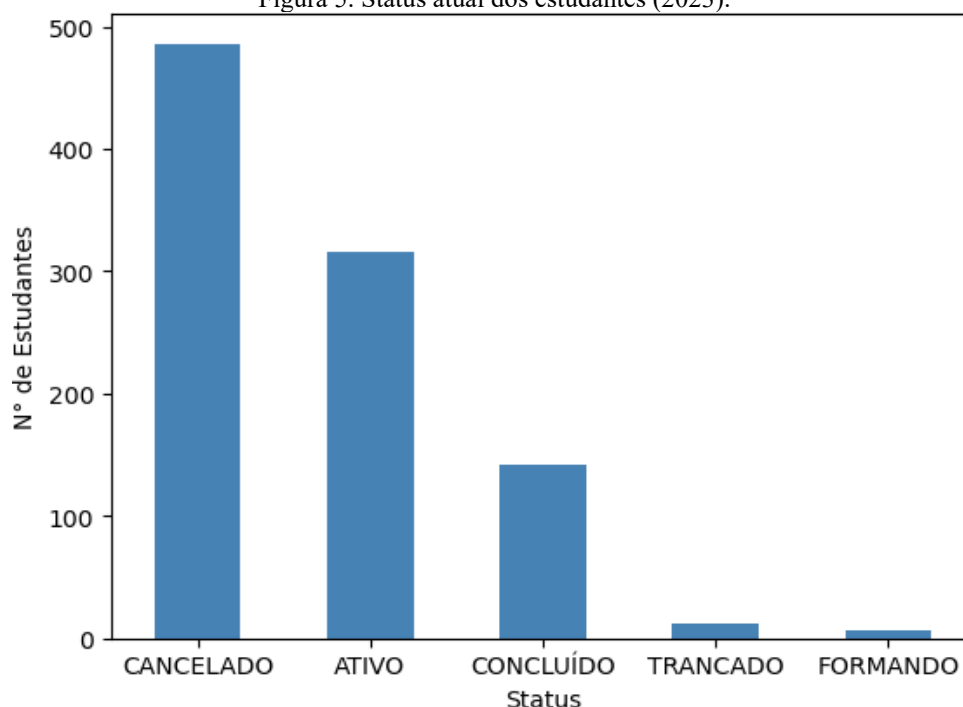
Fonte: Autorial própria.

4.1.2 Em relação ao status no curso

A Figura 5 e Figura 6, onde ambos descrevem o status atual (2023) de cada aluno do curso de Ciência da Computação, ao longo dos 17 anos de fundação do curso, na qual representa as quantidades absolutas e percentuais dos alunos em cada variável. Desses alunos(as), possuem 486 matrículas canceladas (representa 50,5%), 316 matrículas ativas (32%), 142 concluídas (14,7%), 12 trancamentos (1,2%) e 7 formandos (0,7%).

Abaixo temos um gráfico de barras (Figura 5), onde podemos observar os valores absolutos de cada categoria.

Figura 5. Status atual dos estudantes (2023).



Fonte: Autorial própria.

Abaixo temos um gráfico de rosca (Figura 6), onde podemos observar as porcentagens com proporcionalidade, tornando a informação mais legível.

Figura 6. Status atual dos estudantes (2023)



Fonte: Autorial própria.

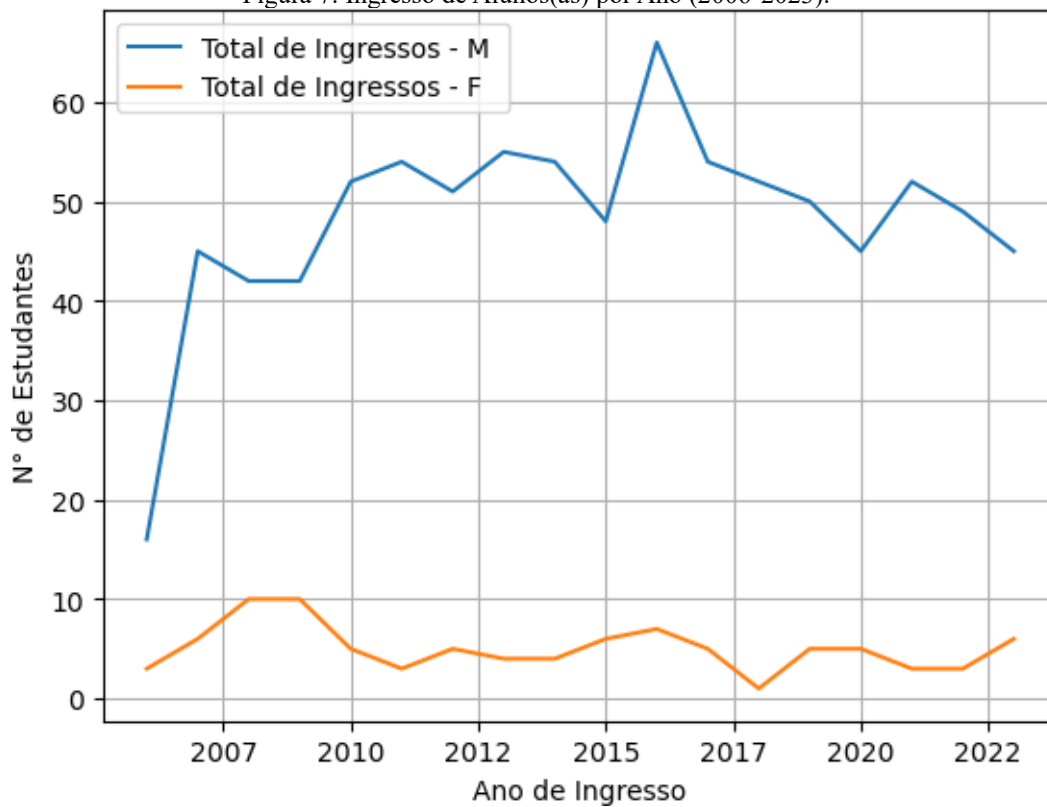
4.1.3 Em relação ao quantitativo de ingressantes no curso por ano

Na Figura 7, temos um gráfico de linhas onde conseguimos observar o ingresso dos alunos através dos anos, separados em dois grupos, sexo masculino e sexo feminino. Analisando o número de ingressos através dos anos, notamos que de 2006 a 2007, o número de ingressos passou de 20 para 50. Ao longo dos anos, a entrada de alunos variou entre 50 e 60 alunos, somando ambos os sexos.

Essa variação explica-se porque, para além das 50 vagas que são anualmente ofertadas para a entrada direta no curso, a Ufersa abre editais de vagas ociosas permitindo entrada para estudantes: portadores de diploma, reingresso e reopção. Além de alunos de outras universidades, que desejam ingressar no curso.

Deve-se notar a diferença entre ingressos quando são destacados o sexo, enquanto o masculino possui maiores ingressos, o sexo feminino possui taxas menores.

Figura 7. Ingresso de Alunos(as) por Ano (2006-2023).



Fonte: Autoria própria.

4.2 ANÁLISE DOS INSUCESSOS EM DISCIPLINAS

A partir deste ponto, iremos direcionar o trabalho para os casos de insucessos nas disciplinas do curso, a base de dados contém informações de 2017 até 2022, possuindo 2936 registros e 7 colunas.

4.2.1 Em relação a quantidade de insucessos por aluno no curso

A Tabela 1 apresenta as estatísticas descritivas, em relação a quantidade de insucessos por alunos. Esses dados foram extraídos a partir da consulta realizada para o desenvolvimento da Figura 8.

Tabela 1. Estatísticas Descritivas dos Insucessos por Aluno

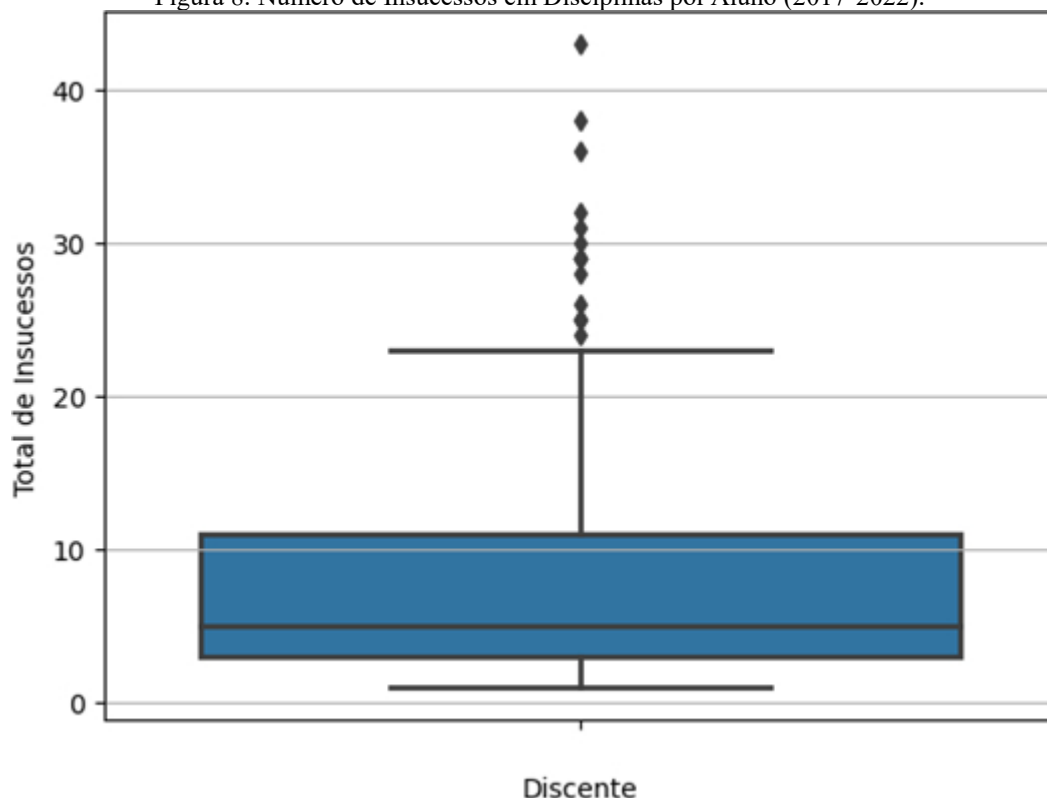
| ESTATÍSTICAS | VALORES |
|---------------------|----------------|
| Contagem | 388 |
| Média | 7.63 |
| Desvio Padrão | 7.07 |
| Mínimo | 1 |
| 1° Quartil | 3 |
| 2° Quartil | 5 |
| 3° Quartil | 11 |
| Máximo | 43 |

Fonte: Autoria própria.

Na Figura 8, foi utilizado um gráfico de BoxPlot na qual compilou os índices de insucessos por aluno durante os anos de 2017-2022 (12 semestres). Utilizando a função describe(), foi constatado que a quantidade de alunos com insucessos durante o período analisado foram 388. De acordo com o cálculo da função describe(), a média de insucessos por estudante foi de 7.63, no entanto esta métrica está enviesada, podemos concluir isso devido ao resultado do desvio padrão, igual a 7.07; ou seja, existe uma alta distribuição entre os dados em relação à média.

Devemos destacar também o intervalo interquartil dos dados, onde mostra que 25% dos alunos (1° Quartil) possuem até 3 insucessos, 50% (2° Quartil) possuem até 5 insucessos, enquanto 75% (3° Quartil) dos alunos possuem até 11 insucessos. Por fim, deve-se comentar a presença dos outliers neste gráfico, na qual ocorre quando o aluno possui um alto número de insucessos, manifestando-se a partir de 24 insucessos registrados, deve-se destacar também que o número de outliers presentes na base de dados contribui para o resultado da média. Abaixo podemos conferir o gráfico.

Figura 8. Número de Insucessos em Disciplinas por Aluno (2017-2022).

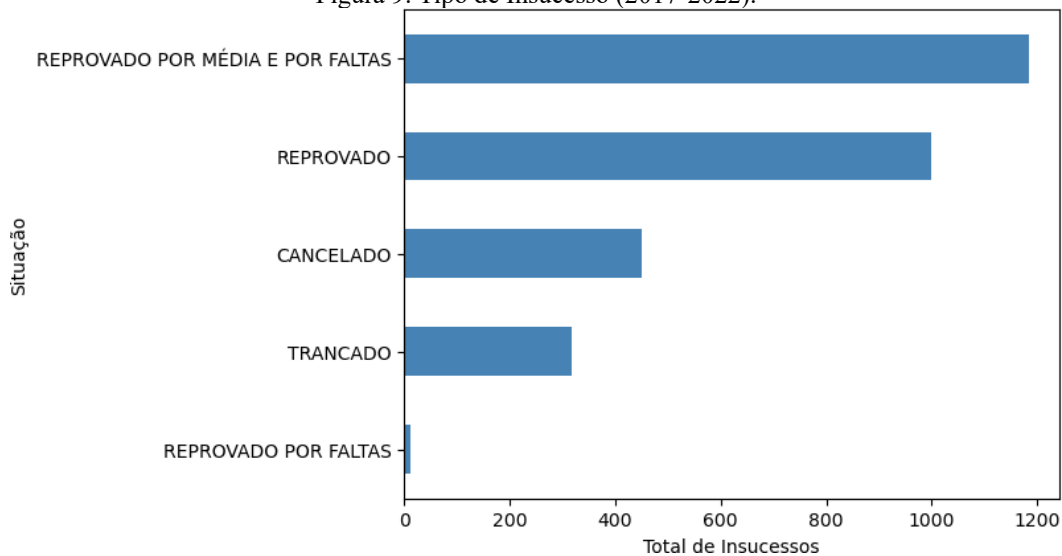


Fonte: Autoria própria

4.2.2 Em relação ao tipo de insucesso dos alunos no curso

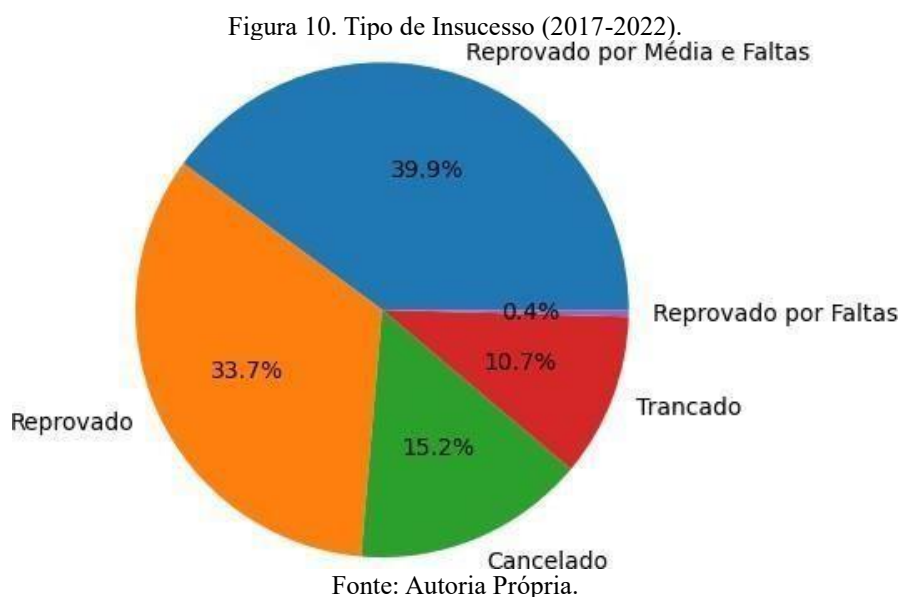
Na Figura 9, temos um gráfico de barras que apresenta os tipos de insucessos dos estudantes, essas categorias são: reprovado por média e por faltas, reprovado por média, cancelado, trancado e reprovado por faltas.

Figura 9. Tipo de Insucesso (2017-2022).



Fonte: Autoria Própria.

Podemos observar que as duas primeiras categorias representam as maiores taxas de insucesso entre as 5, somando 2.182 observações na base de dados, representando 73,6%, podendo ser observado na Figura 10. É de se notar que os alunos raramente reprovam apenas por faltas, representando 0,4%. Isso pode indicar que, quem normalmente reprova por falta também reprova por média, caracterizando um abandono da disciplina. Por fim, as categorias cancelado e trancado representam taxas de 15,2% e 10,7%, respectivamente. A Figura 10 abaixo apresenta através do gráfico de setores, a distribuição dos tipos de insucesso percentualmente.



4.2.3 Em relação a quantidade de insucessos por ano

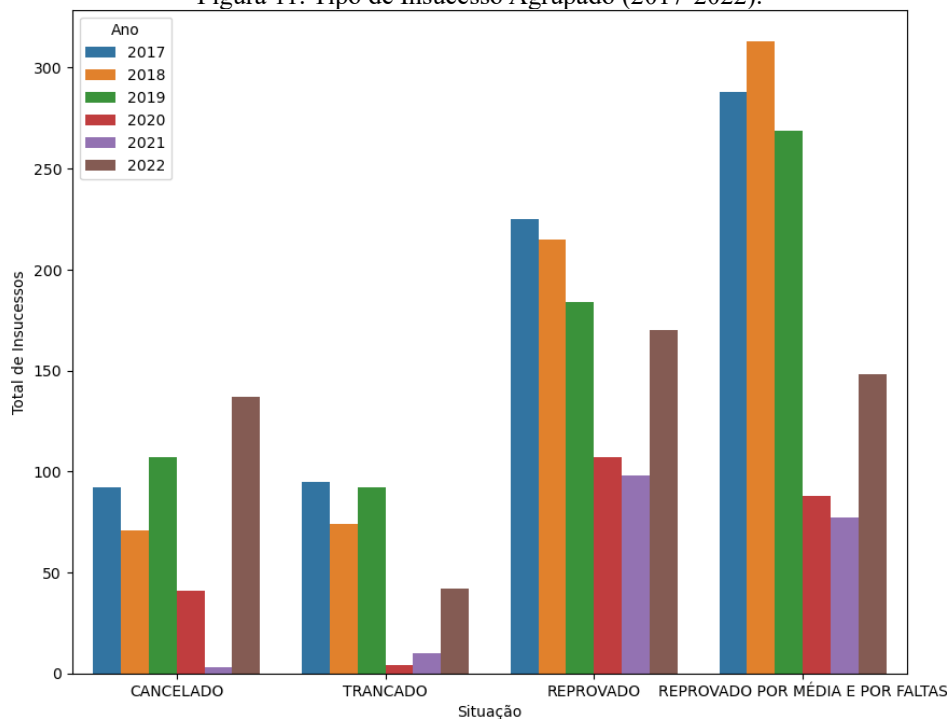
Na Figura 11, temos um gráfico de barras agrupado por insucesso na qual apresenta evolução do número de insucessos a cada ano, onde estão divididos em 4 grupos: cancelado, trancado, reprovado e reprovado por média e faltas.

Os anos de 2017, 2018 e 2019 antecedem a pandemia, e possuem maiores níveis em todos os grupos se comparados aos anos de 2020 e 2021, na qual houve uma redução desses índices. Essa diminuição dos casos de insucesso, deve-se em parte à resolução do CONSEPE/UFERSA N° 005/2020, onde no Capítulo VI - Das Matrículas e Cancelamentos (art. 18), trata sobre a possibilidade de o aluno realizar a exclusão da disciplina até o último dia letivo, sem prejuízo no histórico e nos índices acadêmicos dos discentes. Devido a essa condição, houve a diminuição dos casos de insucessos.

Em 2022, no pós-pandemia, nota-se que todos os grupos tiveram aumentos considerando os anos anteriores. Em especial, no caso de status cancelado, nota-se que os

índices de 2022 supera dados pré-pandemia, fenômeno que não acontece nos demais grupos.

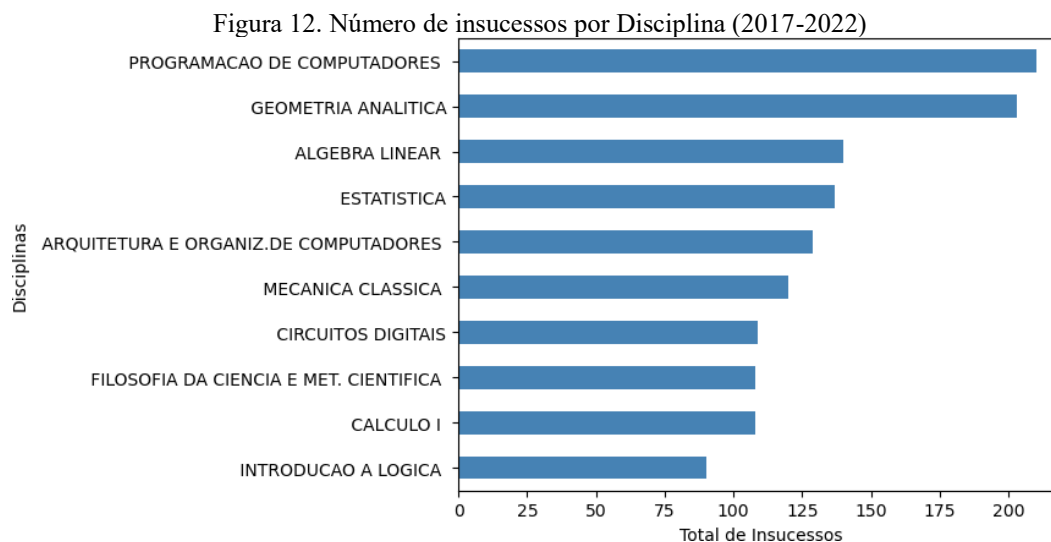
Figura 11. Tipo de Insucesso Agrupado (2017-2022).



Fonte: Autoria Própria.

4.2.4 Em relação às disciplinas com mais casos de insucesso

Na Figura 12, temos um gráfico de barras onde foram extraídas as 10 disciplinas com maiores índices de insucesso durante os anos de 2017-2022, as disciplinas disponíveis no total da base de dados eram de 92. Devido a quantidade de disciplinas disponíveis, foi necessário realizar uma filtragem entre as 10 disciplinas com as maiores taxas de insucesso.

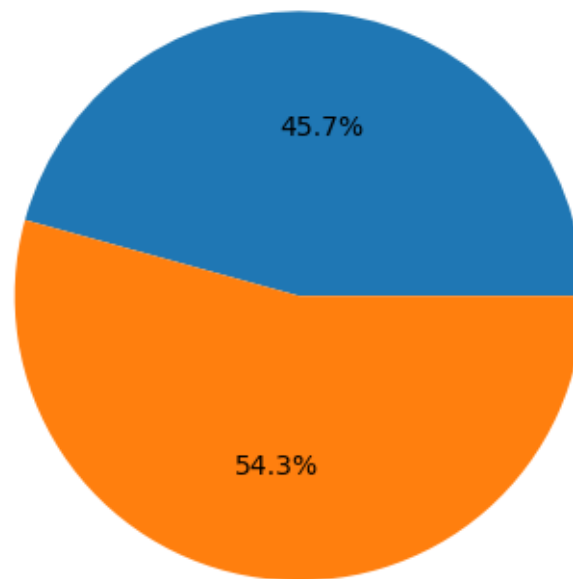


Fonte: Aatoria Própria

Dentre as disciplinas contempladas pelo filtro, devemos observar que mecânica clássica não faz parte da grade obrigatória do curso. Enquanto cálculo 1, é uma disciplina equivalente à disciplina de fundamentos de cálculo. As demais disciplinas são da grade obrigatória do curso. Observando as áreas em que estão contidas essas disciplinas, podemos pontuar que temos quatro disciplinas na área da computação (Programação de computadores, Arquitetura e Organização de Computadores, Circuitos Digitais, Introdução a lógica), quatro disciplinas da área lógico-matemático (Geometria analítica, Álgebra Linear, Estatística, Cálculo 1), uma de física (Mecânica Clássica) e uma de ciências humanas (Filosofia da Ciência e Metodologia científica).

É necessário destacar que 9 das 10 disciplinas que mais reprovaram durante os 6 anos, estão inseridas nos 3 primeiros semestres do curso de Ciência da Computação. A somatória de insucessos dessas 10 disciplinas é igual a 1354, abarcando 45,7% de insucessos de todas as disciplinas, que como dito anteriormente são 92, enquanto as outras 82 disciplinas são responsáveis no total por 54,3% dos casos de insucessos. Na Figura 13 abaixo, temos um gráfico de setores que representa o contraste das 10 disciplinas com maiores casos de insucesso, e o restante das 82 disciplinas agrupadas.

Figura 13. Total de Insucessos (Acumulado).
Acumulado (10 disciplinas)



Acumulado (82 disciplinas)

Fonte: Aatoria Própria

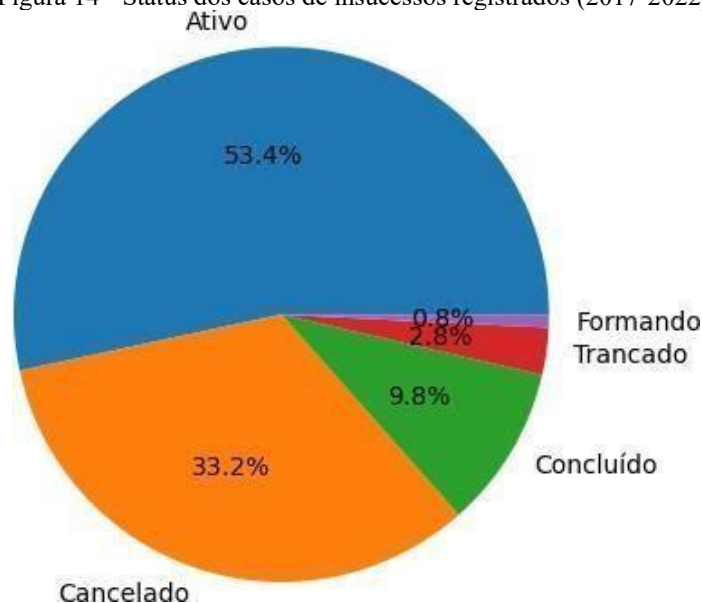
4.3 ANÁLISE DO PERFIL DE INSUCESSOS DOS ESTUDANTES

Nesta parte, este trabalho irá se dedicar em analisar o perfil dos estudantes que obtiveram insucessos durante os anos de 2017-2022.

4.3.1 Análise do status de insucessos

Na Figura 14, temos um gráfico de setores que apresenta o perfil via status dos alunos que possuem registrado algum insucesso. Dito isso, tivemos o seguinte resultado: o número de matrículas ativas no período são 206 (53,23%), enquanto o número de matrículas canceladas são 128 (33,33%), as matrículas concluídas são 38 (9,82%). O número de matrículas trancadas é 11 (2,84%), matrículas formandas são 3 (0,78%).

Figura 14 - Status dos casos de insucessos registrados (2017-2022).



Fonte: Autoria Própria

5 CONCLUSÕES

Compreender a natureza dos dados, em aspectos gerais e específicos, e sob diferentes perspectivas, determina as chances de sucesso na elaboração de intervenções incisivas, que visam sanar os problemas. Os resultados das análises permitiram observar a diferença do quantitativo entre pessoas de ambos os sexos, que ocupam e ocuparam o curso, podendo ser visualizado que pessoas do sexo feminino representam menos de 10%, enquanto pessoas do sexo masculino representam pouco mais de 90%.

Dentre as eventuais soluções para os problemas do baixo ingresso de pessoas do sexo feminino, poderia ser através da elaboração de projetos de extensão, voltados para divulgação do curso para o público feminino, visando o aumento do ingresso dessas estudantes. Essa é uma ação que pode ocasionar o aumento do índice de ingresso dessas pessoas no futuro.

Além disso, foi constatado durante a análise, que 9 das 10 disciplinas que obtiveram maiores taxas de insucessos, estão todas contidas nos primeiros 3 semestres. Representando em conjunto 45,7% do total de insucessos gerados entre os anos de 2017-2022.

Como medida visando reduzir esses casos, poderia se dar foco nas disciplinas que possuem os maiores índices registrados, elaborando novas formas de avaliação, abrindo editais com mais ofertas de monitoria (remunerada ou não), e até mesmo, a divulgação dos resultados obtidos, na qual indicam que aqueles que possuem insucessos nas

primeiras disciplinas do curso, há maiores possibilidades de demora na conclusão. Isso pode fazer com que os alunos matriculados, percebam que precisam se esforçar mais para não reprovar nessas disciplinas.

Por fim, tornar público as descobertas realizadas neste trabalho, é bastante importante, permitindo que seja de conhecimento de todo corpo docente e discente, essa prática é essencial para mudar no futuro os índices que foram apresentados neste trabalho.

REFERÊNCIAS

- CAFARDO, R; 2023. **Por que TI tem um dos maiores abandonos?**. Disponível em: <<https://nace.com.br/mais-da-metade-dos-alunos-desiste-da-faculdade-no-brasil-por-que-ti-tem-um-dos-maiores-abandonos-06-2023/>>. Acesso em: 29 set. 2023.
- CETAX, 2022. **Ciência de Dados: O que é, conceito e definição**. Disponível em: <<https://cetax.com.br/data-science-ou-ciencia-de-dados/>>. Acesso em: 25 set. 2023.
- CETAX, 2022. **Diferença entre Data Science, Big Data & Data Analytics**. Disponível em: <<https://cetax.com.br/data-science-vs-big-data-vs-data-analytics/>> Acesso em: 25 set. 2023.
- EACH USP, 2011. **Análise Exploratória de Dados**. Disponível em: <http://www.each.usp.br/lauretto/SIN5008_2011/aula01/aula1>. Acesso em: 25 set. 2023.
- EDUCA + BRASIL, 2023. **Mapa do ensino superior aponta evasão em cursos de TI**. Disponível em: <<https://www.educamaisbrasil.com.br/educacao/carreira/mapa-do-ensino-superior-aponta-evasao-em-cursos-de-ti>>. Acesso em: 29 set. 2023.
- JUPYTER.ORG, 2022. **About Us - Project Jupyter's origins and governance**. Disponível em: <<https://jupyter.org/about>>. Acesso em: 25 set. 2023.
- PEREIRA, 2019. **Estatística Descritiva!**, Voitto. Disponível em: <<https://www.voitto.com.br/blog/artigo/estatistica-descritiva>>. Acesso em: 25 set. 2023.
- SILVA, Luiz Paulo Moreira. **"O que é gráfico?"**, Brasil Escola. Disponível em: <<https://brasilecola.uol.com.br/o-que-e/matematica/o-que-e-grafico.htm>>. Acesso em 26 de set. 2023.
- SOUZA, T.O.; DE CASTRO, A.F.; DE OLIVEIRA, A.G. **Análise de dados: um estudo do perfil dos participantes do ENEM 2019**. Seven Editora, 2023.
- SOUZA, T. O.; CASTRO, A. F. e OLIVEIRA, A. G. 2023. **Understanding a Profile of the Participants of the Exame Nacional do Ensino Medio (ENEM), Brazil, in the Year 2019, Through Data Analysis**. Journal of Computer Science. <https://doi.org/10.3844/jcssp.2023.888.899>.
- TIBCO, 2023a. **O que é um gráfico de linhas?** Disponível em: <<https://www.tibco.com/pt-br/reference-center/what-is-a-line-chart>>. Acesso em: 25 set. 2023.
- TIBCO, 2023b. **O que é um gráfico de colunas?** Disponível em: <<https://www.tibco.com/pt-br/reference-center/what-is-a-column-chart>>. Acesso em: 25 set. 2023.
- TIBCO, 2023c. **O que é um gráfico de pizza?** Disponível em: <<https://www.tibco.com/pt-br/reference-center/what-is-a-pie-chart>>. Acesso em: 25 set. 2023.
- TIBCO, 2023d. **O que é um gráfico de histograma?** Disponível em: <<https://www.tibco.com/pt-br/reference-center/what-is-a-histogram-chart>>. Acesso em: 25 set. 2023.
- TIBCO, 2023e. **What is a Box Plot?** Disponível em: <<https://www.tibco.com/reference-center/what-is-a-box-plot>>. Acesso em: 25 set. 2023.